ON THE

FAST FREQUENCY SHIFT KEYING

CHARILAOS E. MOUTAFIS

MAJOR TECHNICAL REPORT

in

The Faculty

Of

ENGINEERING

Presented in Partial Fulfillment of the Requirements for
the Degree of Master of Engineering at
Concordia University
Montreal, Quebec, Canada

APRIL, 1979

ABSTRACT

ON THE FAST FREQUENCY SHIFT KEYING

CHARILAOS E. MOUTAPIS

Fast Frequency Shift Keying (FFSK) modulation technique is recently considered for use in non-linear, bandlimited communication channels because it retains a constant envelope IF waveform and a fast spectral roll-off rate while allowing optimum detection performance. This technique is a particular case of Frequency Shift Keying (FSK), where the modulation index h, defined as the ratio of the reference frequencies spacing and the bit rate, is equal to 0.5. The analysis of the FFSK system is the subject of this report and it is shown that such system can be viewed as a quadrature modulation technique where a certain pulse shaping is used on its modulating waveforms such that desirable spectral properties are present on the IF signal. These properties make it very attractive in satellite communication systems where non-linear channels are used. The report is enriched with comparative test results of several simulated modulation techniques (including FFSK).

These results [1] , [2] verify the theoretical predictions that FFSK has a bit error performance closely equivalent to those of Quadrature Phase Shift Keying (QPSK), Offset QPSK (O-QPSK) and

Binary Phase Shift Keying (BPSK) techniques in linear channels. Tests of a particular FFSK modem showed that its spectrum is almost insensitive to non-linearities, a fact that makes it ideal for satellite communication links as far as performance and bandwidth efficiency is concerned.

ACKNOWLEDGEMENTS

The author wishes to express his gratitude and appreciation to Dr. Tiu Le Van for supervising this study, and for his criticism, advice and valuable guidance and suggestions during the preparation of this work.

Appreciations are also extended to Canadian Pacific Telecommunications, Montreal, for the support and facilities given to the author in preparing this report.

The author gratefully acknowledges inspiration and understanding accorded to him during his study by his wife, Mrs. Gianna Moutafis.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

α	: Bit Energies Ratio
ASK	: Amplitude Shift Keying.
B	: Baseband (or Message Signal) Bandwidth
2B	: IF or RF Signal Bandwidth
BT	: Normalized Baseband Bandwidth
2BT	: Normalized IF or RF Bandwidth
BER	: Bit Error Rate
BTR	: Bit Timing Recovery
BPSK	: Binary Phase Shift Keying
CTR	: Carrier Timing Recovery
C(t)	: In-Phase Component of the FFSK Signal
$\Delta\varphi$(coh)	: Phase Difference between the reference FFSK Signals ($S_1(t)$, $S_2(t)$ at Bit Transitions
E	: Bit Energy
f	: Frequency
fb	: Bit Rate
fo	: Carrier Frequency
f_1	: Low FSK Frequency
f_2	: High FSK Frequency
FSK	: Frequency Shift Keying
FFSK	: Fast Frequency Shift Keying
G(f)	: Power Spectral Density
$G_{FFSK}(f)$	: Power Spectral Density of FFSK

$G_{O\text{-}QPSK}(f)$	: Power Spectral Density of O-QPSK'
h	: Modulation Index
I(t)	: Symbol Stream modulating the In-Phase FFSK Component C(t)
IF	: Intermediate Frequency
NRZ	: Non Return to Zero (Data)
O-QPSK	: Offset Quadrature Phase Shift Keying
PSK	: Phase Shift Keying
Q(t)	: Symbol Stream Modulating the Quadrature FFSK Component S(t)
QPSK	: Quadrature Phase Shift Keying
RZ	: Return to Zero (Data)
RF	: Radio Frequency
$R(\tau)$	: Autocorrelation Function
S(t)	: Quadrature Component of the FFSK Signal
S(t)	: Reference signal ($S_2(t)-S_1(t)$)
$S_{dif}(t)$	: The Difference Signal equal to $\cos w_2 t - \cos w_1 t$
Ssim(t)	: The Sum Signal equal to $\cos w_2 t + \cos w_1 t$
T	: Bit Period
t	: Time
φ_k	: Bit Reference Angle of the FFSK Signal
w	: Randian Frequency
w_b	: Radian Bit Frequency
w_o	: Radian Carrier Frequency
X(f)	: Fourrier Transform of $x(t)$

$y_{FFSK}(t)$: The FFSK Signal

$y_{O\text{-}QPSK}(t)$: The O-QPSK Signal

$y_{QPSK}(t)$: The QPSK Signal

$Y_{FFSK}(f)$: Fourrier Transform of $y_{FFSK}(t)$

$Y_{O\text{-}QPSK}(f)$: Fourier Transform of $y_{O\text{-}QPSK}(t)$

$Y_{QPSK}(f)$: Fourier Transform of $y_{QPSK}(t)$

CHAPTER I

INTRODUCTION

Due to the overwhelming development of Data Communications during the recent years, a great deal of effort was necessary, to be put on the improvement of modulation methods so as to overcome problems encountered in crowded communication links.

Although digital communication has been used in various rundimentary forms for centuries, the science of data communications is relatively new and developing rapidly. This development has been spurred by two related facts. The amount of data to be transmitted is increasing almost exponentially and, because of the complex way in which the data are used, higher transmission accuracies are necessary. To meet the above requirements, it is necessary to employ transmission techniques that can handle as high a data rate and an accuracy, as possible. Consequently, the emphasis in the discipline of data communications is on the development of efficient and accurate transmission techniques.

This latter statement implies the necessity of transmission systems which can convey a given data rate along a certain (long) distance, requiring as less bandwidth as possible while achieving a bit error rate as small as possible for a given transmitted power (which is limited in practice).

One factor that rises the error probability in a data communication link, is the noise added to the channel.

The modulation technique, employed in a system, is what determines its bit error rate and bandwidth requirement when the source encoding and the filtering fashion are decided.

In this report it is assumed that the source encoder generates an NRZ* data stream which is to be conveyed through the link and the optimum filtering (using the Nyquist minimum bandwidth) or its approximation, is employed.

The bit error rate performance of the optimum receiver for the infinite bandwidth case and in a White Gaussian Noise environment is shown in [3] to be:

$$Pe = \frac{1}{2}\operatorname{erfc}\left(\sqrt{\frac{E1+E2-2\cdot E1E2\,\rho}{4N_o}}\right) \quad \ldots\ldots\ldots(1.1)$$

where E1 and E2 are the bit energies of the space and mark transmitted signals, $S_1(t)$ and $S_2(t)$, N_o is the power spectral density of the noise (assumed to be constant for all frequencies), and ρ is the correlation co-efficient.

$$E1 = \int_0^T S_1^2(t)\,dt$$

$$E2 = \int_0^T S_2^2(t)\,dt \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1.2)$$

---

*NRZ stands for Non Return to Zero data. A sequence of one's and zero's is represented by a voltage signal that assumes a constant value, say A, as long as a one is sent, and a value -A as long as a zero is sent.

$$\rho = \frac{1}{\sqrt{E1E2}} \int_0^T S_1(t)S_2(t)\,dt \qquad \text{.....(1.3)}$$

By defining

$$E = \frac{E1 + E2}{2} \qquad \text{.....(1.4)}$$

and

$$\alpha^2 = \frac{E2}{E1} \qquad \text{.....(1.5)}$$

as the average bit energy and the bit energies ratio respectively, equation (1.1) can be written as:

$$Pe = \frac{1}{2}\operatorname{erfc}\left(\sqrt{\frac{E}{2N_o}\left(1 - \frac{2}{1+\alpha^2}\rho\right)}\right) \qquad \text{.....(1.6)}$$

From the above equation the optimum value of α (the one that results in the minimum error error rate) is unity ($\alpha = 1$) which implies that the bit energies E1 and E2 be equal (from (1.5):

$$\alpha = 1$$

or $$E1 = E2 \qquad \text{.....(1.7)}$$

The optimum value of the receiver threshold can be shown to be equal to the bit energies difference divided by two ($K_o = (E2-E1)/2$), indicating that, E1 and E2 should be kept constant so as a threshold point stability be achieved (a necessity for reliable decisions). Since the signals $S_1(t)$ and $S_2(t)$ are sinusoids in practical systems and the energies E1 and E2 are related to them through (1.2), a phase-locked

loop (PLL) between each of $S_1(t)$ and $S_2(t)$ oscillators to the data clock (of rate $fb=1/T$) is necessary to produce invariant bit energies (E1, E2) since the integrations in (1.2) are over a bit period. Phase-locked loop circuits require that the frequencies of $S_1(t)$ and $S_2(t)$ are integer multiples of $fb/2N$*. Hence the threshold stability is achieved when:

$$f_i = n\frac{fb}{2N} \quad \text{, where n and N are integers....(1.8)}$$

and the optimum threshold valve, for (1.7) becomes zero:

$$Ko = 0 \quad \text{.....(1.9)}$$

There exist several digital modulation techniques in use today which are identified by their signal parameters α and ρ in equ. (1.6). The frequency constraint of (1.8) is maintained in order that the threshold be stable. Investigation of (1.2), (1.3) and (1.8) results in only three possible values of ρ : 0, 1, -1. Any α such that $0<\alpha<1$ would imply $0<\rho<1$ or $-1<\rho<0$, and the values of α are restricted to 0 or 1. The three basic among the existing digital modulation techniques are:

.....Amplitude Shift Keying (ASK) with $\alpha=0$, $\rho=0$

.....Frequency Shift Keying (FSK) with $\alpha=1$, $\rho=0$

.....Binary Phase Shift Keying (BPSK) with $\alpha=1$, $\rho=-1$

---

*A PLL can generate a frequency that is an integer multiple of the input clock. However, it is always possible that a clock be frequency divided by an odd integer M=2N before it is fed into the PLL. Under such conditions (1.8) holds.

where the choice $\rho = 1$ has been excluded as useless.

The resulting bit error rates for these cases are (from (1.6)).

$$Pe(ASK) = \frac{1}{2}\text{erfc}\left(\sqrt{\frac{E}{2N_o}}\right) = \frac{1}{2}\text{erfc}\left(\sqrt{\frac{A^2T}{8N_o}}\right) \quad \ldots\ldots\ldots(1.10)$$

$$Pe(FSK) = \frac{1}{2}\text{erfc}\left(\sqrt{\frac{E}{2N_o}}\right) = \frac{1}{2}\text{erfc}\left(\sqrt{\frac{A^2T}{4N_o}}\right) \quad \ldots\ldots\ldots(1.11)$$

$$Pe(BPSK) = \frac{1}{2}\text{erfc}\left(\sqrt{\frac{E}{N_o}}\right) = \frac{1}{2}\text{erfc}\left(\sqrt{\frac{A^2T}{N_o}}\right) \quad \ldots\ldots\ldots(1.12)$$

where A is the peak value of the sinnsoids $S_1(t)$ and $S_2(t)$ (note that $S_2(t) = 0$ in ASK, and A is the peack value of $S_1(t)$).

The ASK technique has the poorest performance because α has not been optimized. In BPSK and FSK α is optimized and they differ only in the value of ρ . Their bit error rates can be expressed as a function of ρ once α has been set to unity:

$$Pe(\alpha=1) = \frac{1}{2}\text{erfc}\left(\sqrt{\frac{E}{2N_o}(1-\rho)}\right) \quad \ldots\ldots\ldots\ldots\ldots(1.13)$$

This expression will be used extensively during the course of this report in conncection with equal bit energy signals. It shows that $\rho = -1$ is the optimum choice for such signals while $\rho = o$ is suboptimum.

From the above discussion it becomes evident that BPSK ($\rho = -1$) is the optimum digital modulation technique. Indeed this conclusion

agrees with actual systems performance results. For this reason BPSK is considered the most popular technique and has been used extremely in its Quadrature version (QPSK), a combination of two parallel BPSK branches that results in a halfed RF spectrum signal* compared to the basic BPSK (2B=fb compared to 2B=2fb).

However, the use of either BPSK or QPSK (and with these are ASK) in satellite communication links was found to lead in serious problems related to both, bandwidth conservation and error performance. In the satellite transponder, a saturated power amplifier, usually a Travelling Wave Tube Amplifier (TWTA), is used due to the limited available supply power (from solar cell). The output of this non-linearly operated amplifier is a distorted version of its input (the peaks of the signal are flattened), the distortion being two-fold: an AM to PM (AM/PM) conversion effect and an AM/AM conversion effect are present in the output waveforms [8]. Obviously, both the above effects are proportional to the amplitude variations (AM) of the input signal. The AM/PM effect results in excessive phase shift at bit transitions because envelope variations occur at these instants when the modulated signal is bandlimited. Consequently, phase detection is performed erroneously at the receiver. The AM/AM effect results in the re-

*When reference to the width of a spectrum is made, the main lobe width is meant throughout this report.

generation of the sidelobes of the modulated signal, the net result being that the filtering is undone [22].

The maximum envelop variation in these waveforms is 100% (AM 100%) and it obviously results in a large error performance impairement as well as high degree of side-lobes regeneration when used in a satellite link. The spectrum regeneration is a bandwidth inefficiency.

Both the above problems are partially removed when a modified QPSK technique is used, namely offset QPSK (O-QPSK). In this version of QPSK the maximum envelop variation is about 30%, thus both AM/PM and AM/AM conversion effects are kept to acceptable limits.

The envelop of a bandlimited FSK signal, when it is phase continuous, is nearly constant and the effects of non-linear TWTA devices (AM-PM and AM-AM) are very small [23]. Hence, FSK techniques are ideal for use in satellite links except for the poorer error performance compared to that of BPSK (equ's. (0.11), (0.12)). Also, the bandwidth efficiency of FSK is poor compared to QPSK because of the spectral spacing between the two transmitted tones (mark and space).

When this spacing is equal to half the bit rate, the FSK signal consists of two quadrature components both of which convey the information and have a correlation coefficient equal to -1 [4]. Hence, optimum detection can be achieved, resulting in an error

performance equal to that of BPSK, provided that the two components can be separately demodulated, which is feasible because of their quadrature relation. The spectrum of such a signal is slightly wider than that of QPSK. (The first nulls occur at $f_o \pm 0.75fb$ while in QPSK at $f_o \pm 0.5fb$) but its roll-off rate for $f > fo + 0.75fb$ and $f < fo-0.75fb$ is much higher (proportional to f^{-4} rather than f^{-2} for large $|f-fo|$). Consequently the disadvantage of wider bandwidth can be reduced by proper filter design. A filter with a 3dB cut-off frequency located somewhere between $fo \pm 0.5fb$ (first nulls of QPSK) and $fo \pm 0.75fb$ (first nulls of this FSK) may be designed, through which the signal from the modulator of either system may be processed with roughly equal amount of out-of-band energy. Then, both the Adjacent Channel Interference (ASI) and the amount of information loss will be roughly equal in both cases.

This FSK system is referred to as Fast FSK (FFSK) because it occupies the minimum bandwidth that an FSK system can, thus it permits the highest data rate for a given available IF, bandwidth. It is sometimes called Minimum Shift Keying (MSK) due to the minimum spacing used.

The subject of this report is the complete presentation and analysis of the FFSK technique. It is organized around the principles of generalized FSK, the study of the FFSK waveform, the modulation and demodulation methods for FFSK, and finally

the analysis and discussion of test results obtained from other sources.

Chapter 2 gives a description of generalized FSK along with signal parameters optimization from which the FFSK is extracted as the optimum FSK signal. Chapter 3 deals with the FFSK waveform presentation and analysis, and the modulation-demodulation processes are presented in Chapter 4. Chapter 5 presents test results of a 60 Mb/s modem [5] in both linear and non-linear channel conditions as well as test results of several modulation techniques (including FFSK) for comparison purposes. These tests [1], [2] are useful for a comparative evaluation of FFSK with respect to other techniques.

CHAPTER 2

PRINCIPLES OF FREQUENCY SHIFT KEYING (FSK)

An introduction to the generalized FSK technique is presented in this chapter. The parameters of the FSK signal are monitored and related to those of the information signal. Optimum relations in connection to the bit error rate performance of the system are derived and the FFSK system is isolated.

2.1 Frequency Shift Keying (FSK) - Modulation Index (h)

Frequency Shift Keying is a digital modulation technique in which the information is contained in the frequency of the IF signal.

In this technique a well defined frequency f_1, f_2, f_3...or f_M is assigned to each of the possible levels M_1, M_2, M_3...M_M of the message signal (baseband). In this report only Binary systems are considered which implies only two level message signals. Consequently the following correspondence applies:

Transmission of

$$S_1(t) = A\cos(2\pi f_1 t + \phi_1) \text{ when message} = M_1, kT \leq t \leq (k+1)T \quad (2.1.1a)$$

or transmission of

$$S_2(t) = A\cos(2\pi f_2 t + \phi_2) \text{ when message} = M_2, kT \leq t \leq (k+1)T \quad (2.1.1b)$$

Let $f_1 = f_o - \frac{h}{2} f_b$ .....(2.1.2a)

$f_2 = f_o + \frac{h}{2} f_b$ .....(2.1.2b)

where,

$f_o = \frac{f_1+f_2}{2}$ the carrier frequency ...(2.1.3)

$f_b = \frac{1}{T}$ the bit rate .....(2.1.4)

T is the bit period

$$h=\frac{f_2-f_1}{f_b} \quad \text{the Modulation Index} \qquad (2.1.5)$$

$$h>o, \text{ i.e. } f_2>f_1 \qquad (2.1.6)$$

The reference signals can be written as:

$$S_1(t)=\cos\left\{2\pi(f_o-\frac{h}{2}f_b)t+\phi_1\right\} \qquad (2.1.7)$$

$$S_2(t)=\cos\left\{2\pi(f_o+\frac{h}{2}f_b)t+\phi_2\right\} \qquad (2.1.8)$$

For an FSK signal to be phase continuous, it must be phase coherent. Hence, $S_1(t)$ and $S_2(t)$ should be equal at bit transitions:

$$S_1(kT)=S_2(kT)$$

and

$$2\pi(f_o-\frac{h}{2}f_b)kT+\phi_1=2\pi(f_o+\frac{h}{2}f_b)kT+\phi_2$$

from which the phase difference $\Delta\phi(\text{coh})$ between $S_1(t)$ and $S_2(t)$ at $t=kT$, should be:

$$\Delta\phi(\text{coh})=\phi_2-\phi_1=2h\pi k \qquad (2.1.9)$$

for a phase coherent FSK signal.

2.2 Phase Characteristics of FSK Signals

A generalized FSK waveform can be represented by

$$y_{FSK}(t)=\cos\left\{w_o t+\phi_k+\frac{h\pi t}{T}d_k\right\}, \quad kT\leq t\leq(k+1)T \qquad (2.2.1)$$

where: w_o : radian frequency of an implied carrier, located midway between the mark and space radian frequencies $W_1 = W_o - h\pi/T$ and $W_2 = W_o + h\pi/T$

h : the modulation index

dk : the value (± 1) of the kth bit

φk : an angle whose value depends on conditions of previous intervals and it is constant in an interval $kT \leq t \leq (k+1)T$

T : bit period

Clearly the phase of the FSK signal, $\varphi_k + (h\pi t/T)dk$ is linear in $(kT \leq t \leq (k+1)T)$. The value of φ_k is determined by the imposed conditions at the bit transitions ($t = kT$). To achieve phase continuity at bit transitions (which is a highly desirable feature in non linear channel applications), the following conditon must hold:

$$\left(\varphi_k + \frac{h\pi t}{T} d_k\right)\Big|_{t=kT} = \left(\varphi_{k-1} + \frac{h\pi t}{T} d_{k-1}\right)\Big|_{t=kT} \qquad (2.2.2)$$

from which

$$\varphi_k = \varphi_{k-1} + (d_{k-1} - d_k)h\pi k \qquad (2.2.3)$$

and

$$\varphi_k - \varphi_{k-1} = (d_{k-1} - d_k)h\pi k = \pm 2h\pi k = \Delta\varphi(coh) \qquad (2.2.4)$$

since $d_k = \pm 1$. In fact φ_k is the phase of the currently transmitted reference S_1 or S_2 and (2.2.4) is equivalent to (2.1.9).

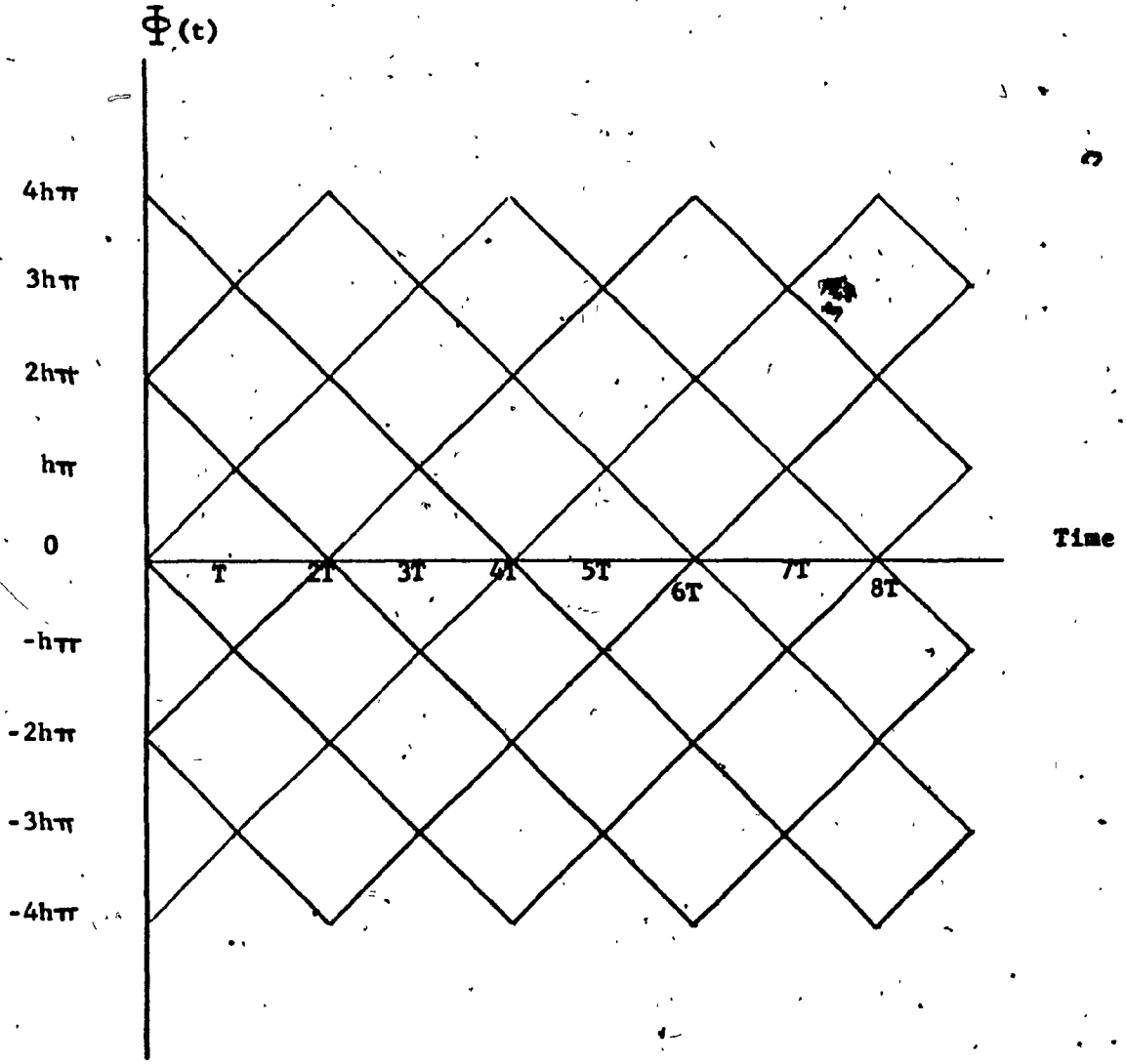


Fig. 2.2.1: Phase of the FSK signal versus time.

For coherent detection, a reference value of φ_k, say φ_o, can be set to zero without loss of generality. This assumption will be used throughout this report, with the result, using (2.2.3), that:

$$\varphi_k = 0 \text{ or } 2h\pi \qquad (2.2.5)$$

The phase of the FSK signal due to the message bits only is from (2.2.1):

$$\Phi(t) = \varphi_k + \frac{h\pi t}{T} dk, \quad kT \leq t \leq (k+1)T \qquad (2.2.6)$$

When (2.2.3) is in effect (phase continuity exists), (2.2.6) may be plotted as in fig. 2.2.1 where the continuous characteristics of $\Phi(t)$ at bit transitions and its linearity during bit periods are evident. From this figure it becomes clear that $\Phi(t)$ simply varies its slope at bit transitions rather than its value. Under such conditions the FSK signal is phase continuous.

From Fig. 2.2.1, the phase change between two successive bits is $h\pi$.

2.3 Optimization Of f1 And f2 With Repsect To fb

One of the requirements for optimum reception [3] is that the bit energies E1 and E2 be equal. Since

$$E1 = \int_o^T S_1^2(t)dt, \qquad E2 = \int_o^T S_2^2(t)dt$$

in the case of FSK, where $S_1(t)$ and $S_2(t)$ are given by (2.1.1), these energies are:

$$E_1 = \int_0^T A^2 \cos^2\left\{2\pi(f_o - h\frac{f_b}{2})t\right\} dt$$

$$= \frac{A^2T}{2}\left\{1 - \frac{\sin 4\pi(fo - hfb/2)T}{4\pi(fo - hfb/2)}\right\} \qquad .....(2.3.1)$$

$$E_2 = \int_0^T A^2 \cos^2\left\{2\pi(f_o + h\frac{f_b}{2})t\right\} dt$$

$$= \frac{A^2T}{2}\left\{1 - \frac{\sin 4\pi(fo + hfb/2)T}{4\pi(fo + hfb/2)}\right\} \qquad .....(2.3.2)$$

Evidently E1 and E2 become equal when the seconet terms in the brackets in (2.3.1) and (2.3.2) are equal. Thus:

$$E1 = E2$$

when

$$4\pi(f_o - \frac{h}{2}f_b)T = \eta\pi, \quad \eta = 0,1,2,.....(2.3.3)$$

$$4\pi(f_o + \frac{h}{2}f_b)T = \lambda\pi, \quad \lambda = 0,1,2,.....(2.3.4)$$

because then:

$$\sin 4\pi(f_o - \frac{h}{2}f_b)T = \sin 4\pi(f_o + \frac{h}{2}f_b)T = 0.....(2.3.5)$$

and E1, E2 become:

$$E1 = E2 = \frac{A^2T}{2} \qquad .....(2.3.6)$$

From (2.3.3) and (2.3.4) the mark and space frequencies f_1, f_2 can be written as:

$$f_1 = f_o - \frac{h}{2}f_b = \frac{\eta\pi}{4\pi T} = \eta\frac{f_b}{4} \qquad .....(2.3.7)$$

$$f_2 = f_o + \frac{h}{2}f_b = \frac{\lambda\pi}{4\pi T} = \lambda\frac{f_b}{4} \qquad .....(2.3.8)$$

The spacing between f_1 and f_2 can thus be given as:

$$f_2 - f_1 = (\lambda - \eta)\frac{f_b}{4} \qquad \text{.....(2.3.9)}$$

Substituting $f_2 - f_1$ from (2.3.9) into (2.1.5), h can be written as:

$$h = \frac{f_2 - f_1}{f_b} = \frac{\lambda - \eta}{4} \qquad \text{.....(2.3.10)}$$

Thus, h should be an integer multiple ofl/4. Finally, the ratio fo/fb can be found by combining (2.1.3) with (2.3.7) and (2.3.8):

$$\frac{f_o}{f_b} = \frac{f_1 + f_2}{2f_b} \quad \frac{(\eta f_b/4) + (\lambda f_b/4)}{2f_b} \quad \frac{\eta+\lambda}{8} \qquad \text{.....(2.3.11)}$$

Expressions (2.3.7), (2.3.8) and (2.3.11) show that both the mark and space frequencies f1 and f2 along with the mean frequency fo (which will be viewed as the carrier frequency in this report) must be in integer multiples of one fourth and one eighth of the bit rate respectively in order that the bit energies E_1 and E_2 be equal (a requirement for optimum performance). As a consequence of the expressions (2.3.7), (2.3.8), the spacing between f1 and f2 must be also an integer multiple of one fourth the bit rate as shown in (2.3.9).

For reference convenience the results of this section are summarized in Table 2.3.1.

From figure 2.2.1, the phase change of the FSK signal between two successive bit transitions is:

$$\Delta\Phi = h\pi \qquad \text{.....(2.3.12)}$$

$$f_1 = \eta \frac{f_b}{4} \qquad \ldots.(2.3.7)$$

$$f_2 = \lambda \frac{f_b}{4} \qquad \ldots.(2.3.8)$$

$$f_2 - f_1 = (\lambda - \eta) \frac{f_b}{4} \qquad \ldots.(2.3.9)$$

$$f_o = (\eta + \lambda) \frac{f_b}{8} \qquad \ldots.(2.3.11)$$

$$h = (\lambda - \eta)(M) \qquad \ldots.(2.3.10)$$

where: $f_1 = f_o - \frac{h}{2} f_b$ ....(2.1.2a)

$$f_2 = f_o + \frac{h}{2} f_b \qquad \ldots.(2.1.2b)$$

$$h = \frac{f_2 - f_1}{f_b} \qquad \ldots.(2.1.5)$$

λ and η are integers

TABLE 2.3.1
Requirements for Optimum Condition:
$E_1 = E_2$

From (2.3.10) and (2.3.12):

$$\Delta\Phi = (\lambda-\eta)\frac{\pi}{4}K \qquad (2.3.13)$$

which shows that $\Delta\Phi$ will be an integer multiple of $\pi/4$ when the bit energies are equal.

The phase difference between $S_1(t)$ and $S_2(t)$ at bit transitions (for phase coherence) similarity becomes (using (2.1.9) and (2.3.10):

$$\Delta\varphi(coh) = K(\lambda-\eta)\frac{\pi}{2} \qquad (2.3.14)$$

thus an integer multiple of $\frac{\pi}{2}$, and the reference signals $S_i(t)$ must be chosen in such a way that they satisfy (2.3.14) when a continuous FSK signal is desired.

2.4 Optimization Of The Modulation Index (h)

Once the optimum condition of equal bit energies is satisfied, the correlation co-efficient, ρ, must now be optimized for the FSK system.

It was mentioned in Chapter 1 that the error probability of an optimum receiver (Correlator Receiver) for equal bit energy signals is given by (1.13) as:

$$Pe = \frac{1}{2}\text{erfc}\left(\sqrt{\frac{E}{2N_o}(1-\rho)}\right)$$

From this expression it is evident that PE is lower when ρ is lower. Therefore, to optimize a modulation system the value of ρ should be minimized.

The co-efficient ρ is defined in (1.1) as:

$$\rho = \frac{1}{\sqrt{E1E2}} \int_0^T S_1(t)S_2(t)dt, \quad -1 \leq \rho \leq 1 \quad .....(2.4.1)$$

where E1 and E2 have been already chosen optimally in the preceding section and are given by (2.3.6) as $E1 = E2 = A^2T/2$.

Equation (2.4.1) can be reduced, using (2.1.7), (2.1.8) and (2.3.6), to:

$$\rho = \frac{\sin(4\pi\, fo/fb)}{4\pi\, fo/fb} + \frac{\sin(2\pi h)}{2\pi h} \quad ....(2.4.2)$$

Letting:

$$\beta = \frac{\sin(4\pi\, fo/fb)}{4\pi\, fo/fb} \quad ....(2.4.3)$$

$$\gamma = \frac{\sin(2\pi h)}{2\pi h} \quad ....(2.4.4)$$

ρ can be written as:

$$\rho = \beta + \gamma \quad ....(2.4.5)$$

The terms β and γ are of the sinx/x form and are plotted in Fig. 2.4.1.

The values of β and γ are minimized for

$$\frac{fo}{fb} = 2h = \frac{3}{2} \quad ...(2.4.6)$$

and their minimum value is:

$$\beta = \gamma = -0.212 \quad ...(2.4.7)$$

Equation (2.4.6) implies (using table 2.3.1) that:

$$f_1 = 0$$

$$f_2 = 3\frac{fb}{4}$$

This means that $S_1(t)$ is a d.c. component and, as such, is not acceptable. Indeed, one of the reasons for employing modulation is to avoide the non-linear phase and group delay characteristics of the low frequency spectrum of transmission channels, especially at $f=0$.

If f_o is chosen much higher than the bit rate, or if it is chosen to be an integer multiple of $fb/4$ as in (2.4.8) below

$$f_o = n\,fb/4, \quad n=1,2,3 \qquad \ldots(2.4.8)$$

β becomes zero from (2.4.3) and ρ becomes exactly equal to γ. This choice of f_o satisfies (2.3.11), a requirement for equal bit energies and it is acceptable. Then (2.4.2) reduces to:

$$\rho = \frac{\sin 2\pi h}{2\pi h} \qquad \ldots(2.4.9)$$

whose plot is identical to that of γ in Fig. 2.4.1. The optimum value of h is then:

$$h = \frac{3}{4} = 0.75 \qquad \ldots(2.4.10)$$

since it minimizes ρ ($\rho=-0.212$). (Note that this value satisfies the requirement for equal bit energies given by (2.3.10).

The error probability, PE, (with White, Gransian Noise as the only impairement and for infinite bandwidth), which is given by (1.13) for the FSK case, becomes:

$$Pe(h=3/4) = \frac{1}{2}\,\mathrm{erfc}\left(\sqrt{0.606\frac{E}{No}}\right) \qquad \ldots\ldots(2.4.11)$$

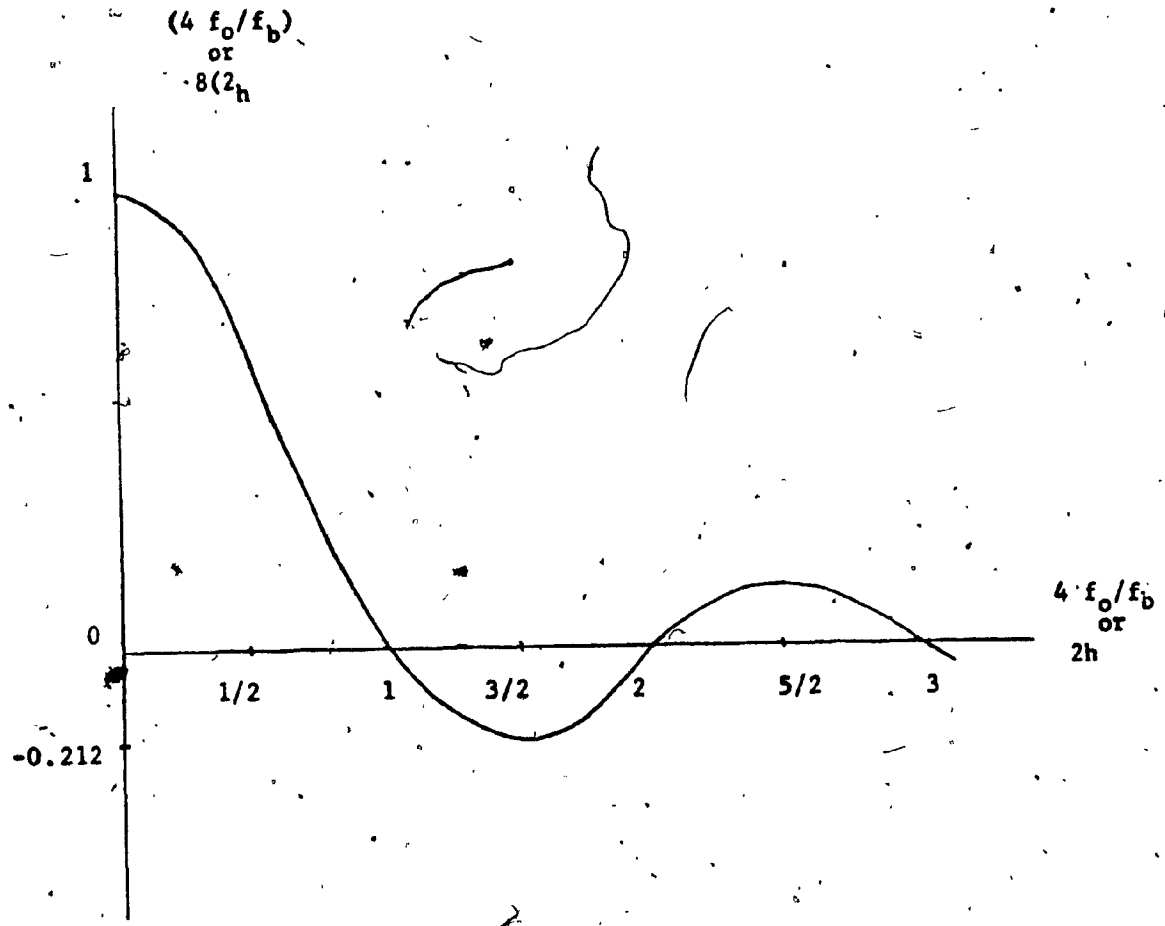


Fig. 2.4.1: Variation of β or γ with 4 f_o/f_b or 2 h

Although this is the optimum bit error rate performance for an FSK system* it is obvioulsy poorer than that of the Binary or Quadrature PSK (BPSK or QPSK) case, where:

$$PE(BPSK)=PE(QPSK)=\tfrac{1}{2}\text{erfc}\left(\sqrt{\frac{E}{N_o}}\right) \quad \ldots(2.4.12)$$

2.5 Spacing And Interference Considerations For The Choice of h

The choices $h=1$ and $h=\frac{1}{2}$ possess advantages in some important areas of signal characteristics. The choice $h=1$, which was analysed by Sunde [6] , provides the reference signals $S_1(t)$ and $S_2(t)$, necessary for coherent detection, contained in the FSK signal. These signals can be easily isolated from the rest of the received signal by two tuned circuits or phase-locked loops (PLL's) at f_1 and f_2 thus eliminating the need for Carrier Timing Recovery (CTR) circuits. It has been shown [6] that it is possible to eliminate Intersymbol Interference (ISI) when $h=1$ by using rectangular modulating pulses of duration T (i.e. NRZ data), a carrier frequency f_o much higher than f_2-f_1 and a bandpass channel with appropriate amplitude/frequency characteristics and with linear phase characteristics.

The choice $h=\frac{1}{2}$ provides with bandwidth conservation due to the shorter spacing and it will be shown in a later section that ISI is not present.

---

*It is implied that there is only one correlator receiver used.

Both these choices give a higher ρ ($=0$) resulting in a poorer theoretical bit error rate performance according to (1.13).

$$PE(h=\tfrac{1}{2},1)=\tfrac{1}{2}\operatorname{erfc}\left(\sqrt{\frac{E}{2No}}\right) \qquad \ldots(2.5.1)$$

The values $\frac{1}{2}$ and 1 for h satisfy table 2.3.1, an optimal requirement for equal E_1 and E_2 and (2.3.14) for phase continuity.

2.6 The h = 1 FSK (Sunde's FM) And The h = ½ FSK

The spectrum of the h = 1 continuous FSK can be evaluated [6] and is given by:

$$W(f)_{h=1}=\frac{1}{8}\delta(f-f_1)+\frac{1}{8}\delta(f-f_2)+\frac{2T}{\pi}\cdot\frac{\cos\left\{\pi(f-fo)/(f_2-f_1)\right\}}{1-\left\{2(f-fo)/(f_2-f_1)\right\}^2} \qquad \ldots(2.6.1)$$

for

$$S_1(t)=\cos\left\{2\pi\left(fo-\frac{fb}{2}\right)t\right\} \qquad \ldots(2.6.2)$$

$$S_2(t)=\cos\left\{2\pi\left(fo+\frac{fb}{2}\right)t\right\} \qquad \ldots(2.6.3)$$

W. R. Bennet and S. O. Rice [7] considered several choices of h and derived the power stectral densities of the associated FSK signals. One of those choices was h = ½ and the spectral density in this case was found for continuous signals, to be:

$$W(w)_{h=\frac{1}{2}}=\frac{\sin^2(wT)}{2T}\left\{\left(\frac{1}{w-w_1}-\frac{1}{w-w_2}\right)^2+\left(\frac{1}{w+w_1}-\frac{1}{w+w_2}\right)^2+2\left(\frac{1}{w-w_1}-\frac{1}{w-w_2}\right)\left(\frac{1}{w+w_1}-\frac{1}{w+w_2}\right)\cos 2\phi\right\} \qquad \ldots(2.6.4)$$

where ϕ is an initial angle.

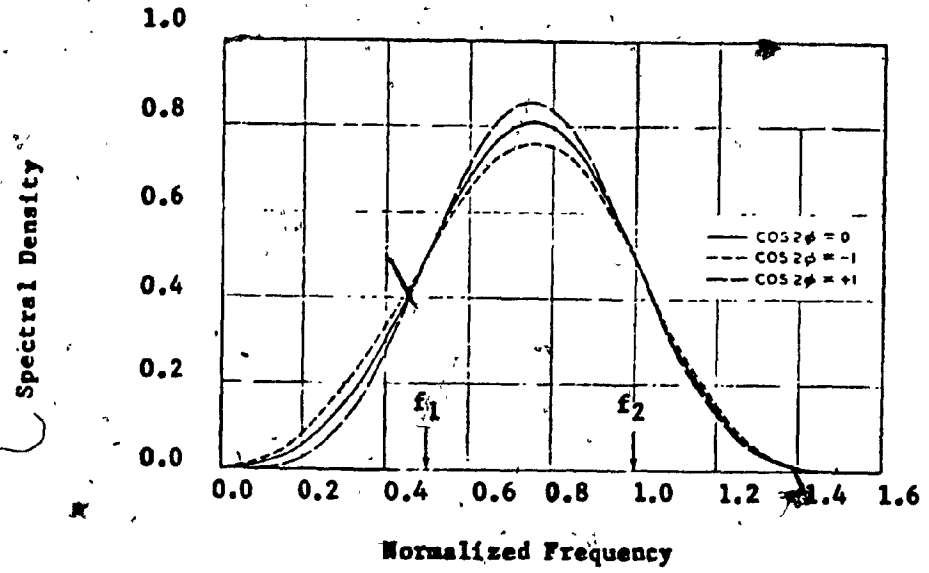
Bennet and Rice derived the spectral density for the case h=1 as:

$$W(f)=\frac{1}{8}\delta(f-f_1)+\frac{1}{8}\delta(f-f_2)$$
$$+\frac{1}{2T}\left\{\cos^2\left(\frac{w-(w_2-w_1)/2}{2}\cdot T\right)\cdot\left(\frac{1}{w-w_1}-\frac{1}{w-w_2}\right)^2\right.$$
$$\left.+\cos^2\left(\frac{w+(w_2-w_1)/2}{2}\,T\right)\cdot\left(\frac{1}{w+w_1}-\frac{1}{w+w_2}\right)^2\right\}$$

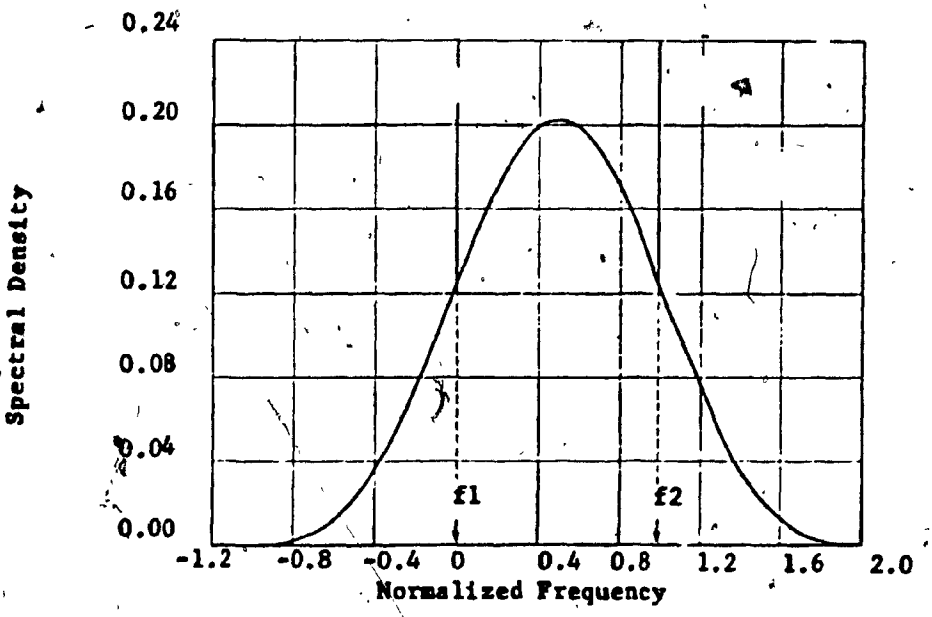
...(2.6.5)

Both, (2.6.4) and (2.6.5) are plotted in Fig. 2.6.1, where it becomes evident that in the case of h=1 there exist two impulses of infinite height at f_1 and f_2, while no impulses are present in the case of h=½. Two more cases are illustrated in Fig. 2.6.2 for h=0.8 and h=1.2, which are representative for $h<1$ and $h>1$. It can be seen from this figure that in both cases finite peaks occur outside and inside the region: $f_1<f<f_2$ for $h<1$ and $h>1$ respectively. Thus the case of h=1 is a threshold case. It is also to be noted that the height of the peaks increase with increasing h for $h<1$ reducing to zero at h=½ (from Fig. 2.6.2 and (2.6.4)) and reaching infinity at h=1, while decreases with increasing h for $h>1$.

The Sunde FM (h=1) offers the facility for easy carrier recovery at the receiver for coherent detection due to the presence of the discrete tones at f_1 and f_2 in its spectrum. However, half of the transmitted energy contributes to the maintenance of these

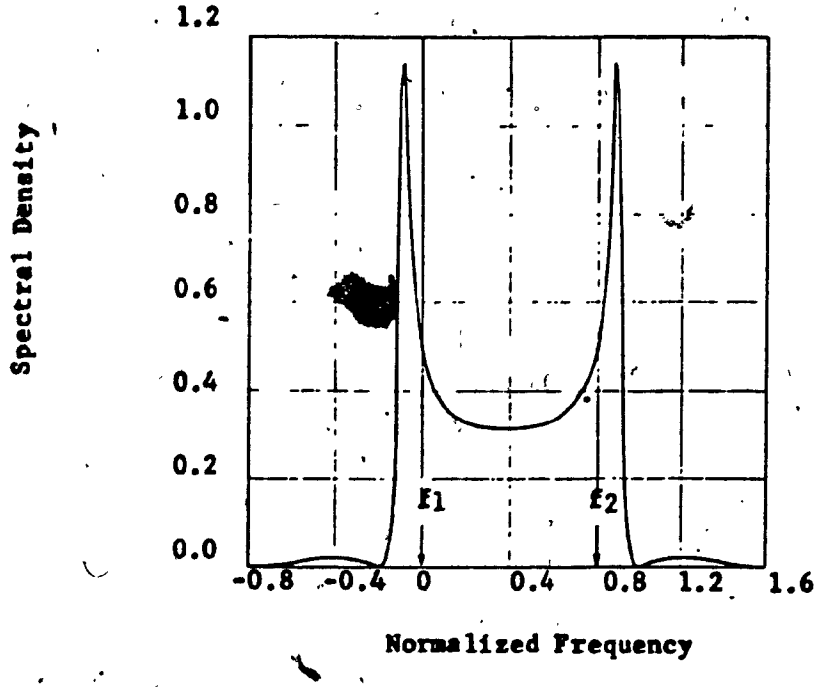


(a) Modulation Index: 0.5

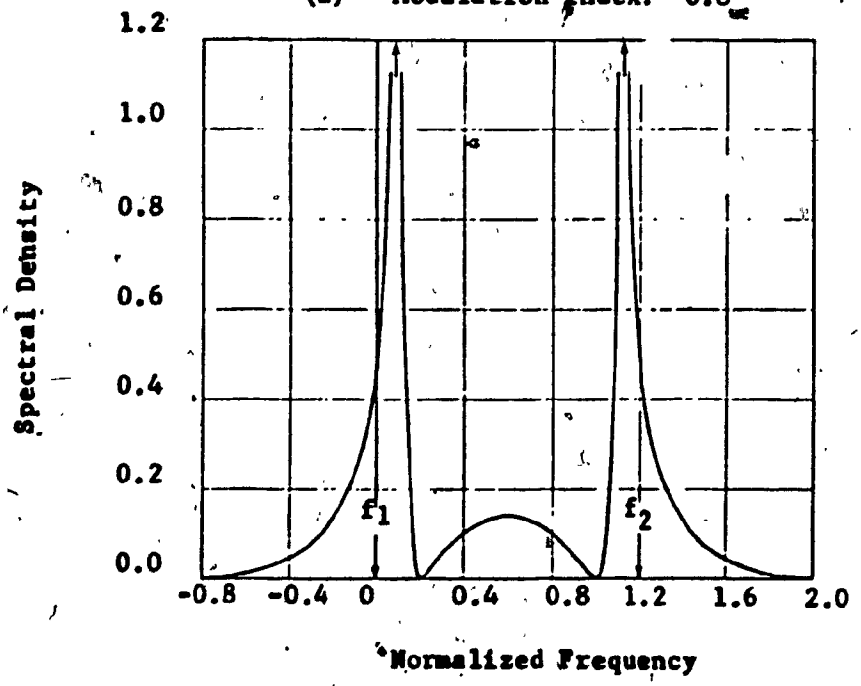


(b) Modulation Index: 1

Fig. 2.6.1: Spectral density of random binary FSK wave with continuous phase at transitions.



(a) Modulation Index: 0.8



(b) Modulation Index: 1.2

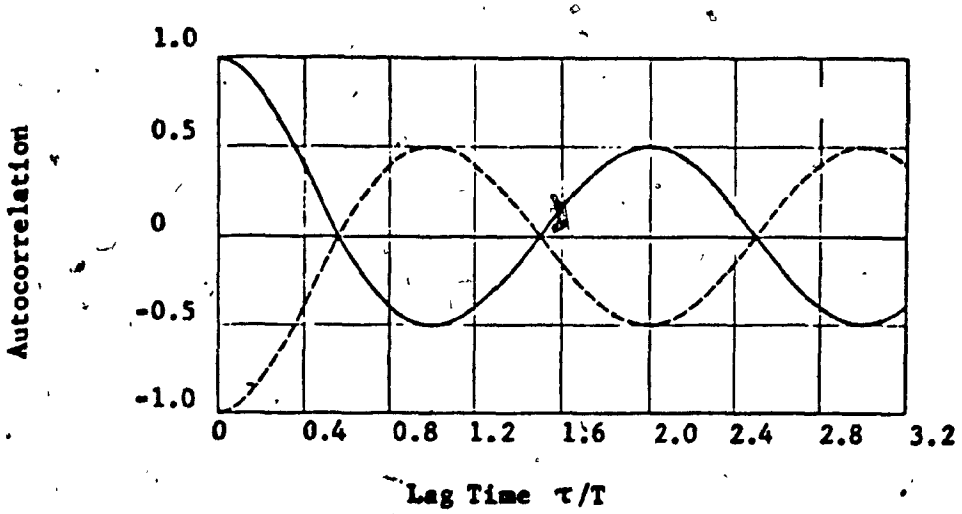
Fig. 2.6.2: Spectral density of random binary FSK with continuous phase at transitions

tones (spikes) without conveying any message. In addition to this the disadvantage of its wider spectrum compared to that of the case $h=\frac{1}{2}$ made it non satisfactory from a practical point of view. Indeed the required bandwidth when $h=1$ is 3fb, while when $h=\frac{1}{2}$ this bandwidth is 1.5fb as can be observed in Fig. 2.6.1.

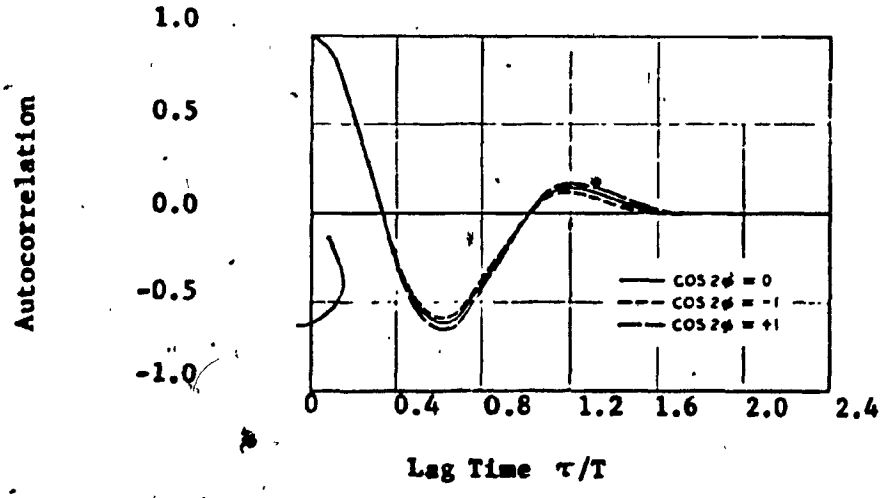
A consideration of the auto-correlation functions of the two continuous cases ($h=1$ and $h=\frac{1}{2}$) will be found to be helpful in making an optimum choice. These functions are plotted in Fig. 2.6.3 [7] . In this figure the envelope of the auto-correlation function for $h=1$ is shown (to avoid making the curve depend on the frequencies f_1 and f_2), while for $h=\frac{1}{2}$ the actual auto-correlation function is shown.

It is important to note that, due to the discrete tones at f_1 and f_2 in the case $h=1$, the envelope for this case maintains a constant value 0.5 for $|\tau|>T$ rather than decaying as it does for $|\tau|<T$. In the case $h=\frac{1}{2}$, however, the envelope of the auto-correlations function continuously decays rapidly and vanishes at $|\tau|=2T$ maintaining its value at zero for all $|\tau|>2T$.

This latter property of the case with $h=\frac{1}{2}$ permits, as will be shown in the next section, to modify the detection procedure in such a way that the optimum bit error rate performance be achieved as given by (2.4.12), thus identical to that of the QPSK technique.



(a) Envelope of Autocorrelation Function when h = 1.



(b) Actual Autocorrelation function when h = 0.5.

Fig. 2.6.3: Normalized Autocorrelation function when phase is continuous.

It is important to note here that, when $h=\frac{1}{2}$ the frequencies f_1 and f_2 relate to the bit rate fb, as follows from (2.1.2) through:

$$f_1=fo-\frac{fb}{4} \qquad ...(2.6.6)$$

$$f_2=fo+\frac{fb}{4} \qquad ...(2.6.7)$$

When $fo=nfb/4$ from (2.4.8), equations (2.6.6) and (2.6.7) agree with (2.3.7) and (2.3.8) for equal E_1 and E_2.

2.7 Examination Of The Autocorrelation Function Of FSK With $h=\frac{1}{2}$ - The Fast Frequency Shift Keying (FFSK) And Its Detection Method

The auto-correlation function of the FSK with $h=\frac{1}{2}$ was shown in the previous section (Fig. 2.6.3) to be zero for time lags exceeding 2T. Thus, if such a signal is transmitted through an infinite bandwidth, linear, additive White Gaussian Noise (WGN) channel, no interference will be present between the pulses representing the $2k_{th}$ bits of information, and the same will hold for the $(2K+1)_{th}$ bits. Consequently no ISI will be produced between odd or even bits (obviously this is not true for successive bits).

The above statement can be easily visualized by referring to Fig. 2.2.1, which shows the phase, $\Phi(t)$ variations of an FSK signal, and letting $h=\frac{1}{2}$. Indeed, it is evident from this figure that:

$$\Phi(t)=\begin{cases}0 \text{ or } \pi & \text{at } t=2kT \\ \pm\frac{\pi}{2} & \text{at } t=(2k+1)T \end{cases} \qquad ...(2.7.1)$$

thus, $\Phi(t)$ assumes two totally orthogonal (between each other) sets of values* at successive bit transitions (two orthogonal sets do not interfere with each other). In addition to this, each of these sets is antipodal since the two possible phase values (within a set) differ by π.

Thus, this signal may be viewed as two orthogonal binary channels, where antipodal symbols of tenth 2T seconds are used in each channel, and a half-symbol time skew of T seconds exists between the orthogonal channels.

For independent binary data bits, the signs of the symbols in each channel are independent from one 2T interval to the next, and the auto-correlation function would be expected to go to zero for lags exceeding 2T.

In general, it can be shown, by expanding (2.2.1), that any binary FSK signal is composed of two orthogonal (quadrature) components as was mentioned above:

$$y_{FSK}(t) = \cos(w_o t + \varphi_k + \frac{h\pi t}{T} d_k)$$

$$= \cos(\varphi_k + \frac{h\pi t}{T} d_k)\cos w_o t - \sin(\varphi_k + \frac{h\pi t}{T} d_k)\sin w_o t$$

$$= \left[\cos\varphi_k \cos(\frac{h\pi t}{T} d_k) - \sin\varphi_k \sin(\frac{h\pi t}{T} d_k) \right] \cos w_o t$$

$$- \left[\sin\varphi_k \cos(\frac{h\pi t}{T} d_k) + \cos\varphi_k \sin(\frac{h\pi t}{T} d_k) \right] \sin w_o t$$

$$\text{for } kT \leq t \leq (k+1)T \quad \text{.....(2.7.2)}$$

*A phase difference of $\pi/2$ exists between the two sets.

where use was made of the trigonometric identities:

$$\sin(u+v) = \sin u \cos v + \cos u \sin v$$

$$\cos(u+v) = \cos u \cos v - \sin u \sin v$$

For $h=\frac{1}{2}$, the above generalized FSK expression becomes

$$y_{FSK}(t)_{[h=\frac{1}{2}]} = \cos\varphi_k \cos\left(\frac{\pi t}{2T}\right) \cos w_o t - d_k \cos\varphi_k \sin\left(\frac{\pi t}{2T}\right) \sin w_o t$$

$$\text{for } kT \leq t \leq (k+1)T \qquad \ldots(2.7.3)$$

Taking into account the fact that $k=0,\pi$ (i.e. $\cos\varphi k=\pm 1$) modulo 2π according to (2.2.5), evaluated for $h=\frac{1}{2}$. The information, d_k, is conveyed on both carrier components (as $\cos\varphi k$ and $dk\cos\varphi k$) on a fifty-fifty basis. This requires two correlator receivers for the defection of the data. Each receiver will detect the data $\cos\varphi_k$ or $d_k\cos\varphi_k$ optimally (since $\cos\varphi_k=\pm 1$ and $d_k=\pm 1$, thus $d_k\cos\varphi_k=\pm 1$).

Investigation of (2.7.3) shows that it agrees with the nature of the $h=\frac{1}{2}$FSK signal, given earlier in this section.

The quadrature relation between its two components permits to detect them separately, as in a QPSK system, using two correlator receivers. Because of the antipodal nature of each component, the optimum performance of the BPSK and QPSK system will then be achieved as given in (1.12) (since $\rho=-1$ for each component)

$$Pe = \frac{1}{2}\operatorname{erfc}\left(\sqrt{\frac{E}{No}}\right) \qquad \ldots(2.7.4)$$

if two groups of sampling times, $2kT$ and $(2k+1)T$, are used, one for each channel (i.e. one for each receiver). The initial data, dk, will

be recovered finally by interleaving the detected symbols from each channel.

Such an FSK system, with $h=\frac{1}{2}$, is referred to as the Fast Frequency Shift Keying (FFSK) [8] and is the subject of this report. It is sometimes called Minimum Shift Keying (MSK).

It would be of interest to evaluate (2.7.2) for $h=1$ in order to visualize in what aspects a FFSK system differs from a Sunde's FM, and why. Letting $h=1$ in (2.7.2) and using (2.2.5), then taking into account that $\phi k = 0$ or 2π modulo 2π according to (2.2.5) evaluated for $h=1$, we obtain:

$$y_{sunde}(t) = \cos(\frac{\pi t}{T})\ \cos w_o t - d_k \sin(\frac{\pi t}{T}) \sin w_o t$$

$$= \frac{1}{2}\cos(w_o + \frac{\pi}{T})t + \frac{1}{2}\cos(w_o - \frac{\pi}{T})t - d_k \sin(\frac{\pi t}{T}) \sin w_o t$$

$$\text{for } kT \leq t \leq (k+1)T \qquad \ldots(2.7.5)$$

This expression shows that one half of the bit energy is used to convey the information while the other half maintains the spectral tones at $f_1 = fo - fb/2$ and $f_2 = fo + fb/2$. The information is conveyed in the optimum fashion, (the correlation coefficient of the component that carries the information is -1 since antipodal data (d_k) is used to modulate it) but only half the bit energy is used. The error performance then must be expected to be outperformed by 3dB by an antipodal modulation scheme in which the whole bit energy is used (as in BPSK or QPSK). This 3dB offset corresponds to a correlation coefficient equal to zero and the total Sunde's FM signal has such a coefficient. Its error performance is evaluated

by (1.11) or by letting $\rho=0$ in (1.13). The fact that the auto-correlation function of Sunde's FM never goes to zero (it stays at $\frac{1}{2}$ for $|\tau| \geq T/2$), as shown in Fig. 2.6.3a, shows a close correspondence with the above. This function also shows that the minimum ISI (i.e. the maximum detection margin) occurs at $t=kT/2$ and this is the proper sampling time. This agrees with Fig. 2.2.1, from which it is evident that $\Phi(t)=\pm\pi/2$ at $t=kT/2$ (at $t=kT$, $\Phi(t)=0,2\pi$ i.e. no information present at these times), which implies optimum phase distance ($=\pi$) and the information component is modulated by antipodal data as mentioned earlier.

Turning back to the FFSK case, we can conclude that is is possible to detect a bit by observing it for $2T$ seconds before a decision is made. This can be done since all adjacent bits are conveyed on different components of the signal. A data encoding has been physically made as shown by (2.7.3). Indeed, there are two data streams, $\cos\phi_k$ and $d_k\cos\phi_k$, which modulate the two quadrature components and they relate to the initial data d_k through a certain encoding law which will be derived in Chapter 3.

Although the correlation coefficient of the composite FFSK signal is zero (Fig. 2.4.1 at $2h=1$ gives $\rho=0$), the individual correlation coefficients of its components are -1 and one must detect these components separately to obtain the optimum performance.

Summary

In this chapter a general treatment of FSK signals was done towards the distinction of one which can be considered as the optimum. The criterion of optimality was the correlation coefficient, which is the only deciding factor in the evaluation of the error probability, once the receiver optimization has been done and a correlator receiver is used.

It was shown that the correlation coefficient depends exclusively on the modulation index when the carrier frequency is an integer multiple of one fourth the bit rate. For bandwidth efficiency, small values of the modulation index appeared to be preferable. Although the value 3/4 for the index appeared as the optimum for FSK signals (as giving the minimum correlation coefficient), the choice of the value ½ was investigated through the resulting autocorrelation function (thus, considering ISI), and the expected bandwidth efficiency. From this investigation, it was shown that the property of the autocorrelation function to be zero for $|\tau| \geq 2T$ rises the capability of detecting the FSK signal with $h=\frac{1}{2}$ in a fashion similar to that used in QPSK systems, thus achieving the optimum performance of BPSK or QPSK systems. This FSK system (FFSK) is to be analysed in detail through the following chapters.

CHAPTER 3

FFSK WAVEFORM

The analysis and study of the FFSK signal is presented in this chapter. Conditions for phase coherence, thus for phase continuity, are determined and its orthogonal nature is derived along with the physical encoding of the initial data into the data that modulate its two orthogonal components. Then the spectrum and autocorrelation functions are derived and their influence on the signal characteristic properties is discussed on a comparative basis against the offset QPSK signal.

3.1 Formulation Of The FFSK Signal

The FFSK signal has $h=\frac{1}{2}$ and using a carrier frequency, fo, according to (2.4.8), $fo = nfb/4$, the required mark and space frequencies, from (2.1.2), are:

$$f_1 = n\frac{fb}{4} - \frac{fb}{4} \quad (n-1)\frac{fb}{4} \qquad \ldots(3.1.1a)$$

$$f_2 = n\frac{fb}{4} - \frac{fb}{4} \quad (n+1)\frac{fb}{4} \qquad \ldots(3.1.1b)$$

For phase coherence (i.e. phase continuity), equation (2.3.14) should hold:

$$\Delta\varphi(coh) = k(\lambda-\eta)\frac{\pi}{2} = k\pi \qquad \ldots(3.1.2)$$

Since both f_1 and f_2 are integer multiples of fb/4, from (3.1.1), two PLL's may be used to derive them from the data clock divided by 4, where care should be taken so that (3.1.2) be in effect.

Using two signals $S_1(t)$ and $S_2(t)$ of frequencies f_1 and f_2 respectively with f_1, f_2 satisfying (3.1.1), the spacing becomes:

$$f_2 - f_1 = \frac{fb}{2} \qquad \ldots(3.1.3)$$

thus equal to half the bit rate.

Figure 3.1.1 shows a possible signal choice, satisfying all

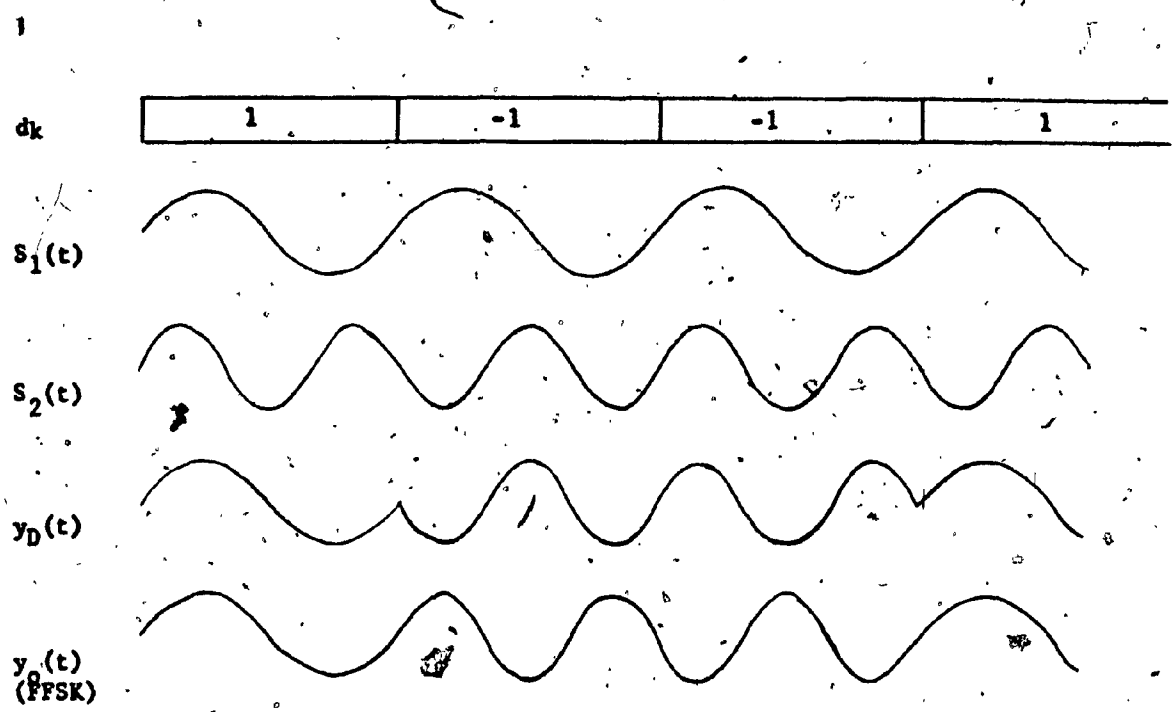


Fig.: 3.1.1: Waveforms for Fig. 3.1.2

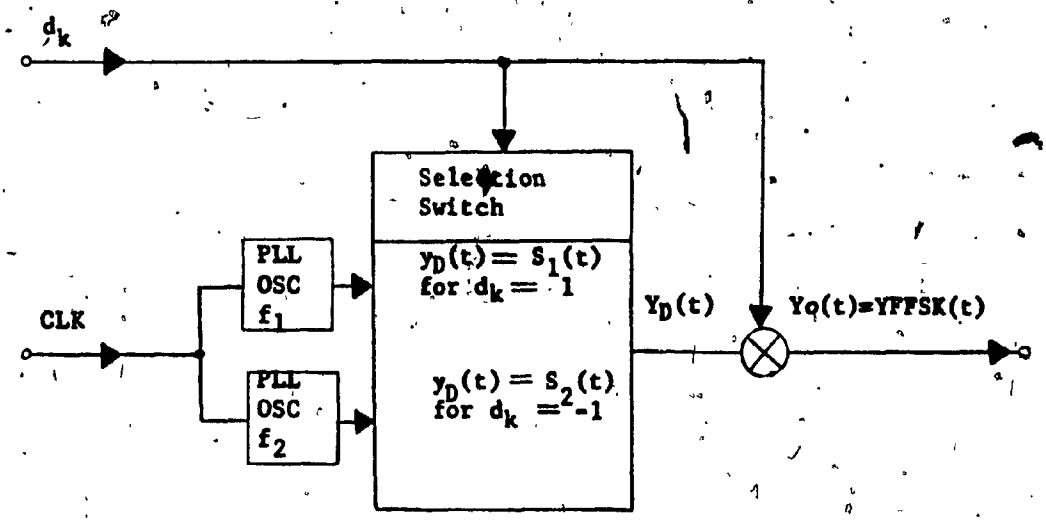


Fig. 3.1.2: Generation of FFSK by interswitching between two oscilatiors.

requirements for FFSK. Note that the resulting FFSK waveform after modulation by a bit stream differs by $\pi/2$ (as expected from (2.3.12)) from the center of one bit interval to the center of the next. In this figure, $y_D(t)$ is the signal resulting by using a switch which selects the output of the $S_1(t)$ oscillator when $d_k=+1$ and the output of the $S_2(t)$ oscillator when $d_k=-1$. Apparently, $y_D(t)$ is not a continuous phase waveform when f_1 and f_2 are fb/2 apart. For this reason a switched inverting system is required whose inverting or non-inverting function can be controlled by the bit stream d_k as shown in Fig. 3.1.2. The function of this system is to process the $y_D(t)$ signal as it is when $d_k=+1$ and to invert it when $d_k=-1$. Therefore, this inverting system can be implemented as multiplier (BPSK modulator).

Apparently, the phase difference between $S_1(t)$ and $S_2(t)$ at bit transitions, $\Delta\varphi(coh)$ is 0 or π. This agrees with (3.1.2).

3.2 The FFSK Signal

The FFSK signal was derived in section 2.7 (equ. 2.7.3) as:

$$y_{FFSK}(t)=\cos\varphi_k \cos\left(\frac{\pi t}{2T}\right) \cos w_o t - d_k \cos\varphi_k \sin\left(\frac{\pi t}{2T}\right) \sin w_o t$$

$$\text{for } kT \leq t \leq (k+1)T \qquad \text{.....(3.2.1)}$$

where φ_k is given by:

$$\varphi_k=\varphi_{k-1}+(d_{k-1}-d_k)\frac{\pi k}{2}=\varphi_{k-1}\pm\pi k \qquad \text{.....(3.2.2)}$$

in order to meet the phase continuity requirement (obtained from (2.2.3) for $h=\frac{1}{2}$ or from (3.1.2)).

The phase $\Phi(t)$ of the FFSK signal is plotted in Fig. 3.2.1, which is identical to Fig. 2.2.1, with h replaced by $\frac{1}{2}$ [4] .

The thick path in this figure represents $\Phi(t)$ for a given data sequence: +1, -1, +1, +1, -1, +1, -1, -1, +1.

Evidently $\Phi(t)$ is piece-wise linear during bit intervals and it differs by $\pi/2$ between times kT and (k+1)T. At bit transitions it changes only slope, thus being continuous. This is due to the phase continuity constraint in (3.2.2).

Only four phase values are permitted at bit transitions: 0 or π after an even bit interval, and $+\pi/2$ or $-\pi/2$ after an odd. This can be seen also from the associated state diagram [9] in Fig. 3.2.2, where it is assumed that an increase (by $\pi/2$) in phase results for $d_k=+1$ and a decrease (by $\pi/2$) for $d_k=-1$.

To simplify the notation for the FFSK signal expression, given in (3.2.1), let:

$$\left.\begin{aligned} &\cos\varphi_k = I(t), \quad \text{the I-channel data} \\ &d_k\cos\varphi_k = Q(t), \quad \text{the Q-channel data} \\ &\cos(\pi t/2T) = C(t), \quad \text{the I-channel Symbol Weighting Sinusoid} \\ &\sin(\pi t/2T) = S(t), \quad \text{the Q-channel Symbol Weighting Sinusoid} \end{aligned}\right\} \quad \ldots(3.2.3)$$

Then (2.2.1) reduces to:

$$y_{FFSK}(t) = I(t)C(t)\cos w_o t - Q(t)S(t)\sin w_o t \quad \ldots(3.2.4)$$

for $kT \leq t \leq (k+1)T$

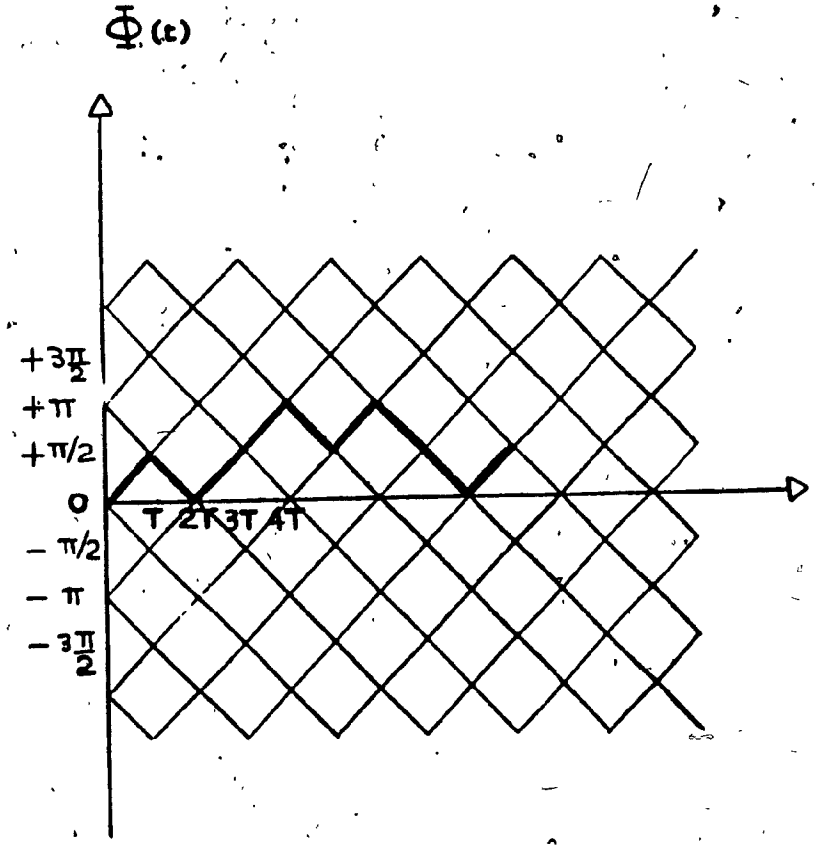


Fig. 3.2.1: The phase of FFSK versus time.

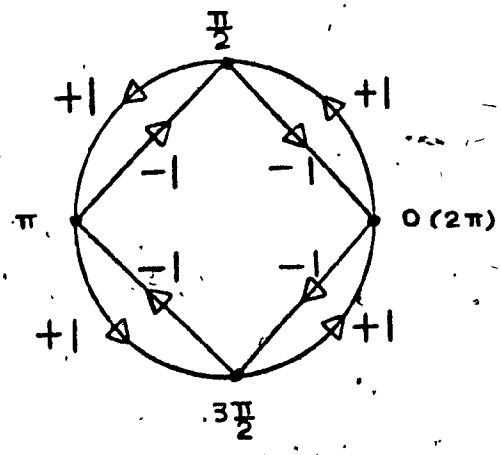


Fig. 3.2.2: State diagram for $\Phi(t)$.

where, from (2.2.5) with $h=\frac{1}{2}$, φ_k is 0 or π modulo 2π. This representation of the FFSK signal shows that it consists of two quadrature carrier components, both phase modulated (BPSK) by the sinusoids, C(t) or S(t), whose phases alternate between $+\pi/2$ and $-\pi/2$ or 0 and π according to the current value of $I(t)=\cos\varphi_k=\pm 1$ or $Q(t)=d_k\cos\varphi_k=\pm 1$, the data dependent terms (symbols). C(t) and S(t) are referred to as the symbol weighting sinusoidal functions.

The data dependent terms, I(t) and Q(t), possess an important property. Although the data d_k may change sign every T seconds, these terms can change sign only every 2T seconds. The possible instants for such a change, relative to their corresponding symbol weighting sinusoids, are the zero crossings of C(t) for I(t) and of S(t) for Q(t). This can be derived [4] based on the recursive phase constraint (3.2.2) and the fact that

$$\varphi_k = 0 \text{ or } \pi \text{ modulo } 2\pi \qquad \ldots(3.2.5).$$

This derivation is repeated here for the sake of convenience.

Substitude (3.2.2) into $\cos\varphi_k$ and expand by trigonometric identities:

$$\begin{aligned}\cos\varphi_k &= \cos\left\{\varphi_{k-1}+(d_{k-1}-d_k)\frac{\pi k}{2}\right\}\\ &=\cos(\varphi_{k-1})\cos\left\{(d_{k-1}-d_k)\frac{\pi k}{2}\right\}\\ &\quad-\sin(\varphi_{k-1})\sin\left\{(d_{k-1}-d_k)\frac{\pi k}{2}\right\}\end{aligned}$$

We now note that:

$\sin(\varphi_{k-1})=0$ for $\varphi_{k-1}=0,\pi$ modulo 2π.

$d_{k-1}-d_k=0,\pm 2$ for $d_k=\pm 1$

$\sin\left\{(d_{k-1}-d_k)\frac{\pi k}{2}\right\}=0$ for k any integer

$$\cos\left\{(d_{k-1}-d_k)\frac{\pi k}{2}\right\}=\begin{cases}+1 \text{ for k even or for k odd and } d_k=d_{k-1}\\ -1 \text{ for k odd and } d_k \text{ differentthan } d_{k-1}\end{cases}$$

Therefore,

$$\cos(\varphi_k)=\cos(\varphi_{k-1}) \quad \text{for k even or for k odd and } d_k=d_{k-1}$$

$$\cos(\varphi_k)=-\cos(\varphi_{k-1}) \quad \text{for k odd and } d_k\neq d_{k-1}.$$

Equivalently, replacing k even by 2k and k odd by 2k+1 we have:

$$\cos(\varphi_{2k})=\cos(\varphi_{2k-1}) \quad \text{always}$$

$$\cos(\varphi_{2k+1})=\cos(\varphi_{2k}) \quad \text{for } d_{2k+1}=d_{2k}$$

$$\cos(\varphi_{2k+1})=-\cos(\varphi_{2k}) \quad \text{for } d_{2k+1}=-d_{2k}$$

Letting $\cos(\varphi_k) = I_k$ from (3.2.3), the above can be written as:

$$I_{2k} = I_{2k-1}, \quad \text{always} \qquad (3.2.6)$$
$$I_{2k+1} = \pm I_{2k} = \pm I_{2k-1}, \quad \text{for } d_{2k+1} = \pm d_{2k}$$

In a similar manner it can be shown that:

$$Q_{2k+1} = Q_{2k} \quad \text{always} \qquad \ldots(3.2.7)$$
$$Q_{2k} = \pm Q_{2k-1} = \pm Q_{2k-2}, \quad \text{for } d_{2k} = \pm d_{2k-1}$$

Equations (3.2.6) and (3.2.7) give the law by which the initial data d_k are physically encoded into the I and Q Symbols. These symbols have a rate equal to half the bit rate as it can be visualized from (3.2.6) and (3.2.7).

At this point it would be of interest to compare the FFSK quadrature component representation with that of the Offset QPSK, given by:

$$y_{O\text{-}QPSK}(t) = d_{2k+1}\cos w_o t + d_{2k}\sin w_o t \qquad \ldots(3.2.8)$$

where d_{2k} and d_{2k+1} represent the symbol streams consisting of the even and odd bits of the source data stream d_k respectively (each of these symbols lasts 2T seconds). The FFSK and O-QPSK waveforms are of the same nature except for the different shapes in their symbol weighting functions. In O-QPSK, these functions are rectangular pulses while in FFSK sinusoidal. This shape is in fact responsible for the continuous phase of the FFSK wave, as well as for the fast roll-off rate of its spectral side-lobes as will be evident in the next section.

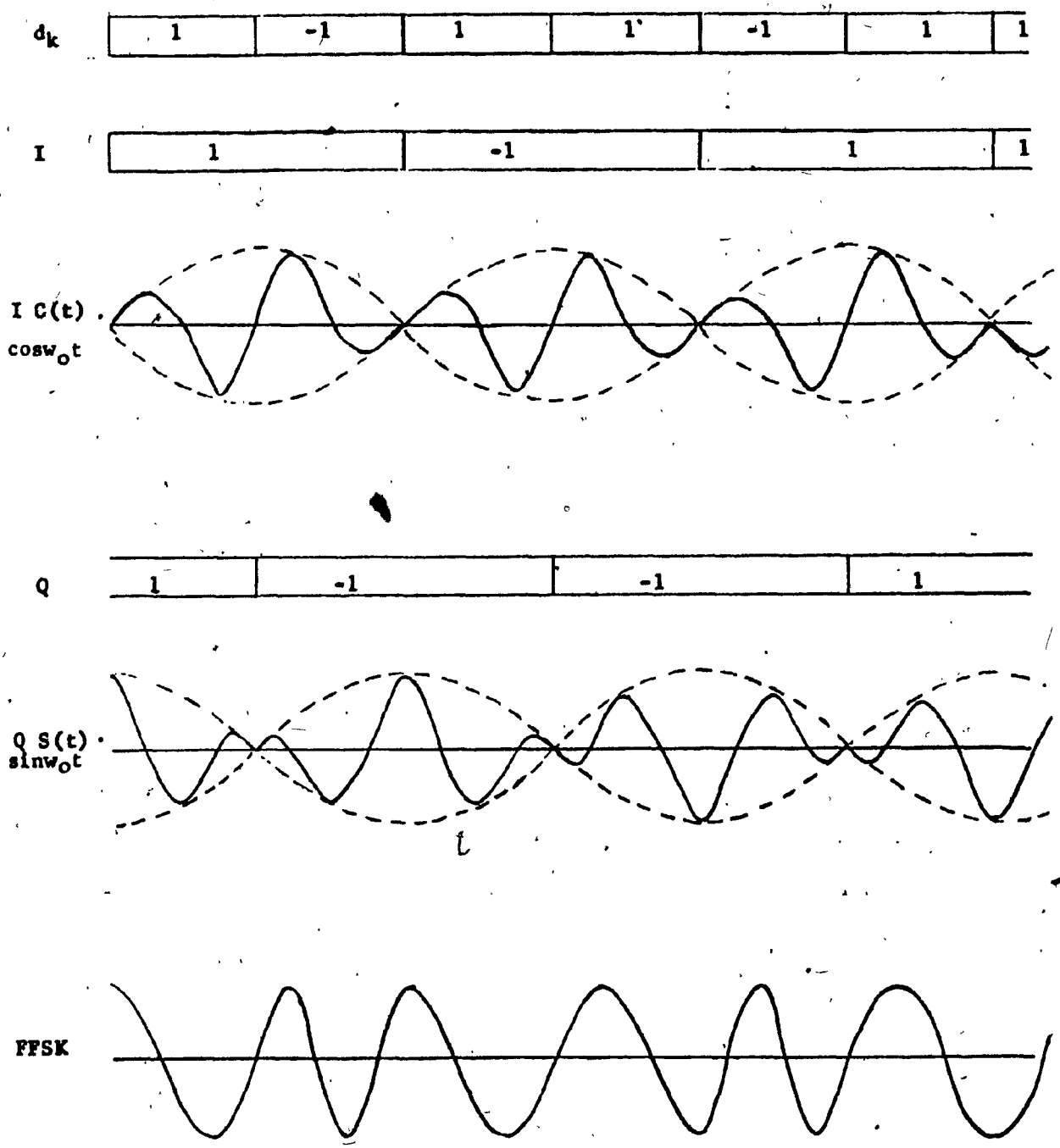


Fig.: 3.2.3: Data and symbol timing relationships in FFSK waveform.

Figure 3.2.3 illustrates the FFSK waveform for a specific input data stream. It is the linear combination of the two quadrature components and it has a constant envelope, which is a highly important property in applications, where non-linear channels are involved as it will be shown in section 4.7 of Chapter 4.

It is possible to generate such a signal by establishing its components (I and Q channels) and the equivalent I(t) and Q(t) data streams rather than interswitching two oscillators, which is the direct procedure. Chapter 4 of this report deals with implementation techniques of this type.

In Figure 3.2.3 a T seconds overlap between I(t) and Q(t) data is obvious. It results from the fact that each of these data may change value at the zero crossings of its symbol weighting sinnsoid and these sinnsoids C(t) and S(t) differ by $\pi/2$ in phase.

3.3 Spectrum And Autocorrelation Function Of FFSK Signals

In this section the power spectral density of the FFSK signal will be derived using its quadrature nature along with the sinusoidal symbol weighting functions C(t) and S(t), which actually modulate the two quadrature carriers coswot and sinwot respectively, producing two pseudocarriers at $f_o \pm f_b/4$.

Equation (3.2.4) can be expanded as:

$$y_{FFSK}(t)=\frac{I(t)+Q(t)}{2}\cos(w_o+\frac{\pi}{2T})t+\frac{I(t)-Q(t)}{2}\cos(w_o-\frac{\pi}{2T})t$$

$$=(\frac{1+d_k}{2}\cos\phi_k)\cos(w_o+\frac{\pi}{2T})t+(\frac{1-d_k}{2})\cos\phi_k)\cos(w_o-\frac{\pi}{2T})t$$

$$\text{for } kT\leq t\leq(k+1)T \qquad \ldots\ldots\ldots\ldots\ldots\ldots(3.3.1)$$

and gives a new appearence to the FFSK signal which is very helpful in determining its spectrum in a more or less intuitive fashion based on relatively practical grounds.

This representation shows that the FFSK waveform can be also viewed as two binary-phase modulated (BPSK) carriers at fo-fb/4 and fo+ fb/4, the modulating signals being I(t)-Q(t) and I(t) + Q(t) repsectively.

These signals are Polar Return to Zero data (Polar RZ) since they assume the values ±1 and 0, as can be easily varified by assigning different values to d_k and cos ϕk in the terms:

$$\frac{1+d_k}{2}\cos\varphi_k \qquad \text{and} \qquad \frac{1-d_k}{2}\cos\varphi_k$$

The Fourier transform of a Polar RZ data stream of rate fr is [10]:

$$V(f)=2T\ \frac{\sin(\pi f/2f_r)}{\pi f/2fr} \qquad \ldots\ldots\ldots\ldots\ldots\ldots(3.3.2)$$

The rate of I(t)+Q(t) and I(t)-Q(t) polar RZ data that modulate the carriers cos(wo-π/2T)t and cos(wo+π/2T)t respectively is fb/4. Thus the Fourier transform of these data using (3.3.2) is:

$$V(f)=2T\ \frac{\sin(2\pi f/f_b)}{2\pi f/f_b} \qquad \ldots\ldots\ldots\ldots\ldots\ldots(3.3.3)$$

which is exactly the same as that of Non-Return to Zero (NRZ) data of rate fb/2 (first nulls at $\pm fb/2$).

Applying the basic transform pair [3] :

$$x(t)\cos(2\pi f_o t) \longleftrightarrow \frac{1}{2}X(f-f_o)+\frac{1}{2}X(f+f_o)$$

where $x(t) \longleftrightarrow X(f)$

to both channels of (3.3.1), the one sided transform of the FFSK waveform is obtained explicitely as:

$$Y_{FFSK}(f)=\frac{2T\ \sin\{2\pi T[f-(f_o+f_b/4)]\}}{2\pi T[f-(f_o+f_b/4)]}+\frac{2T\sin\{2\pi T[f-(f_o-f_b/4)]\}}{2\pi T[f-(f_o-f_b/4)]}$$

for $0 \leqslant f \leqslant \infty$ .....(3.3.4)

Evidently the FFSK spectrum is the sum of two BPSK spectra centered at fo-fb/4 and fo+fb/4 and having their nulls at $(fo-fb/4)\pm fb/2$ and $(fo+fb/4)\pm fb2$ respectively. Thus, it is the sum of the spectra of two BPSK signals modulated by a bit stream of rate fb/2. The nulls of the composite (FFSK) spectrum occur at:

$$f_o \pm 3\frac{f_b}{4}\ ,\quad f_o \pm 5\frac{f_b}{4}\ ,\quad f_o \pm 7\frac{f_b}{4}\ ,\ldots\quad \ldots(3.3.5)$$

Figure 3.3.1 shows the two BPSK spectra corresponding to the two terms of (3.3.4) along with their linear combination, thus the FFSK spectrum. The K_{th} side lobe of the spectrum centered at fo + fb/4 and the $(K+1)_{th}$ side lobe of that centered at fo-fb/4 partially cancel out due to their opposite magnitudes. This results in a fast roll-off rate of the side lobes, a very desirable property of FFSK. It is important to note that this cancellation effect,

and consequently the fast roll-off rate, owes its presence to the existance of the sinusoidal pulses I(t)C(t) and Q(t)C(t) Indeed, the resulting spectra of the carriers coswot and sinwot, modulated by the sinusoidal pulses I(t)C(t) and Q(t)C(t) (which is an equivalent representation of the FFSK signal) have a lower roll-off rate and are slightly wider, since the Fourier transform of a sinusoidal pulse possesses these characteristics (the Fourier transform of a sinusoidal pulse has a lower roll-off rate and a slightly wider spectrum compared to that of a rectangular pulse of equal duration). To show this, consider the Fourier transforms of I(t)C(t) and Q(t)S(t):

$$v_c(t)=\cos(\pi t/2T)\longleftrightarrow V_c(f)=\int_{-T}^{T}\cos(\pi t/2T)\exp(-j2\pi ft)dt$$

$$=\frac{4T\cos(2\pi fT)}{\pi(1-16f^2T^2)}$$

$$=V(f) \qquad \text{.....(3.3.6)}$$

$$v_s(t)=\sin(\pi t/2T)\longleftrightarrow V_s(f)=\int_{0}^{2T}\sin(\pi t/2T)\exp(-j2\pi ft)dt$$

$$=\int_{-T}^{T}\cos(\pi t/2T)\exp(-j2\pi ft)dt$$

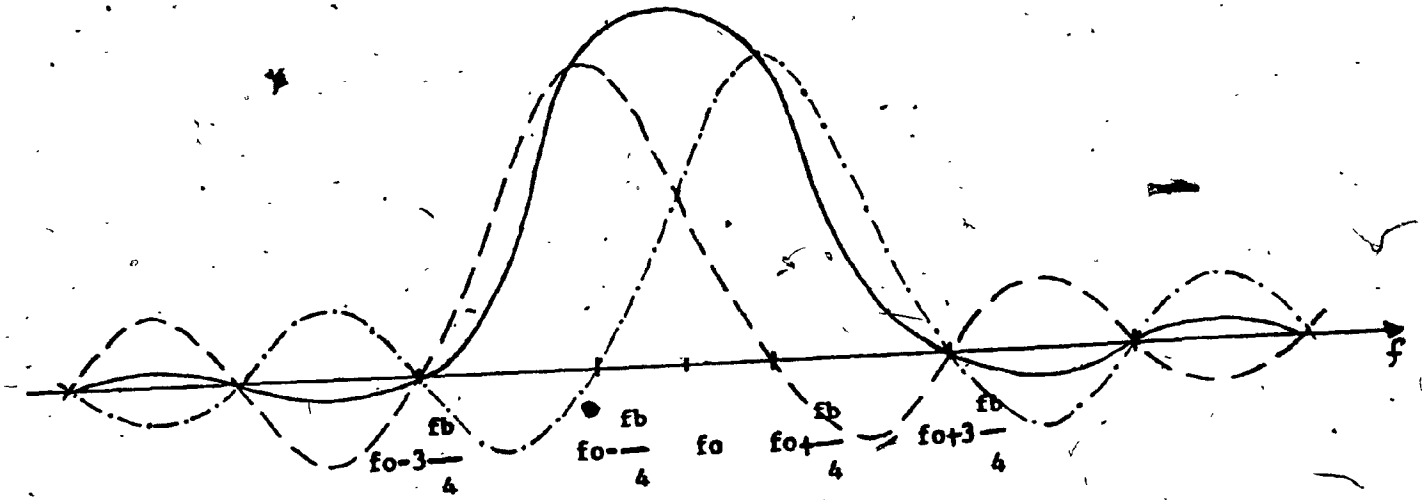
$$=V(f) \qquad \text{.....(3.3.7)}$$

Thus,

$$V_c(f-fo)=V_s(f-fo)=\frac{2T\cos\{2\pi(f-fo)T\}}{\pi\{1-16(f-fo)^2T^2\}} \qquad \text{.....(3.3.8)}$$

and the one sided FFSK spectrum is:

$$Y_{FFSK}(f)=\frac{4T\cos\{2\pi(f-fo)T\}}{\pi\{1-16(f-fo)^2T^2\}} \qquad \text{.....(3.3.9)}$$



_____ Spectrum of FFSK

- - - - -Spectrum of $(I-Q)/2 \cdot \cos(W_o - \pi/2T)t$

-.-.-.-.-Spectrum of $\{(I+Q)/2\}.\cos(Wo+\pi/2T)t$

Fig. 3.3.1: The FFSK Spectrum as a Linear Combination of two BPSK Spectra.

Equations (3.3.6) and (3.3.7) show that the spectra of the modulating pulses I(t)C(t) and Q(t)S(t) have their nodes at

$$f=(2k\pm1)\frac{fb}{4}$$

with k=1,2,3.....(i.e.k≠0). For K=0, that is at f=fb/4, the numerator and denominator in (3.3.6) become both equal to zero and V(f) reduces to πT (≠0) using L'Hospital rule. Thus, the nodes of V(f) are at 3fb/4, 5fb/4, 7fb/4....For f > fb/4, V(f) becomes a $\sin(x)/x^2$ type function and has a faster roll-off rate than that of a sin(x)/x function (proportional to $\{f/fb\}^{-2}$ rather than $\{f/fb\}^{-1}$). Correspondingly, the FFSK spectrum goes to zero at fo±3fb/4, fo±5fb/4...having the fast roll-off rate proportional to $\{|f-fo|/fb\}^{-2}$ (from equation 3.3.9) because of the sinnsoidal modulating pulses I(t)C(t) and Q(t)S(t).

The power spectral density can be found using (3.3.9):

$$G_{FFSK}(f)=\frac{8T^2\{1+\cos[4\pi(f-f_o)T]\}}{\pi^2\{1-16(f-f_o)^2T^2\}^2} \qquad .....(3.3.10)$$

The autocorrelation function is the inverse Fourier transform of the power spectral density and is given [4] by:

$$R_{FFSK}(\tau)=\begin{cases}\frac{1}{\pi}\left[\pi\left[1-\frac{|\tau|}{2T}\right]\cos\left[\frac{\pi|\tau|}{2T}\right]+\sin\left[\frac{\pi|\tau|}{2T}\right]\right], & |\tau|\leq 2T\\ 0, & |\tau|>2T\end{cases} \qquad ...(3.3.11)$$

Both GFFSK(f) and RFFSK(τ) can be derived alternatively by using the Markov process [11], [12].

The power spectral density and autocorrelation function of Offset QPSK (O-QPSK) are given by [4] :

$$G_{O\text{-}QPSK}(f) = 2T^2\left\{\frac{\sin[2\pi(f-fo)T]}{2\pi(f-fo)T}\right\}^2 \qquad \text{.....(3.3.12)}$$

$$R_{O\text{-}QPSK}(\tau) = \begin{cases} 1-\frac{|\tau|}{2T}, & |\tau| \leq 2T \\ 0, & |\tau| > 2T \end{cases} \qquad \text{.....(3.3.13)}$$

and their plottes are shown along with those of FFSK in Fig. 3.3.2 and 3.3.3 for comparison.

The agreement between the autocorrelation for FFSK in Fig. 3.3.2 and 2.6.3b is evident and the discussion of section 2.7 about the detection procedure for FFSK signals is valid. Indeed, the encoding discussed in that section exists in the nature of the FFSK signal, as it consists of two quadrature components (I and Q channels) as shown by (3.2.1), each of which encodes the data bits dk according to (3.2.6) and (3.2.7) into I(t) and Q(t) data ($\cos\phi k$ and $dk\cos\phi k$) whose rate is half that of dk and are skewed by T seconds, permitting a full observation of one encoded bit for 2T seconds. The similarity between the FFSK and Offset QPSK signals is responsible for their equal zero crossing time-length auto-correlation functions (Fig. 3.3.2) and, consequently, of the equivalent bit error rate performances.

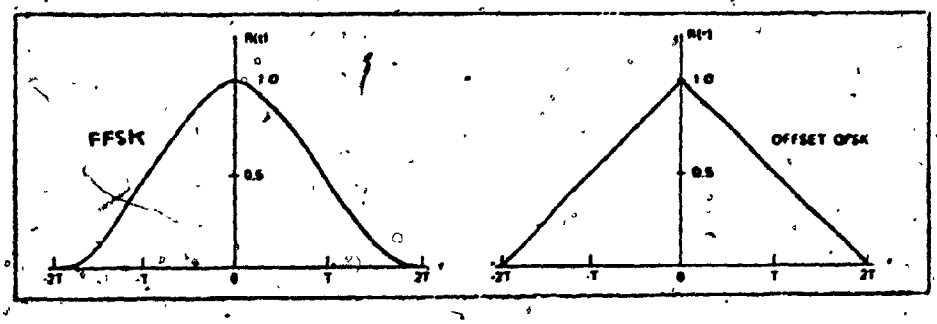


Fig. 3.3.2: Autocorrelation functions for FFSK and O-QPSK.

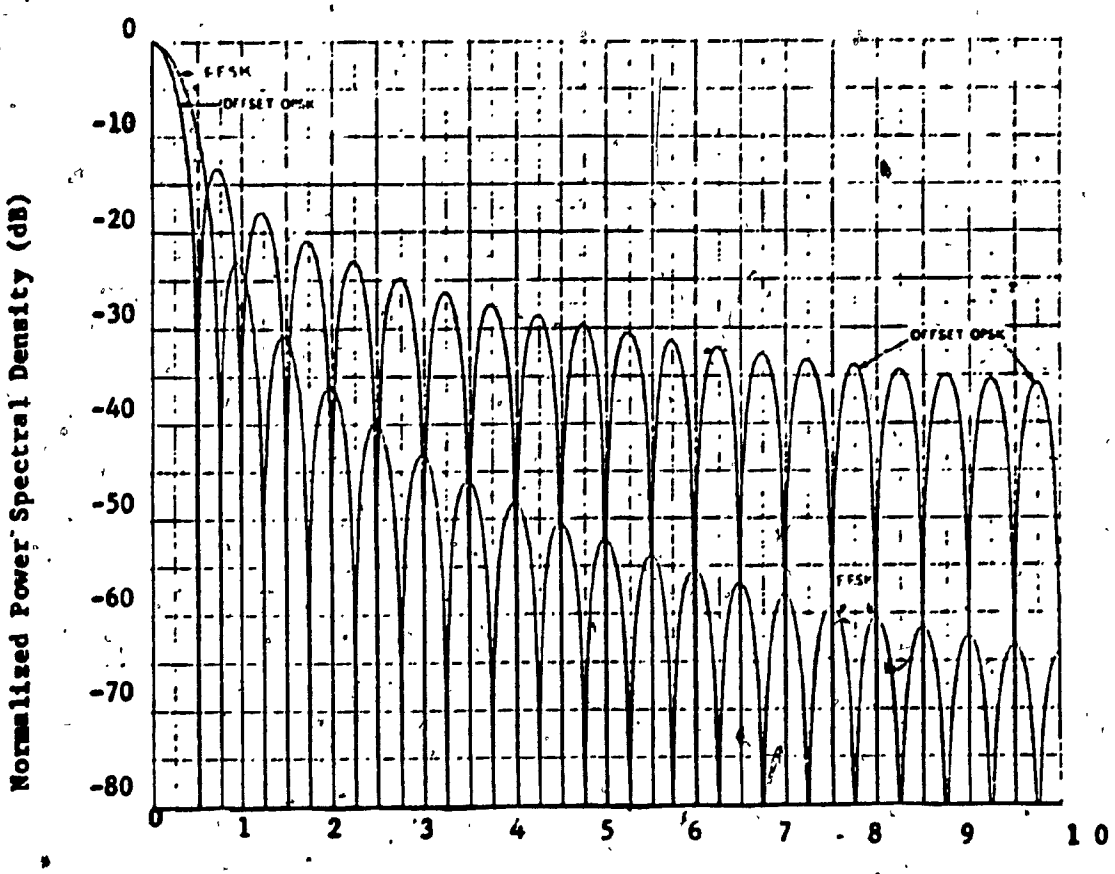


Fig. 3.3.3: Power Spectral Densities for FFSK and O-QPSK.

The roll-off rates of their power spectral densities are different (proportional to $[(f-fo)/fb]^{-4}$ for FFSK, and $[(f-fo)/fb]^{-2}$ for O-QPSK for large values of (f-fo)/fb). However, the main lobe of the FFSK spectrum is somewhat wider than that of O-QPSK. (The nulls in O-QPSK occur at (f-fo)= fb/2). Both of these differences result from the different shapes of the modulating pulses.

In practical transmission systems the modulated waveform is band-limited prior to transmission to a bandwidth 2B, chosen in such a manner that no distortion will be introduced in the conveyed data [13] . Under this condition a portion of the transmitted energy, the out-of-band energy, does not enter the receiver and the energy E in (1.13) does not represent the bit energy any more but a portion of it (note that (1.13) holds for infinite bandwidth conditions). Thus, the actual error probability is always higher than the theoretical.

The out-of-band energy is defined as:

$$E_{ob} = 1 - \frac{\int_{-B}^{B} G(f)df}{\int_{-\infty}^{\infty} G(f)df} \qquad \ldots\ldots (3.3.14)$$

and in FFSK it is less than that in O-QPSK when the available RF normalized bandwidth 2BT is greater than 1.5 due to the lower roll-off rate of the latter. This can be visualized from Fig. 3.3.4 [4] . As a result, the actual bit error rate performance of FFSK is superior to that of O-QPSK for 2BT > 1.5. However, when 2BT

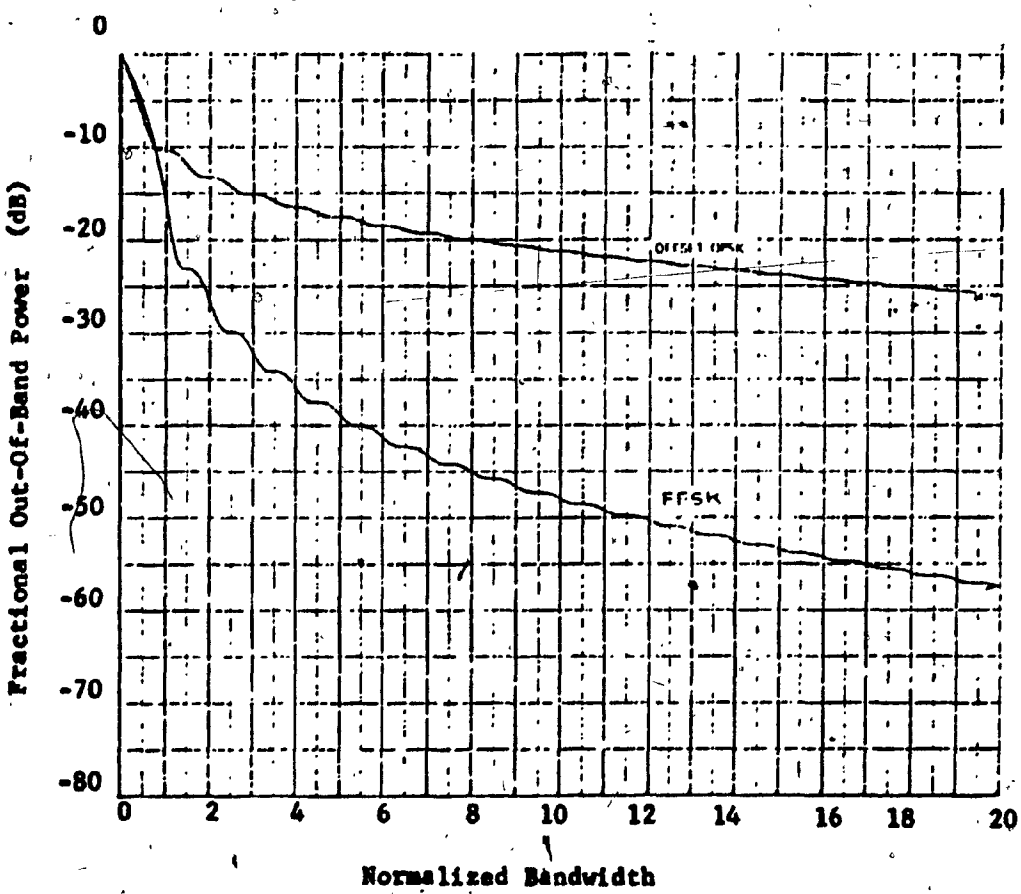


Fig. 3.3.4: Fractional Out-Of-Band Power for FFSK and O-QPSK. (Normalized two-sided bandwidth = 2BT)

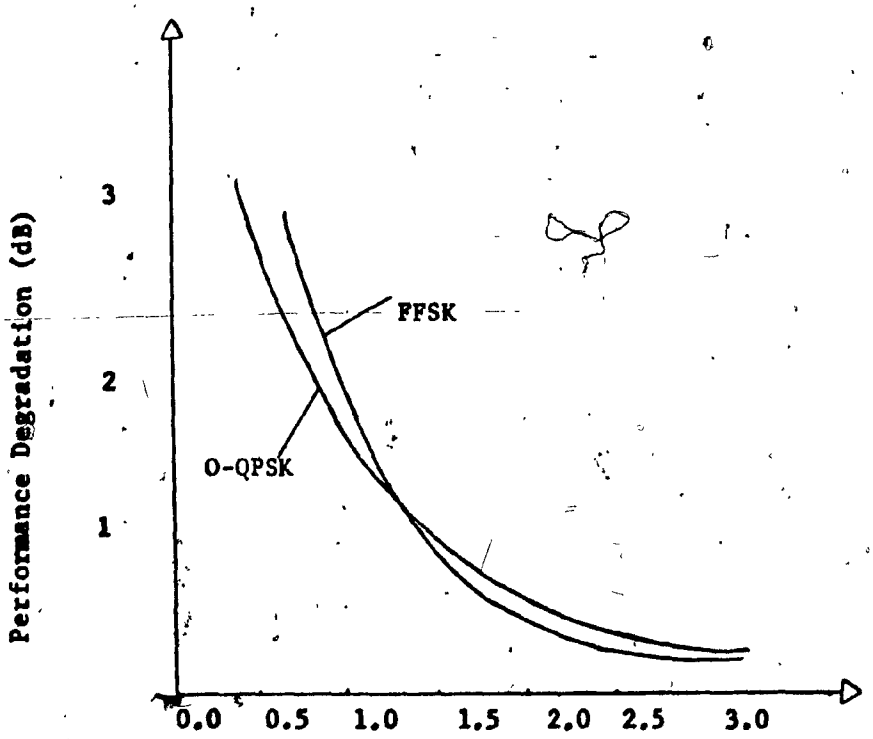


Fig. 3.3.5 : FFSK and O-QPSK performance degradation.

is reduced beyond 1.5 the out-of-band energy in FFSK is higher (see Fig. 3.3.4) and increases faster than that in O-QPSK as 2BT is reduced. At 2BT = 1 a significant portion of the main lobe in FFSK is lost while the whole main lobe in O-QPSK is transmitted (Fig. 3.3.3). As an expected result, the bit error rate performance of QPSK should be superior for $2BT < 1$. A cross-over point must be reached between 2BT = 1.5 and 2BT = 1. This point can be determined only when the detailed parameters of the channels are known. The performances of the two systems converge to the infinite bandwidth case as 2BT is increased beyond 1.5. Figure 3.3.5 shows the expected comparative performance degradation versus the available bandwidth 2BT, curves for a specified bit error rate.

Summary

The FFSK signal was analysed in this chapter. It was shown that it consists of two quadrature components of frequency fo with quadrature sinusoidal envelopes, C(t) and S(t), and binary-phase modulated by two new data streams. The I (t) and Q(t) data. In other words, the quadrature carriers are modulated by the sinusoidal pulses I(t)C(t) and Q(t)S(t). The FFSK signal can also be regarded as two pseudocarriers at fo-fb/4 and fo + fb/4, binary-phase modulated by the I(t) + Q(t) and I(t)-Q(t) data. With either of the above interpretations (which are equivalent) all spectral properties of this signal, such as the fast roll-off rate and

the wider width were explained.

The power spectral density with the autocorrelation function and the out-of-band energy of the signal were presented, and the effect of the latter on the error probability was pointed out on a comparison basis with the O-QPSK signal.

CHAPTER 4

GENERATION, DETECTION AND PROCESSING OF FFSK SIGNALS

From the analysis of the FFSK signal in Chapter 3, the methods of producing and detecting it have already become somewhat clear. This chapter deals with the presentation of these methods, covering every single step of processing. A discussion of the effects of bandlimiting and of non-linear channels on the FFSK signal is also presented.

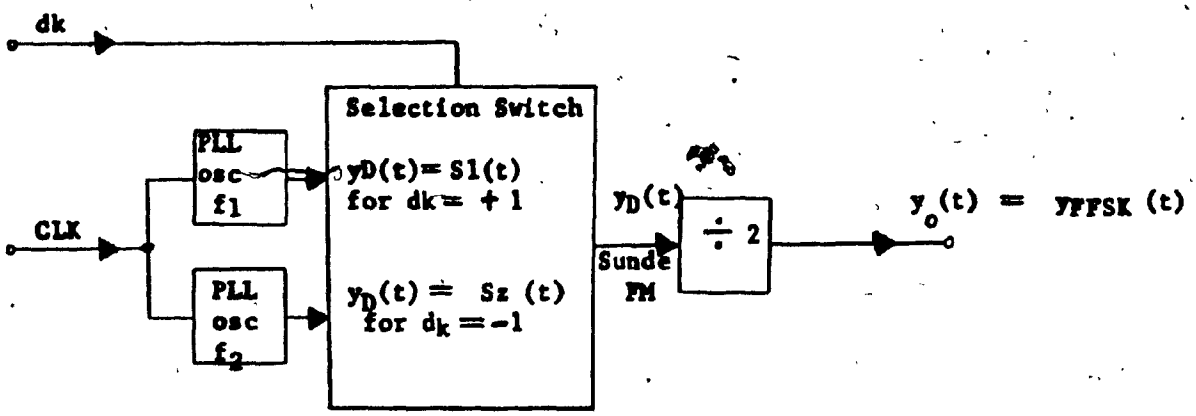
4.1 Generation Of The FFSK Signal - Modulation

In principle, the FFSK signal can be generated by applying frequency division by 2 on a Sunde FM signal shown in Fig. 4.1.1a. The associated waveforms are shown in Fig. 4.1.1b. Alternatively, the FFSK signal can be generated in the manner described in Section 3.1 (Fig. 3.1.2) where the PLL's used operate at the FFSK frequencies f_1, f_2 rather than at $2f_1$, $2f_2$ and an inverter controlled by the input data is used instead of the frequency divider.

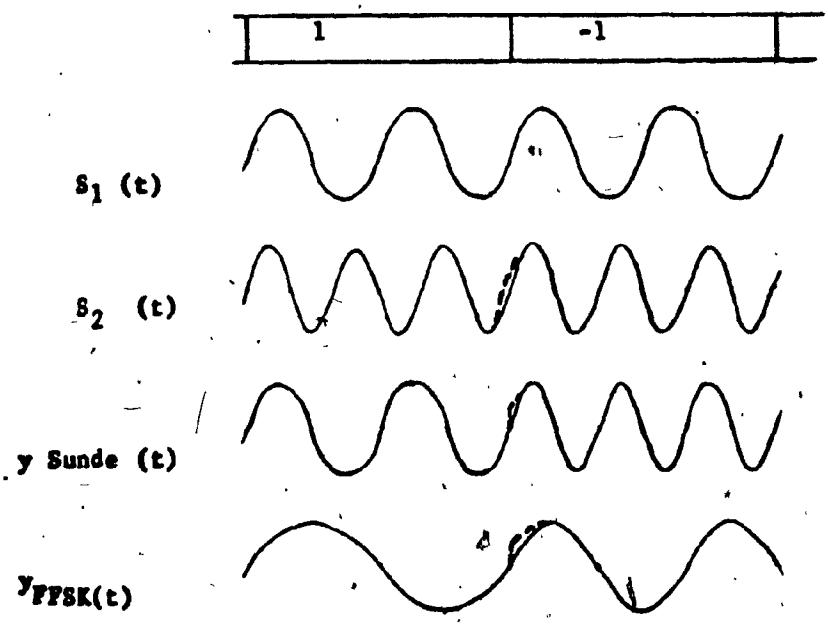
Both the described methods suffer from a serious disadvantage associated with the phase coherence of the signal. According to (2.1.9) the phase difference $\Delta\phi(coh)$ between $S_1(t)$ and $S_2(t)$ at bit transitions, should be $2h\pi k$ ($k=0,1,2,....$). Thus, it depends on the value of h, and any variation in this value (produced by variations of f_1 or f_2 due to instability of the carrier oscillators or due to any change in bit rate) is reflected to $\Delta\phi(coh)$. The immediate result is the loss of phase coherence [5] and, consequently, phase discontinuities in the ysunde(t) and the yFFSK waveforms will be introduced as shown by the dashed lines in Fig. 4.1.1b.

The so called orthogonal methods of generation of the FFSK waveform eliminate this problem by generating its components in synchronism with the data clock.

One such method is illustrated in Fig. 4.1.2. In this method the tones coswot and sinwot are generated from the data clock through the use of a PLL and a 90° phase shifting network.



(a) Circuit



(b) Waveforms

Fig. 4.1.1: Generation of FFSK from Sunde FM.

The symbol weighting sinusoids $C(t)=\cos\pi t/2T$ and $S(t)=\sin\pi t/2T$ are similarly produced. Two multipliers form the two orthogonal carriers $C(t)\cos\omega_o t$ and $S(t)\cdot\sin\omega_o t$, and a data encoder produces the I and Q-data by implementing the natural data encoding, described by (3.2.6) and (3.2.7). The orthogonal carriers $C(t)\cos\omega_o t$ and $S(t)\sin\omega_o t$ are then modulated by the I and Q-data respectively and the products are combined linearly in the output summer to produce the FFSK signal.

Figure 4.1.3 shows the associated waveforms for the generation of the orthogonal carriers.

The phase shift produced by a phase shifting network is a function of the signal frequency. But slight frequency deviations due to the instability of the oscillators do not affect the amount of phase shift considerably because a frequency variation is not linearly reflected on the shift according to the expression $\tan(\text{phase shift})=R/2\pi fL$ that holds for an elementary RL phase shifting network.

However, in applications where flexibility in the use of carrier frequency is required (where widely different frequencies may be necessary to be used according to current assignments) this implementation is not appropriate.

An alternative configuration [5] which avoids the use of the 90^o phase shifting networks and it is therefore flexible, is shown in Fig. 4.1.4. Here the mark and space signals $S_1(t)=\cos(\omega_o-\omega_b/4)t$ and $S_2(t)=\cos(\omega_o+\omega_b/4)t$ are produced by modulating

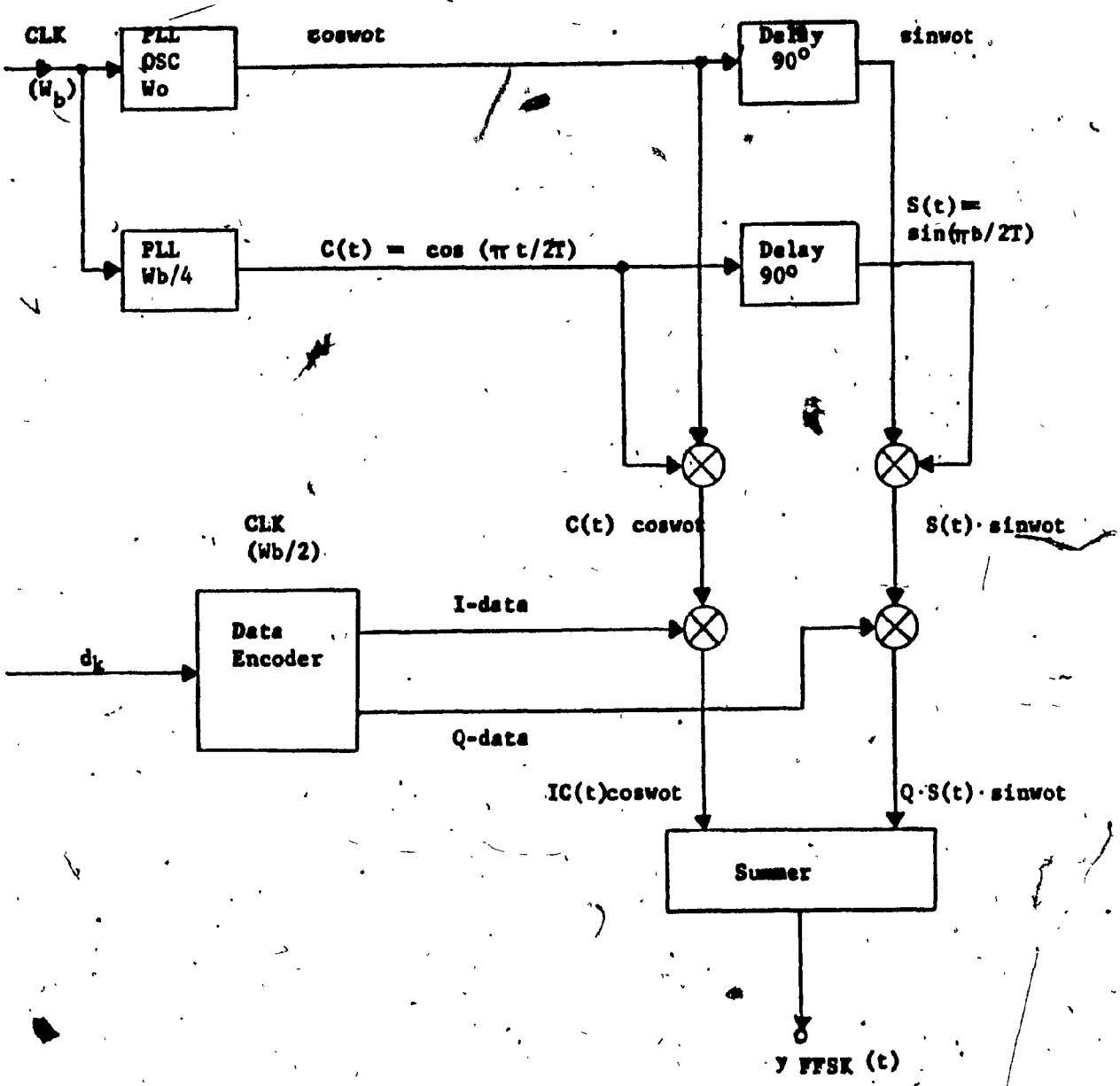


Fig. 4.1.2: Orthogonal method of generating FFSK signal (using phase shifting networks.)

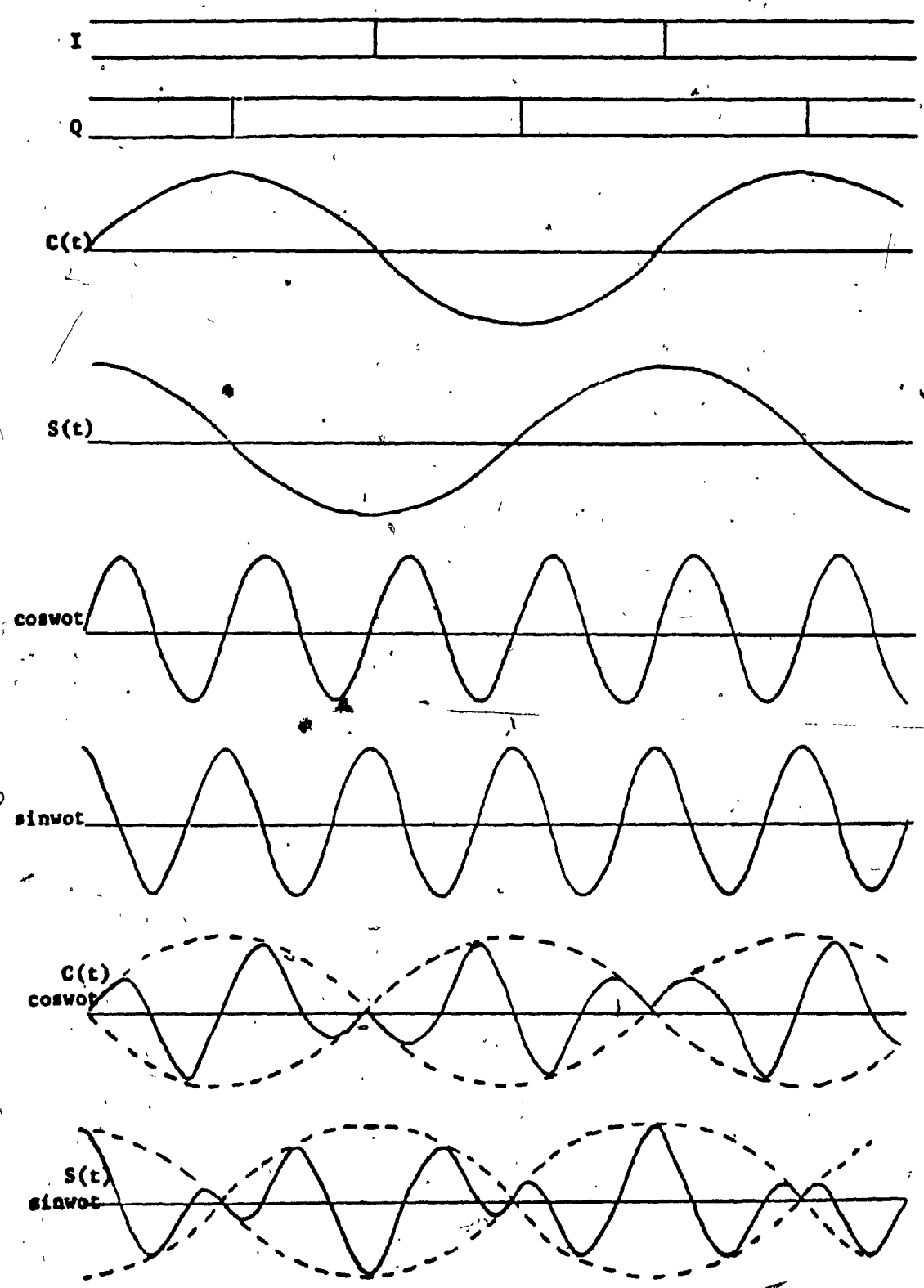


Fig. 4.1.3: Waveforms for the Configuration of Fig. 4.1.2

a master carrier coswot by the data clock, frequency divided by 4, and by selecting the modulation products at (wo-wb/4) and (wo+wb/4) using two bandpass filters centered at these frequencies. The generation of these tones as a result of such a modulation can be visualized if we consider the Fourier transform of a generalized pulse train with period To and pulse duration τ [14] :

$$V(f)=\sum_{-\infty}^{\infty}\frac{\sin(n\pi\tau/To)}{n\pi\tau/To}\quad,\quad n=\frac{f}{fo}=fT_o \qquad \ldots\ldots(4.1.1)$$

where V(f) is the Fourier transform of the pulse train.

In the case of the clock signal of rate equal to one fourth the bit rate, which is used to modulate the carrier coswot, To=4T and τ=To/2=2T and the Fourier transform of this clock is:

$$V(f)=\sum_{-\infty}^{\infty}\frac{\sin(2\pi f/fb)}{2\pi f/fb} \qquad \ldots\ldots(4.1.2)$$

Figure 4.1.5a depicts the one-sided spectrum, represented by (4.1.2). Clearly, there exist spectral lines at multiples of fb/4. The output of the modulator (multiplier), using the transform pair [3] :

$$v(t)\cos(2\pi fo)t \longleftrightarrow \frac{1}{2}V(f-fo)+\frac{1}{2}V(f+fo)$$

where: $v(t)\longleftrightarrow V(f)$

has the Fourier transform:

$$Y(f)=\sum_{-\infty}^{\infty}\frac{\sin\{2\pi(f-fo)/fb\}}{2\pi(f-fo)/fb} \qquad \ldots\ldots(4.1.3)$$

which is shown in Fig. 4.1.5b. From this figure it is evident that,

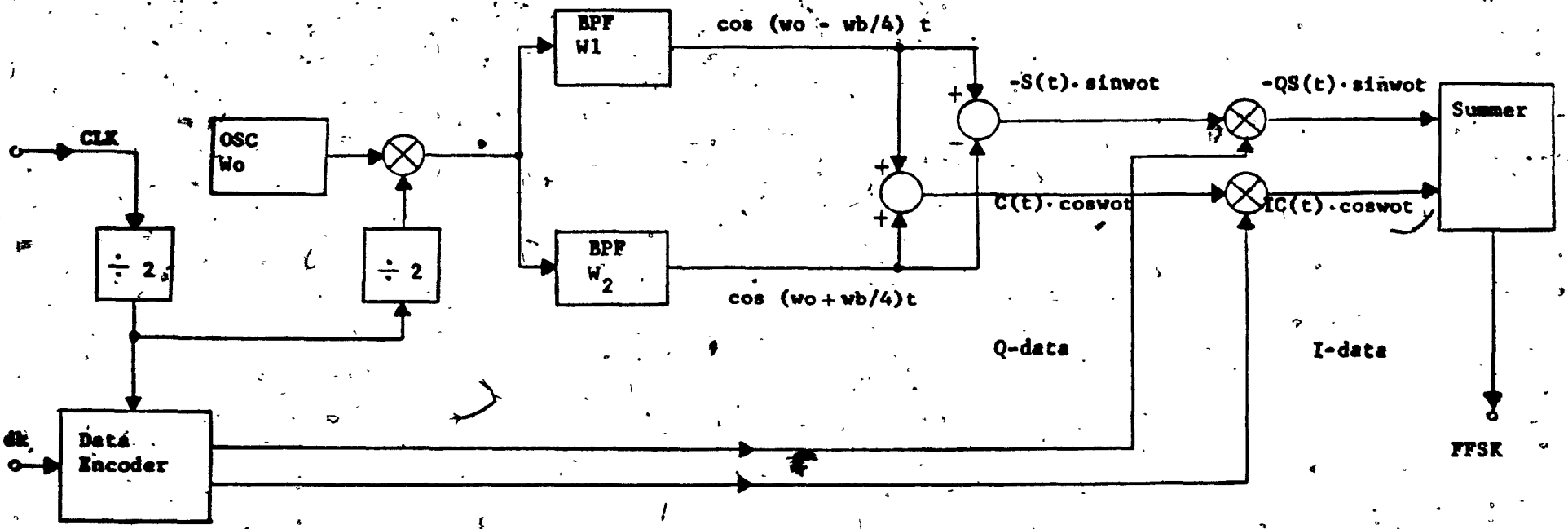
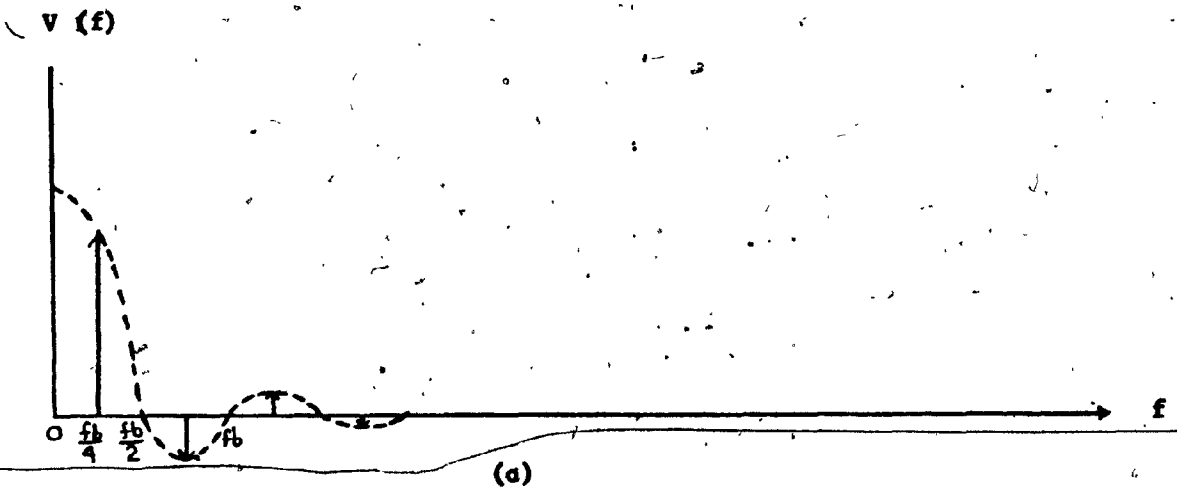
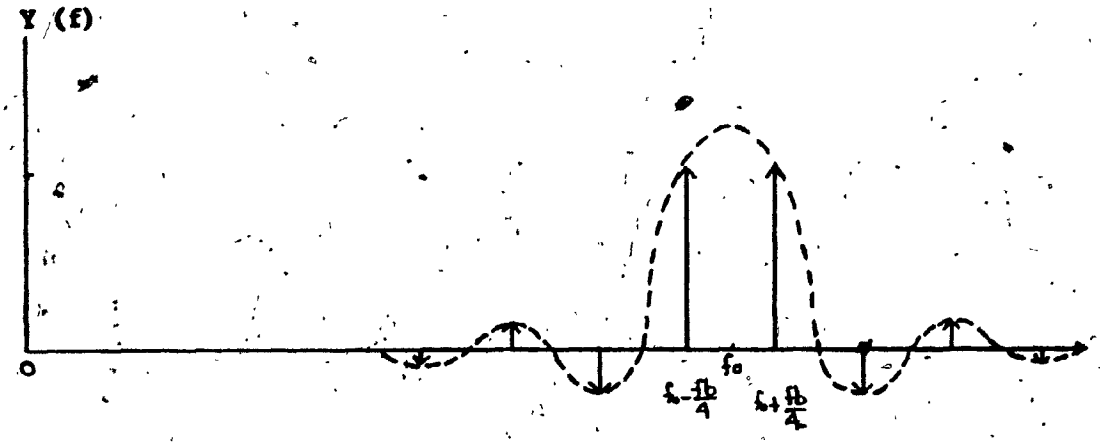


Fig. 4.1.4: Orthogonal method of FFSK generation using sum and difference devices.



(a)



(b)

Fig. 4.1.5: Generation of the pseudocarriers at fo + fb/4 and fo - fb/4.
(a) Spectrum of a clock with rate fb/4
(b) Spectrum of the modulation product of coswot and the clock in (a).

among others, there exist two spectral lines at (fo-fb/4) and (fo + fb/4) which can be selected by the two band-pass filters (or by two PLL's).

By adding and substructing the so produced signals $S_1(t)$ and $S_2(t)$, the carrier components C(t)coswot and S(t)sinwot are available respectively:

$$S_I(t)=S_2(t)+S_1(t)=\cos(w_o+w_b/4)t+\cos(w_o-w_b/4)t$$

$$=2\cos(\pi t/2T)\cos w_o t \qquad \ldots\ldots(4.1.4)$$

$$S_Q(t)=S_2(t)-S_1(t)=\cos(w_o+w_b/4)t-\cos(w_o-w_b/4)t$$

$$=-2\sin(\pi t/2T)\sin w_o t \qquad \ldots\ldots(4.1.5)$$

These carriers are then binary-phase modulated by the I and Q-data generated from the data dk by the data encoder. This encoder implements the physical encoding described by (3.2.6) and (3.2.7).

Finally, the modulated carriers are linearly combined in the summing network to produce the FFSK waveform.

4.2 Optimum Reception For The FFSK Signal

The maximum phase openings in the phase diagram of Fig. 3.2.1 occur at times t=(2K+1)T and t=2KT and detection-decisions must be taken for these times:

$$\Phi(t)=\begin{cases}\pm\pi/2 & , \quad t=(2k+1)T\\ 0,\pi & , \quad t=2kT\end{cases} \qquad \ldots\ldots(4.2.1)$$

The phase distance at these instants is π rads which implies that a correlation coefficient $\rho = -1$ is effective and optimum detection can be obtained by sampling at $(2K+1)T$ and $2kT$.

Indeed, the signal can be evaluated at these instants as:

$$y_{FFSK}(2k+1)T = \begin{cases} -d_k \cos\varphi_k \sin(\pi t/2T) \sin w_o t = -Q(t)S(t)\sin w_o t, & t = T, 5T, 9T, .. \\ +d_k \cos\varphi_k \sin(\pi t/2T) \sin w_o t = +Q(t)S(t)\sin w_o t, & t = 3T, 7T, .. \end{cases}$$

.....(4.2.2)

$$y_{FFSK}(2kT) = \begin{cases} +\cos\varphi_k \cos(\pi t/2T) \cos w_o t = +I(t)C(t)\cos w_o t, & t = 0, 4T, 8T, .. \\ -\cos\varphi_k \cos(\pi t/2T) \cos w_o t = -I(t)C(t)\cos w_o t, & t = 2T, 6T, 10T, .. \end{cases}$$

.....(4.2.3)

where $\varphi k = 0, \pi$ (i.e. $\cos\varphi k = \pm 1$) and $dk = \pm 1$. Hence $\cos\varphi k = I(t) = \pm 1$ and $dk\cos\varphi k = Q(t) = \pm 1$.

Clearly, the FFSK signal reduces to only one of its components at multiples of T. Each of these components are antipodal and the FFSK signal at these times is such. As a consequence, optimum detection can be obtained by taking detected samples for $t = (2k+1)T$ (Q-data) and for $t = 2kT$ (I-data).

From the basic theory of the optimum detection of binary signals [3] a correlator receiver is required to detect a modulated signal of the generalized form:

$$y(t)=\begin{cases} S_1(t), & kT \leq t \leq (k+1)T \\ \\ S_2(t), & kT \leq t \leq (k+1)T \end{cases} \qquad \ldots.(4.2.4)$$

and the reference signal must be:

$$S(t)=S_2(t)-S_1(t) \qquad \ldots(4.2.5)$$

The two components of the FFSK signal are orthogonal and a correlator receiver operation using the reference signal for one of its components will delete the effect of the other component because the latter is orthogonal to the reference and the integral of the product of two orthogonal signals is zero. This property of orthogonality permits the use of two parallel correlator receivers (one for each component), both fed by the FFSK signal but one using as reference:

$$S_I(t)=S_{I2}(t)-S_{I1}(t)=2\cos(\pi t/2T)\cos w_o t \quad \ldots(4.2.6)$$

and the other:

$$S_Q(t)=S_{Q2}(t)-S_{Q1}(t)=-2\sin(\pi t/2T)\sin w_o t \quad \ldots(4.2.7)$$

The integration cycles of the correlator receivers must be (2K-1)T, (2K+1)T for the I-channel and 2kT,(2K+2)T for the Q-channel so that the midpoint be the maximum opening instant for which the sample is to be taken. The integrators outputs will be:

$$y_I(t) = \int_{(2k-1)T}^{(2k+1)T} \left\{ I(t)C(t)\cos w_o t - Q(t)S(t)\sin w_o t \right\} 2C(t)\cos w_o t$$

$$= I(t)T \text{ , the I-data} \qquad \ldots\ldots\ldots\ldots\ldots\ldots (4.2.8)$$

$$y_Q(t) = \int_{2kT}^{(2k+2)T} \left\{ I(t)C(t)\cos w_o t - Q(t)S(t)\sin w_o t \right\} (-2S(t)\sin w_o t)$$

$$= Q(t)T \text{ , the Q-data} \qquad \ldots\ldots\ldots\ldots\ldots\ldots (4.2.9)$$

The pulse shape of the detected symbols is curved due to the IF filtering and they should be hard limited (required for detection decisions in the presence of noise and of bandlimiting) and sampled at (2K+1)T (the I(t)) and at 2kT (the Q(t)). These samples should be hold for 2T seconds in order that the symbols I(t) and Q(t) be regenerated (their duration is 2T seconds). There will be of course a 2T seconds delay. Such an optimum FFSK receiver is shown in Fig. 4.2.1. The samples are taken by triggering the sample-hold devices with a trigger train of repetition rate equal to half the bit rate and occuring exactly at the nodes of

C(t) and S(t) respectively. The same trigger train is used to reset the integrators for the next symbol.

In practical applications low-pass filters are usually employed instead of integrators because there may not be enough time available for the integrators to go to zero before the next symbol proceeds too far [5] when the bit rate used is relatively high.

In such a case the high frequency products of the multipliers are greately attenuated by the filters and the I and Q-data appear at the outputs. The multiplier outputs $m_I(t)$ and $m_Q(t)$ are:

$$m_I(t)=\left\{I\cos(\pi t/2T)\cos w_o t - Q\sin(\pi t/2T)\sin w_o t\right\}\left(2\cos(\pi t/2T)\cos w_o t\right)$$

$$= \frac{I}{2}+\frac{I}{2}\cos(\pi t/T)+\frac{I}{2}\cos 2w_o t+(1+\frac{Q}{8})\cos(2w_o+\frac{\pi}{T})t$$

$$+ (1-\frac{Q}{8})\cos(2w_o-\frac{\pi}{T})t$$

$$m_Q(t)=\left\{I\cos(\pi t/2T)\cos w_o t - Q\sin(\pi t/2T)\sin w_o t\right\}\left(-2\sin(\pi t/2T)\sin w_o t\right)$$

$$= \frac{Q}{2}-\frac{Q}{2}\cos(\pi t/T)-\frac{Q}{2}\cos 2w_o t+(1+\frac{I}{8})\cos(2w_o+\frac{\pi}{T})t$$

$$+ (1-\frac{I}{8})\cos(2w_o-\frac{\pi}{T})t$$

If the low-pass filters eliminate the components at $f > fb$, the final outputs $m_{OI}(t)$, $m_{OQ}(t)$ will be:

$$m_{OI}(t)=\frac{I}{2} \qquad \text{and} \qquad m_{OQ}(t)=\frac{Q}{2}$$

and the I and Q-data will be recovered.

The detected I and Q symbols must be sampled directly at their midpoints $t = 2kT$ (the $I(t)$) and $t=(2k+1)T$ (the $Q(t)$). These points are the optimum sampling times for bandlimited reception in the presence of noise because the values of the detected symbols are then maximum and the noise effect is minimized.

The recovered I and Q-data must next be decoded in order that the initial data dk be obtained.

4.3 Carrier And Bit Timing Recovery (CTR, BTR)

A common method of recovering the carrier reference in most modulation techniques is to use non linear signal processing [14] . The incoming signal is fed into a square law device. If the FFSK signal be squared the resulting output will be a Sunde FM-type signal:

$$y^2_{FFSK}(t) = \cos^2(w_o t + \varphi_k + d_k \frac{\pi t}{2T})$$

$$= \frac{1}{2} + \frac{1}{2}\cos(2w_o t + 2\varphi_k + \frac{\pi t}{T})$$

where $2\varphi_k = 0, 2\pi$ (note that $\varphi_k = 0, \pi$).

Finally,

$$y^2_{FFSK}(t) = \frac{1}{2} + \frac{1}{4}\cos(2w_o + \frac{\pi}{T})t + \frac{1}{4}\cos(2w_o - \frac{\pi}{T})t$$

$$+ \frac{d_k}{4}\cos(2w_o + \frac{\pi}{T})t - \frac{d_k}{4}\cos(2w_o - \frac{\pi}{T})t \quad \ldots\ldots\ldots(4.3.1)$$

The second and third terms of (4.3.1) are two spectral lines (free of information) at $(2w_o - \pi/T) = 2w_1$ and $(2w_o + \pi/T) = 2w_2$ and can be isolated by two narrow bandpass filters or PLL's

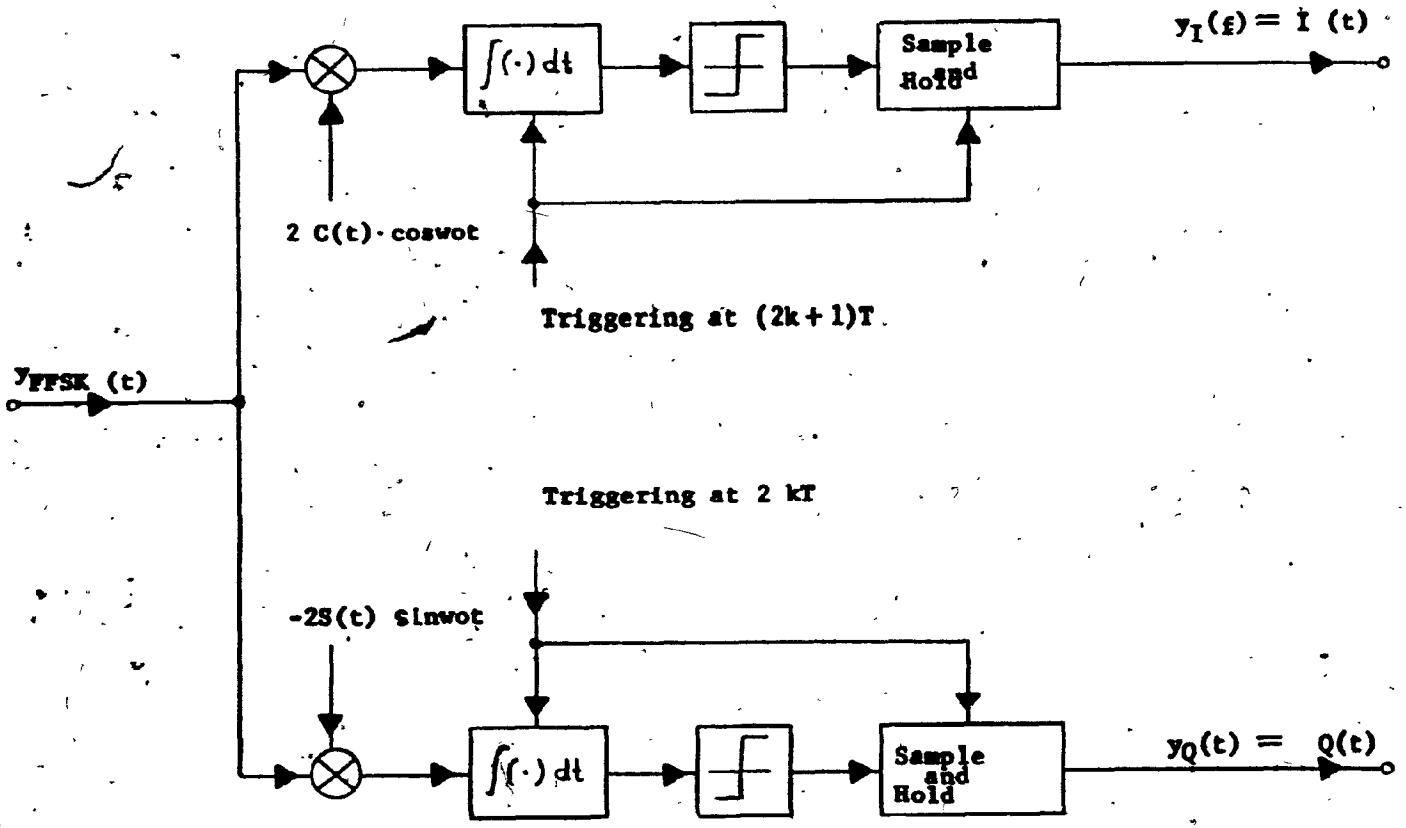


Fig. 4.2.1: Optimum receiver for FFSK.

centered at these frequencies as in Fig. 4.3.1.

To obtain the required references, 2C(t)coswot and -2S(t)sinwot, these tones can be multiplied (Fig. 4.3.2) to produce two new tones at $4w_o$ and w_b and then separated by a low-pass and a high-pass filter [8] .

The output of the high-pass filter (the tone at 4wo) can be divided by 4 to obtain coswot.

The output of the low-pass filter (the tone at w_b) can also be divided by 4 to obtain cos(wb/4) (one fourth the bit rate, wb/4). A 90° phase-shift is required to produce the quadrature components of these tones, and the reference signals $S_I(t)$ and $S_Q(t)$ can be constructed by proper multiplications combination (Fig. 4.3.2).

A phase ambiguity equal to $2\pi/N$ is introduced by a divider by N. In the case of Fig. 4.3.2, N=4, and the phase ambiguity is $\pi/2$. This ambiguity can be converted to π if the spectral tones at $2w_1$ and $2w_2$ are divided by 2, producing w_1 and w_2, and the sum and difference of the new tones $\cos w_1 t$ and $\cos_2 t$ are formed [8] , as shown in Fig. 4.3.3 (in which case the phase shift networks have been omitted).

$$S_{sum}(t) = \cos w_1 t + \cos w_2 t = \cos(w_o - \frac{\pi}{2T})t + \cos(w_o + \frac{\pi}{2T})t$$

$$= 2\cos(\pi t/2T)\cos w_o t = S_I(t) \qquad \ldots\ldots(4.3.2)$$

$$S_{dif}(t) = \cos w_2 t - \cos w_1 t = \cos(w_o + \frac{\pi}{2T})t - \cos(w_o - \frac{\pi}{2T})t$$

$$= -2\sin(\pi t/2T)\sin w_o t = S_Q(t) \qquad \ldots\ldots(4.3.3)$$

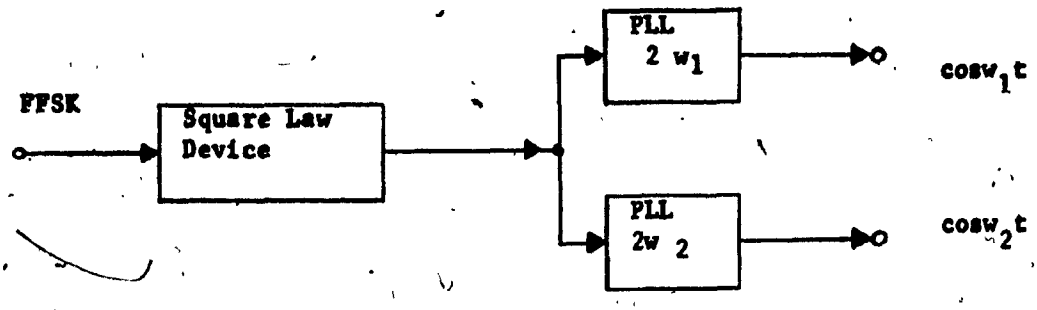


Fig. 4.3.1: Extraction of the tones at 2_{w1} and 2_{w2}.

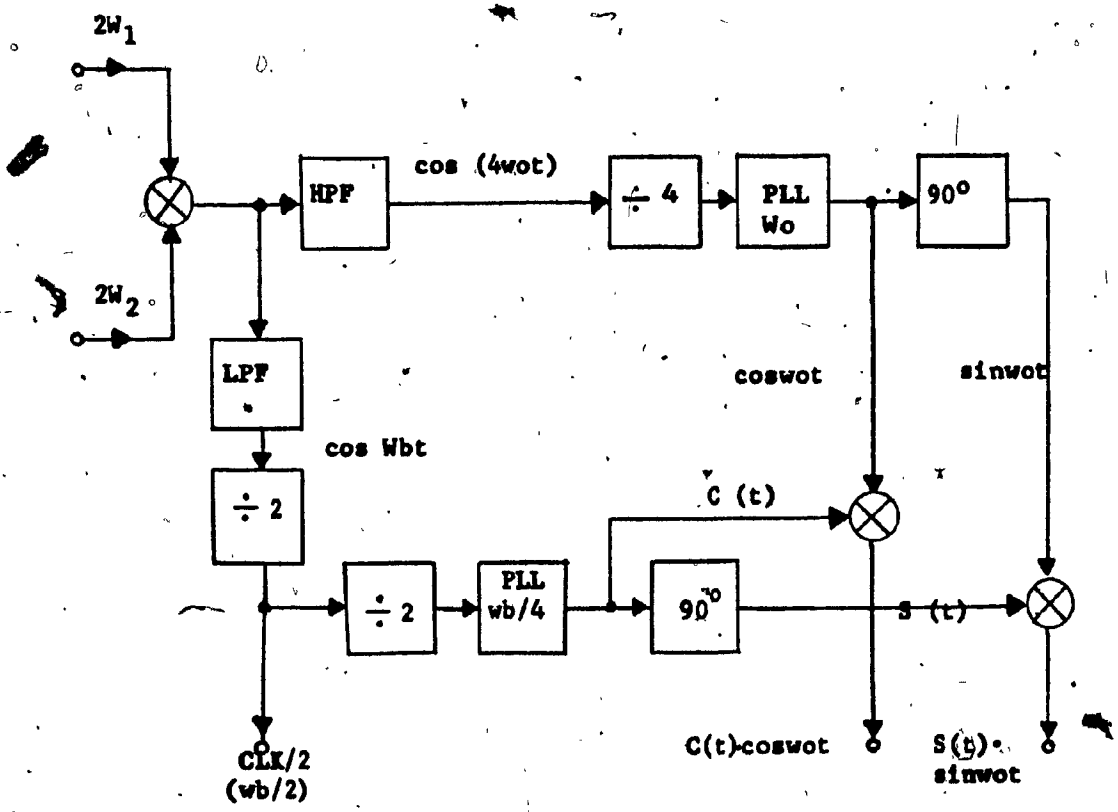


Fig. 4.3.2: Method of reference recovery.
(Phase ambinguities of 90° are introduced.)

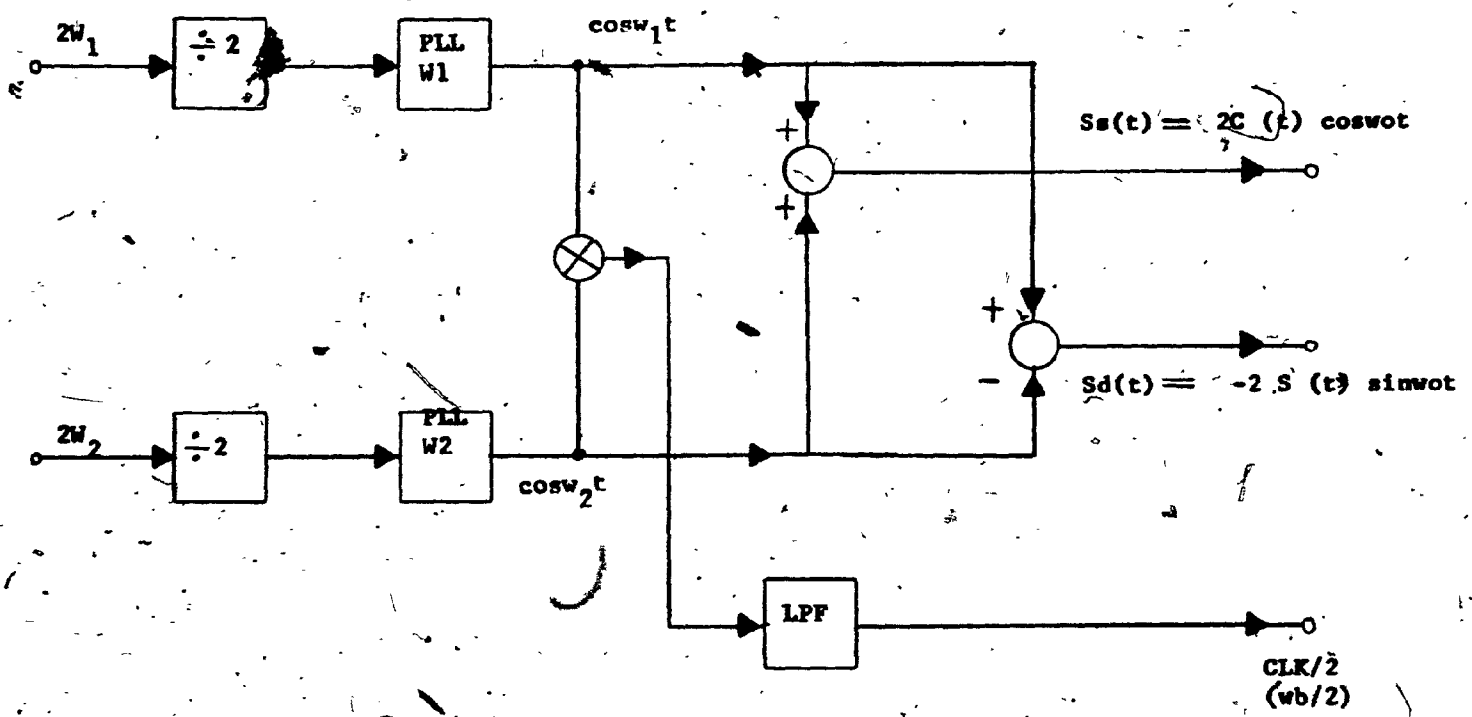


Fig. 4.3.3: Method of reference recovery. Phase ambiguities of 90° are eliminated.

Since the phases of the dividers outputs are ambiguous by π ($\pm\cos(\omega_o t + \pi t/2T)$, $\pm\cos(\omega_o t - \pi t/2T)$) the output of the sum device may give the difference and that of the difference device may give the sum. This creates an uncertainty about which detection branch represents the I-channel and which the Q-channel. However, this uncertainty is not harmful if the triggering of each integrator is kept in synchronism with its reference. To achieve this, the triggering pulses must be recovered in a manner that ensures the introduction of the same uncertainty. One way to obtain such triggers is to multiply the outputs of the PLL's $\cos\omega_1 t$ and $\cos\omega_2 t$, and low-pass filter the product. The resulting waveform is:

$$\cos\omega_1 t \cdot \cos\omega_2 t = \cos(\omega_o t - \frac{\pi t}{2T}) \cdot \cos(\omega_o t + \frac{\pi t}{2T}) = (\tfrac{1}{2})\cos 2\omega_o t + (\tfrac{1}{2})\cos(\pi t/T)$$

where the first term is greatly attenuated by the filter. The second term is a tone at $\omega_b/2$, always in phase with the envelope of Ssum(t) because its phase ambiguities are introduced by the same dividers (Fig. 4.3.3).

Alternatively, the triggering pulses can be obtained by using circuits that can detect the nulls of the envelopes of Ssum(t), Sdif(t) [9].

The fact that each detection branch does not represent a specified channel (I or Q) requires a method of data decoding that it is insensitive to which input is the I-data and which is the Q-data. This matter will be covered in the two following sections.

4.4 Data Encoding

In the orthogonal methods of generation of the FFSK signal a data encoder is always used to encode the initial data d_k into the I and Q-data. This encoder has the task to excecute the encoding laws given by (3.2.6) and (3.2.7) as:

$I_{2k+1} = +I_{2k} = +I_{2k-1}$ when $d_{2k+1} = +d_{2k}$

$I_{2k+1} = -I_{2k} = -I_{2k-1}$ when $d_{2k+1} = -d_{2k}$

$Q_{2k} = +Q_{2k-1} = +Q_{2k-2}$ when $d_{2k} = +d_{2k-1}$

$Q_{2k} = -Q_{2k-1} = -Q_{2k-2}$ when $d_{2k} = -d_{2k-1}$

Using 0 and 1 logic levels and assigning logic 1 for +1 and logic 0 for -1, these laws can be written as:

$$\begin{aligned} I_{2k+1} &= I_{2k} = I_{2k-1} && \text{when } d_{2k+1} = d_{2k} \\ I_{2k+1} &= \overline{I_{2k}} = \overline{I_{2k-1}} && \text{when } d_{2k+1} = \overline{d_{2k}} \end{aligned} \quad \ldots.(4.4.1)$$

$$\begin{aligned} Q_{2k} &= Q_{2k-1} = Q_{2k-2} && \text{when } d_{2k} = d_{2k-1} \\ Q_{2k} &= \overline{Q_{2k-1}} = \overline{Q_{2k-2}} && \text{when } d_{2k} = \overline{d_{2k-1}} \end{aligned} \quad \ldots.(4.4.2)$$

These expressions can be equivalently written, using exclusive OR operations ($\oplus$), as:

$$I_{2k+1} = (d_{2k+1} \oplus d_{2k}) \oplus I_{2k} = (d_{2k+1} \oplus d_{2k}) \oplus I_{2k-1}$$

$$= d_{2k+1} \oplus (d_{2k} \oplus I_{2k}) \quad \ldots.(4.4.3)$$

$$Q_{2k} = (d_{2k} \oplus d_{2k-1}) \oplus Q_{2k-1} = (d_{2k} \oplus d_{2k-1}) \oplus Q_{2k-2}$$

$$= d_{2k} \oplus (d_{2k-1} \oplus Q_{2k-1}) \quad \ldots.(4.4.4)$$

A comparison between (4.4.3), (4.4.4) and the output expressions of a typical differential encoder for even and odd bits

$$b_{2k+1}=d_{2k+1}\ b_{2k} \qquad \ldots\ldots(4.4.5)$$

$$b_{2k}=d_{2k}\ b_{2k-1} \qquad \ldots\ldots(4.4.6)$$

(where b_k is the output and d_k the input streams) shows that I_{2k+1} and Q_{2k} can be regarded as the $(2k+1)$th and $2k$th outputs of a differential encoder whose input stream is d_k, and its preceding outputs are $d_{2k}\oplus I_{2k}$ and $d_{2k-1}\oplus Q_{2k-1}$ respectively. Indeed, such an encoder output is:

$$b_{2k+1}=d_{2k+1}\oplus b_{2k}=d_{2k+1}\oplus(d_{2k}\oplus I_{2k})$$

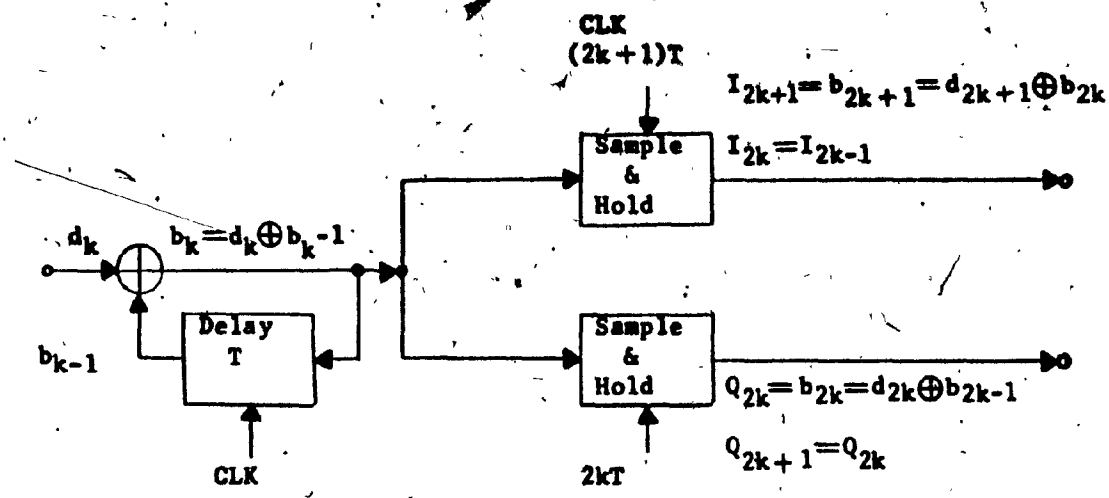
$$=d_{2k+1}\oplus(d_{2k}\oplus I_{2k-1})=I_{2k+1} \qquad \ldots\ldots(4.4.7)$$

$$b_{2k}=d_{2k}\oplus b_{2k-1}=d_{2k}\oplus(d_{2k-1}\oplus Q_{2k-1})$$

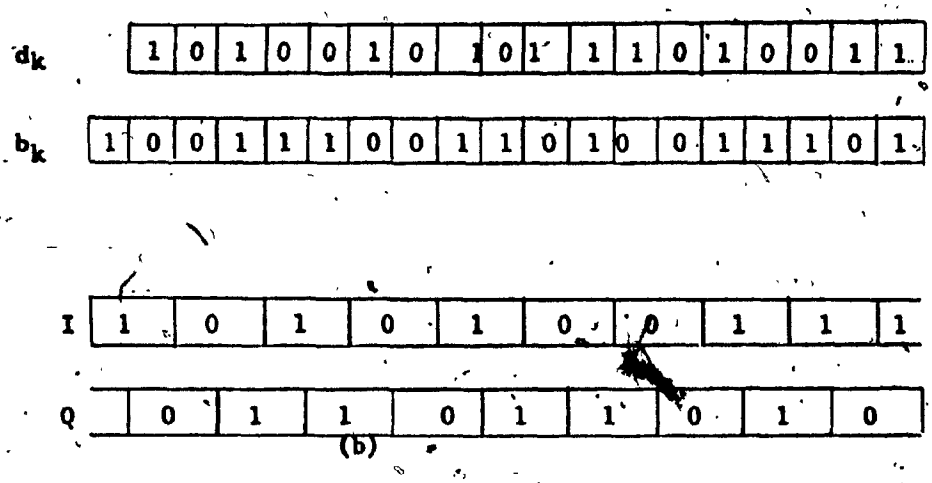
$$=d_{2k}\oplus(d_{2k-1}\oplus Q_{2k-2})=Q_{2k} \qquad \ldots\ldots(4.4.8)$$

and (4.4.7) and (4.4.8) are identical to (4.4.3) and (4.4.4) respectively.

The physical encoding laws given by (3.2.6), (3.2.7) can then be implemented by a differential encoder whose output is sampled and hold for 2T seconds at times $t=(2k+1)T$ and $t=2kT$ into two separate branches one for I_{2k+1} and one for Q_{2k} output bits (this later operation is required to realize the equalities $I_{2k-1}=I_{2k}$ and



(a)



(b)

Fig. 4.4.1: Data Encoder
(a) Block diagram
(b) Encoding of d_k into I and Q

$Q_{2k-2}=Q_{2k-1}$ that appear in (3.2.6) and (3.2.7).

Such a data encoder along with related wave forms is shown in Fig. 4.4.1.

4.5 Data Decoding

The I and Q-data are detected by the demodulation process described in section 4.2. To obtain the initial data d_k from the I and Q-data, (4.4.7), (4.4.8) may be solved for d_{2k+1} and d_{2k}. Then, using the following logic identities:

$$a \oplus 1 = \bar{a} \qquad \ldots(i)$$

$$a \oplus 0 = a \qquad \ldots(ii)$$

$$a \oplus b = d \longleftrightarrow a = b \oplus d \qquad \ldots(iii)$$

we have:

$$d_{2k+1} = I_{2k+1} \oplus b_{2k} = I_{2k+1} \oplus Q_{2k} = I_{2k+1} \oplus Q_{2k+1} \qquad \ldots\ldots(4.5.1)$$

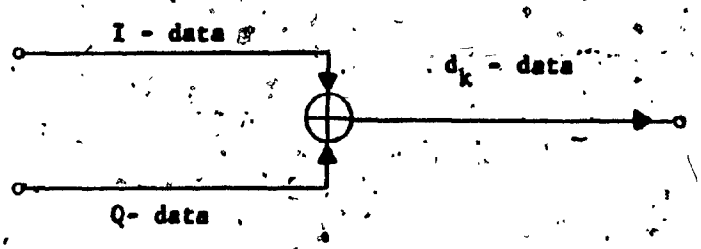
$$d_{2k} = Q_{2k} \oplus b_{2k-1} = Q_{2k} \oplus I_{2k-1} = Q_{2k} \oplus I_{2k} \qquad \ldots\ldots(4.5.2)$$

Clearly from (4.5.1), (4.5.2):

$$d_k = I_k \oplus Q_k \qquad \ldots\ldots(4.5.3)$$

and a single exclusive-OR gate (modulo-two adder) can decode the I and Q-data into the desired d_k data as in Figure 4.5.1, where related waveforms are also shown.

The 180° ambinguity due to the carrier recovery circuits is removed by the differential encoding-decoding which takes place in



(a)

d_k	1	0	1	0	0	1	0	1	0	1	1	1	0	1	0	0	1	1
I		1		0		1		0		1		0		0		1	1	
Q			0		1		1		0		1		1		0		1	0
y=I⊕Q		1	0	1	0	0	1	0	1	0	1	1	1	0	1	0	0	1

(b)

Fig. 4.5.1: Data Decoder
(a) Implementation
(b) Decoding of I an Q. The initial data d_k are reproduced.

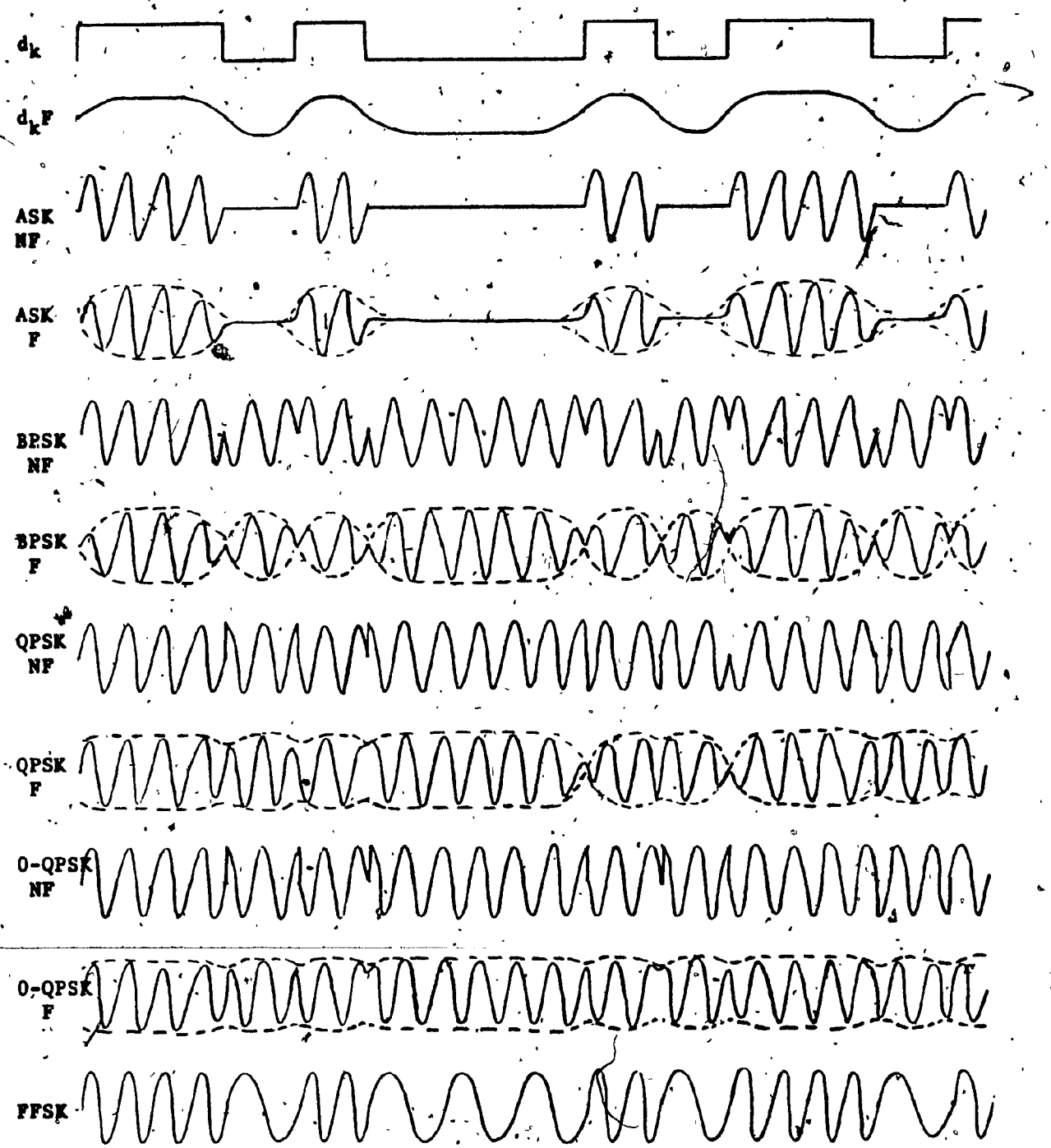
the data encoding-decoding process. Clearly, the exclusive-OR gate is not sensitive to which input is the I and which the Q-data and the channel identification uncertainty does not introduce any problem.

4.6 Effects Of Bandlimting

The abrupt changes of the information signal are reflected on the amplitude of the IF waveform in all linear modulation techniques as ASK and PSK (PSK can be regarded as a special case of linear ASK in which the message signal takes on the values ± 1 rather than 1, 0). However, the envelope of PSK is constant when bandlimiting is not applied. The effect of bandlimiting is the introduction of envelope variations near the abrupt signal changes, thus near bit transitions. Such variations are present in ASK and PSK signals (100% or 30%).

In an FSK signal the information is carried on the frequency and the process is non linear. The abrupt changes of the information signal are reflected on the frequency while the amplitude is not affected if the phase continuity condition (equation 2.2.2) is satisfied. Hence, the continuous phase FSK and, in particular the FFSK, signal has no abrupt amplitude changes [20] and, when bandlimited, is nearly free of envelope variations compared to the PSK techniques. The existing AM in FFSK is around 15% when $2BT=1.1$ and 24% worst case when $2BT=0.71$ [23] which is a very hard filtering.

Figure 4.6.1 shows the typical effect of bandlimiting on the



Note: NF stands for "Non Filtered".
F stands for "Filtered".

Fig. 4.6.1: Envelope variations of several Binary Modulation Techniques.

envelopes of several signal types.

There are two types of interference [8] on FFSK introduced by bandlimiting: ISI within each of the I and Q channels, and interference between the two channels. The latter can be avoided by proper filter design: there will be no such interference when the overall filter action is that of a symmetrical bandpass. The ISI with one channel (I or Q) is not easily removed but it is much smaller than that occurring in QPSK systems due to a slight difference in the shape of the autocorrelation functions of the two systems (Fig. 3.3.2): in the FFSK this function reaches very close to zero at $|\tau| < 2T$. This interference is negligible when $2BT \geq 1.5$ and starts increasing with decreasing B (or T) for $2BT < 1.5$, which limits the permissible bit rate for a given available bandwidth. However, it is possible to increase the spead at which we can transmit data by correcting the ISI by transversal filters at the baseband [15] , [16] , [17] .

Ideally, the 3dB IF bandwidth must be equal to the main lobe bandwidth (1.5fb) but even more severe bandlimiting may be applied (3dB bandwidth somewhere between 1.5fb and fb) without significant ISI and information loss to be experienced. Because of the fast spectral roll-off rate the out-of-band energy does not increase rapidly with decreasing available bandwidth. According to the comparative discussion in Chapter 3 between FFSK and O-QPSK, there exist a point at around $2BT = 1.2$ where the performances of both systems are equal although more energy than that of the main lobe of O-QPSK and less than that of FFSK is then used.

4.7 Effects Of Non Linear Channels

In satellite communications high power Travelling Wave Tube Amplifiers (TWTA) are used at the satellite transponder to amplify the signal. These amplifiers operate in their saturation region to achieve high efficiency. When a device operates in saturation, non-linearities are introduced.

A quadrature model of a general non-linearity in a TWTA was analysed [18] . If the input signal is

$$v_i(t) = r(t)\cos(w_o t + \phi) \qquad \ldots\ldots(4.7.1)$$

where r(t) and ϕ are the envelope and phase functions respectively, the output of the TWTA may be represented as

$$v_o(t) = R\{r(t)\}\cos(w_o t + \Theta\{r(t)\} + \phi) \qquad \ldots\ldots(4.7.2)$$

where $R\{\cdot\}$ and $\Theta\{\cdot\}$ represent the amplitude and the phase non-linearities of the TWTA. Equ. (4.7.2) can be expanded as:

$$v_o(t) = \left[R\{r(t)\}\cos\Theta\{r(t)\}\cos\phi - R\{r(t)\}\sin\Theta\{r(t)\}\sin\phi\right]\cos w_o t - \left[R\{r(t)\}\sin\Theta\{r(t)\}\cos\phi + R\{r(t)\}\cos\Theta\{r(t)\}\sin\phi\right]\sin w_o t \quad ..(4.7.3)$$

The output complex baseband signal of TWTA can be written in terms of the in-phase and quadrature components as:

$$V_{ob}(t) = \left[R\{r(t)\}\cos\Theta\{r(t)\}\cos\phi - R\{r(t)\}\sin\Theta\{r(t)\}\sin\phi\right] + j\left[R\{r(t)\}\sin\Theta\{r(t)\}\cos\phi + R\{r(t)\}\cos\Theta\{r(t)\}\sin\phi\right] \quad ..(4.7.4)$$

The terms $\cos\Theta\{r(t)\}$, $\sin\Theta\{r(t)\}$ represent an amplitude-to-phase conversion effect (AM-PM), while the terms $R\{r(t)\}$ represent an amplitude-to-amplitude conversion (AM-AM). Clearly, these non-linearity effects are present only when envelop varitations r(t) exist, being proportional to them.

The AM-PM effect is responsible for excessive phase errors in detections, producing a performance degradation of the form shown in Fig. 4.7.1 [21] . The AM-AM effect is an amplitude distortion which results in regeneration of out-of-band harmonics with a spectral distribution similar to that of the original signal. As a result, newly generated side lobes are now present which added to the low-level side lobes of the bandlimited signal sum up to a total out-of-band spectrum of about the same magnitude and shape as that of the signal at the output of the modulator (the unfiltered signal). Hence, the bandlimiting is partially cancelled-out (regeneration of side lobes).

Both the above effects are proportional to the envelope variations, r(t), as shown by (4.7.4) and signals with small such variations are desired for use in non-linear channels. The envelope variations in FFSK is much smaller than those in PSK as outlined in the preceding section (15-24% compared to 30 or 100%). As a result, the contributions of AM-PM and AM-AM to the SNR degradation and the side-lobe regeneration respectively is very small in FFSK. According to experimental results, the AM-PM effect adds only a 0.3 to 0.5 dB SNR

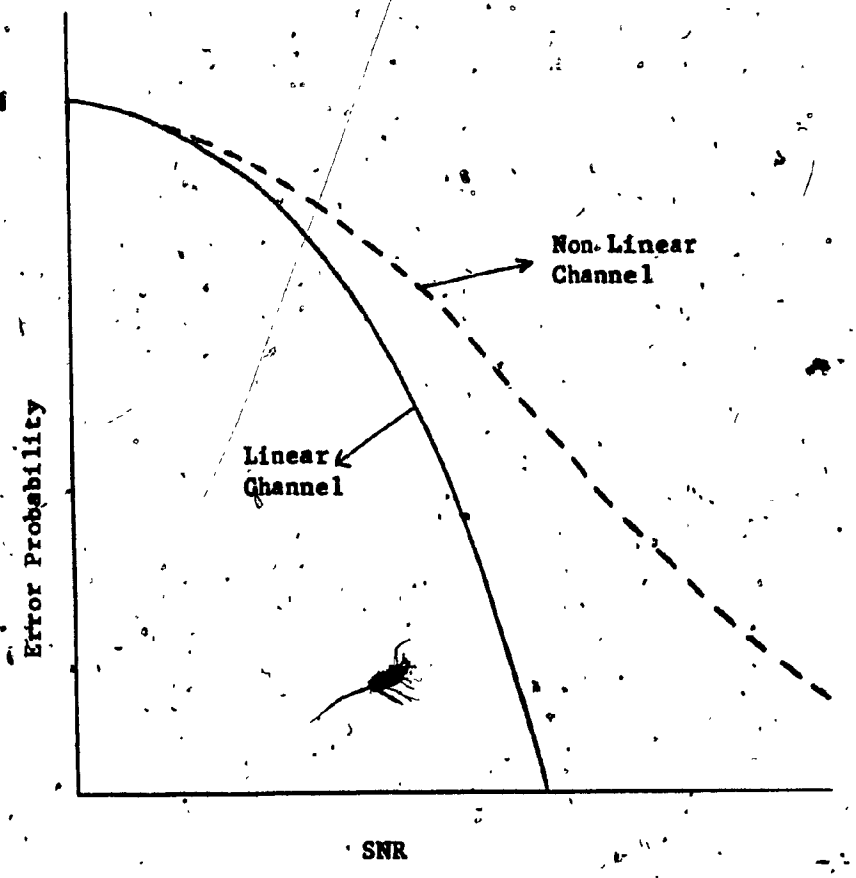


Fig. 4.7.1: Performance degrantation due to the AM - PM effect in non linear channels.

degradation in the FFSK performance [23] and the AM-AM effect increases the level of the second order FFSK side lobes by only 10dB while this increase is about 17 and 40 dB in O-QPSK and QPSK respectively when $2BT \simeq 1$ [22], [23].

Summary

Methods of generating, detecting, encoding and decoding FFSK signals as well as their reference recovery methods were presented in this chapter. It was shown that orthogonal generation of these signals is insensitive to data rate and carrier frequency variations. The detection is similar to that of QPSK signals and the data encoding and decoding is implemented with fairly simple and inexpensive hardware. The phase ambiguity due to the carrier recovery circuits is removed by the data encoding-decoding circuits and no additional effort is required. This is an advantage over QPSK systems.

The effect of bandlimiting is negligible on the FFSK signal shape and consequently the non-linearities of a satellite communication channel do not degrade the performance while, on the other hand, do not alter the transmitted spectrum significantly.

The required IF bandwidth for FFSK is closely equal to that for O-QPSK (or QPSK) and this, along with the negligible effect of non-linearities, gives the FFSK technique a significant advantage for use in satellite links over any other binary modulation.

CHAPTER 5

FFSK PERFORMANCE - COMPARISONS WITH OTHER MODULATION TECHNIQUES

In this chapter the performance of FFSK technique will be presented and compared with the performances of other techniques. Theoretical expectations will be discussed in connection with the test results.

The first part of this chapter presents actual tests for a FFSK modem [2].

The second part presents results obtained from the simulation of eight digital modulation techniques [1], namely:

Coherent Two-Phase PSK (Co.BPSK)

Coherent Four-Phase PSK (Co.QPSK)

Coherent Eight-Phase PSK (Co 8-PSK)

Differential Two-Phase PSK (DBPSK)

Differential Four-Phase PSK (DQPSK)

Offset Coherent Four-Phase PSK (O-QPSK)

Fast Frequency Shift Keying (FFSK)

Amplitude-Phase Shift Keying (A-PSK)

The results of these tests are presented in a fashion which permits a direct performance comparison between the above techniques.

5.1: EXPERIMENTAL RESULTS FOR A 60Mb/s FFSK MODEM

In this part some of the test results [2] of a 60 Mb/s FFSK modem are presented. The FFSK signal at the output of the modem transmitter has a 3 dB RF bandwidth of 66 MHz. The bit rate was kept at 60 Mb/s and all tests that involved variation of the bandwidth product 2BT were conducted by varying the system available bandwidth by means of additional filtering external to the modem.

First a back-to-back test was conducted to evaluate the performance of the modem under several conditions but always in linear channel environment. Next, several tests were conducted in which a simulation of the satellite transponder (containing a TWTA) was used to evaluate the modem performance in a satellite link. The TWTA was adjusted to operate at the linear and non-linear region in order to evaluate the non-linearity effects on the performance and the spectrum.

5.1.1 Back To Back Tests

Figure 5.1.1.1 shows the test set up used to obtain the back-to-back performance of the 60 Mb/s modem in linear channel conditions. The obtained performance is shown in Fig. 5.1.1.2 (curve labeled Baseband) where the level at the receiver input is +2.5 dBm.

The effect of bandlimiting was evaluated using the test set up of Fig. 5.1.1.3 where an additional bandpass filter of a 3 dB RF bandwidth of 36 MHz was inserted following the transmitter. The

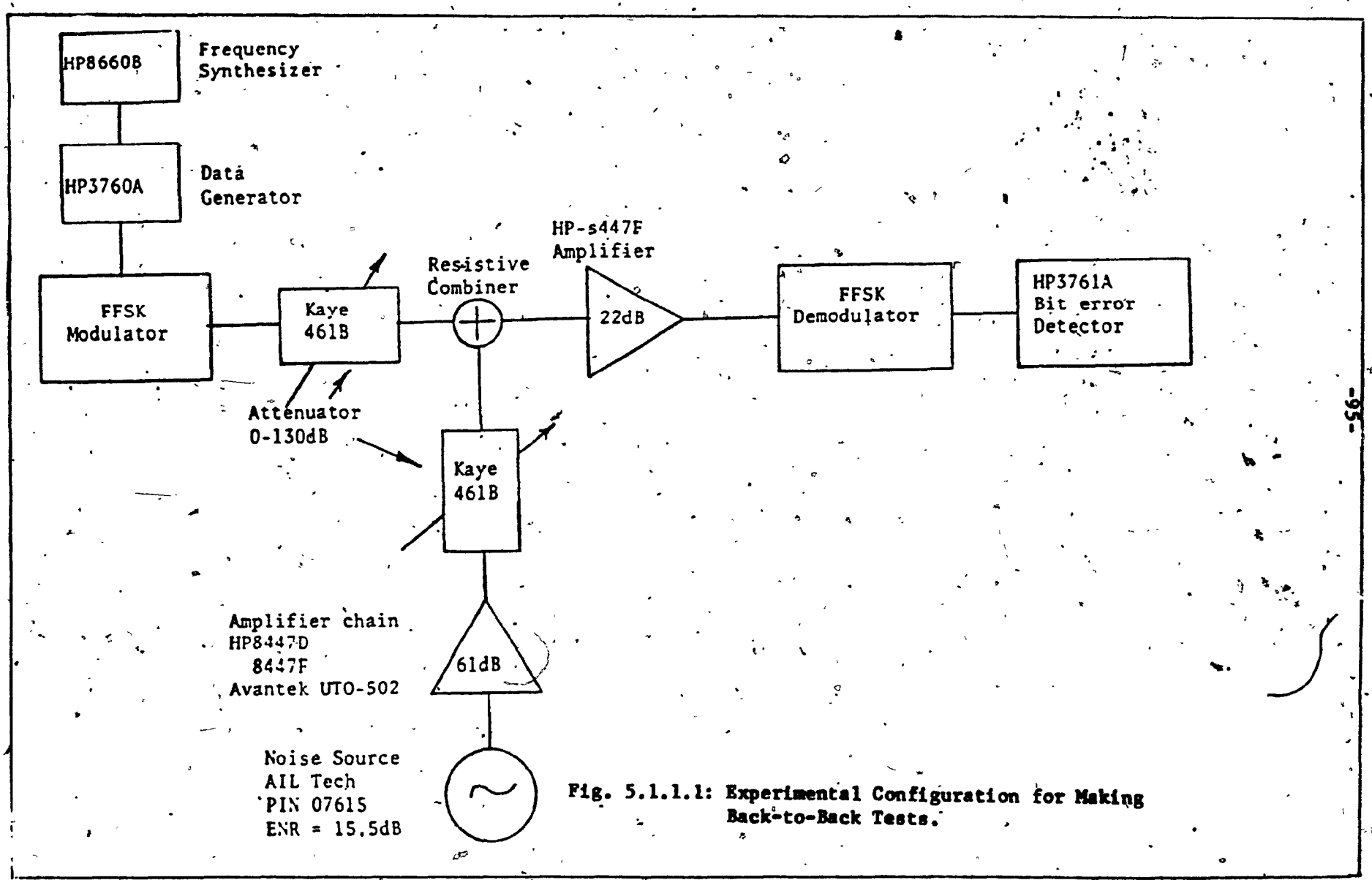


Fig. 5.1.1.1: Experimental Configuration for Making Back-to-Back Tests.

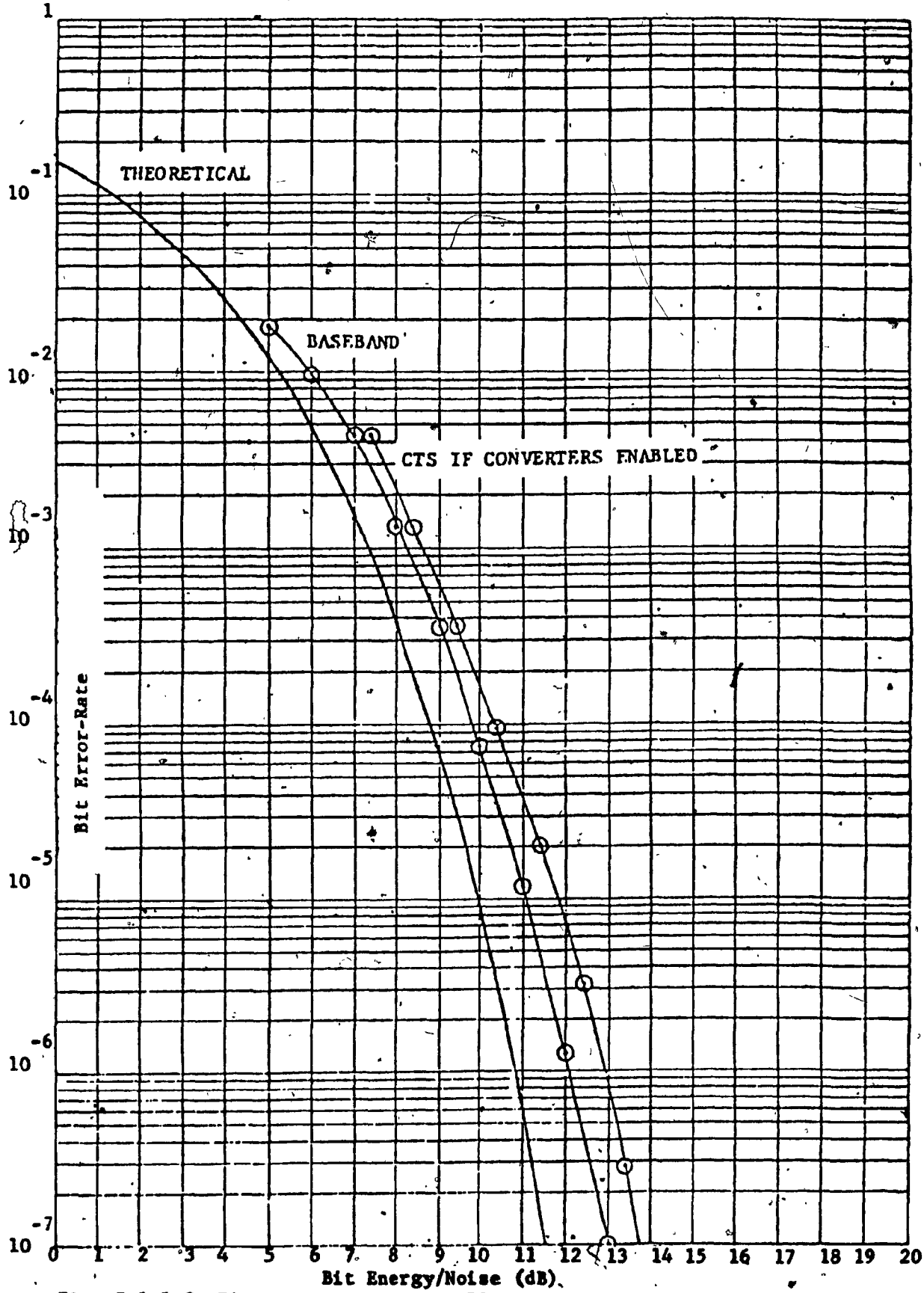


Fig. 5.1.1.2: Bit error-rate versus Bit energy to noise spectral density.

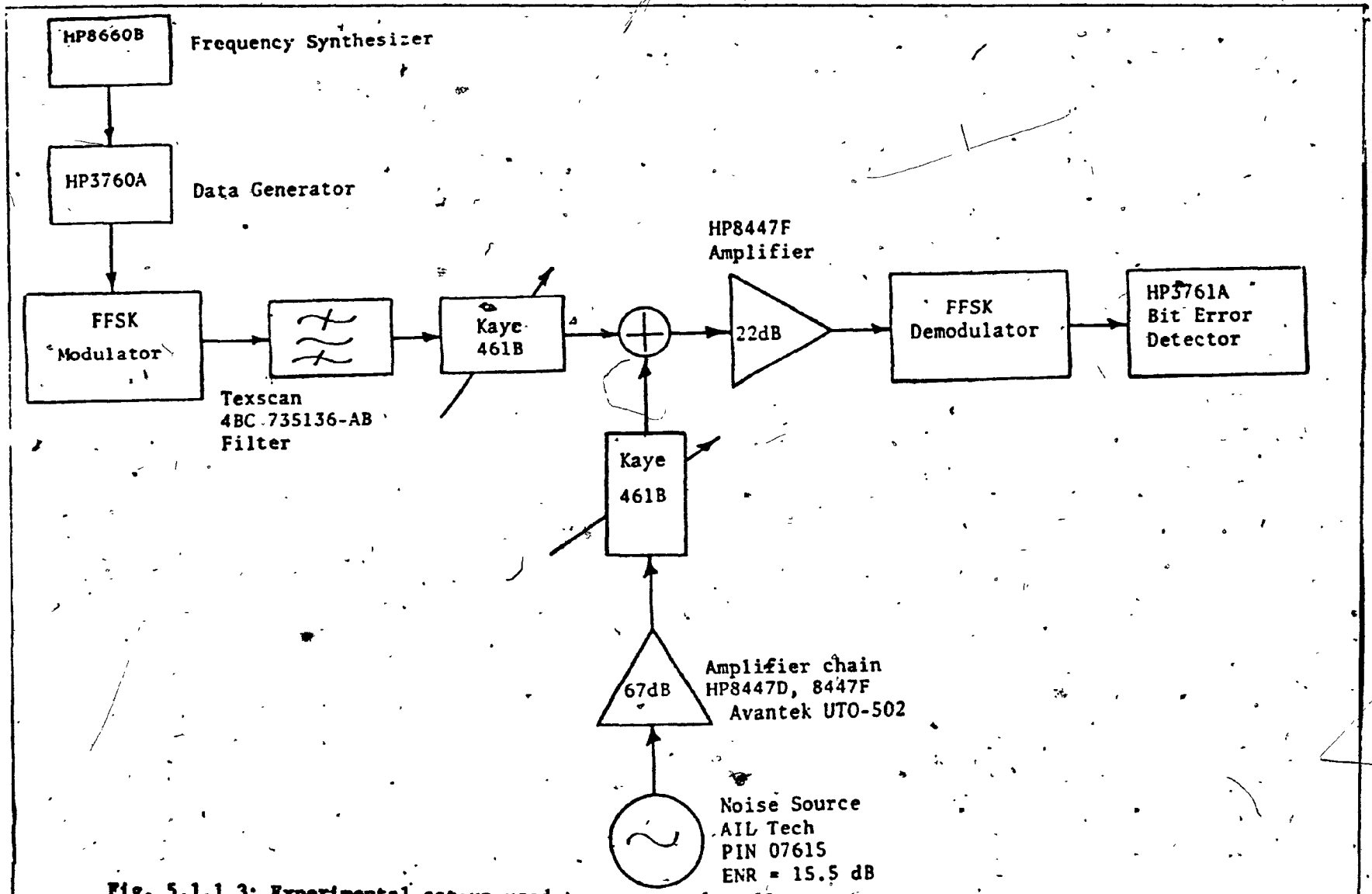


Fig. 5.1.1.3: Experimental set-up used to measure the effect of bandlimiting on error-rate performance.

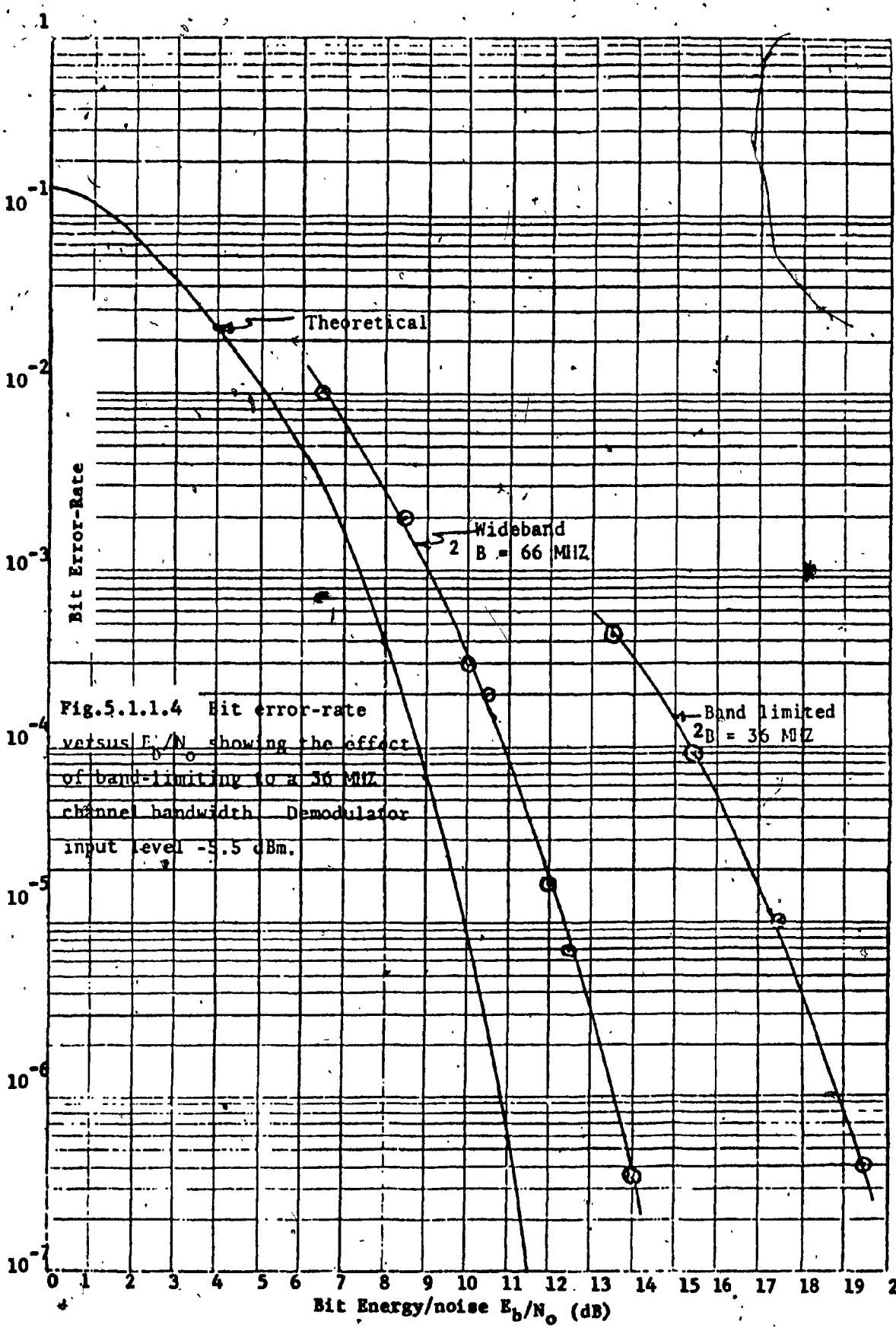


Fig.5.1.1.4 Bit error-rate versus E_b/N_o showing the effect of band-limiting to a 36 MHZ channel bandwidth. Demodulator input level -5.5 dBm.

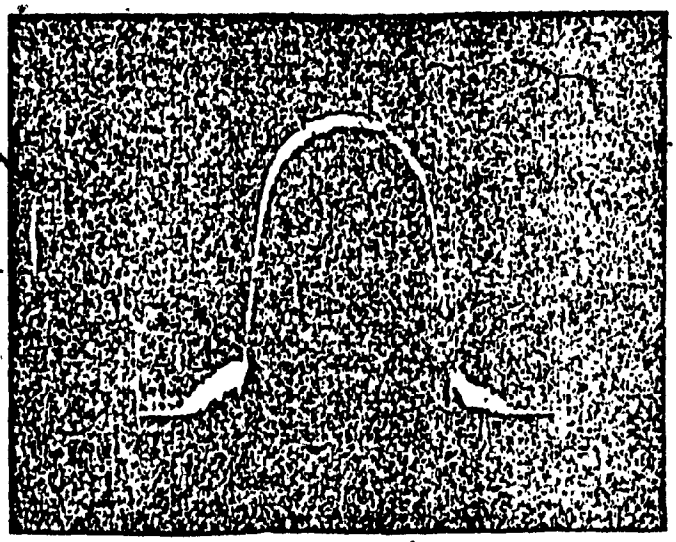


Fig. 5.1.1.5: Received signal spectrum at the demodulator input after bandlimiting to 36 MHz 3 dB bandwith.
Vertical Scale 10 dB/div.
Horizontal Scale 20 MHz/div.
Centre Frequency 735 MHz

level at the receiver input was set to -5.5 dBm. Figure 5.1.1.4 shows the results without this filter (in which case the effective bandlimiting is 66 MHz, thus $2BT = 1.1$) and those with this filter (which reduces the normalized RF bandwidth product to $2BT \cong 0.6$). The performance degradation due to bandlimiting is approximately 5 dB at an error rate of 10^{-5} and 5.4dB at 10^{-6} and it is due to the out-of-band energy and probably due to interference since the available bandwidth is less than the minimum theoretical.

Figure 5.1.1.5 shows the spectrum of the signal as it reaches the receiver when bandlimiting is used at the transmitter output with the 36 MHz filter. Clearly the side lobes are almost completely eliminated. The reason for the significant elimination of the side lobes is their fast roll-off rate.

The differences in the performance curve of Fig. 5.1.1.4 labeled $2B = 66$MHz and that of Fig. 5.1.1.2 labeled Baseband, are due to the different levels at the receiver input (+2.5dB and -5.5dB respectively).

5.1.2 Tests With The Transponder Simulator

The transponder hardware simulation used in these tests is shown in Fig. 5.1.2.1 and consists of an up-converter which converts the FFSK signal centered at 735 MHz into the microwave band centered at 11.265GHz, a TWTA, a microwave filter and a down-converter which

converts the microwave band back to the 735 MHz frequency.

Fig. 5.1.2.2 shows the measured performance with the TWTA operated in saturation with the 36 MHz filter enabled and disabled, and the performance obtained without the transponder and with the 36 MHz filter disabled (this latter curve is that of Fig. 5.1.1.4, labelled 2B=66MHz, obtained from the back-to-back tests). These curves were all obtained with an input level at the receiver of -5.5 dBm and at the nominal bit rate of 60 Mb/s. A comparison between the performance curves of Fig. 5.1.2.2 shows that the effect of saturation is a degradation of 1.3 dB at a bit error rate of 10^{-5} and 1.7 dB at 10^{-6} (by comparing with the back-to-back curve). Also the effect of the bandlimiting, when saturation is in effect, is a degradation of 3.3 dB at 10^{-5} and 3.6 dB at 10^{-6} (when the 36 MHz filter is enabled). Thus the total degradation introduced by the combination of bandlimiting and saturation of the TWTA is about 4.6 dB at 10^{-5} and 5.3 dB at 10^{-6}.

The spectral effect of the bandlimiting/saturation conditions can be seen in figures 5.1.2.3 and 5.1.2.4a/b. These figures show the spectrum at the receiver input when the 36 MHz filter is not used and when this filter is used respectively. The bandlimited spectrum is somewhat narrower (as expected) having a slight side lobe elevation which is negligible. The insignificant elevation of the side lobes is due to the small envelope variations of the FFSK signal.

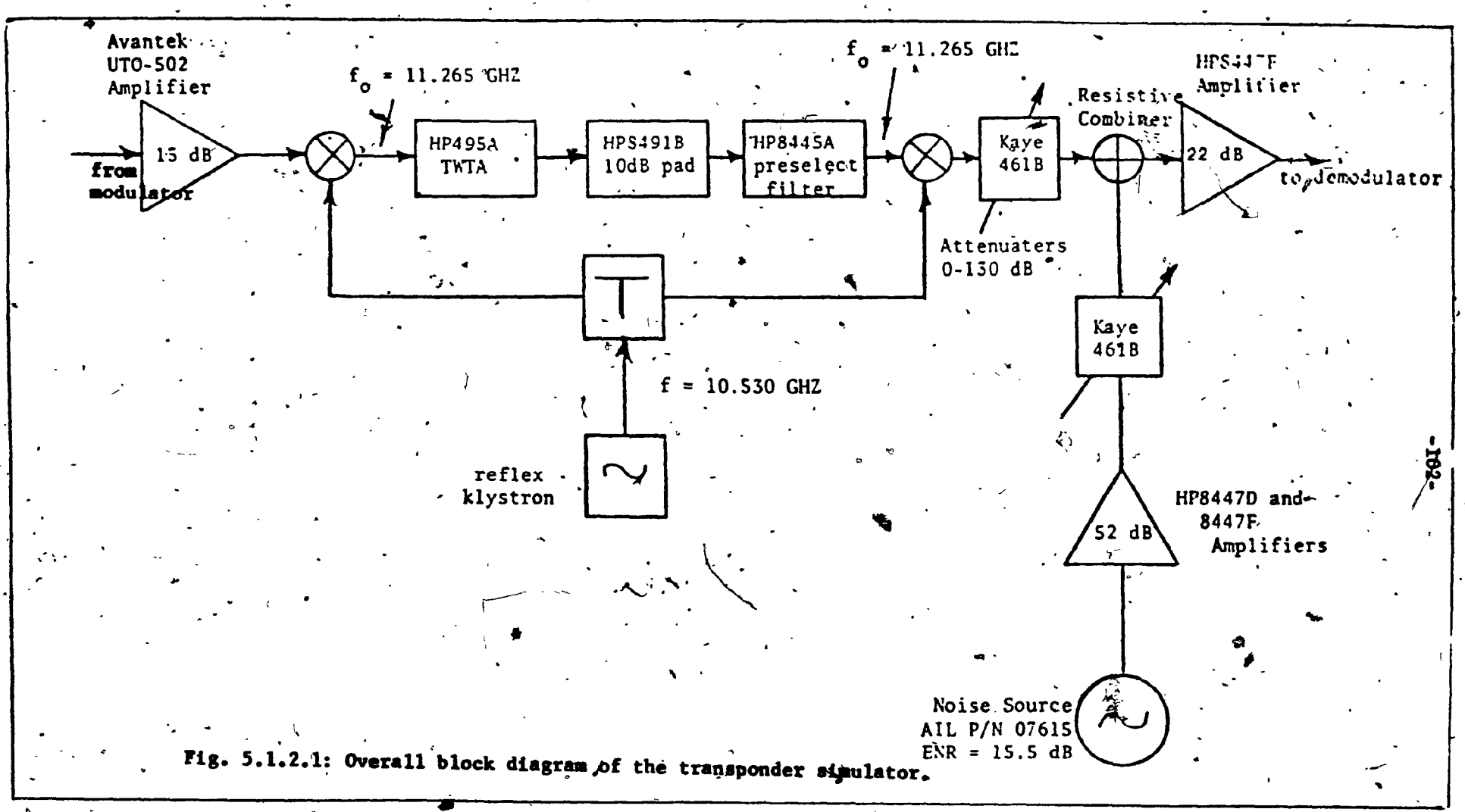


Fig. 5.1.2.1: Overall block diagram of the transponder simulator.

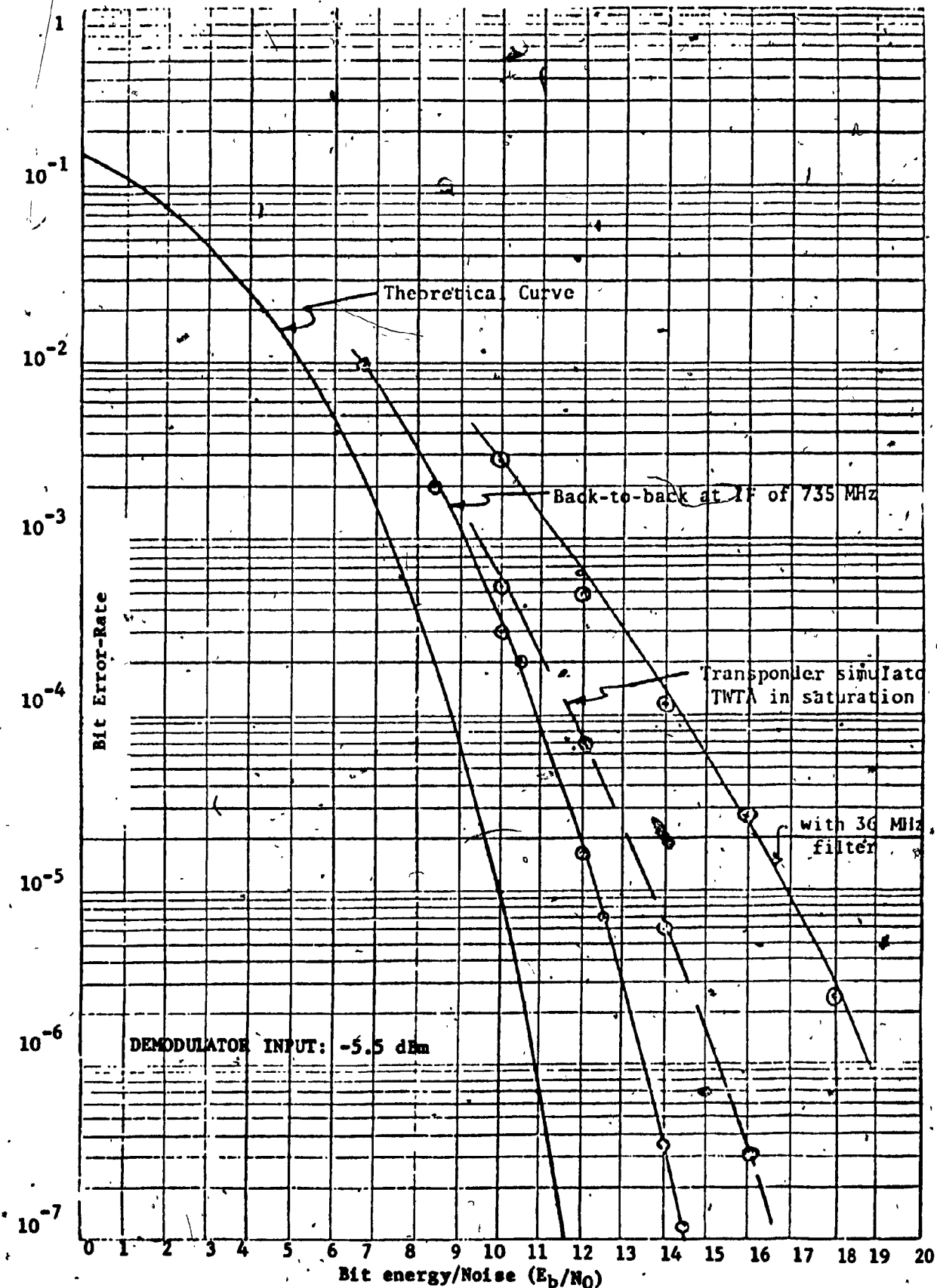


Fig. 5.1.2.2: Bit error-rate versus E_b/N_0 using wideband transponder simulator.

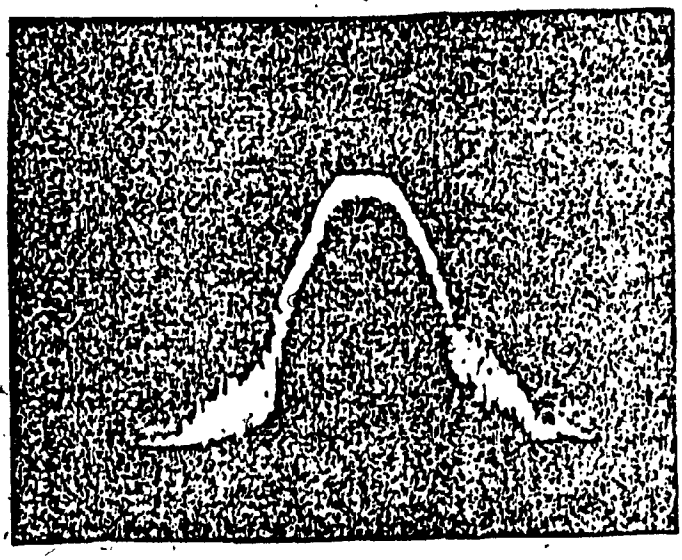


Fig. 5.1.2.3: Signal spectrum at the demodulator input after passage through the transponder simulator in the wideband mode.
Vertical Scale IO dB/div.
Horizontal Scale 20 MHz/div.
Centre frequency is 735 MHz.

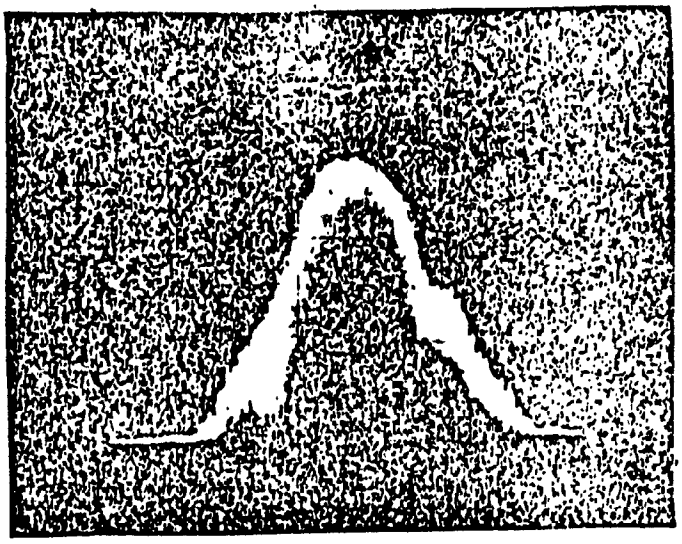


Fig. 5.1.2.4a: Vertical scale 10 dB/div.
Horizontal scale 20 MHz/div.

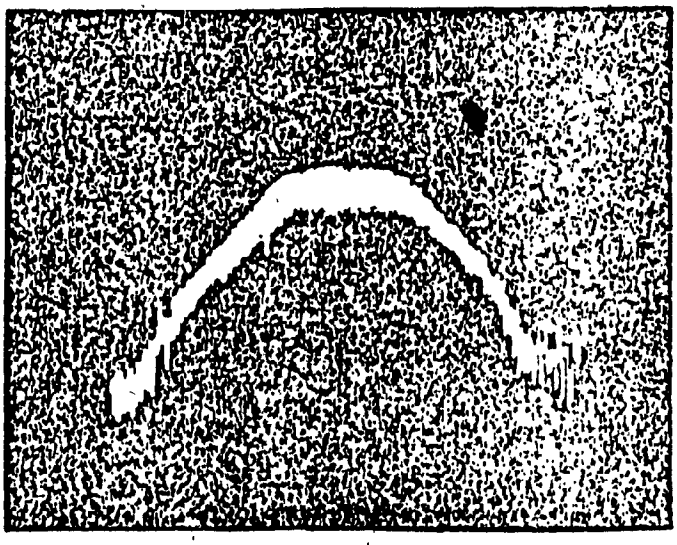


Fig. 5.1.2.4b: Vertical scale 10 dB/div.
Horizontal scale 10 MHz/div.

Fig. 5.1.2.4 ; Signal spectrum at demodulator input after passage through the 36 MHz filter and the transponder simulator. Center frequency is 735 MHz.

Other tests conducted with the TWTA operated in the linear region showed that the performance degradation of the modem is about 0.6 dB compared to that obtained in the back-to-back results at an error rate of 10^{-6} for the same input level to the receiver (-5.5 dBm). This degradation must be the effect of the up-link noise of the transponder.

Summary

The test results of this part show that the sensitivity of the FFSK signal to the non-linearities is very low. The degradation due to saturation of the TWTA was found 1.7 dB at a bit error rate of 10^{-6} compared to the performance obtained in the back-to-back test. A portion of 0.6 dB, which is due to the up-link path noise should be subtracted. Then the net degradation due to saturation comes to 1.1 dB only at the considered bit error rate.

The spectrum of the FFSK signal was found to be very insensitive to the non-linearities.

Both the above slight impairments result from the small envelope variations of the FFSK signal.

5.2: SIMULATION RESULTS FOR DIFFERENT MODULATION TECHNIQUES

The following results were obtained from the simulation of a satellite communication link including a Travelling Wave Tube Amplifier (TWTA), which was driven in its linear and non-linear (saturation) operation region for comparison purposes. The model of such a link is as shown in Fig. 5.2.

The transmit and receive filters in these simulations are 4-pole, 0.5 dB ripple Chebyscheff filters with a fixed 85 MHz 3 dB RF bandwidth (2B=85 MHz).

The effects of up-link noise are not included in these simulations, however, a feeling of these effects can be gained by degrading the down-link SNR by about 1 dB.

5.2.1 Bit Error Rate (BER) Performance Tests

Bit error rate performance tests were conducted [1] for different techniques. The TWTA of Fig. 5.2 was operated both at its linear (12 dB input power back-off) and non-linear or saturated region (1 dB input power back-off).

The BER was found to be degraded with increasing bit rate when the available bandwidth was kept fixed, as it would be expected from the theory. The 3dB RF bandwidth was kept at 85MHz with the transmit and receive filters in Fig. 5.2.

The effect of the increase in the input power to the TWTA was again a BER degradation which is also expected from theory since an AM-PM conversion takes place due to the non-linearities introduced in the TWTA operation and envelope variations are reflected as phase changes in the received signal.

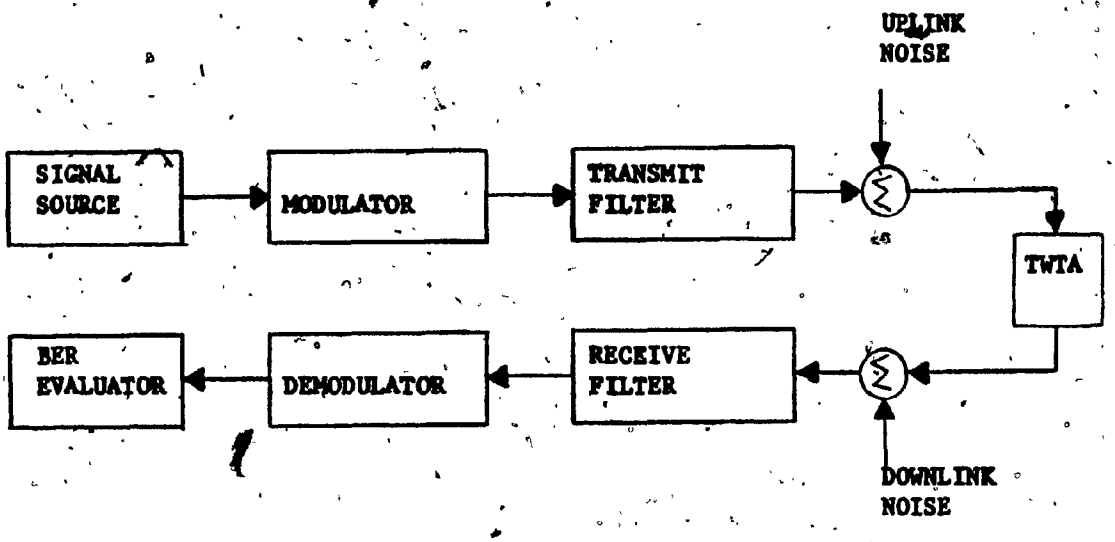


Fig. 5.2: Block diagram of system simulation.

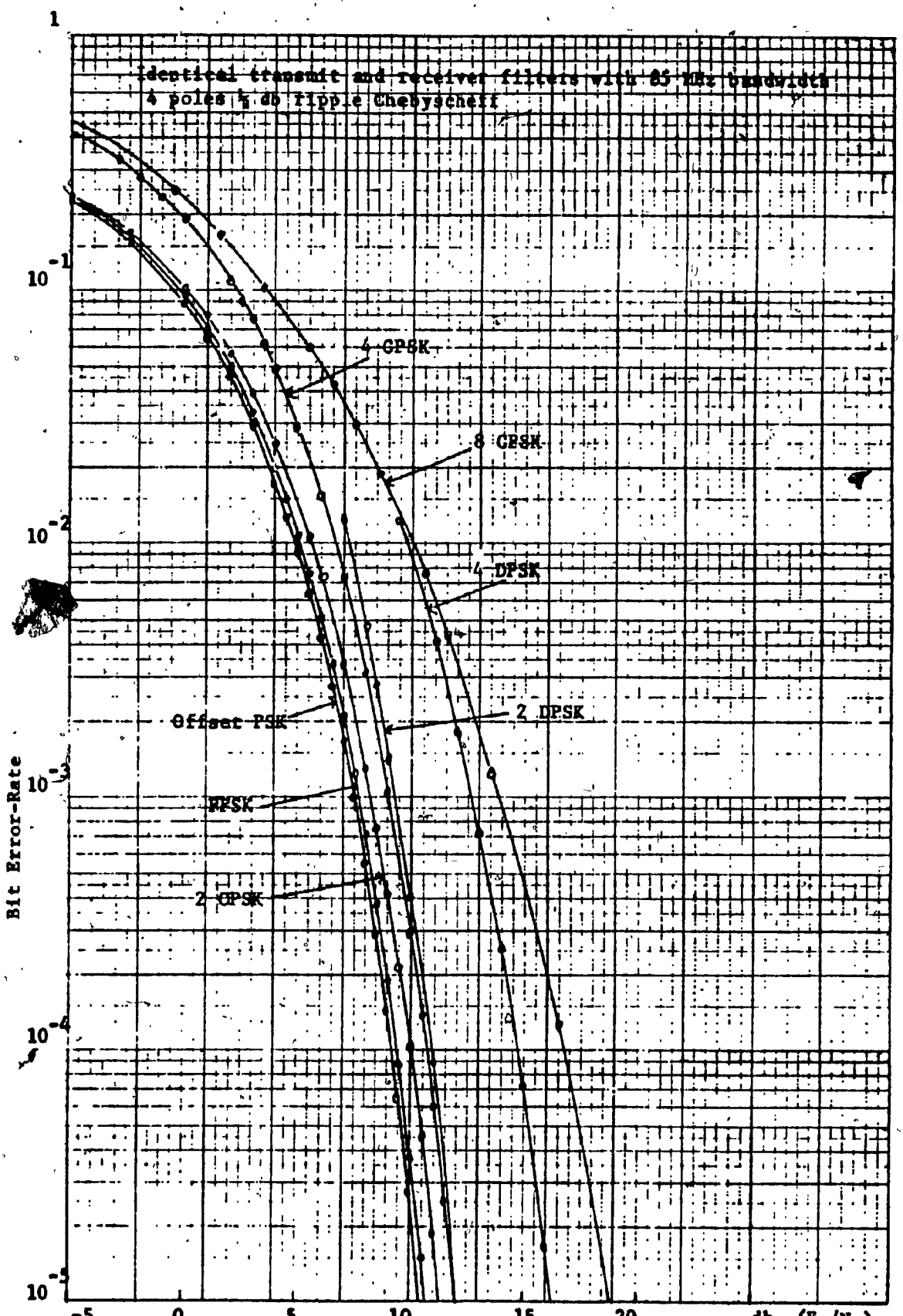


Fig. 5.2.1.1: Comparative error-rates at 65.5 megabauds with a TWT input backoff of 1 dB.

According to these test results the performance of O-QPSK and FFSK was found to be closely insensitive to the TWTA non-linearities. This agrees with the theoretical expectations because the minimum envelope fluctuation occur in the bandlimited signal of these techniques compared to the others. The worst performance with non-linearities was found to be that of APK.

The performances of all coherent PSK techniques along with this of FFSK with the TWTA in its linear operation are closely identical to that of the BPSK (thus the optimum performance). This is expected from the fact that their waveforms are composed of BPSK-type components.

Figure 5.2.1.1 shows the BER performance curves at 65.5Mb/s with the available RF bandwidth of 85MHz (thus $2BT \simeq 1.3$) and with the TWTA operated in its saturated region (1dB power back-off). The APK technique has been excluded due to the overwhelming deterioration of its performance in non-linear channels. The advantage of FFSK and O-QPSK is evident from these curves.

5.2.2 SNR Degradation As A Function Of Available Bandwidth And Bit Rate

A measure of performance of a digital modulation technique is the SNR degradation with decreasing product 2BT for a given BER to be achieved.

In these tests, the available RF bandwidth was kept constant with

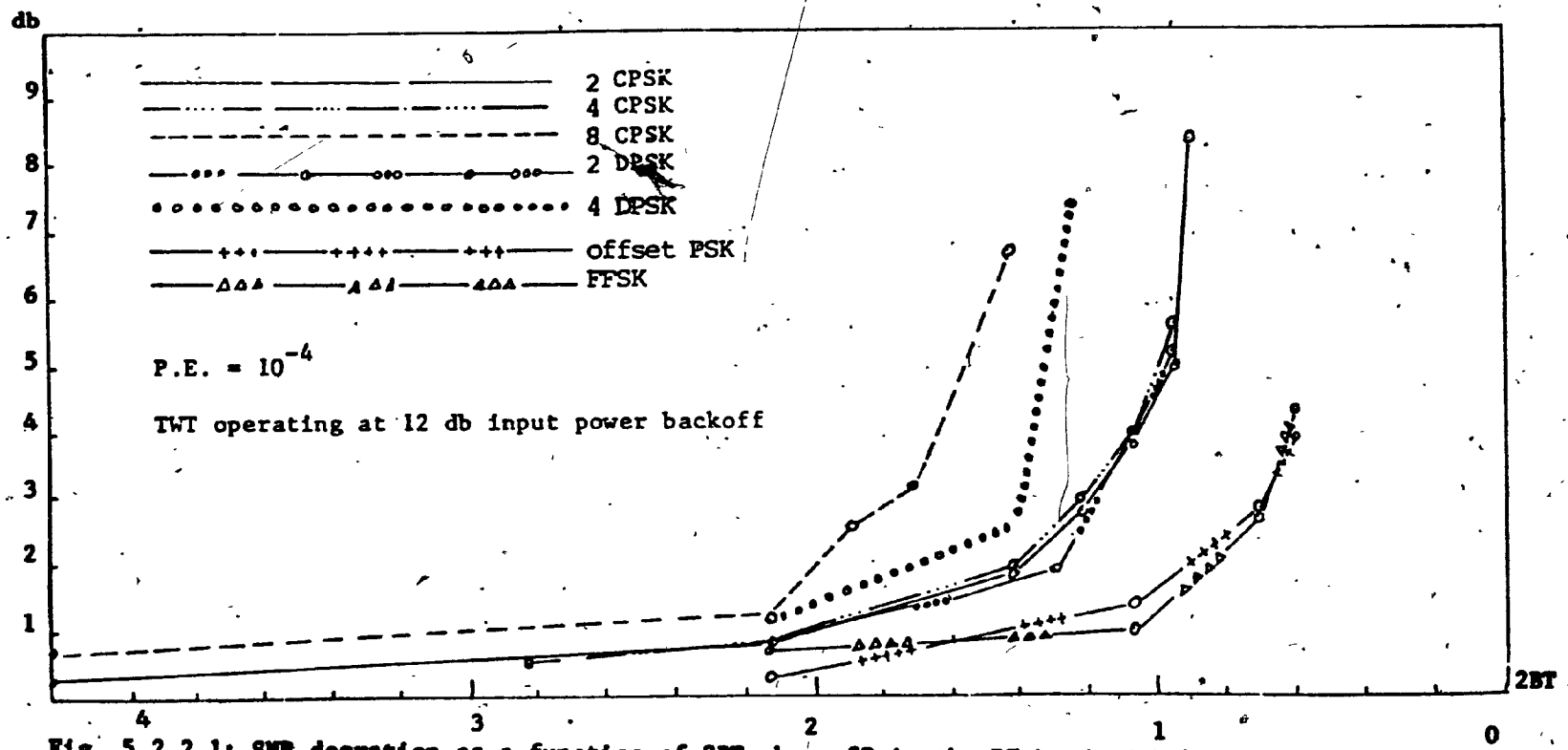


Fig. 5.2.2.1: SNR degration as a function of 2BT where 2B is the RF bandwidth and T is bit period. For these diagrams 2B=85MHz.

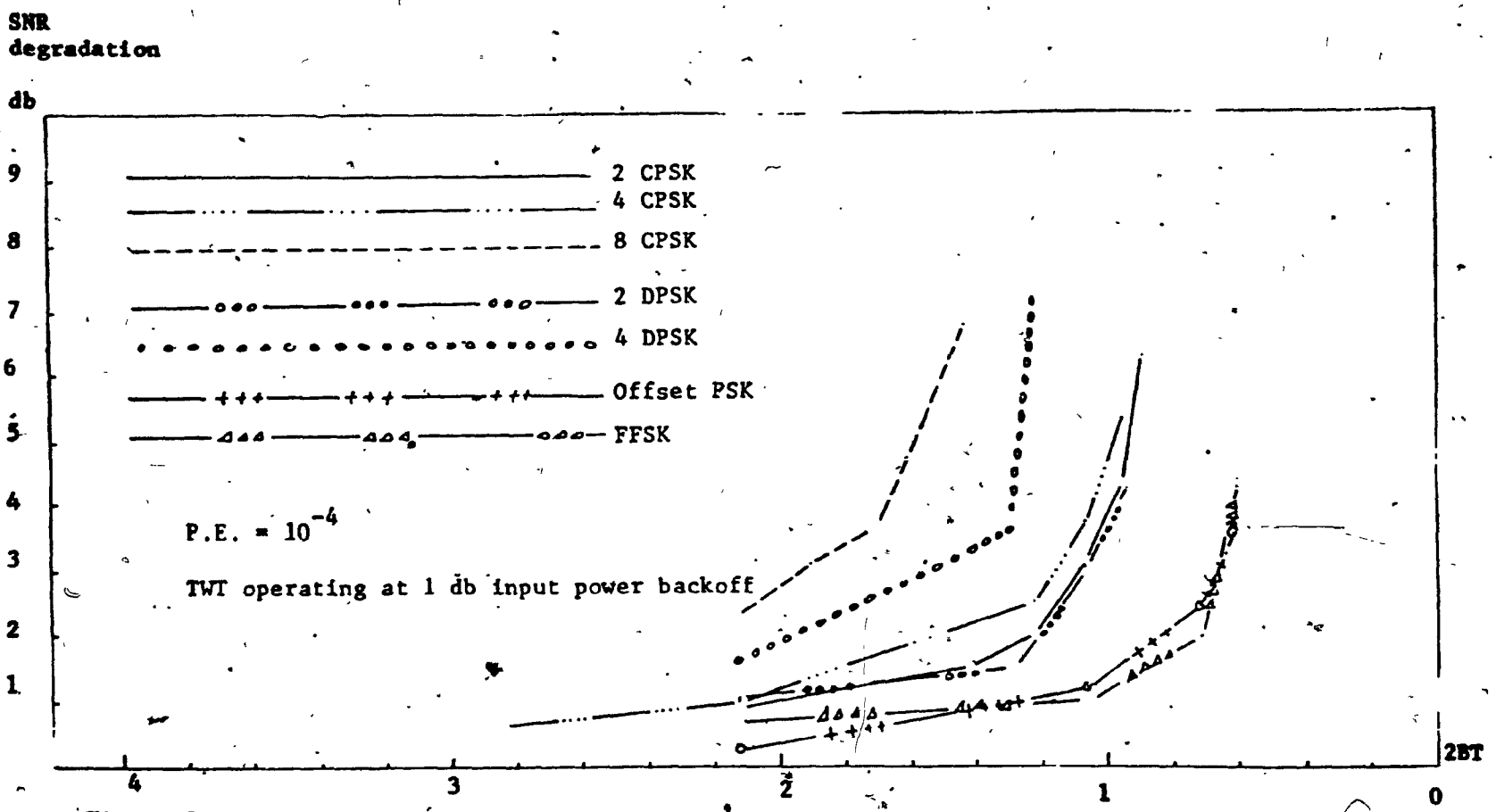


Figure 5.2.2.2: SNR degration as a function of 2BT where 2B is the RF bandwidth and T is bit period. For these diagrams 2B=85MHz.

the fixed 85MHz Chebyscheff transmit and receive filters, and the bit rate was varied. Fig. 5.2.2.1 shows the SNR degradation versus the product 2BT when the TWTA is linearly operated, and Fig. 5.2.2.2 shows the same results when the TWTA is saturated. The superiority of FFSK and O-QPSK over the other techniques is evident in these figures.

Summary

The test results presented at this part show clearly the advantage of FFSK and O-QPSK over all other binary techniques in both linear and non-linear channels.

A comparison between these two techniques shows that O-QPSK has a slight bandwidth efficiency advantage.

In these tests the spectral side lobe regeneration due to the AM-AM effect of the saturated TWTA was not examined. But it is known that the side lobes spectrum of bandlimited O-QPSK signals is much more sensitive to the AM-AM effect compared with FFSK [22], [23].

CHAPTER 6

CONCLUSIONS

In this report, the FFSK modulation technique has been presented analytically and comparisons between this and other techniques (O-QPSK mainly, due to their similarities in nature and performance) have been made, based on both, theoretical and experimental grounds.

It was shown that the FFSK signal, although a member of the FSK family, is of a nature very similar to that of the O-QPSK signal. They both consist of two quadrature carrier components which are binary phase modulated by antipodal time overlapping pulses of rate equal to half the data rate and with a time overlap equal to the data bit period (i.e. half the pulse period). The difference is in the pulse shaping in O-QPSK the pulses are rectangular while in FFSK they are half-cycle sinusoids.

The performances of the two techniques were theoretically found equivalent in linear channels due to the similarities in their nature. The bandwidth efficiencies were found slightly different in favour of O-QPSK due to the different pulse shaping. There exists a cross-over point in available RF bandwidth (in the domain between fb and 1.5fb) where the performances of both techniques are identical.

The FFSK signal is phase continuous as a result of the sinusoidal shape of the modulating pulses, and its envelope when bandlimited is nearly constant. This makes it relatively insensitive

to non-linearities compared to the O-QPSK signal which has a more variable envelope. However, both signals are nearly insensitive to non-linearities when the available RF bandwidth is fairly wide ($2B \geq 1.5fb$). The non-linearities start to be effective in the presence of severe bandlimiting. The cross-over point then shifts to lower values of B because the SNR degradation in bit error performance due to non-linearities is only some parts of a dB in FFSK while some dB's in O-QPSK. The sidelobes regeneration is greater in O-QPSK than in FFSK.

All the above theoretical outcomes were verified in this report through experimental results presentation [1] ,[2] or discussions [22], [23].

Methods of generating FFSK signals were also presented. The conventional method, in which switching between two oscillators (one for the mark, and the other for the space frequency) is used, was shown to be inefficient due to its tendency of loss of coherence with data rate variations. The orthogonal method ensures constant coherence and is the one widely used today.

The detection process is similar to that used in QPSK techniques, thus the use of a parallel pair of correlator receivers or multipliers followed by low-pass filters.

The carrier recovery technique in FFSK is based on squaring the received signal and selecting and dividing the so produced tones to further process them in addition and subtraction operations.

Bit timing recovery is accomplished by multiplying the obtained

mark and space tones and low-pass filtering them.

The data encoding used to obtain the data for the two quadrature components is implemented by means of differential encoding (and sample-hold-operations) which take care of the 180° phase ambiguities of the recovered carrier references. The data decoding is accomplished by simply interleaving the detected data from the two correlator receivers into a single modulo-two (exclusive OR) gate.

From the FFSK signal analysis and its behaviour in bandlimited linear and non-linear channels as compared with other signals, conclusions about the relative significance of the FFSK technique may be drawn. It can be said that it outperforms all binary techniques under any channel conditions except O-QPSK to which it is equivalent in terrestrial communication applications where usually linear channels are used. However, it is superior to O-QPSK in most satellite communication applications where Travelling Wave Tube Amplifiers are widely used.

REFERENCES

1. H. C. Chan, D. P. Taylor and S. S. Haykin "Comparative Evaluation of Digital Modulation Techniques: Simulation Study", CRL Internal Report Series No. CRL-18 Part III, April 1974.

2. D. P. Taylor, G. Horvai and S. S. Haykin "Laboratory Test Program for the CRL 60 Mb/s Fast Frequency Shift Keying Modem", CRL Internal Report Series No. CRL-36, May 1976.

3. Ziemer and Tranter "Principles of Communication, Systems, Modulation and Noise", Houghton Mifflin,1976.

4. S. A. Gronemeyer and A. L. Mc Bride "MSK and Offset QPSK Modulation", IEEE Transactions on Communications, Vol. COM-24, No. 8, P. 809-819, Aug. 1976.

5. S. T. Ogletree, D. P. Taylor and S. S. Haykin "Design and Implementation of a 60Mb/s Fast Frequency-Shift Keying Modem", CRL Internal Report Series, No.CRL-30, March 1975.

6. E. D. Sunde "Ideal Binary Pulse Transmission by AM and FM", The Bell System Technical Journal, pp 1357-1426, Nov. 1959.

7. W. R. Bennett and S. O. Rice "Spectral Density and Autocorrelation Functions Associated with Binary Frequency-Shift Keying", The Bell System Technical Journal, pp. 2355-2385, Sept. 1963.

8. Rubi de Buda "Coherent Demodulation of Frequency-Shift Keying With Low Deviation Ratio", IEEE Transactions on Communications, June 1972, pp 429-435.

9. Rubi de Buda "Fast FSK Signals and Their Demodulation", Canadian Electrical Eng. Journal, Vol. 1, No. 1, pp 591-596, 1976.

10. "Seminar Notes on Digital Comm. & Signal Processing", May 19-21, 1976, Center of Continuing Education and Electrical Engineering, Concordia University.

11. M. Hecht and A. Guida, "Delay Modulation", Proc. IEEE (Lett.) Vol. 57, pp 1314-1316, July 1969.

12. R. C. Titsworth and C. R. Welch, "Power Spectra of Signals Modulated by Random and Pseudorandom Sequences", Jet Propulsion Lab., Tech. Report 32-140, Oct. 10,1961.

13. H. Nyquist, "Certain Topics on Telegraph Transmission Theory", Trans. AIEE, Vol. 47, pp 617-644, April 1928.

14. Taub and Schilling "Principles of Communication Systems", Mc Graw-Hill, 1971.

15. R. W. Lucky, "Techniques for Adaptive Equalization of Digital Communication System " Bell Syst. Tech. J., Vol. 45, pp 255-286, Feb. 1966.

16. D. A. George, D. C. Coll, A. R. Kaye, and R. R. Bowen, "Channel Equalization for Data Transmission", Eng. J., Vol. 53, pp 20-31, May 1970.

17. J. P. Pearce and P. H. Wittke, "Optimum Reception of Digital FM Signals", Dig. 1970 IEEE Symp. Communications, Montreal, Que., Canada, Nov. 12, 13, 1970.

18. M.J. Eric, "Intermodulation Analysis of Non-Linear Devices for Multicarrier Inputs", Communication Research Center Report No. 1234, Ottawa, Nov. 1972.

19. K. Feher "Digital Modulation Techniques in an Interference Environment", EMC Encyclopedia, Vol. IX.

20. M. K. Simon "A Generalization of Minimum-Shift-Keying (MSK) - Type Signaling Based Upon Input Data Symbol Pulse Shaping" IEEE Transactions on Communications, Vol. COM-24, No. 8, pp 845-856, Aug. 1976.

21. Toshio Mizuno, Notihiko Morina, Toshihiko Bamekawa, "Transmission Characteristics of an M-any Coherent PSK Signal via a Cascade of N Bandpass Hard Limiters" IEEE Trans. on Com. Vol. COM-24, No. 5, May 1976, pp 540-545.

22. S. A, Rhodes:"Effects of Hardlimiting on Bandlimited Transmissions with Conventional and Offset QPSK Modulation" IEEE National Telecommunication Conference, Houston, Dec. 1972.

23. H. R. Mathwich, J. F. Balcewicz and M. Hecht, "Effect of Tandem Band and Amplitude Limiting on the Eb/N$_o$ Performance of Minimum (Frequency) Shift Keying (MSK)", IEEE Trans. on Com., Vol. COM-22, No. 10, Oct. 1974, pp 1525-1974.

24. W. R. Bennett and J. J. Salz, "Binary Data Transmission by FM over a Real Channel", Bell System Technical Journal, Vol. 42, pp 2387-2426, Sept. 1963.

25. R. W. Lucky, J. Salz and E. J. Weldon, Jr. "Principles of Data Communications", Mc Graw-Hill Book Company, 1968.

26. R. de Buda "The Fast FSK Modulation System" Conf. Record of the International Conference on Communications, Montreal, June 1971, pp 41-25 to 41-27.

27. M. L. Doelz et al. "Binary Data Transmission Techniques for Linear System", Proc IRE Vol. 45. pp. 656-661, May 1957.

28. W. P. Osborne and M. B. Luntz "Coherent and Non-coherent Detection of CPFSK", IEEE Trans. on Comm. Vol. COM-22, Number 8, Aug. 1974, pp 1023-1036.

29. R. de Buda "About Optimal Properties of Fast Frequency Shift Keying", IEEE Trans. on Comm., Vol. COM-22, pp 1726-1728, Oct. 1974.

30. R. W. Lucky "A Survey of Communications Theory Literature 1968-73", IEEE Trans. on Information Theory, Vol. 17-19, pp 725-739, Nov. 1973.

31. S. Stein and J. J. Jones, "Modern Communication Principles with Application to Digital Signaling", N. Y.:Mc Graw-Hill, 1967, ch. 3.