A DETECTOR FOR FSK SIGNALS USING DIGITAL PHASE-LOCK LOOPS

BY

ROBERT G. YOUNGER, B.Eng.

A DETECTOR FOR FSK SIGNALS USING DIGITAL PHASE-LOCK LOOPS

By

ROBERT GEORGE YOUNGER, B.Eng.

A Thesis

Submitted to the Faculty of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Master of Engineering

McMaster University

(May) 1970

MASTER OF ENGINEERING (1970) McMASTER UNIVERSITY
(Electrical Engineering) Hamilton, Ontario.

TITLE: A Detector For FSK Signals

Using Digital Phase-Lock Loops

AUTHOR: Robert George Younger, B.Eng. (McMaster University)

SUPERVISOR: Professor S.S.Haykim

NUMBER OF PAGES: 140

ABSTRACT:

A new detector for use in a frequency-shift keying communication system is described. Digital phase-lock loops replace correlators in the optimum detector implimentation. Results of a working system, for a range of input signal to noise ratios are presented.

PREFACE

This work has aimed at the continuation of research on the phase-lock loop, a signal tracking system. Attempts have been made to apply digital phase-lock loops as part of a detection scheme for FSK signals. Much of the work of the thesis was of a practical nature. A working model of the detector was constructed and the results of tests performed on the system are presented in the last chapter. To provide a better understanding of the operation of the detector, the first two chapters have been devoted to a thorough discussion of analogue and digital phase-lock loops. Digital phase-lock loops were constructed, and tested to determine how closely a practical loop agreed with theory. The detector configuration is based on that of the so-called optimum coherent detector which is derived and discussed in the third chapter.

ACKNOWLEDGEMENTS

The author would like to express his gratitude to the research director, Dr. S.S. Haykim, for his guidance throughout the duration of this work. A good deal of unclear thinking was redirected by Dr. Haykim and a number of students, too numerous to mention. Their patience with the author was appreciated.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

page

Chapter Four:

Appendix (A):

CHAPTER ONE

THE PHASE-LOCK LOOP

1-1 INTRODUCTION

Since the advent of electrical communications, engineers have directed their attention towards solutions to the optimum detection of signals in additive noise. The detection process generally involved a filter to maximize the signal to noise ratio. During the 1940's D.O. North,[1] Van Vleck and Middleton[2] independently developed a theory of filters now called "matched filters".[3] The basic result in the theory presented was that in the case of white additive noise, the signal-to-noise ratio is maximized by a filter whose impulse response has the form of the image of the signal to be detected. In radar detection, banks of matched filters were used, with each filter designed for a different delay and Doppler shift of the specific transmitted signal. The output of the filters were monitored to determine the most likely received message.

In 1950, Lee, Cheathem and Wiesner[4] described a system of detection of a periodic signal buried in additive noise by an electronic correlator as in figure (1.1). In the publication, they drew the parallel between filtering in the frequency domain with matched filters and filtering in the time domain with correlators. The application of the correlators for signal detection presupposes that at the receiver, the frequency and phase of the transmitted signal are known. For a system such as radar in which the transmitter and receiver are located at the same geographical location, a system as illustrated in figure (1.2) might be used. However, if the transmitter and receiver are located great distances apart, it is unlikely that a suitable reference signal will be readily available for correlation. The problem is further complicated if the transmitter is in motion relative to the receiver, giving rise to a Doppler shift in the frequency of the received carrier. In Chapter 3, we introduce the optimum detector for the reception of digital information, and as one might expect, the matched filter or correlator provides a basis for detection. Again, the difficulty in the implimentation of the detector, is the availability of a suitable reference signal. In this Chapter, we will discuss a device called a phase-lock loop that can be used for tracking the phase of the carrier component of the received signal. This device thus generates a signal suitable for synchronous detection.

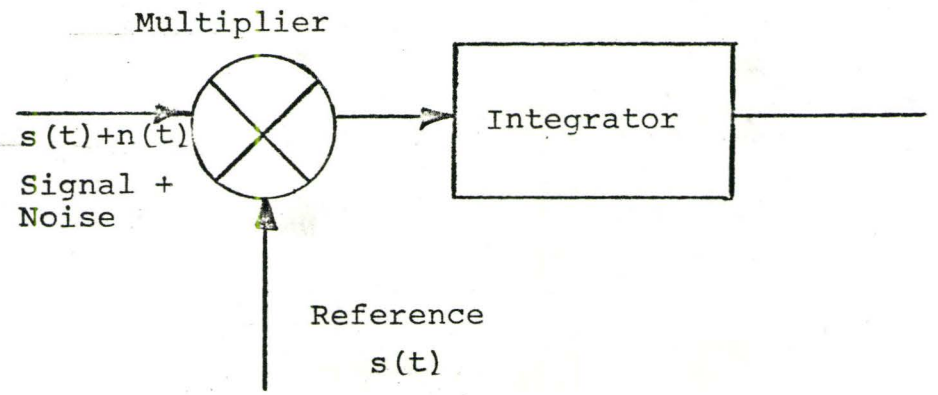


fig.1.1 Correlator

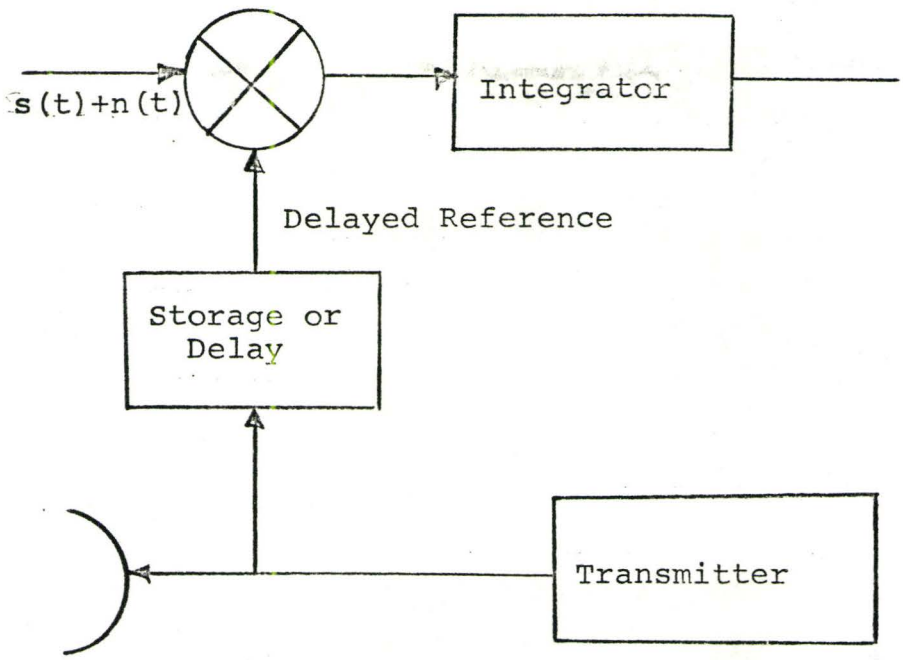


fig.1.2 Correlator For Radar System

1-2 NATURE OF THE PHASE-LOCK LOOP

The phase-lock loop contains three basic components: a phase detector or multiplier; a voltage controlled oscillator, and a low pass filter, as in figure (1.3).

The multiplier compares the phase of a periodic input signal with the phase of the voltage controlled oscillator output. The output of the multiplier is a measure of the phase difference between the two signals. This output is applied to a low pass filter in the loop and then applied to the voltage controlled oscillator. This control voltage changes the frequency of the oscillator in a direction which reduces the phase difference between the two signals.

When the loop is "locked", the control voltage is such that the frequency of the VCO is equal to the average frequency of the input signal. For each cycle of input; there is one, and only one cycle of oscillator output.

Suppose that the incoming signal carries information in its phase or frequency; this signal will inevitably be corrupted by additive noises. The function of the phase-lock loop is to reproduce the original signal while removing as much of the additive noise as possible.

By beating the incoming signal with the output of the local oscillator and applying a filtered version of the beat signal to the voltage controlled oscillator, much of

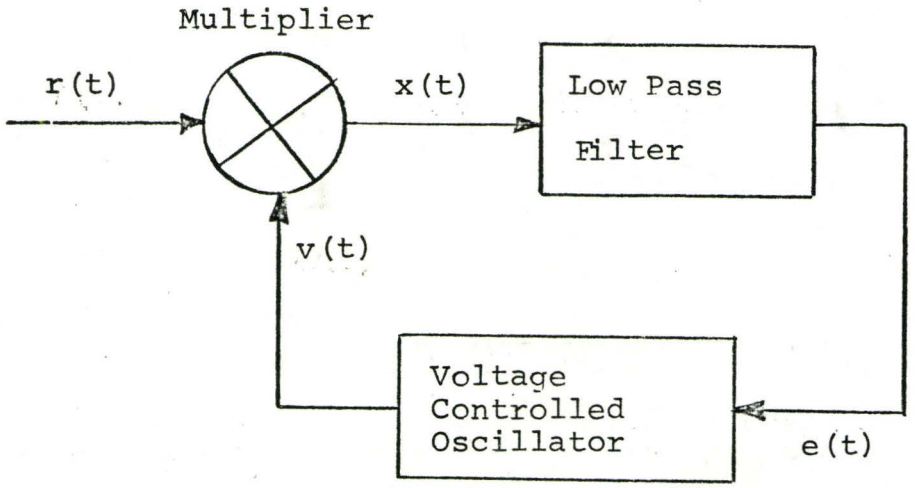


fig.1.3 Phase-Lock Loop

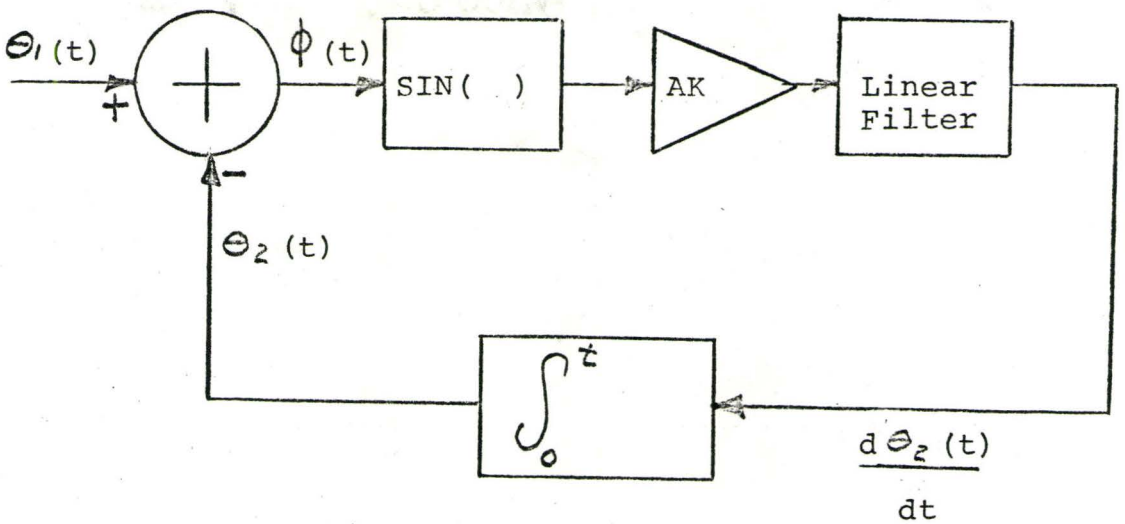


fig.1.4 Phase Model of Phase-Locked Loop

the noise can be suppressed. Since the input to the loop is a noisy signal, whereas the output is a cleaned up version of the input; it is reasonable to consider the loop as a kind of filter that passes signals and rejects noise.

Two important characteristics of such a filter are that the bandwidth can be made very small and the filter automatically tracks the signal frequency. These two features; automatic tracking, and narrow bandwidth, account for the major uses of the phase-lock loop.

1-3 HISTORY OF THE PHASE-LOCK LOOP

The first description* of the phase-lock loop was published by, de Bellescize[5] in 1932. This paper dealt with the synchronous reception of radio signals (homodyne receiver). For the receiver to operate, the local oscillator has to be adjusted to the frequency of the incoming signal, which is then converted to an intermediate frequency of zero Hertz. The output of the mixer contains the demodulated information.

* The existence of the original paper was brought to light by T.J. Rey.[6]

Since any interference is not synchronous with the local oscillator, the output of the mixer, due to the interfering signal, can be suppressed by audio filtering. For correct operation, the oscillator output must have the same frequency as the input; and the phase difference must be minimal. In other words, the local oscillator must be phase-locked to the incoming signal. The superheterodyne receiver however, became more popular with radio engineers and very little use has been made of the homodyne receiver.

The first widespread use of the phase-lock loop was in the synchronization of the horizontal and vertical scan in television receivers[7]. The construction of the scan raster on the television tube is basically as follows: the synchronization pulses are stripped off the received signal and individually used to trigger a pair of oscillators which in turn drive a pair of sweep generators. This scheme will work in an acceptable fashion in the absence of noise. Of course, noise is always present, and it is possible that a noise spike may be confused with a "sync" pulse causing incorrect triggering of an oscillator and therefore, a sweep generator. The noise problem may be eliminated by phaselocking the two oscillators to the "sync" pulses. The oscillators in this approach are not triggered by each pulse and because the phase-lock loop looks at many pulses, it is unlikely that it will be perturbed by the occassional noise spike. The so-called "fly wheel synchronizers" used in present-day tele-

vision receivers are really just phase-lock loops.

The analysis and application of phase-lock loops began in earnest with the launching of the first artificial satellites. These vehicles carried low power transmitters; the received signals being correspondingly weak. The problem of receiver design was further complicated by a Doppler shift, and a drift of the transmitting oscillator, which caused considerable uncertainty about the exact frequency of the received signal. At the 108 MHz frequency originally used, the Doppler shift could range over $\pm$ 3KHz. Thus, with ordinary fixed tuned receivers, the bandwidth would have to be in the order of 6KHz. The signal itself however, occupied a bandwidth of only 6Hz. The noise penalty suffered would be 1000 times (30dB). As technology has progressed, the numbers involved have become even more dramatic. The noise penalties involved are intolerable with the fixed tuned receivers and that is why narrow-band, phase-locked tracking receivers are used in space communications.

1-4 ANALYSIS OF PHASE-LOCK LOOP

Referring again to figure (1.3) let the received signal be denoted by;

$$r(t) = \sqrt{2} A \sin \theta(t) \tag{1.1}$$

and the output of the VCO by

$$v(t) = \sqrt{2} K_1 \cos \theta'(t) \tag{1.2}$$

where A and K, are root-mean-square amplitudes. Assume, that if the error voltage e(t) is removed, the VCO oscillates at a quiescent frequency of ω_o radians per second. When the control signal is applied, the VCO frequency becomes $W_o + K_2 e(t)$ radians per second, where K_2 is the VCO proportionality constant. Thus, we can write:

$$\frac{d\theta'(t)}{dt} = \omega_o + K_2 e(t) \tag{1.3}$$

The output of the multiplier or phase detector is the product of (1.2) and (1.1), obtaining,

$$x(t) = A K_1 \left\{ \sin[\theta(t) - \theta'(t)] + \sin[\theta(t) + \theta'(t)] \right\} \tag{1.4}$$

Since the linear filter in figure (1.3) takes the form of a low pass filter, the sum frequency term of (1.4) may be discarded.

Applying the convolution integral to the linear time-invariant filter, its output may be written as:

$$e(t) = e_o(t) + \int_o^t x(t-u) f(u)\, du = e_o(t) + \int_o^t x(u) f(t-u)\, du \tag{1.5}$$

where we have assumed that the input is applied at t=0 and $e_o(t)$ is the zero input response of the filter. Generally, the initial conditions of the filter are set to zero, so that $e_o(t)=0$ for all times. The weighting function is, of course, the impulse response of the filter. We can now use equations (1.3), (1.4), and (1.5) and write:

$$\frac{d\theta'(t)}{dt} = \omega_o + K_2 \int_o^t f(t-u) A K_1 \sin[\theta(u)-\theta'(u)]\, du \qquad (1.6)$$

Now defining the phase error,

$$\phi(t) = \theta(t) - \theta'(t) \qquad (1.7)$$

and the loop gain,

$$K = K_1 K_2 \qquad (1.8)$$

we have:

$$\frac{d\phi(t)}{dt} = \frac{d\theta(t)}{dt} - \omega_o - AK \int_o^t f(t-u) \sin\phi(u)\, du \qquad (1.9)$$

For any input phase $\theta(t)$ the solution $\phi(t)$ to this integro-differential equation describes exactly, the operation of the phase-locked loop. Since the quiescent frequency of the VCO serves only as a reference, we can, for convenience eliminate it from the analysis by writing:

$$\theta_1(t) = \theta(t) - \omega_0 t \tag{1.10}$$

$$\theta_2(t) = \theta'(t) - \omega_0 t \tag{1.11}$$

Thus, from equation (1.9) we have:

$$\frac{d\phi(t)}{dt} = \frac{d\theta_1(t)}{dt} - AK\int_0^t f(t-u)\sin\phi(u)\,du \tag{1.12}$$

The equation (1.12) suggests the model illustrated in figure (1.4).

We can see that the multiplier or phase detector has been replaced by a summing point and a sinusoidal nonlinearity. The replacement of the VCO by an integrator is valid since the phase of the VCO output signal is proportional to the integral of the control signal. It should also be noted that the gain of the loop is augmented by the root-mean-square amplitude of the input signal. This often leads to the use of a hard limiter at the input to a phase-lock loop implimentation, in order to maintain the loop gain constant.

1-5 LINEAR MODEL OF THE LOOP

It is apparent from the preceeding discussions and diagrams that there must at all times be a non-zero control signal; otherwise, the VCO will oscillate at its quiescent frequency. This further implies that the loop must operate with a finite phase error even when it is in a locked state; that is, when the frequency of the VCO is equal to the average frequency of the input signal. However, when the phase difference $\phi(t)$ is small (less than thirty degrees) we may use the approximation;

$$\sin \phi(t) \doteq \phi(t) \tag{1.13}$$

and the sinusoidal nonlinearity may be removed from the model. The operation of the loop can now be described by the linear differential equation:

$$\frac{d\phi(t)}{dt} = \frac{d\theta_1(t)}{dt} - AK\int_0^t f(t-u)\,\phi(u)\,du \tag{1.14}$$

By using Laplace transforms, equation (1.14) may be transformed to;

$$s\tilde{\phi}(s) + AKF(s)\tilde{\phi}(s) = s\tilde{\theta}_1(s) \tag{1.15}$$

where F(s) is the transfer function of the linear filter and $\tilde{\phi}(s)$ and $\tilde{\theta}_1(s)$ are the Laplace transforms of $\phi(t)$ and $\theta_1(t)$ respectively. Equation (1.15) can be represented by the

block diagram illustrated in figure(1.5). The basic loop equations may be written as;

$$\frac{\tilde{\phi}(s)}{\tilde{\theta}_1(s)} = \frac{1}{1+AKF(s)/s} \tag{1.16}$$

and

$$\frac{\tilde{\theta}_2(s)}{\tilde{\theta}_1(s)} = H(s) = \frac{AKF(s)/s}{1+AKF(s)/s} \tag{1.17}$$

where H(s) is known as the closed-loop transfer function. Much of the analysis done on phase-lock loops has been based on the use of this linear model[8].

1-6 LOOP FILTERS

It is apparent from the linear model that the performance of the loop is very much dependent upon the choice of filter function F(s). In general, the order of a control system is equal to the number of finite poles in the open-loop transfer function, which in the case of the phase-lock loop is the number of poles of AKF(s)/s.

Thus, to impliment a first order phase-lock loop, we require that F(s)=1, which means that we require no filter within the loop. The inclusion of a first order filter results in a second order loop and so on.

To demonstrate the performance of the loop with different filters, let us consider the response of the loop

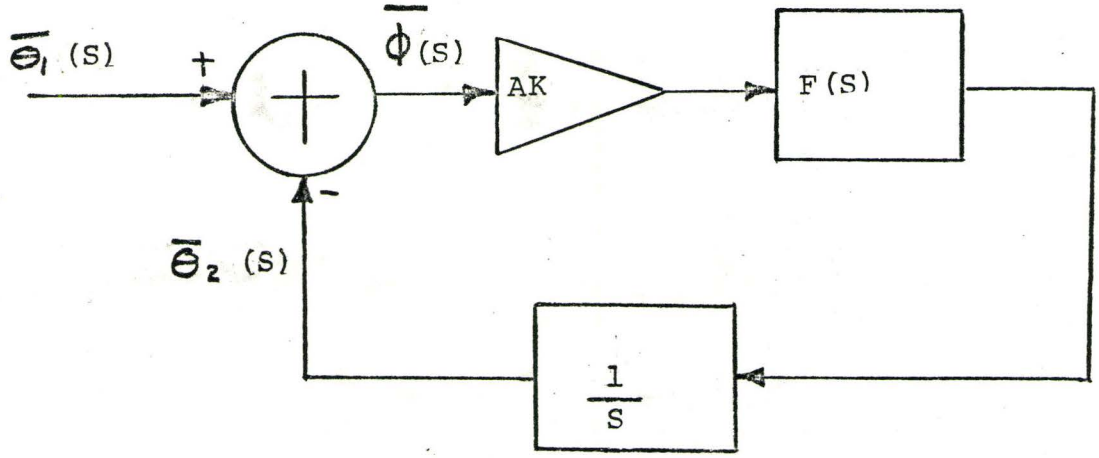


fig.1.5 Linear Model of Loop

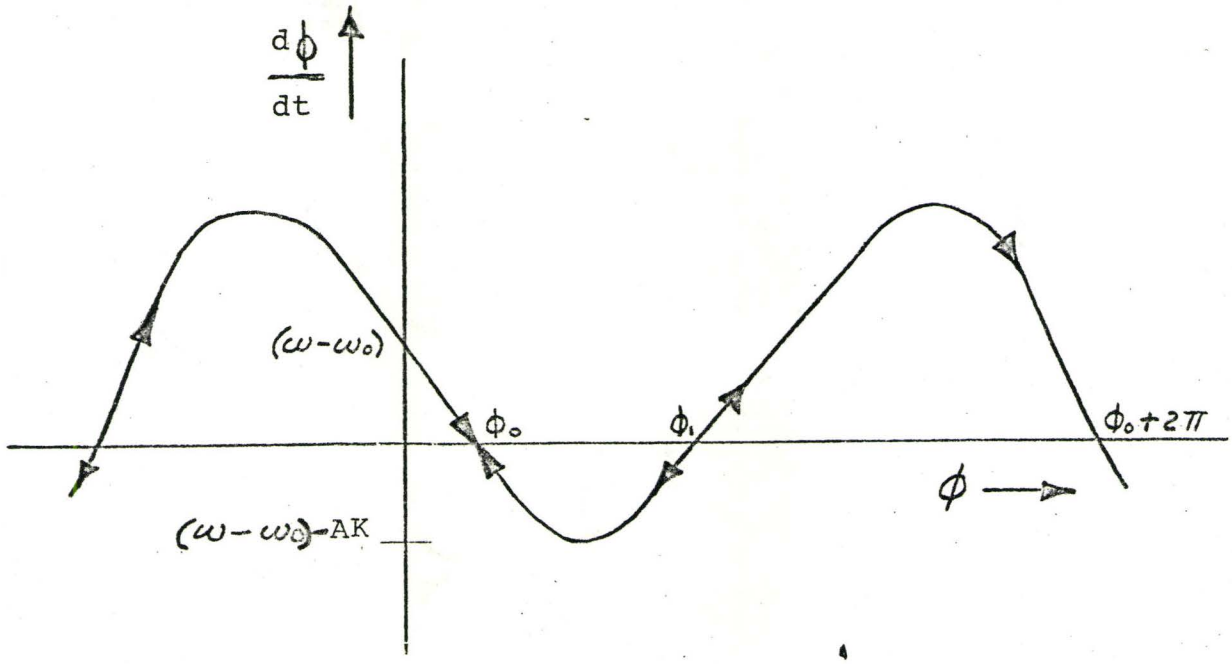


fig.1.6 Phase-Plane Trajectory of 1st Order Loop

to a constant offset in frequency. The phase angle of the received signal can then be written as $\omega_1 t + \theta$ and we can write;

$$\theta_1(t) = u(t)[(\omega_1-\omega_0)t+\theta] \tag{1.18}$$

where;

$$u(t) = 1 \quad , \quad t \geq 0$$
$$= 0 \quad , \quad t < 0$$

and

$$\tilde{\theta}_1(s) = \frac{\omega_1-\omega_0}{s^2} + \frac{\theta}{s} \tag{1.19}$$

Combining equations 1.16 and 1.19 we have;

$$\tilde{\phi}(s) = \frac{(\omega_1-\omega_0)+\theta s}{s(s+AKF(s))} \tag{1.20}$$

Let us now consider several choices of filters and investigate the behaviour of $\phi(t)$. We are not interested in finding an explicit expression for $\phi(t)$ for all times, because, as pointed out earlier, the linear model gives an accurate description of the loop only when the phase error is small. We will thus confine ourselves to finding under what conditions and for what choices of filters the asymptotic value of error, $\lim_{t\to\infty} \phi(t)$ becomes small. If we consider first, the loop with no filter, equation 1.20 becomes,

$$\tilde{\phi}(s) = \frac{(\omega_1-\omega_0)+\theta s}{s(s+AK)} \tag{1.21}$$

and applying the final value theorem:

$$\lim_{t\to\infty} \phi(t) = \frac{\omega_1 - \omega_0}{AK} \tag{1.22}$$

Thus, only if the initial frequency offset is sufficiently small, will the loop with no filter settle to a satisfactory steady state error. If we consider a filter with an integrator, that is,

$$F(s) = \frac{s+a}{s} \tag{1.23}$$

we find that the steady state phase error is:

$$\lim_{t\to\infty} \phi(t) = 0 \tag{1.24}$$

A system with an integrating filter can thus asymptotically track a constant frequency offset with zero asymptotic error.

In reality, a perfect integrator is impossible to realize. With actual components, we might build a filter which approximated the integrator, with a transfer function of the form:

$$F(s) = \frac{s+a}{s+\epsilon} \tag{1.25}$$

The final value theorem predicts that the steady state phase error for this filter is:

$$\lim_{t\to\infty} \phi(t) = \frac{(\omega_1 - \omega_0)\epsilon}{AKa} \tag{1.26}$$

Thus, although the steady state phase error is not zero, it

can be made to approach zero by appropriately choosing α, ϵ and making the gain of the loop large in comparison to the offset frequency.

Similar investigations can be made for other input phase conditions and the best filter transfer functions can be chosen. So called, optimum filters, for the linear model, have been derived, based on the Wiener[9] criterion.

1-7 HOLD IN PERFORMANCE

The preceding material on tracking and phase error, is based on the assumption that the error is sufficiently small; thus allowing the loop to be considered linear in its operation. This assumption becomes progressively worse as the error increases until finally the loop drops out of lock and the assumption no longer holds. In the previous section, we showed using the final value theorem that the linear approximation of phase error due to a frequency offset is:

$$\phi = \frac{\Delta\omega}{AKF(0)}$$

However, for a sinusoidal phase detector, the true expression[10] should be:

$$\sin\phi = \frac{\Delta\omega}{AKF(0)} \qquad (1.27)$$

The sine function cannot exceed unit magnitude, therefore, if $\Delta\omega > AKF(o)$ there is no solution to this equation. Instead, the loop falls out of lock and the phase-detector voltage becomes a beat-note rather than a dc level. The hold in range of a loop therefore may be defined as:

$$\Delta\omega_H = \pm AKF(o) \tag{1.28}$$

Equation 1.28 implies that the hold-in range can be made arbitrarily large, simply by using very high loop gain. Of course, this cannot be entirely correct because some other component in the loop will then saturate before the phase detector; thus, equation (1.28) applies only theoretically.

Assuming the loop is locked onto a frequency and suddenly a step change in input frequency takes place, one might ask, can the transient error pull the loop out of lock, even if the static error is within the hold in range? At worse, the loop will unlock, skip cycles for a while and then lock up once again. There is some frequency step limit below which the loop does not skip cycles, but remains in lock. This limit is often referred to as the "pull-out frequency". Viterbi[11] has performed analogue computer simulations of various loops and from his results, estimates of "pull-out frequency" can be made.

1-8 ACQUISITION

In previous sections it has been tacitly assumed that the loop was initially in lock. In fact, a loop is initially in an unlocked state and must go through an acquisition period. Experimental evidence indicates that there are a number of ways by which lock may be acquired; each depending on the loop parameters and input conditions. If for some reason, the frequency difference between input and VCO is less than the loop bandwidth, the loop will lock up almost instantaneously without skipping cycles. There are some loop types in which the VCO frequency will slowly walk in toward the input frequency, despite the fact that the initial frequency difference may greatly exceed the loop bandwidth. The maximum difference frequency for which the loop eventually comes into lock is called the "pull in frequency", and its value for any particular loop can only be estimated from experiment or computer simulation.

It is instructive to consider the acquisition of lock for a first order loop. Equation (1.12) becomes:

$$\frac{d\phi(t)}{dt} = \frac{d\theta_1(t)}{dt} - AK \sin\phi(t) \qquad (1.29)$$

The system trajectory can be plotted, according to equation (1.29), as in figure (1.6). We have assumed that the input to the loop is a constant offset in frequency, that is;

$\theta_1(t) = (\omega - \omega_0)t + \theta$

We know that when the loop is locked, $\frac{d\phi}{dt}$ is zero and the phase error ϕ has reached its steady state value. If for the initial value of phase error, the derivative is positive, equation (1.29) indicates that ϕ will increase as a function of time. In fact, the system follows the trajectory illustrated above, moving toward the right until it reaches a value of ϕ for which $d\phi/dt = 0$.

Similarly, $\phi(t)$ decreases if the derivative corresponding to the initial phase error is negative, moving from right to left until a point is reached at which the derivative is zero.

It is clear from the figure that $d\phi/dt$ is zero at any of the following values of ϕ :

$$\phi_n = 2n\pi - \sin^{-1}\frac{\omega-\omega_0}{AK} \qquad (1.30)$$

$$\phi_n' = (2n-1)\pi - \sin^{-1}\frac{\omega-\omega_0}{AK}$$

However, the stable points occur only for the points ϕ_n , while the points ϕ_n' are unstable since perturbations of ϕ in either direction will cause the system to migrate until it reaches the next value at which the derivative is zero, which will necessarily be a stable point ϕ_n.

The figure also serves to illustrate the lock-in range of the loop. If $\omega - \omega_0 > AK$ the trajectory does not cut the ϕ axis, no stable points exist and the loop never achieves lock. $\phi(t)$ continues to increase or decrease forever along the sinusoidal trajectory. Thus, lock-in range for the first order loop is the same as the hold-in range described in a previous section.

Trajectories for higher order loops can be obtained in a similar manner by solving the appropriate non-linear differential equation. For an extensive discussion of these so-called phase plane plots, reference can be made to Viterbi.[11,12]

1-9 ADDITIVE NOISE IN PHASE-LOCK LOOP

As mentioned at the beginning of this chapter, the most important and unavoidable disturbance in the majority of radio communication systems is additive thermal noise, which is a zero-mean wideband Gaussian process whose spectral density is nearly flat over the frequency range of the receiver.

Under certain assumptions it can be shown[13] that a stationary Gaussian process n(t) with zero mean can be

expressed as;

$$n(t) = \sqrt{2}\left[n_1(t)\,\sin\omega_o t + n_2(t)\,\cos\omega_o t\right] \tag{1.31}$$

where one of the assumptions is that the noise process n(t) has been passed through a symmetric wideband band pass filter with center frequency ω_o, such that the two sided spectral density of the noise is $N_o/2$ over a sufficiently wide frequency range centered about $\pm\omega_o$.

To examine how the additive white Gaussian noise affects the operation of the loop, let the received signal be:

$$\sqrt{2}\,A\sin\theta(t) + n(t) = \sqrt{2}\left\{A\sin(\omega_o t + \theta_1(t)) + n_1(t)\sin\omega_o t + n_2(t)\cos\omega_o t\right. \tag{1.32}$$

The VCO output signal can again be written as;

$$\sqrt{2}\,K_1\cos\theta'(t) = \sqrt{2}\,K_1\cos\left[\omega_o t + \theta_2(t)\right] \tag{1.33}$$

But in this case $\theta'(t)$ is not only a function of signal modulation but also of the noise process n(t).

The output of the multiplier can, of course, be

written as,

$$x(t) = AK_1 \sin[\theta_1(t) - \theta_2(t)] - K_1 n_1(t) \sin\theta_2(t) + K_1 n_2(t) \cos\theta_2(t) \tag{1.34}$$

where we have eliminated all high frequency terms because of the presence of the low pass filter following the multiplier.

Continuing, in a fashion similar to that used in the development of equation (1.12), we can write,

$$\frac{d\phi(t)}{dt} = \frac{d\theta_1(t)}{dt} - K\int_0^t \{A\sin\phi(u) + n'(u)\} f(t-u)\, du \tag{1.35}$$

which is the equation of operation of the loop when signal plus additive noise are applied at its input and where,

$$n'(t) = -n_1(t)\sin\theta_2(t) + n_2(t)\cos\theta_2(t) \tag{1.36}$$

Equation 1.35 is represented by the block diagram in figure (1.7). We note, in particular, that for the sake of analysis, the noise enters the loop after the sinusoidal non-linearity. We can again make the approximation

$$\sin\phi \doteq \phi$$

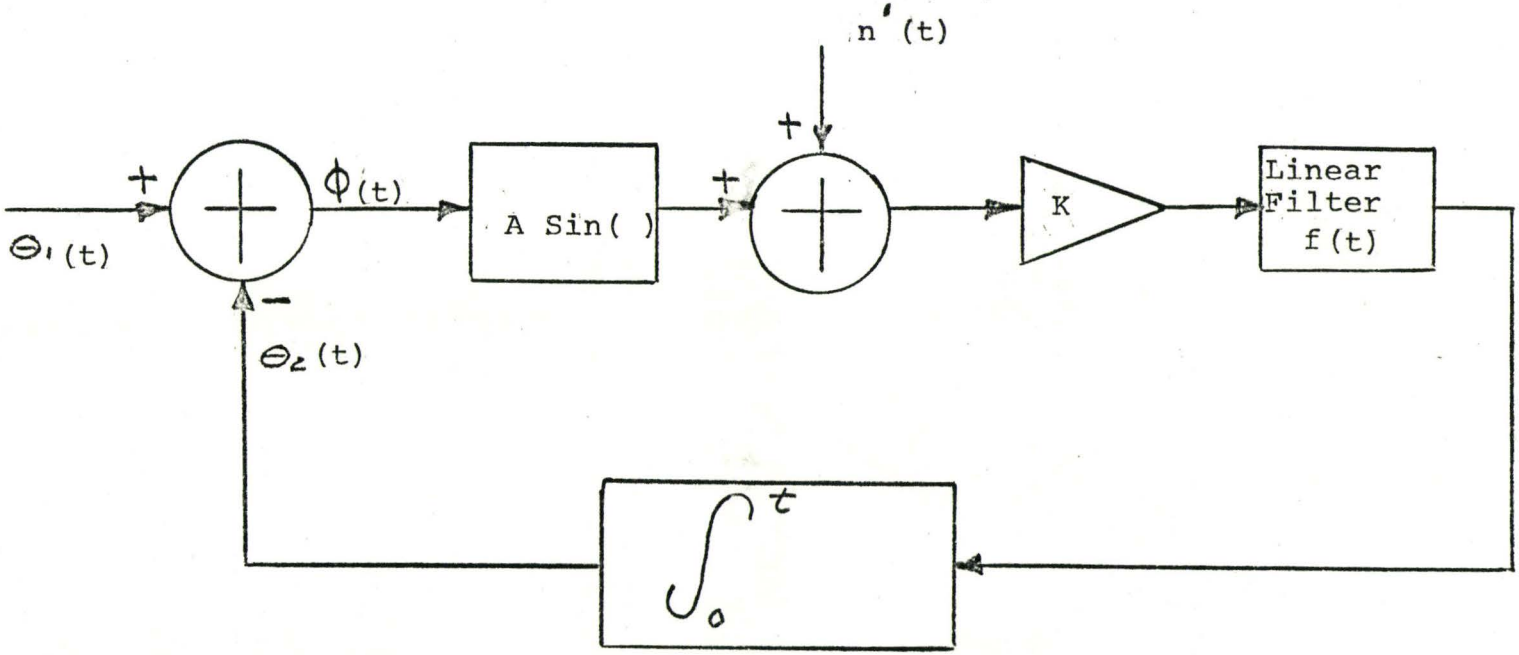


fig.1.7 Model of Loop in the Presence of Noise

where $\phi(t)$ now depends on n'(t) as well as signal, and then A sin( ) in figure (1.7) can be replaced by A.

If we assume that noise alone is present at the loop input then $\theta_1(t)=0$ and $\phi(t)=-\theta_2(t)$; their means are both zero and they are stationary in the steady state. Their spectral densities can be written as;

$$S_{\theta_2}(\omega) = S_{\phi}(\omega) = \frac{1}{A^2}\left|H(j\omega)\right|^2 S_n(\omega) \tag{1.37}$$

where $S_n(\omega)$ is the spectral density of the noise and H(s) is the closed loop transfer function defined in equation (1.17). On the assumption that the noise process is white with one-sided spectral density N_o, equation (1.37) can be written,

$$S_{\phi}(\omega) = \frac{N_o}{2A^2}\left|H(j\omega)\right|^2 \tag{1.38}$$

and the variance of the phase error or phase jitter can be written as:

$$\sigma_{\phi}^2 = \frac{N_o}{2A^2}\int_{-\infty}^{\infty}\left|H(j\omega)\right|^2\frac{d\omega}{2\pi} = \frac{N_o}{A^2}\int_{0}^{\infty}\left|H(j\omega)\right|^2\frac{d\omega}{2\pi} \tag{1.39}$$

The loop noise bandwidth is defined as,

$$B_L = \int_0^\infty |H(j\omega)|^2 \frac{d\omega}{2\pi} \tag{1.40}$$

and the phase jitter, in the steady state is:

$$\sigma_\phi^2 = \frac{N_o}{A^2} B_L \tag{1.41}$$

Thus, the loop noise bandwidth is defined as the bandwidth of an ideal low pass filter whose output variance is σ_ϕ^2 when the input is a white noise process with one-sided spectral density N_o/A^2. Of course, loop noise bandwidths for various loop filters can be determined by evaluating the integral in equation (1.40).

It should be remembered, that the results above are only approximations since they are based on a linear model and will hold only for values of variance less than approximately 0.25.

Attempts have been made to obtain results for loops in the presence of noise without the assumption of linearity, the most successful being, work done by Viterbi.[14]

CHAPTER TWO

THE DIGITAL PHASE-LOCK LOOP

2-1 INTRODUCTION

In this chapter, a new type of loop will be described. Unlike the conventional phase-lock loop, in which analogue circuits are used, the loop discussed below is implimented with logic circuits (gates, storage elements). When excited by appropriate clock signals the loop will exhibit properties very much like those of an analogue loop and as such is referred to as a digital phase-lock loop. The discussion and analysis which is made in this chapter is a condensed version of the analysis which appears in a paper published by Pasternack and Whalin[1]. The block diagram of the so-called "n th" order loop is first presented. Based on this block diagram, an equation of operation is developed using difference equations and the z-transform. The equations are then reduced to those of an n=1 digital loop.

2-2 GENERAL DIGITAL PHASE-LOCK LOOP

In figure (2.1) the general "n th" order digital phase-lock loop is illustrated in functional block diagram form. Among its basic components are an "exclusive or" comparator which develops a gating function dependent upon the phase relation of its two inputs, and transmission gates operating on a number of clock signals to provide inputs to register circuitry. Unlike the analogue phase-lock loop, there is neither a voltage controlled oscillator nor a low pass filter within the loop; although, we will find, that for most applications it is necessary to include a low pass filter at the output of the loop.

When the loop is locked, shift circuitry periodically transfers the contents of i th register to (i+1)th register. The transfer period is one half that of the input waveform. The shift is controlled by the n th register which provides an output pulse after M clock pulses are counted. A flip-flop will convert the output pulse train to a square wave which is used as the second input to the phase comparator.

In order to determine some of the properties of the loop, its operation is described in terms of a linear difference equation. In figure (2.2) we have illustrated some waveforms which will aid in the formulation of this difference equation.

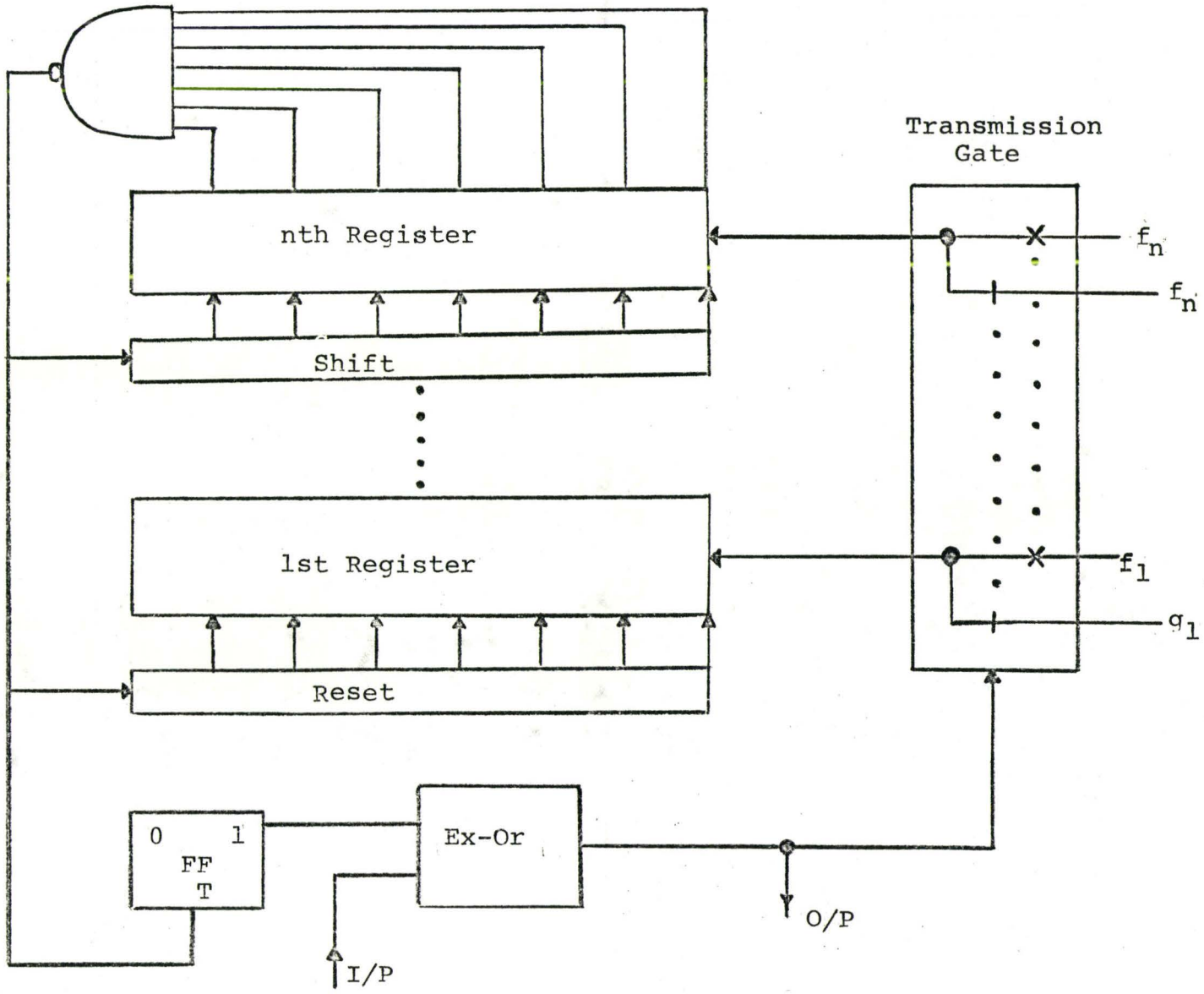
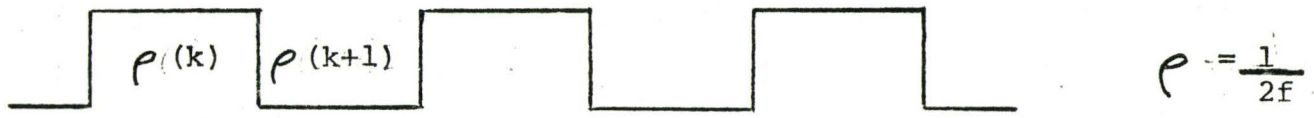
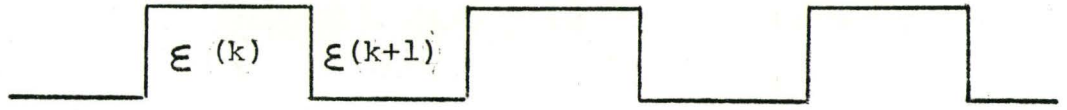


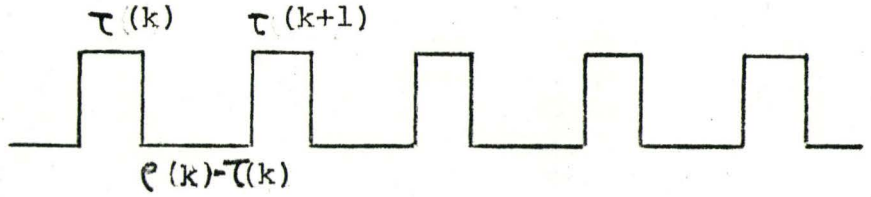
fig.2.1 General, nth Order Digital Phase-Lock Loop



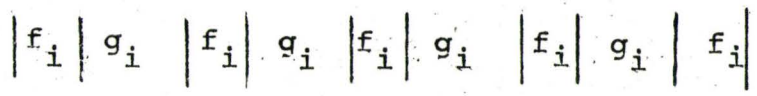
a) Input Signal



b) Feedback Signal



c) Gating Signal



i=1,2,....n

fig.2.2 Waveforms for nth Order Digital Phase-Lock Loop

The waveforms are for a phase-lock loop which is in a locked condition with constant input frequency. We shall assume that clocks g_i, i=1,2··n will be enabled during a "o" level of the gating signal (output of the exclusive-or) and the clocks f_i, i=1,2··n will be enabled during the"1" level of the gating waveform. If we further assume that the first register is initially cleared then its count at the end of the (K+1)th period is :

$$g_1[\rho(k)-\tau(k)]+f_1\tau(k+1)$$

For each successive period the count will be shifted into the i th register and increased by a count of:

$$g_i[\rho(k+i-1)-\tau(k+i-1)]+f_i\tau(k+i) \qquad i=2,3\cdots n$$

This count will propogate through the n registers until the number M is reached in the n th register and the count is reinitiated. It is of interest to note that although $n\rho$ seconds are required for a complete count cycle, new count cycles will be initiated every ρ seconds.

After $n\rho$ seconds the count in the n th register will be the sum of the pulses counted during the previous n/2 periods of the input waveform. Equating this sum to M we get:

$$g_1[\rho(k)-\tau(k)]+g_2[\rho(k+1)-\tau(k+1)]+\cdots\cdots$$
$$+g_n[\rho(k+n-1)-\tau(k+n-1)+f_1\tau(k+1)+\cdots\cdots$$
$$+f_n\tau(k+n)=M \qquad (2.1)$$

Rewriting the above equation:

$$f_n\tau(k+n)+(f_{n-1}-g_n)\tau(k+n-1)+\cdots\cdots-g_1\tau(k)$$
$$=M-[g_n\rho(k+n-1)+g_{n-1}\rho(k+n-2)+\cdots$$
$$+g_2\rho(k+1)+g_1\rho(k)] \qquad (2.2)$$

which is the n th order general difference equation relating the response $\tau(k+n)$ to the excitation $\rho(k+n)$.

Recalling, that for a general analogue system describable by a linear differential equation, we use the Laplace transform to arrive at a system equation in which polynomials in s multiplying the response and excitation functions, result in poles and zeros respectively, in the s-plane. By the appropriate choice of the coefficients in the differential equation we are able to locate the poles and zeros in a

position in the s-plane so as to give a desired frequency response. In a similar manner, we may apply the Z-transform to the difference equation 2.2 and since there is a one to one correspondence between the s and z planes with regard to poles and zeros, it can be expected that by properly choosing coefficients, equation 2.2 can be synthesized to provide a desired frequency response (high-pass, low-pass, band pass etc) possessing specified critical frequencies. In most applications of phase-lock loops and in particular, for the application we shall introduce later, we are interested in exploiting the low-pass properties of 2.2. To achieve this, we set the coefficients g_i , i=2, 3..n to zero thus locationg the z plane zeros at infinity and obtaining maximum high frequency attenuation. Equation 2.2 becomes:

$$f_n\tau(k+n)+f_{n-1}\tau(k+n-1)+\cdots\cdots-g_1\tau(k) = M-g_1e(k) \qquad (2.3)$$

Further, we shall normalize equation 2.3 by making the following substitutions:

$$F_j=\frac{f_j}{f_n} \quad , \quad j=1,\cdots n-1$$

$$G_1=\frac{g_1}{f_n}$$

$$v(k+j)=\tau(k+j)/e(k+j)$$

Equation (2.3) then becomes,

$$v(k+n) + F_{n-1}\, v(k+n-1) + \cdots F_1 v(k+1) - G_1 v(k)$$

$$= 2M F(k) - G_1 \qquad (2.4)$$

where;

$$F(k) = 1/(2 f_n e(k))$$

Notice, that $v(k+j)$ is the cycle-by-cycle average voltage of the $\tau(k+j)$ interval expressed as a fraction of the maximum possible voltage. In figure 2.3 we have redrawn the gating waveform from figure 2.2. The complex Fourier series for the gating waveform using the symbols from the diagrams is

$$f(t) = 1/e \sum_{n=-\infty}^{\infty} c_n e^{j\omega_n t}$$

where

$$c_n = \tau A [\sin(\omega_n \tau/2)] / \omega_n \tau/2$$

and

$$\omega_n = 2\pi n/e$$

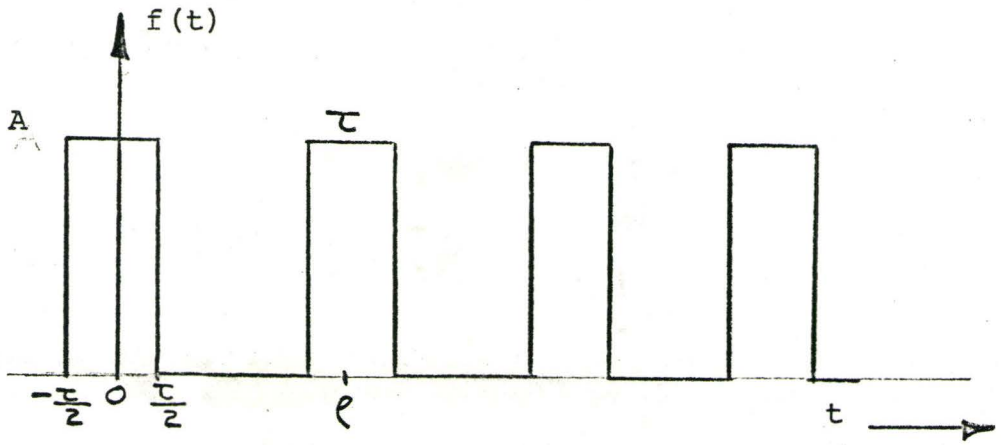


fig.2.3 Gating Waveform

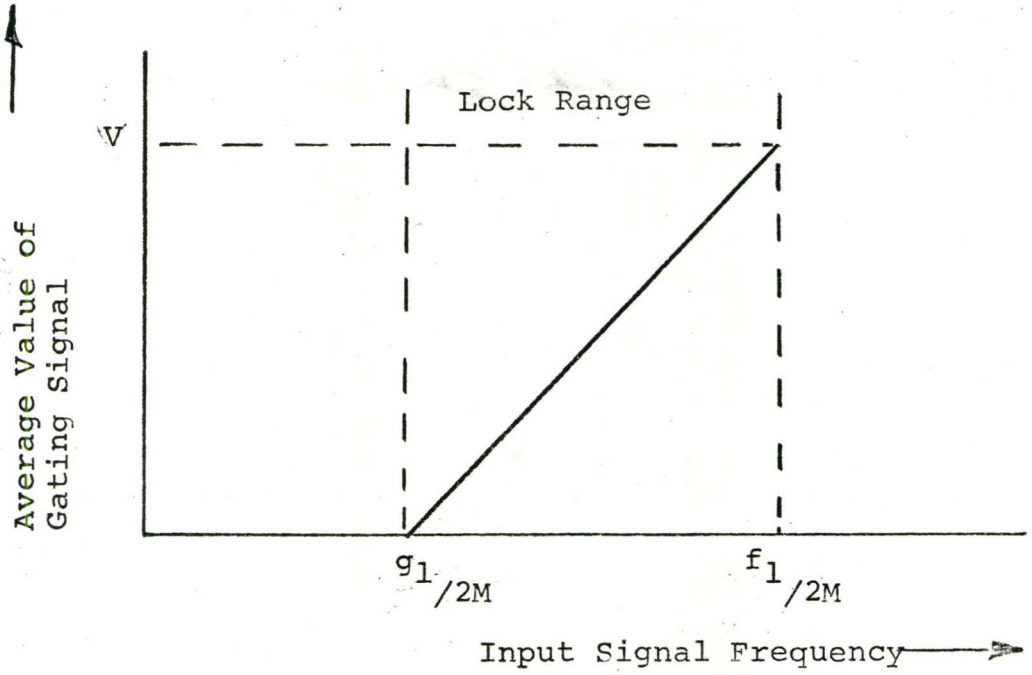


fig.2.4 Static Response of 1st Order Loop

The D.C. component of this series is $A\tau/\ell$ and therefore, $v(k+j)$ is proportional to the D.C. component of the gating waveform. It is this D.C. component, which is a measure of the phase difference between the phase comparator inputs, that we shall make use of for the application described in chapter 3.

2-3 FIRST ORDER DIGITAL PHASE-LOCK LOOP

Equation (2.4) which was developed in the previous section is the system equation for the n th order digital phase-lock loop with a low pass frequency response. In this section we shall reduce equation (2.4) to the difference equation for the first order phase-lock loop and derive some of the important parameters of the loop. The approach can just as well be applied to the design of higher order loops, but as might be expected, the higher the order of the loop, the more tedious the mathematics becomes.

The difference equation for the phase-lock loop with n=1 is easily written from equation (2.4),

$$v(k+1) - G_1 v(k) = 2MF - G_1 \tag{2.5}$$

where $M=2^{N-1}$ and N is the number of counter stages in the register in the system. Under steady state conditions $v(k+1)=v(k)$ so that,

$$v(k)(1-G_1) = 2MF-G_1$$

or

$$v(k) = \frac{2MF-G_1}{1-G_1} \tag{2.6}$$

We can de-normalize equation (2.6) by multiplying the numerator and denominator by $f_n=f_1$ to give;

$$v(k) = \frac{2Mf-g_1}{f_1-g_1} \tag{2.7}$$

where f is the fundamental frequency of the input signal and f_1 and g_1 are the fundamental frequencies of the two clocks.

Since $0 \leq v(k) \leq 1$ we can define the end points of the lock range by:

$$f[v(k)=0] = g_1/2M \quad \text{and} \tag{2.8a}$$

$$f[v(k)=1] = f_1/2M \tag{2.8b}$$

Thus, the loop will lock onto any signal whose fundamental frequency lies between $g_1/2M$ and $f_1/2M$ and will remain locked onto the input signal as long as it remains within this range. In figure (2.4) we have illustrated the so-called static response of the loop. The maximum voltage V is equal to the voltage magnitude of the high logic level of the elements used. In the diagram, we have assumed that the low logic level is at zero volts. It is important to note that in the lock range the voltage output versus frequency characteristic is perfectly linear. This is exactly the relationship required for frequency demodulation without distortion. Thus, the digital phase-lock loop is an ideal device to use as a frequency discriminator.

Applying the z- transform to equation (2.5) we get;

$$(z-G_1)V(z) = (2MF-G_1)\frac{z}{z-1} + z\,v(0) \tag{2.9}$$

A partial fraction expansion yields;

$$V(z) = \left[\frac{2MF-G_1}{1-G_1}\right]\frac{z}{z-1} + \left[v(0) - \frac{2MF-G_1}{1-G_1}\right]\frac{z}{z-G_1} \tag{2.10}$$

Assuming that for $t < o$ the normalized input frequency is F_a we can write from equation (2.6),

$$v(o) = \frac{2MF_a - G_1}{1 - G_1}$$

If we further assume that at t=o the input steps to F_b, we can write the response for $t \geq o$ as,

$$V(z) = \left[\frac{2MF_b - G_1}{1 - G_1}\right] \frac{z}{z-1} + \left[\frac{2M(F_b - F_a)}{1 - G_1}\right] \frac{z}{z - G_1} \qquad (2.11)$$

The inverse z-transform can then be written as;

$$v(k) = \frac{2MF_b - G_1}{1 - G_1} - \frac{2M(F_b - F_a)}{1 - G_1} (G_1)^k$$

$$= \frac{2Mf_b - g_1}{f_1 - g_1} - \frac{2M(f_b - f_a)}{f_1 - g_1} \left(g_1/f_1\right)^k \qquad (2.12)$$

In order to obtain the time constant of the loop we shall assume that v(k) is a continuous function of time and let v(k)=v(t) and $t=k\rho=k/2f_b$. Thus,

$$v(t)=\frac{2Mf_b-g_1}{f_1-g_1}-\frac{2M(f_b-f_a)}{f_1-g_1}\exp\left[-2f_bt\left(\ln f_1/g_1\right)\right] \tag{2.13}$$

The time constant can then be written as;

$$T=\frac{1}{2f_b\ln(f_1/g_1)} \tag{2.14}$$

Unlike most other physical systems, we find that the time constant, and therefore the response of the digital phase-lock loop is dependent upon the input frequency f_b. That is, the loop will respond somewhat differently for each input frequency within its lock range. If we recall that the count cycle of the register is equal to one half the period of the input signal, it is intuitively obvious that the overall time constant of the loop should depend on the period of the input signal.

2-4 IMPLIMENTATION OF THE FIRST ORDER DIGITAL PHASE-LOCK LOOP

In figure (2.5) we have illustrated a logic circuit, whose performance is describable by difference equations (2.5) and (2.12). The "exclusive or" and N stage binary counter are standard logic devices and their operation needs no explanation. However, the transmission gate does require a brief description. The configuration of nand gates is such that for a low logic level ("o") at the output of the "exclusive or," the output of the low frequency clock g_1 will be transmitted unaltered through the nand gates and applied to the triggering point on the first stage of the counter. At the same time the transmission gate will act as an open circuit to the output of the high frequency clock f_1. When a high logic level ("1") appears at the output of the "exclusive or," the situation is reversed, that is, the transmission gate acts as a short circuit to the output of the high frequency clock f_1 and as an open circuit to the output of the low frequency clock g_1. Thus at any time during the operation of the loop, the output of one of the clocks but not both, will appear at the input to the counter.

In figures (2.6), (2.7), and (2.8) we have illustrated waveforms at various points in the loop for three different input conditions. In the figure (2.6), we have assumed that

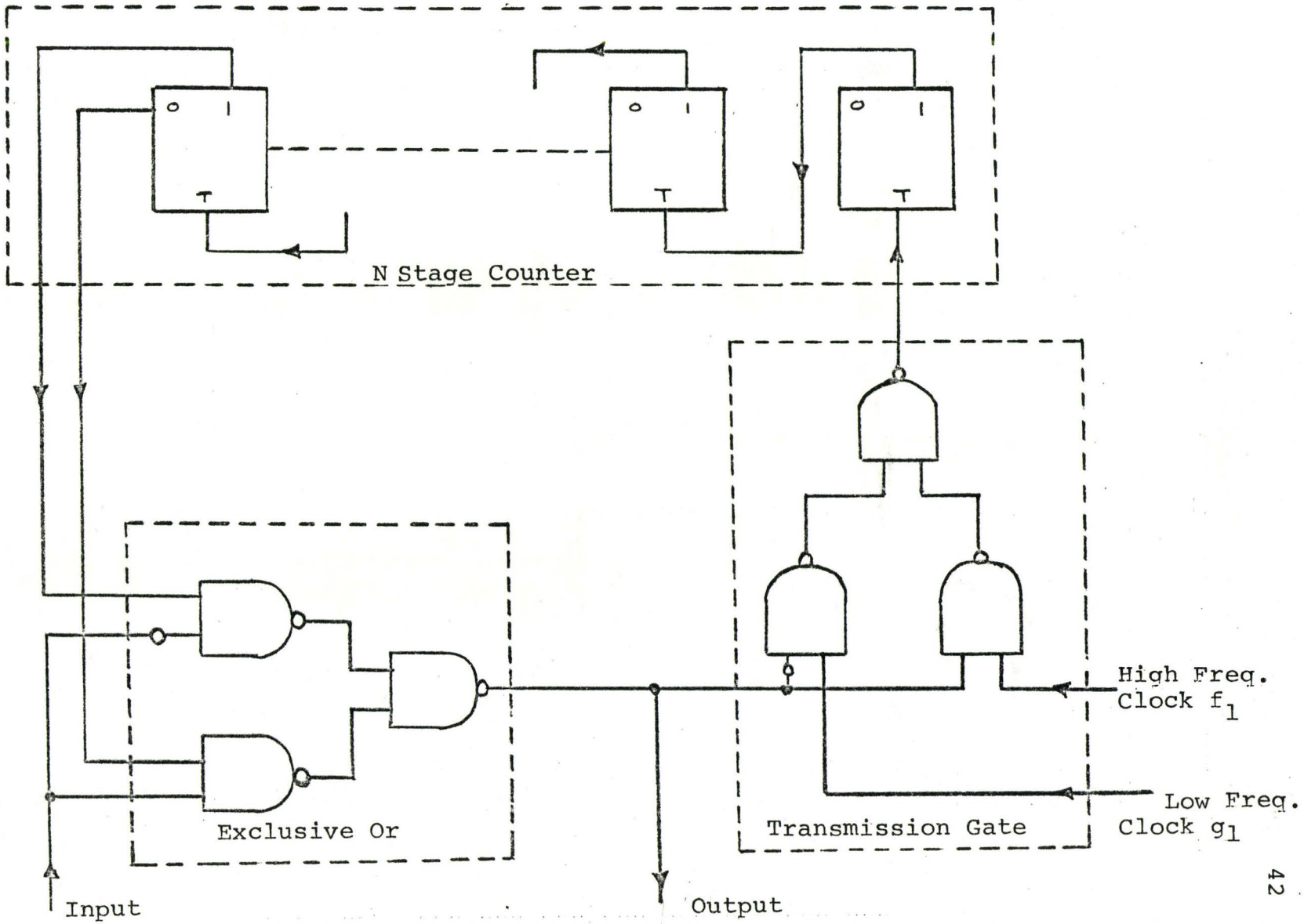


fig.2.5 1st Order Digital Phase-Lock Loop

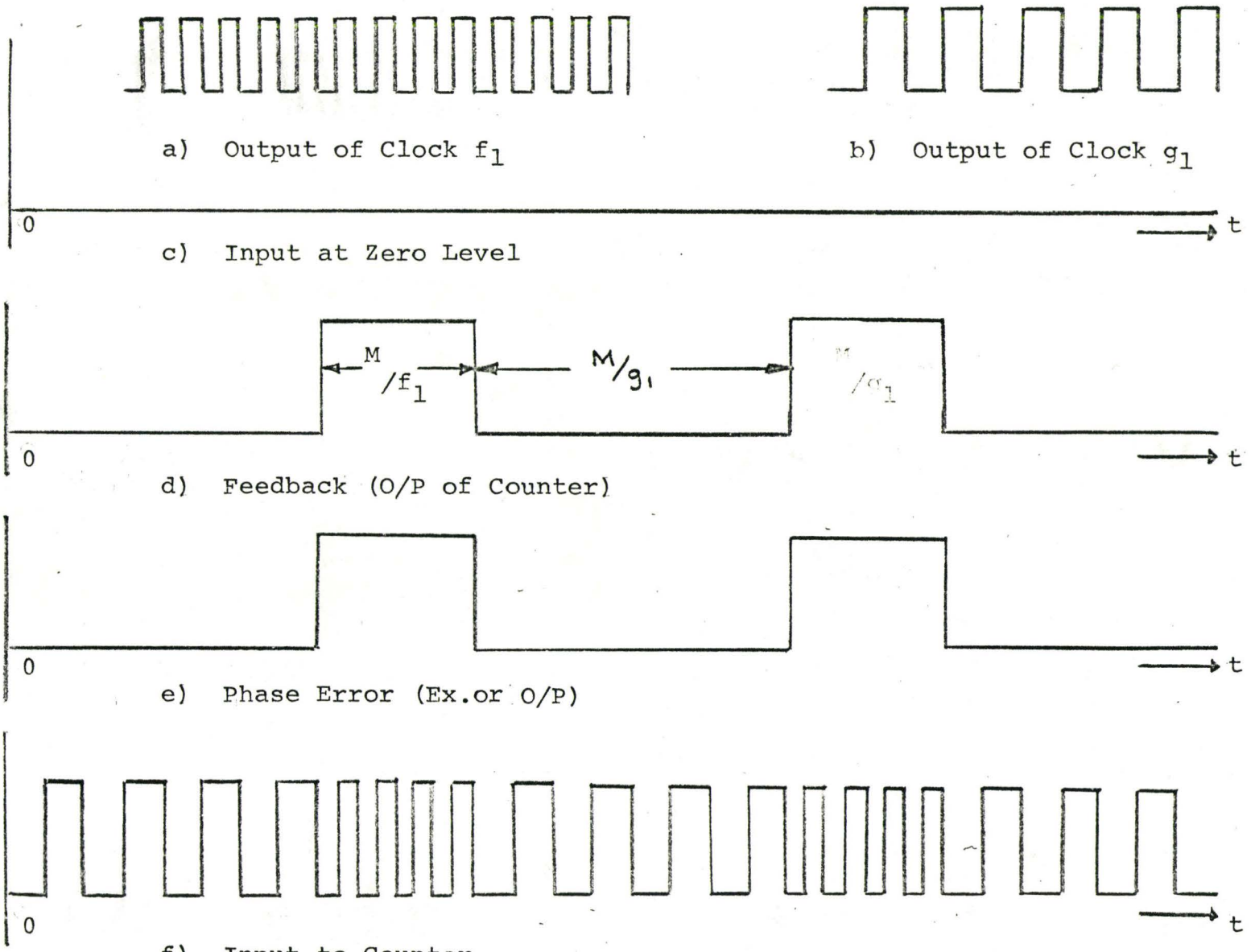


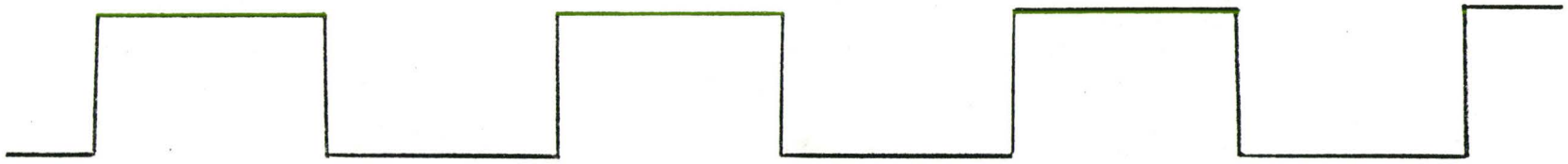
fig.2.6 Waveforms for No Input Signal

there is no input to the loop and that at time t_1 the feedback signal is at a low logic level ("o"). The output of the "exclusive or" is therefore also a "o" and clock g_1 is enabled. For simplicity, we have further assumed that the counter has only three stages and therefore when the fourth (M=4) pulse is applied to the first stage of the counter, the feedback signal makes the transition from a "o" to a "1". Subsequently, the output of the "exclusive or" makes a similar transition and clock f_1 is enabled. After four pulses from clock f_1 another set of transitions occur and cycle is reinitiated. From the illustration, it is evident that the quiescent frequency of the loop is,

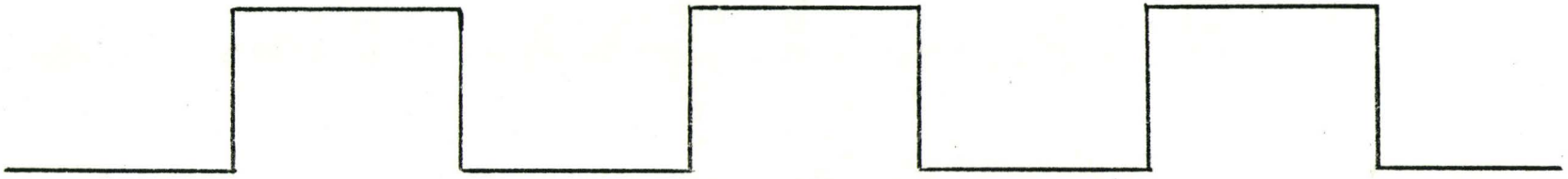
$$f_q = \left(M/f_1 + M/g_1\right)^{-1} = \frac{1}{M}\,\frac{g_1 f_1}{g_1 + f_1} \qquad (2.15)$$

But, one half cycle will be at $g_1/2M$ and the next at $f_1/2M$.

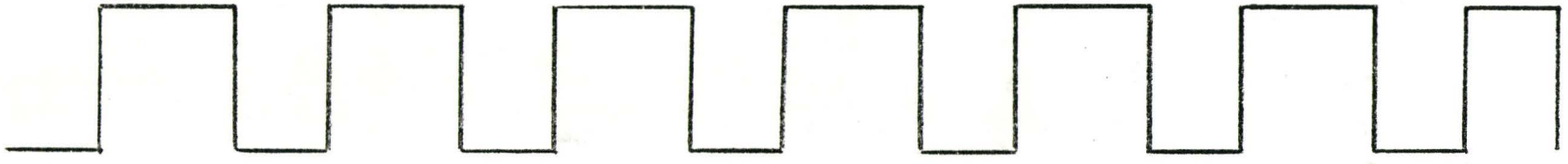
In figure (2.7), we have illustrated similar waveforms for the case of an input signal whose frequency lies within the lock range of the loop; that is within the frequency interval $(g_1/2M,\ f_1/2M)$. We have assumed that at $t=t_1$ both the input and feedback signals are "o"'s. Therefore, the output of the "exclusive or" is a "o" and clock g_1 is enabled. However, before four clock pulses from g_1



a) Input Waveform



b) Feedback Waveform



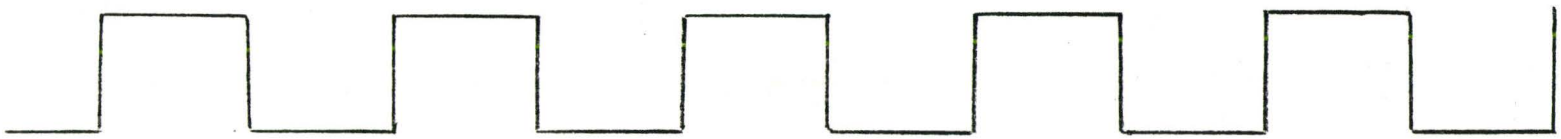
c) Phase Error Waveform

d) Input to Counter

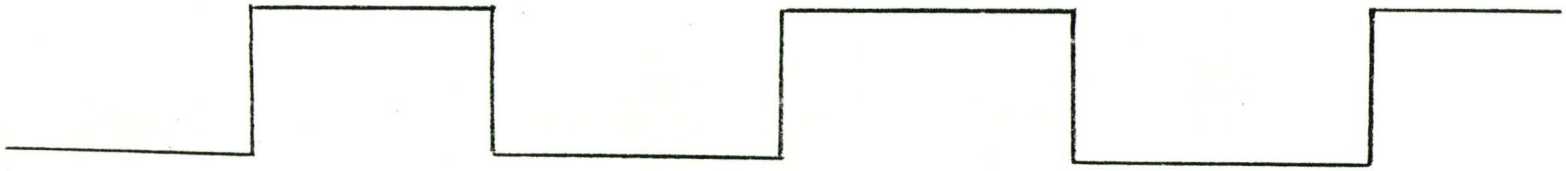
fig.2.7 Waveforms for Input Signal in Lock Range of loop

can be transmitted to the counter, the input signal makes a transition from a "o" to a "1" and as a result, the output of the "exclusive or" makes a similar transition, thus enabling clock f_1. After three pulses are transmitted from clock f_1, the count reaches four, causing a transition in the feedback signal and subsequently, a second transition in the output of the "exclusive or"; clock g_1 is re-enabled and the cycle is reinitiated. We note that the frequencies of the feedback and input signals are the same and that the phase difference between the two signals remains constant; the loop is said to be locked to the incoming signal. Also the output of the "exclusive or" is a measure of this phase difference and as mentioned earlier, the D.C. component of the waveform is directly proportional to the phase difference.

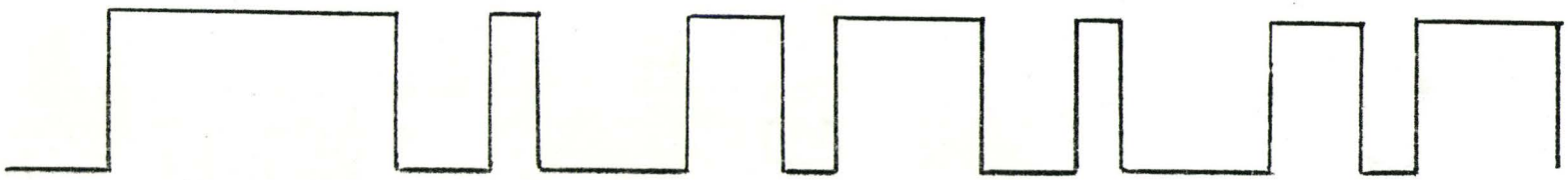
Figure (2.8) illustrates the loop waveforms for the last of the three possible input conditions; that is, an input signal which lies outside the lock range of the loop. The loop reaches a steady state condition where all the waveforms are cyclic. However, the frequency of the feedback signal is not equal to that of the incoming signal; the loop is unable to lock onto the input. Although it may not be evident from the illustration, the phase error waveform is actually a beat note whose fundamental frequency equals the frequency difference between the input and feedback signal. If this beat note were passed through a low pass



a) Input



b) Feedback



c) Phase Error

d) Input to Counter

fig.2.8 Waveforms for Input Signal Outside Lock Range of Loop

filter with cut off frequency of the order of the frequency difference, the output of the filter would be as in figure (2.9).

2-5 CONSTRUCTION OF THE DIGITAL LOOP

The first order Digital Phase-Lock Loop illustrated in figure (2.5) was constructed and tested. A block diagram of the test arrangement is illustrated in figure (2.10). The loop itself was built using Texas Instruments logic integrated circuits. Two four-stage binary counters were cascaded to form an eight-stage counter; thus, from equation (2.5) M=128. The clock* frequencies were arbitrarily chosen at 363.5 KHz and 262.5 KHz. Thus, from equation 2.8 it was expected that the lock range would be from 1025 Hz to 1425 Hz. To insure that the input signal to the loop could assume only voltage levels equal to the two logic levels of the integrated circuits, a limiter* with an input-output characteristic as illustrated in figure (2.11) was included between the oscillator and loop.

* For a description of clock and other circuits used, refer to appendix (A).

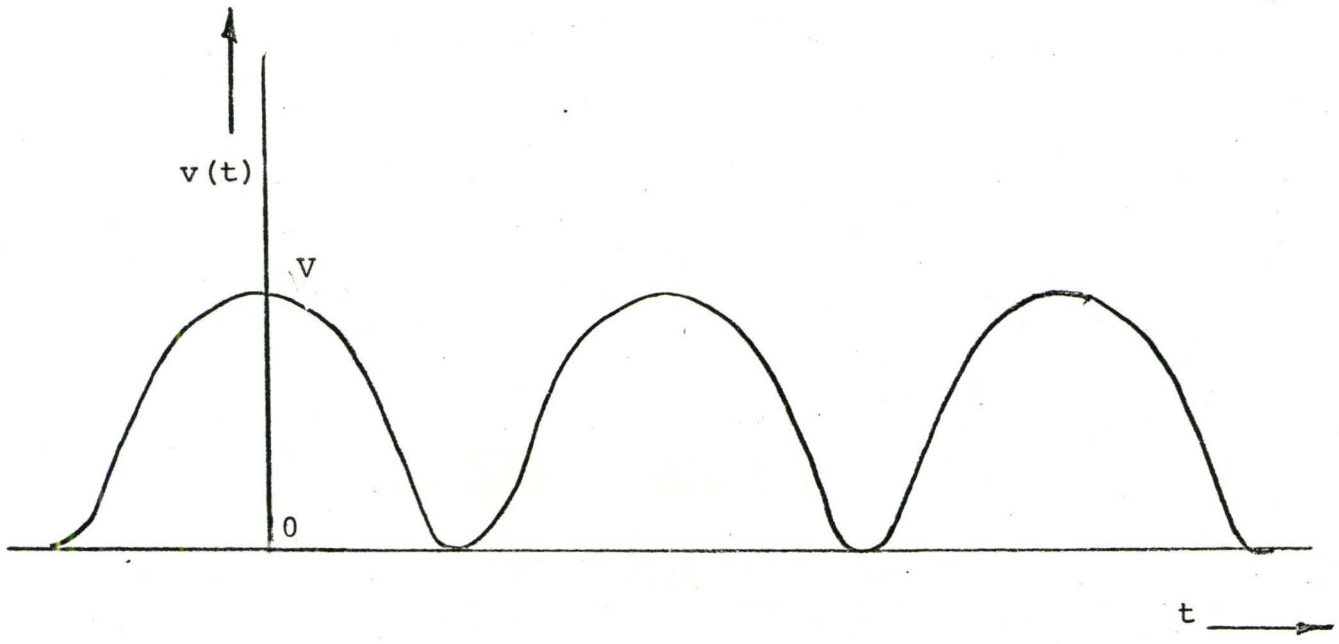


fig.2.9 Beat Signal

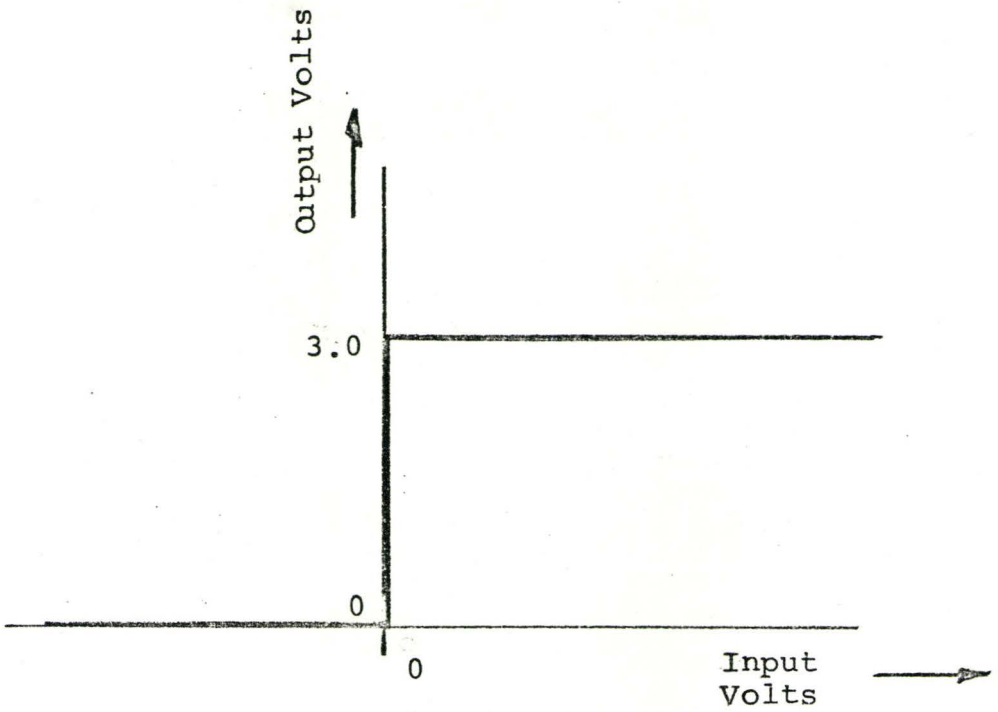


fig.2.11 Characteristic of the Limiter

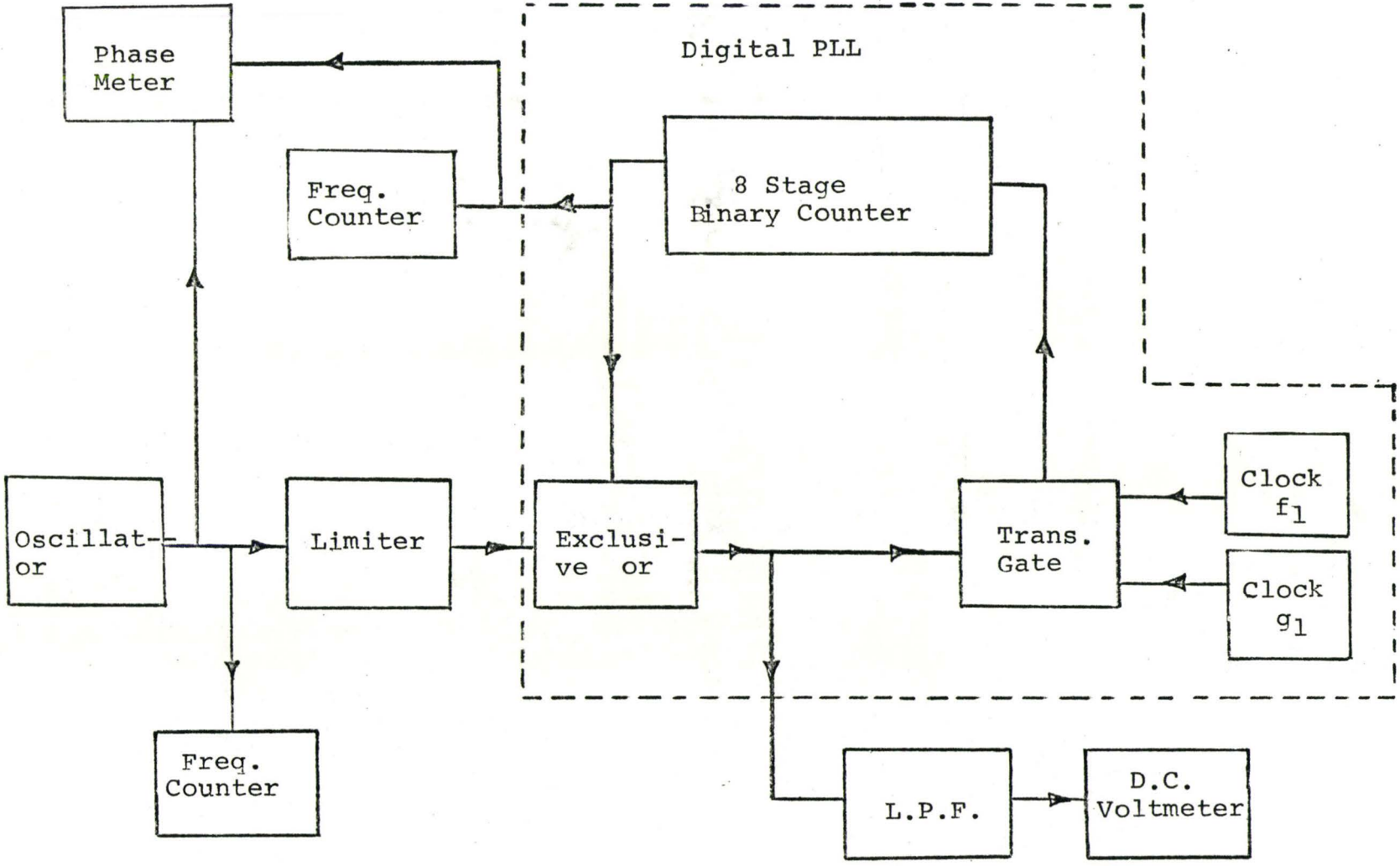


fig.2.10 Block Diagram of Test Arrangement

In the first test, the oscillator frequency was varied over and beyond the expected lock range of the loop. For each input frequency, we measured the phase error, D.C. output of the low pass filter and frequency of the feedback signal, the results are plotted in figure (2.12). The loop performed as theory predicts. Over the predicted lock range the D.C. voltage output of the filter varies linearly with frequency. Over the same range, the phase difference between the input and feedback signals varies from o to π radians.

In the next test, a second oscillator was connected with the first, in such a way that we could electronically switch between the two. Initially, the oscillators were adjusted such that the frequency of each would be within the lock range of the loop. The frequency of switching between the oscillators was 120 Hz. That is, for time equal to $^1/_{120}$ seconds the output of the one oscillator was applied to the loop and at the end of that interval the output of the second oscillator was applied for $^1/_{120}$ seconds. A picture of the waveform at the output of the filter is shown in figure (2.13a).

The frequency of one oscillator was then adjusted such that it would lie outside the lock range. The corresponding picture is shown in figure (2.13b). For each

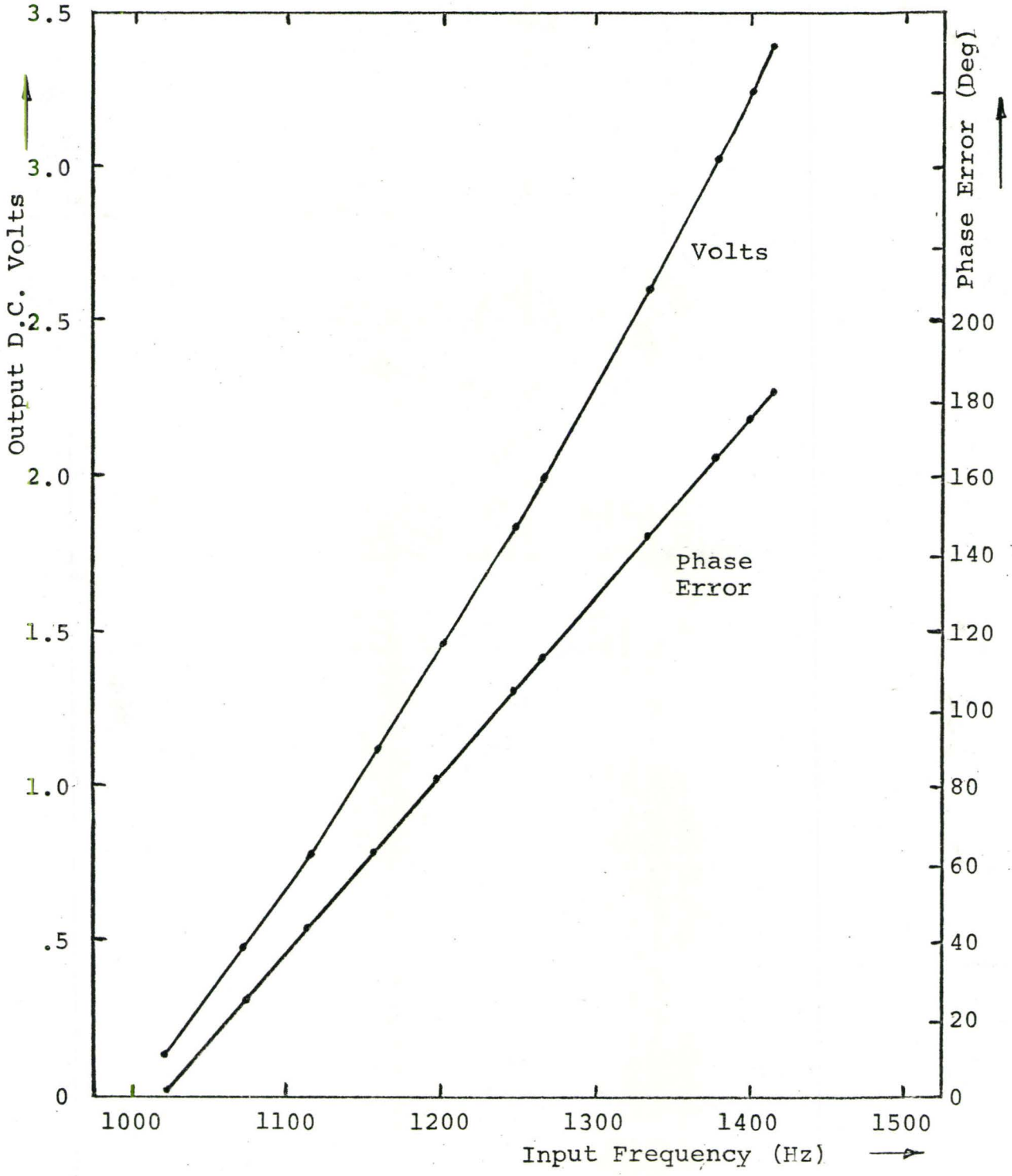
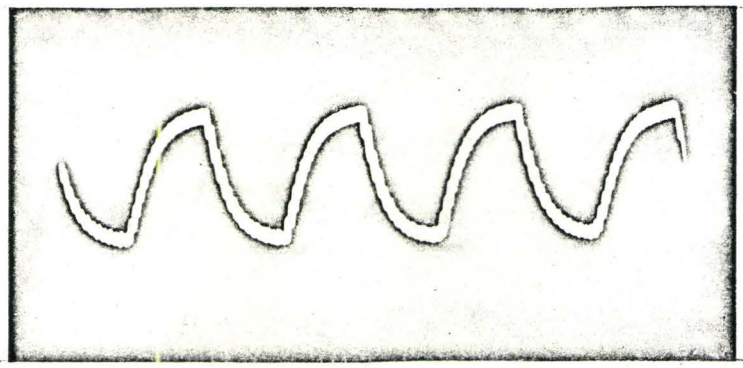


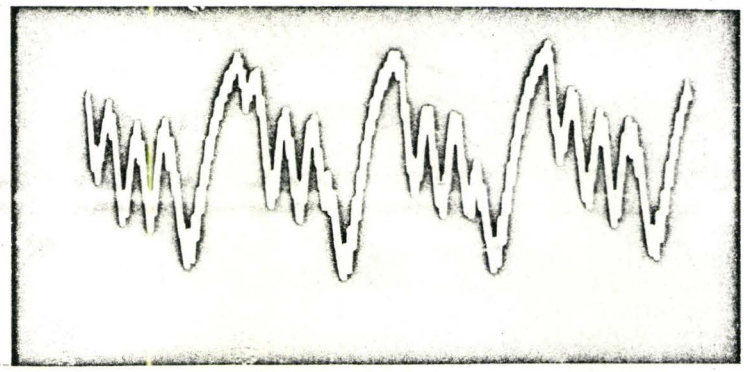
fig.2.12 Input Frequency Versus Phase Error and Output D.C.Volts



Horizontal
5 msec./cm.

Vertical
1V/cm.

(a) Both input frequencies within lock range of loop.



Horizontal
5 msec./cm.

Vertical
1V/cm.

(b) One input frequency within lock range of loop; the second outside the lock range.

fig.2.13 Output Waveforms

successive interval of 1/120 seconds, the loop is unable to lock onto the input and the result is a beat note at the output of the filter. From this picture, we are able to estimate that the time taken for the loop to go from an unlocked to a locked state is approximately 1/360 seconds. This rough estimate of lock-in time will be of some value in the last chapter when the digital loops are used as part of a binary signal detector scheme.

We can conclude from the experiments performed on the digital loop that its performance is similar to that of the analogue loop and, as such, can be used in similar applications. However, it is evident from the construction of the digital loop, that it has certain advantages;

1) It is completely integrable using one or more monolithic chips and can therefore have a small size.

2) Since it is digital, it is more reliable.

3) The stability of the loop is solely dependent on the stability of the clocks, which may be as good as required.

4) It includes, in effect, an ideal voltage-controlled oscillator, the frequency of which is linearly related to voltage.

CHAPTER THREE

DETECTORS

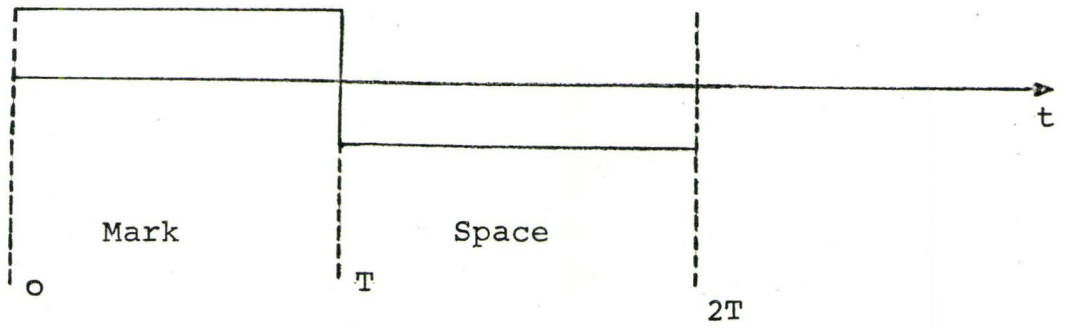
3-1 INTRODUCTION

In this chapter we shall discuss some aspects of communications systems in which digital signaling appears. We shall briefly outline the techniques used for transmitting digital data and then illustrate the receivers which may be used for the detection of such digital signals. The theoretical analysis of these detectors will be based on the principles of statistical decision theory. The problem normally encountered in digital communication involves deciding on a particular signal transmitted and since this problem is of a statistical nature, then decision theory will enable us to find the optimum means of making this decision. After demonstrating the theoretical derivation of these optimum detectors we shall, based on physical intuition, propose a detector which utilizes phase-lock loops as its basic components.

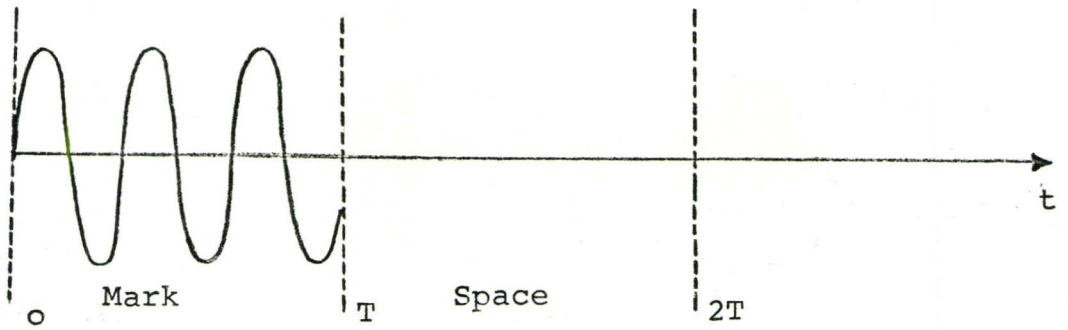
3-2 CARRIER TELEGRAPHY

Telegraphy means the transmission of a message to a remote location. Although the term originally meant the transmission of a language message, it has more recently come to mean the electrical transmission of a message encoded in binary form. In its simplest form, such a signal may utilize on-off (1 or 0), or bipolar (+1 or -1) voltage waveforms. More generally, the binary signal is a train of pulses each having associated with it, two voltages, the choice of which depends on the message to be transmitted.

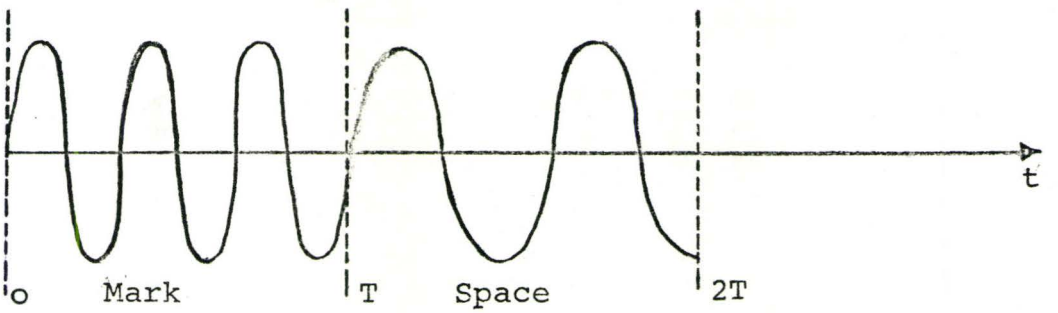
Carrier telegraphy is the use of a telegraph signal to modulate a carrier. The common forms of modulation are of course, amplitude (AM), frequency (FM), and phase (PM). The AM form corresponds to 100 per-cent modulation with a bipolar waveform, the carrier is keyed on and off by the telegraph waveform. This system is referred to as on-off keying (OOK). The FM corresponds to the bipolar modulation of the frequency of the carrier. For a rectangular modulating signal, which is the type we are interested in, the states of the modulating signal are described by a pair of discrete frequencies. This system is commonly referred to as frequency-shift keying (FSK). For the third, a choice of a 0^0 or 180^0 phase shift in the carrier determines the state of the modulating signal and is referred to as phase-



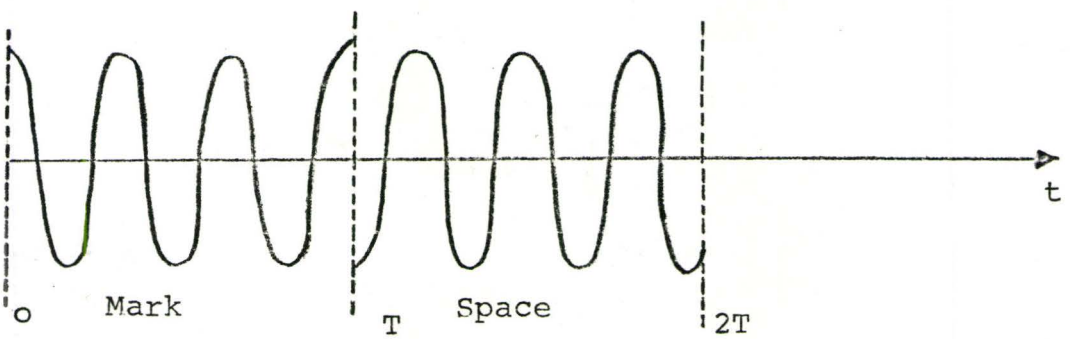
a) Telegraph Waveform



b) OOK



c) FSK



d) PSK

fig.3.1 Carrier Telegraph Signals

shift keying (PSK). Figure (3.1) illustrates the three types of carrier telegraphy.

Extensions from simple carrier telegraphy arise when one considers the multiplexing, or the simultaneous transmission of a number of telegraph waveforms by use of a single radio carrier.

Time division multiplexing (TDM) involves sequenced sampling, at a uniform rate, of each telegraph waveform to be transmitted. The composite waveform which results is then transmitted by standard techniques of carrier telegraphy.

Another multiplexing technique is the so-called frequency division multiplexing (FDM). In FDM, each of the telegraph waveforms is modulated in a standard fashion onto a different subcarrier, with the subcarriers normally in the audio range. This set of modulated subcarriers is then added together to form a composite waveform which, in turn, is modulated onto a radio carrier. At the receiving end, after demodulation down to baseband, the subcarriers are separated by filtering, and subsequently demodulated individually.

In terms of the results presented in this chapter, the analysis of the performance (error rate) of any one channel of a multiplex system is just the problem of determining the probabilty of error for each pulse in the channel.

Thus, the following analysis will be for a non-multiplexed telegraph signal, the results of which may be applied to a multiplex system using the same carrier-modulation technique.

3-3 STATISTICAL DECISION THEORY

In the last section, we described three modulation techniques for transmitting a binary message. In general, the signals for a mark or space (1 or 0) may be described by

$$S_j(t) = C_j(t) \sin\left[\omega_0 t + \theta_j(t) + \phi\right] \qquad (3.1)$$

for $0 < t < T$, $j = 0, 1$

where T is the time duration of one baud (mark or space).

However, in a practical communication system these received signals will be corrupted by additive noise and can therefore only be estimated. The techniques of decision theory will enable us to find the best way to perform these estimates.

Let us consider the simple problem of detecting the absence or presence of a signal pulse. Given a statistical sample of value v, we must select between one of two

alternative hypothesis. Hypothesis H_o may correspond to saying the voltage v measured represents noise, or a "o" transmitted. Hypothesis H_1, the alternative hypothesis, corresponds to saying a pulse, or a "1" is present. Of course, either choice on our part may be in error. Let us divide the one-dimensional space of v into two parts, V_o and V_1; V_o corresponding to our choice of hypothesis H_o and V_1 corresponding to our choice of hypothesis H_1. We must then choose V_o (or V_1) to minimize the overall probability of error.

We assume that the a priori probabilities P_o and P_1 of transmitting a "o" and a "1", respectively, are given. We are also given the conditional probability densities $p_1(v)\,dv$ and $p_o(v)\,dv$, corresponding to the probability of receiving v, given a "1" transmitted or a "o" transmitted, respectively. For Gaussian noise, and pulses of values o and A, these are just the probability densities illustrated in figure (3.2). Thus the error probability that v will fall in region V_o, although a "1" was transmitted, is

$$\int_{V_o} p_1(v)\,dv$$

and likewise that v will fall in region V_1, although a "o" was transmitted, is

$$\int_{V_1} p_o(v)\,dv$$

The overall error probability P_e, that is to be minimized by the appropriate choice of V_o, is given by

$$P_e = P_1 \int_{V_o} p_1(v)\,dv + P_o \int_{V_1} p_o(v)\,dv$$

Since V_o plus V_1 include all possible values of v then we can write:

$$\int_{V_o+V_1} p_o(v)\,dv = 1$$

Now we can eliminate V_1 from the probability of error equation and write:

$$P_e = P_o + \int_{V_o} \left[P_1 p_1(v) - P_o p_o(v) \right] dv$$

Since P_o is assumed to be a known constant, the probability of error can be minimized by choosing the region V_o which minimizes the integral; we can do this by choosing a

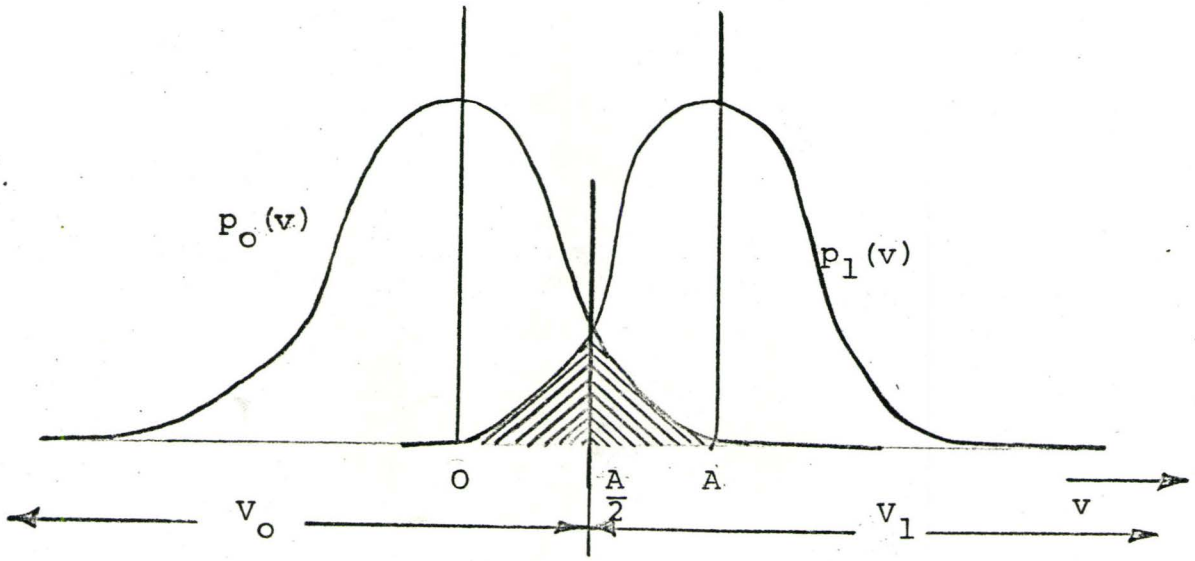


fig.3.2 Probability Densities for Pulses Plus Gaussian Noise

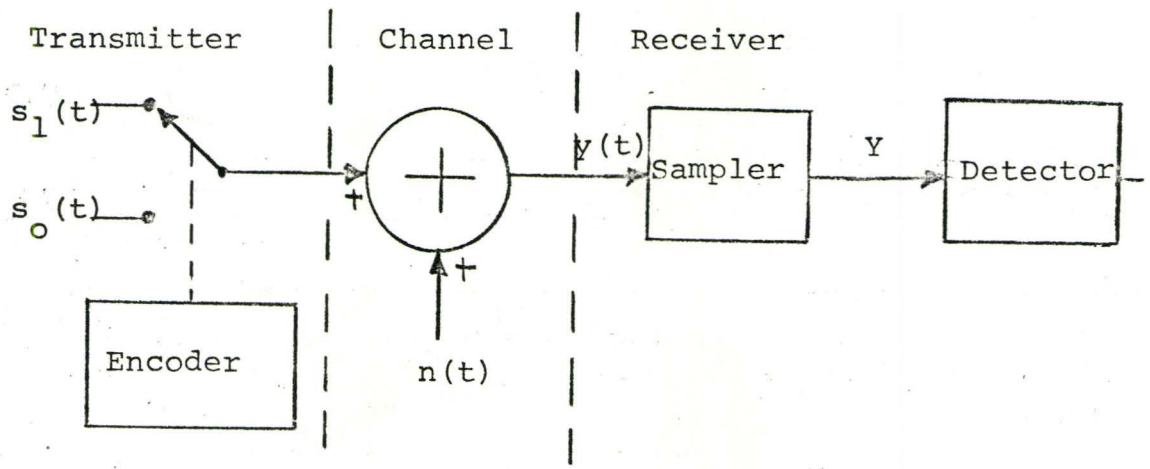


fig.3.3 The Model of a Communication System

region in which

$$P_o p_o(v) > P_1 p_1(v)$$

This will ensure that the integral is always negative and therefore P_e will be as small as possible.

For the case of Gaussian noise which was illustrated in figure (3.2), V_o corresponds to all values of $v < A/2$ and V_1 to all values $v > A/2$. Thus in the implimentation, a threshold level at $A/2$ would be established such that if v exceeded this level we would decide in favour of a "1" being received and if v was below the level, we would make the alternative decision.

The rule for choosing H_o, which we have just developed on the basis of minimum overall probability of error, is called the Bayes decision rule. It is stated usually as

$$p_o(v)/p_1(v) > P_1/P_o \qquad (3.2)$$

where the ratio of probability densities

$$p_o(v)/p_1(v)$$

is called the likelihood ratio.

3-4 DETECTION OF KNOWN SIGNALS IN NOISE (COHERENT RECEPTION)[2,3]

In this section we shall consider the problem in which the binary information is to be carried by either one of two signals $s_1(t)$ and $s_o(t)$ of arbitrary and different shape. Examples would include phase-shift-keyed and frequency-shift-keyed signals which we discussed earlier in this chapter and which may be described by equation 3.1.

The important assumption, necessary for this analysis is that the two signals must be known exactly. That is, we must have precise knowledge of the amplitude, frequency and phase functions of the two signals transmitted. It is in this sense, that we refer to the receiver as being coherent.

The model of the communication system is illustrated in figure (3.3), $s_o(t)$ and $s_1(t)$ are deterministic signals of duration T seconds and n (t) is assumed to be a sample function from a random Gaussian process of known autocorrelation function.

Let us assume that observations of y (t) are available at the instants $t_1, t_2, \ldots t_k$ where $t_k = k\Delta t$ and $T - \Delta t < k\Delta t \leq T$. Then we may use vector notation to represent the pertinent waveforms :

$$\mathbf{Y} = [y_1, y_2 \cdots\cdots y_k]$$

$$\mathbf{S}_j = [s_{j1}, s_{j2} \cdots\cdots s_{jk}] \qquad j = 0, 1 \tag{3.3}$$

$$\mathbf{N} = [n_1, n_2 \cdots\cdots n_k] \qquad \bar{N} = 0, \quad \phi_{nn} = \overline{NN^T}$$

We can then write:

$$Y = S_j + N \tag{3.4}$$

Since S_j is known exactly and the noise is Gaussian with zero mean and covariance matrix ϕ_{nn} we have,

$$\begin{aligned} p(Y|S_j) &= p_j(Y) = p(Y - S_j = N) \\ &= \left[1/\left((2\pi)^{k/2}|\phi_{nn}|^{1/2}\right)\right] \exp\left[-\tfrac{1}{2}(Y-S_j)(\phi_{nn}^{-1})(Y-S_j)^T\right] \end{aligned} \tag{3.5}$$

We can now write the likelihood ratio as

$$L(Y) = \frac{p(Y|S_1)}{p(Y|S_0)} = \frac{\exp\left[-\tfrac{1}{2}(Y-S_1)^T\phi_{nn}^{-1}(Y-S_1)\right]}{\exp\left[-\tfrac{1}{2}(Y-S_0)^T\phi_{nn}^{-1}(Y-S_0)\right]} \tag{3.6}$$

which according to Bayes rule must be compared to $K=P_o/P_1$ where P_o and P_1 are the a priori probabilities of $s_o(t)$ and $s_1(t)$ respectively. For convenience we will choose to compare ln L(Y) with ln(K):

$$\ln(L(Y)) = -\tfrac{1}{2}\left(-Y^T\phi_{nn}^{-1}S_1 - S_1^T\phi_{nn}^{-1}Y + S_1^T\phi_{nn}^{-1}S_1 + Y^T\phi_{nn}^{-1}S_o + S_o^T\phi_{nn}^{-1}Y - S_o^T\phi_{nn}^{-1}S_o\right) \tag{3.7}$$

where $Y^T\phi_{nn}^{-1}S_i$ is a quadratic form* so that:

$$Y^T\phi_{nn}^{-1}S_i = (Y^T\phi_{nn}^{-1}S_i)^T = (\phi_{nn}^{-1}S_i)^T Y = S_i^T\phi_{nn}^{-1}Y \tag{3.8}$$

Equation (3.7) may be written as,

$$\ln L(Y) = Y^T\phi_{nn}^{-1}S_1 - Y^T\phi_{nn}^{-1}S_o - \tfrac{1}{2}S_1^T\phi_{nn}^{-1}S_1 + \tfrac{1}{2}S_o^T\phi_{nn}^{-1}S_o$$

$$= Y^T\phi_{nn}^{-1}(S_1 - S_o) - \tfrac{1}{2}(S_1^T\phi_{nn}^{-1}S_1 - S_o^T\phi_{nn}^{-1}S_o) \tag{3.9}$$

* See Ogata[1].

Since the last term in equation (3.9) does not depend on the received data and is composed of known factors, we can replace the comparison of ln L(Y) to ln K by the comparison of $Y^T \phi_{nn}^{-1} (S_1 - S_0)$ to the constant

$$\ln K + \tfrac{1}{2} (S_1^T \phi_{nn}^{-1} S_1 - S_0^T \phi_{nn}^{-1} S_0)$$

Therefore the decision boundary equation can be written in the form,

$$Y^T \phi_{nn}^{-1} (S_1 - S_0) = \ln K + \tfrac{1}{2} (S_1^T \phi_{nn}^{-1} S_1 - S_0^T \phi_{nn}^{-1} S_0) \qquad (3.10)$$

The constant on the right side of equation (3.10) is often referred to as the bias and we shall denote it by b, that is

$$b = \ln K + \tfrac{1}{2} (S_1^T \phi_{nn}^{-1} S_1 - S_0^T \phi_{nn}^{-1} S_0) \qquad (3.11)$$

The operation of this receiver is now clear from equations (3.10) and (3.11) and is illustrated in figure (3.4). The receiver is to sample the received waveform y(t) to obtain the vector Y and then calculate the statistic

$$v = Y^T \phi_{nn}^{-1} (S_1 - S_0) - b \qquad (3.12)$$

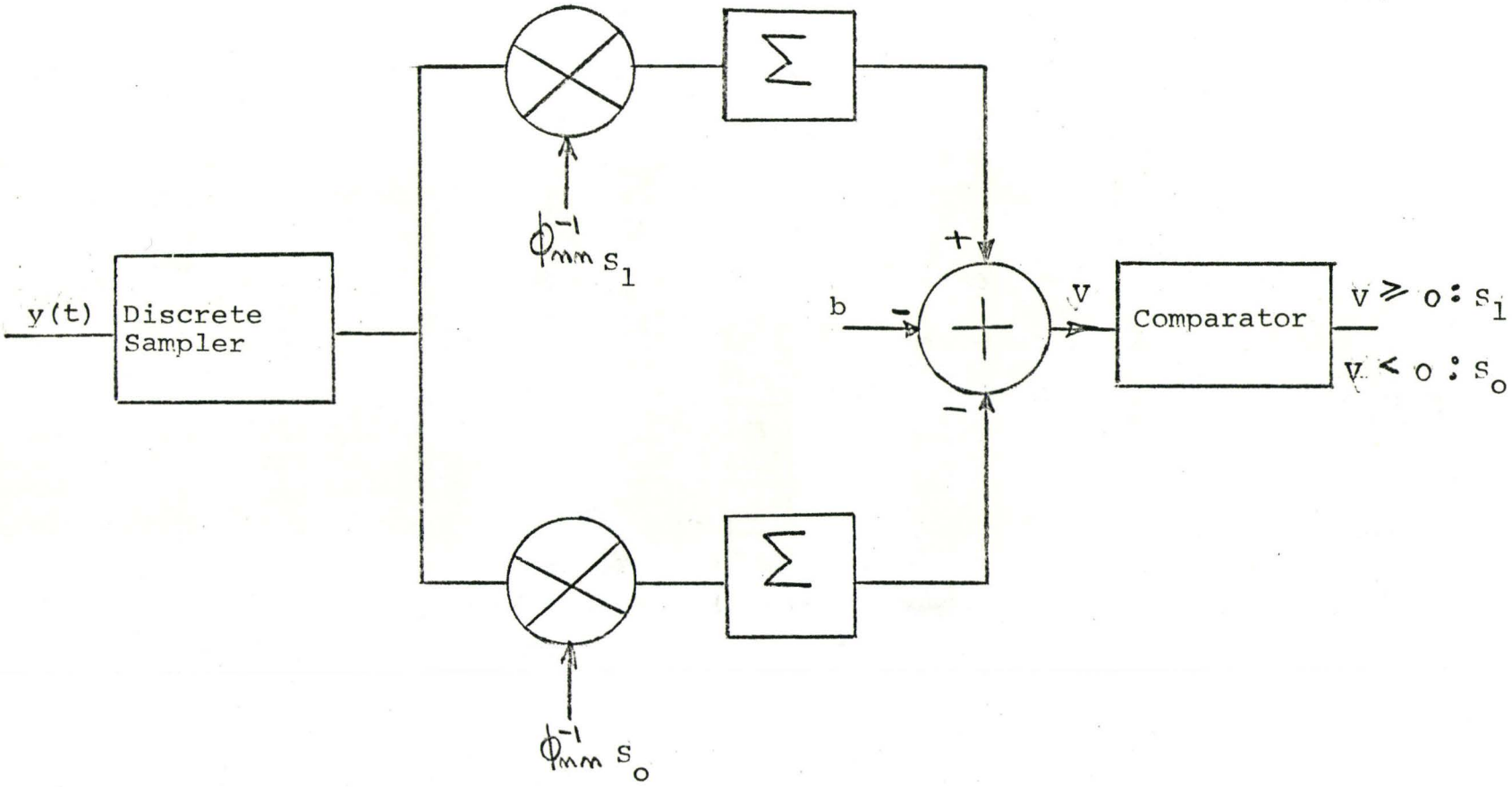


fig.3.4 Receiver Structure

The decision as to which signal was transmitted is then made as follows; if $v<o$ then $s_o(t)$ was transmitted; if $v>o$ then $s_1(t)$ was transmitted.

In most cases we are justified to assume the a priori probabilities equal, the noise white and the signal energies equal. Thus $P_1=P_o=\frac{1}{2}$, $\ln K=o$, $\phi_{nn}=\sigma^2 I$ where I is the identity matrix and σ^{-2} is the mean-squared value of the noise samples, and

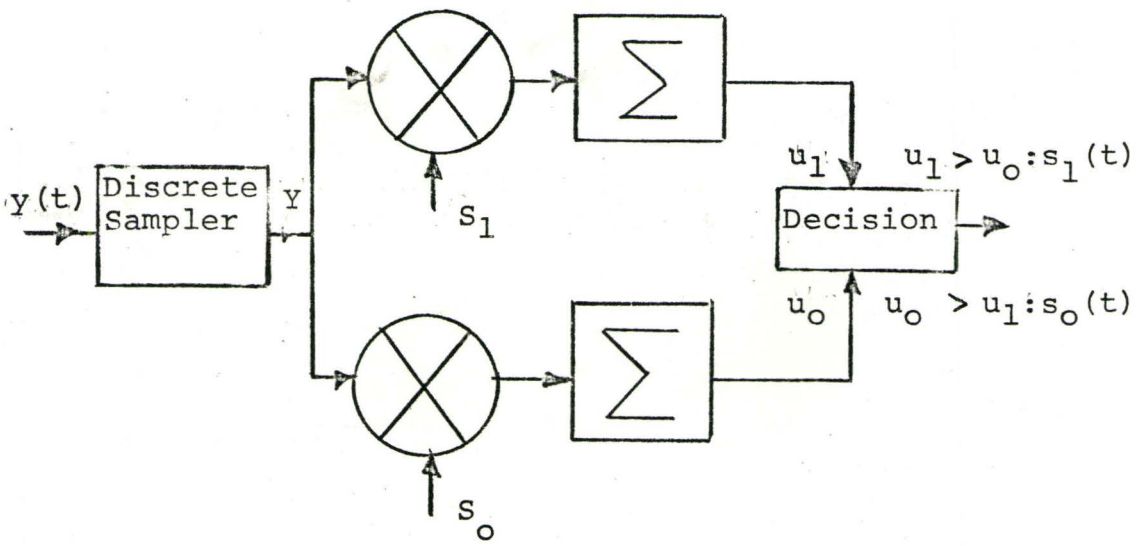
$$S_i^T S_j = E \qquad i=j$$
$$S_i^T S_j = \rho E \qquad i\neq j$$

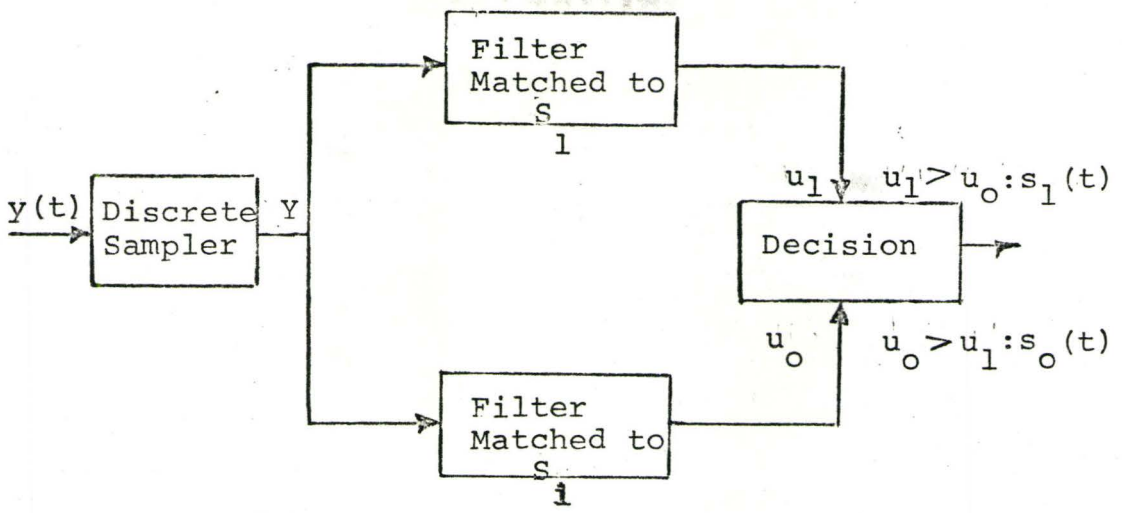
where ρ is the cross-correlation coefficient. Thus, the bias term (equation 3.11) reduces to zero and the decision boundary equation 3.10 becomes,

$$V = Y^T (S_1 - S_o) = o$$

Hence, the receiver must perform a discrete correlation of the received data with stored replicas of each of the two signals S_o and S_1. In figure (3.5) we have two possible discrete systems for implimenting equation (3.13). In the second of two implimentations the discrete correlator has been replaced by a discrete matched filter.

a) Correlation Detector





b) Matched Filter Detector

fig.3.5 Receiver Structures

3-5 PROBABILITY OF ERROR

The measure of performance of the optimum receiver derived in the previous section is the probability that it will make an error, that is, the probability that the receiver will decide in favour of $s_1(t)$ when, in fact, $s_o(t)$ was transmitted, and the probability that the receiver will decide in favour of $s_o(t)$ when, in fact, $s_1(t)$ was transmitted. This overall probability of error can be written as

$$P_E = P_o P_E(0) + P_1 P_E(1) \tag{3.14}$$

where P_o and P_1 are the a priori probabilities and $P_E(j)$ is the probability of error when $s_j(t)$ is transmitted. Referring to the boundary decision equation

$$v = Y^T(S_1 - S_o) = 0$$

we see that

$$P_E(1) = P(v<0/S_1) = \int_{-\infty}^{0} p(v/S_1)\,dv \tag{3.15}$$

Since the random variable N is Gaussian, first and second order statistics of v are necessary to compute $p(v/s_1)$. Defining v_j to be the random variable v under the hypothesis that the j th signal was transmitted, we have,

$$\overline{v_j} = \overline{(S_j + N)^T (S_1 - S_0)}$$

$$= S_j S_1 - S_j S_0$$

$$= \begin{cases} E(1-\rho) & j=1 \\ E(\rho-1) & j=0 \end{cases} \qquad (3.16)$$

Since the variance of v_j is due only to the noise and is therefore independent of the signal, we can compute the variance under the assumption that Y=N. Therefore,

$$\text{var } v_j = \text{var } N^T (S_1 - S_0)$$

$$= \overline{N^T (S_1 - S_0) N^T (S_1 - S_0)}$$

$$= 2\sigma^2 E(1-\rho)$$

$$= N_0 E(1-\rho) \qquad (3.17)$$

where N_0 is the one-sided spectral density of the white noise.

Therefore,

$$P_E(1) = \int_{-\infty}^{0} \exp\left\{-\left[v - E(1-\rho)\right]^2 / 2N_0 E(1-\rho)\right\} dv \Big/ \sqrt{2\pi N_0 E(1-\rho)}$$

$$= \int_{-\infty}^{-\sqrt{E(1-\rho)/N_0}} \frac{\exp(-y^2/2)}{\sqrt{2\pi}}\, dy = erfc\sqrt{\frac{E(1-\rho)}{N_0}} \qquad (3.18)$$

where we have made the substitution,

$$y^2 = \left[v - E(1-\rho)\right]^2 / N_0 E(1-\rho)$$

and we defined

$$\int_{-\infty}^{-z} \frac{e^{-y^2/2}}{\sqrt{2\pi}}\, dy = erfc(z)$$

In a similar manner we can show that $P_E(1) = P_E(0)$

and so,

$$P_E = \tfrac{1}{2}\,\mathrm{erfc}\sqrt{\frac{E(1-\rho)}{N_o}} + \tfrac{1}{2}\,\mathrm{erfc}\sqrt{\frac{E(1-\rho)}{N_o}}$$

$$= \mathrm{erfc}\sqrt{\frac{E(1-\rho)}{N_o}} \qquad (3.19)$$

Since the complimentary error function is a monotonic decreasing function of its arguement, we see that the error probability for coherent reception in the presence of white noise is a decreasing function of the ratio of signal energy to the noise density E/N_o and is an increasing function of the normalized inner product ρ. In particular, the error probability does not depend on the signaling waveforms. In figure (3.6) we have plotted probability of error curves for two different values of cross-correlation coefficient. It is obvious from the expression for probability of error (equation 3.19) that the best choice for ρ is $\rho = -1$. Physically we might expect this result, since we have shown that the optimum receiver decides in favour of the signal which most resembles the received data. For $\rho = -1$, the signals are as dissimilar as possible. The most often used system for generating anticorrelated signals is phase-shift-keying.

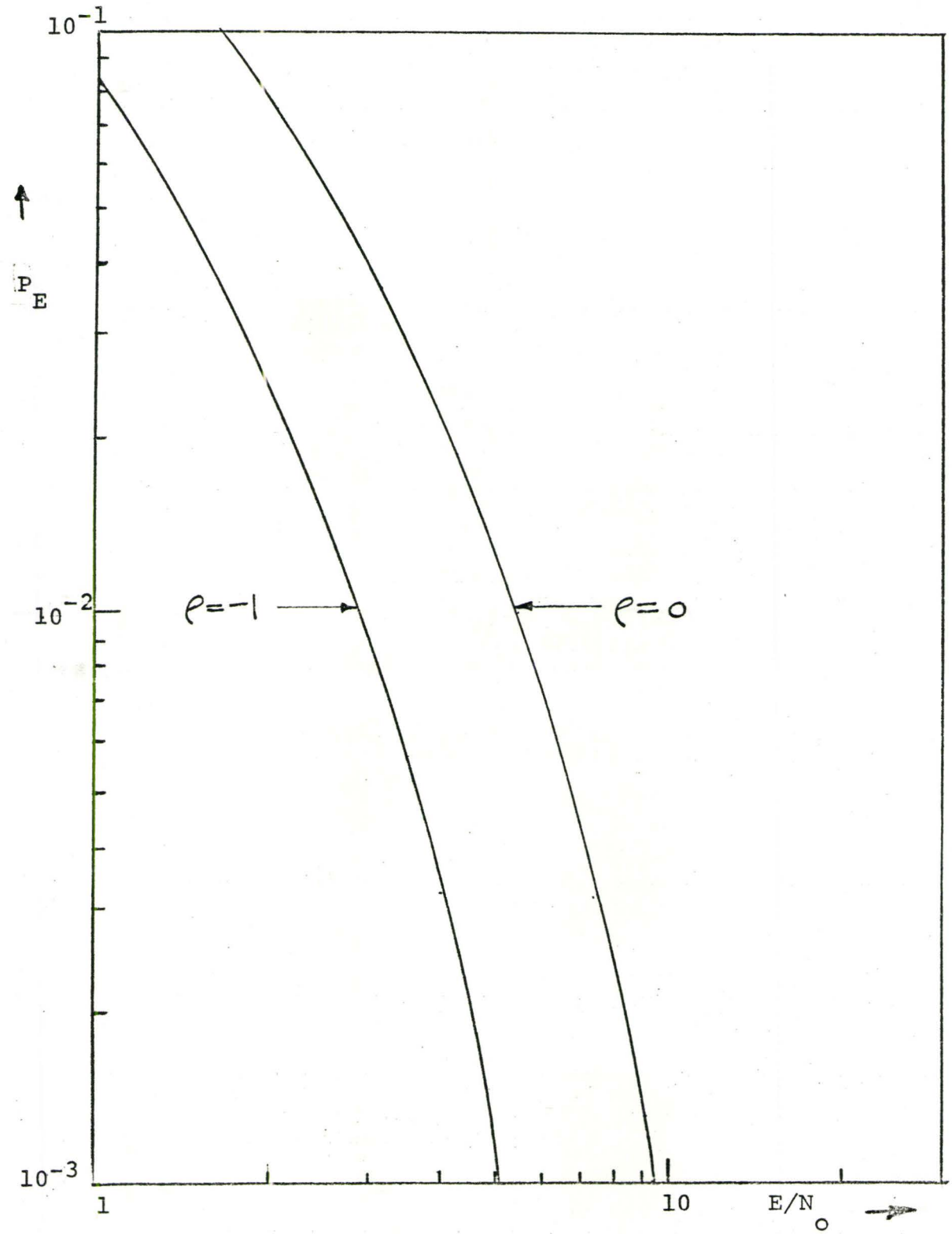


fig.3.6 Probability of Error Curves for Coherent Detector

The second curve of figure 3.6 is for $\rho = 0$ which corresponds to a system in which the signals are orthogonal, that is,

$$\int_0^T s_1(t) s_0(t)\, dt = 0$$

The most often used system for generating orthogonal signals is frequency-shift-keying, where the carrier frequencies are chosen such that the signal spectra do not overlap.

3-6 NON COHERENT RECEPTION[2,3]

In the last section we considered the optimum receiver implimentation for two arbitrary but known signals (coherent reception). In most practical communication systems, however, it is unlikely that precise knowledge of the signal phase is available at the receiver. As a result, we can no longer say that the signals are known and the implimentations of the last section will not be the optimum.

In this section we will present the optimum non-coherent detector. The techniques used in the derivation of the detector are similar to those of the last section and therefore will not be repeated.

Since in the experimental work, which is outlined in the next chapter, we considered only the case of orthogonal signals (frequency-shift-keying), we will present the non-coherent detector for the particular case of orthogonal signals. Thus, the signals may be written as,

$$\begin{aligned} S_j(t,\phi) &= m(t)\cos(\omega_j t+\phi) \qquad j=0,1 \\ &= m(t)\cos\omega_j t(\cos\phi)-m(t)\sin\omega_j t(\sin\phi) \\ &= m_c(t)\cos\phi - m_s(t)\sin\phi \end{aligned} \tag{3.20}$$

Again, using the vector notation,

$$\begin{aligned} Y_j &= S_j(\phi)+N \\ S_j(\phi) &= M_{c,\omega_j}\cos\phi - M_{s,\omega_j}\sin\phi \end{aligned} \tag{3.21}$$

Under the assumption that the a priori probability density function for the phase of the signal is,

$$p(\phi) = 1/2\pi \quad , \quad -\pi \le \phi \le \pi$$

the likelihood ratio can then be written as

$$L(Y)=\frac{I_o\left(\sqrt{(Y^T\phi_{nn}^{-1}M_{c,\omega_1})^2+(Y^T\phi_{nn}^{-1}M_{s,\omega_1})^2}\right)}{I_o\left(\sqrt{(Y^T\phi_{nn}^{-1}M_{c,\omega_o})^2+(Y^T\phi_{nn}^{-1}M_{s,\omega_o})^2}\right)}=\frac{I_o(\epsilon_1)}{I_o(\epsilon_o)} \tag{3.22}$$

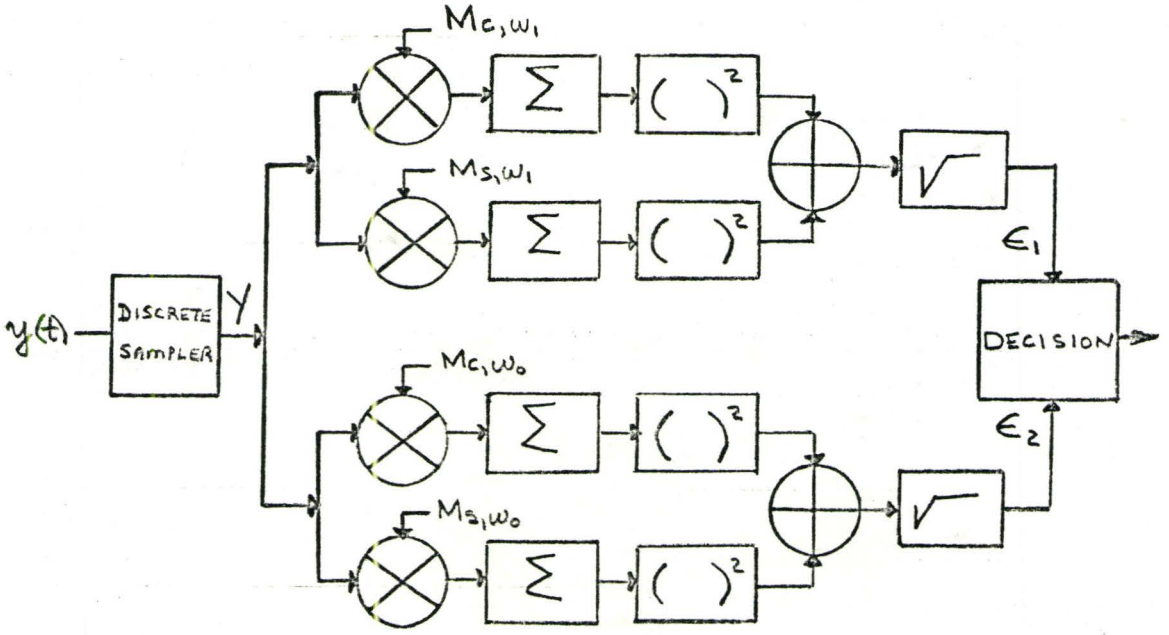
where I_o is the modified Bessel function of the first kind of order zero. But since the Bessel function is a monotonically increasing function of its arguement, it is necessary only to compute and then compare the arguements ϵ_o and ϵ_1. The decision rule will be:

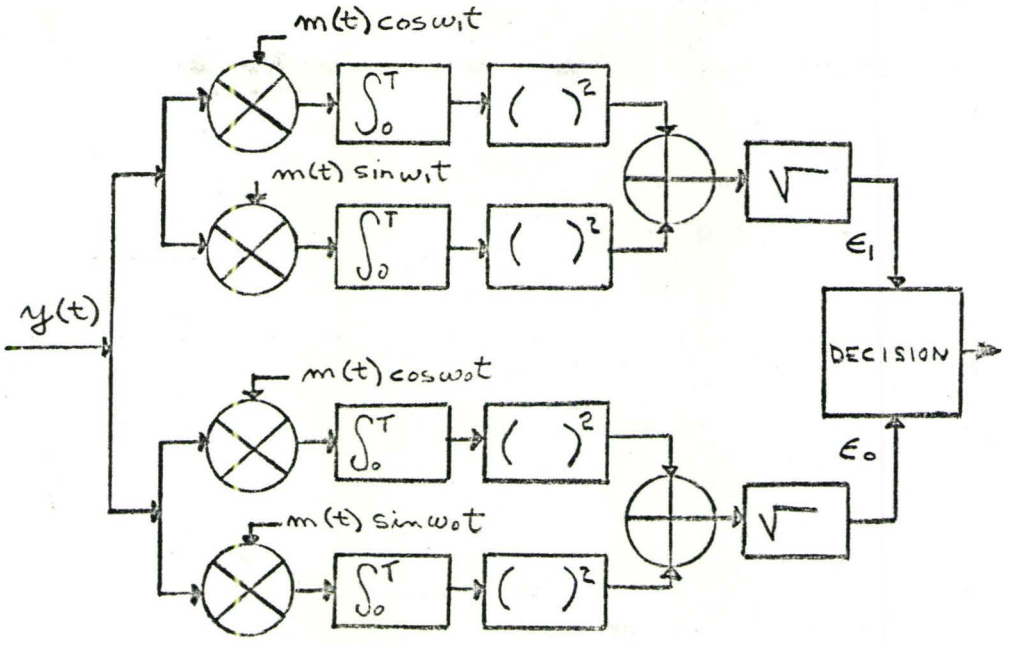
$s_1(t)$ was transmitted if $\epsilon_1 > \epsilon_o$

$s_o(t)$ was transmitted if $\epsilon_o > \epsilon_1$

Under the assumption that the noise is white, the inverse covariance matrix of the noise equals the mean-squared value $\sigma^{-2}I\left(\overline{n^2(t)}=\sigma^2\right)$ and the receiver structures necessary to compute and compare ϵ_1 and ϵ_o is illustrated in figure (3.7). The first implimentation is obvious from equation (3.22) above and the second implimentation follows from the first under the assumption that orthonormal basis functions are used in the discrete representation. However, the third implimentation illustrated requires a brief explanation. Based on the theory presented in the previous sections, we

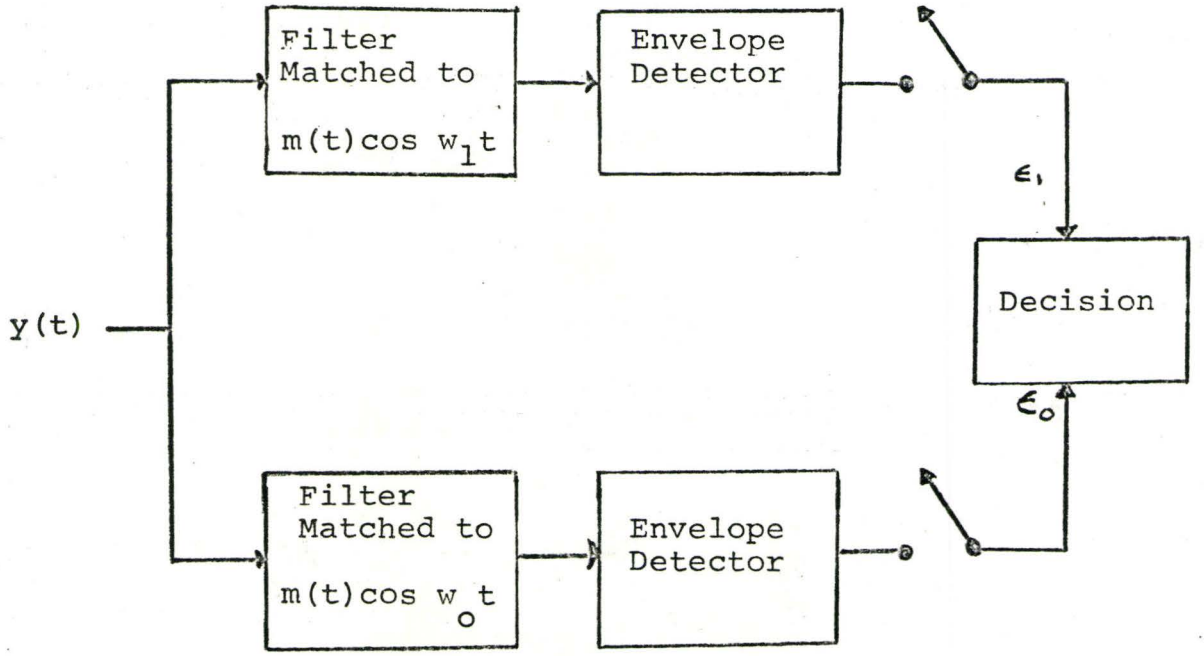
a) Discrete Optimum Noncoherent Detector





b) Continuous Optimum Noncoherent Detector

fig.3.7 Implimentation of Optimum Non-Coherent Receivers



c) Optimum Noncoherent Detector

fig.3.7 (continued)

see that the upper branch of the second implimentation (which is an optimum detector for the signal $m(t)\cos(\omega_1 t+\phi)$, ϕ unknown) can be interpreted as an optimum detector for the signal $m(t)\cos\omega_1 t$ in parallel with an optimum detector for the signal $m(t)\sin\omega_1 t$. The outputs of these two quadrature detectors are squared and summed, indicating that $\epsilon_1(t)$ is the envelope of the output of the optimum detector for $m(t)\cos(\omega_1 t+\phi)$ where ϕ is an angle between 0 and 2π. It is important to note that the receiver destroys the useless carrier phase information by basing the decision on the envelope of the output of a filter matched to $m(t)\cos\omega_1 t$. That is, the filter is matched to the modulation waveform $m(t)$ and carrier frequency, but no attempt is made to match to the carrier phase. Of course, a conventional linear bandpass filter centered on ω_1 and with a bandwidth equal to the bandwidth of $m(t)$ is the required matched filter.

3-7 PROBABILITY OF ERROR FOR INCOHERENT RECEPTION

The mathematical techniques used in the derivation

of the probability error for the incoherent receiver are similar to those used in the previous section on coherent receivers. Thus, it may be shown that,

$$P_E = \tfrac{1}{2} \exp(-E/N_o) \tag{3.23}$$

where E is the signal energy and N_o is the one-sided spectral density of the additive white noise.

3-8 COMPARISION OF COHERENT AND INCOHERENT DETECTORS

The expression for, probability of error for the coherent and incoherent detectors, are plotted in figure (3.8). For the sake of the illustration, we have let the cross correlation coefficient be zero which corresponds to a frequency-shift-keying system. Evidently the coerent receiver performs considerably better (approximately 3dB) for the low values of signal to noise ratio. However, at high signal to noise ratios, the knowledge of carrier phase contributes little (approximately 1dB) to system performance. We can conclude, that the slight improvement in performance of the coherent system does not

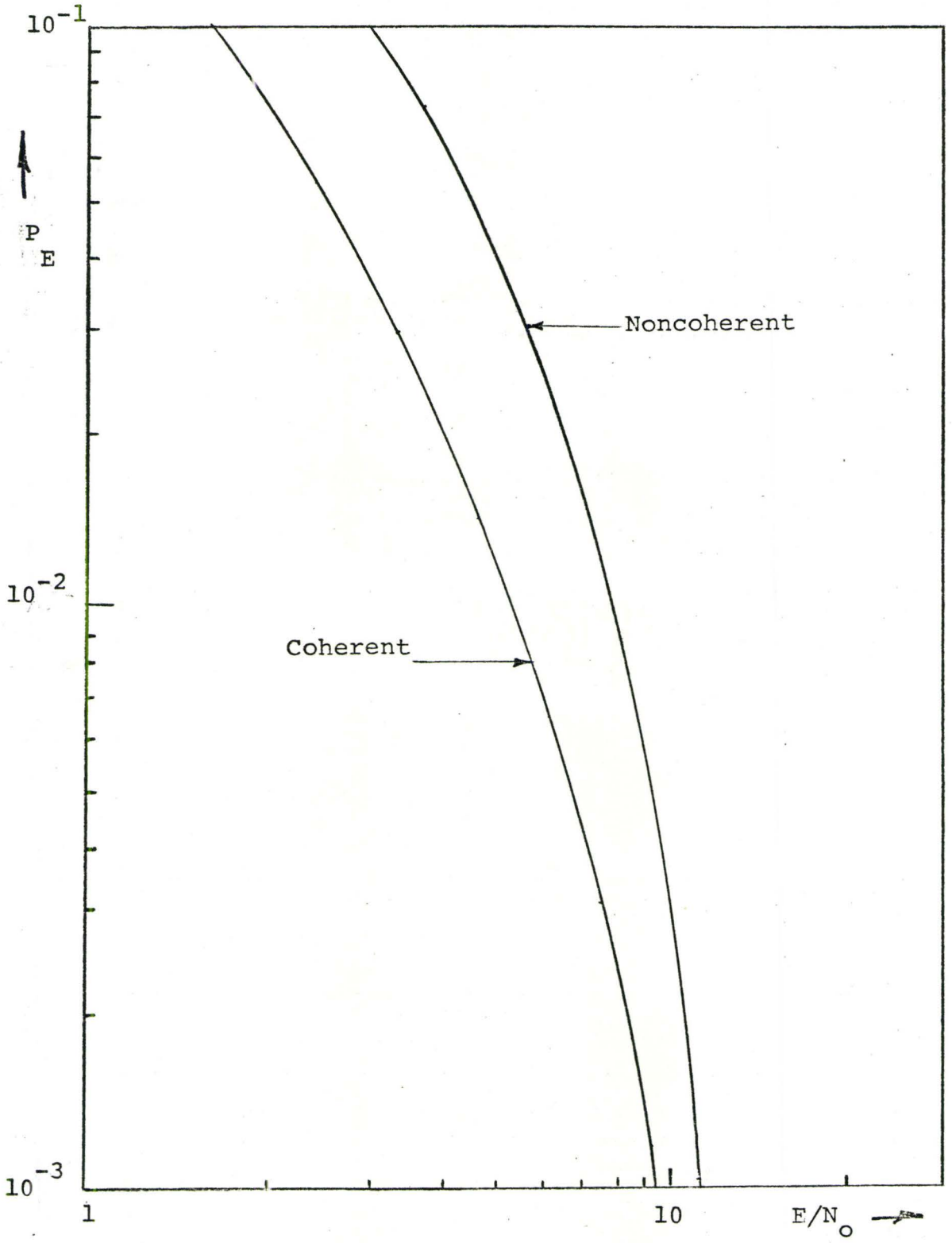


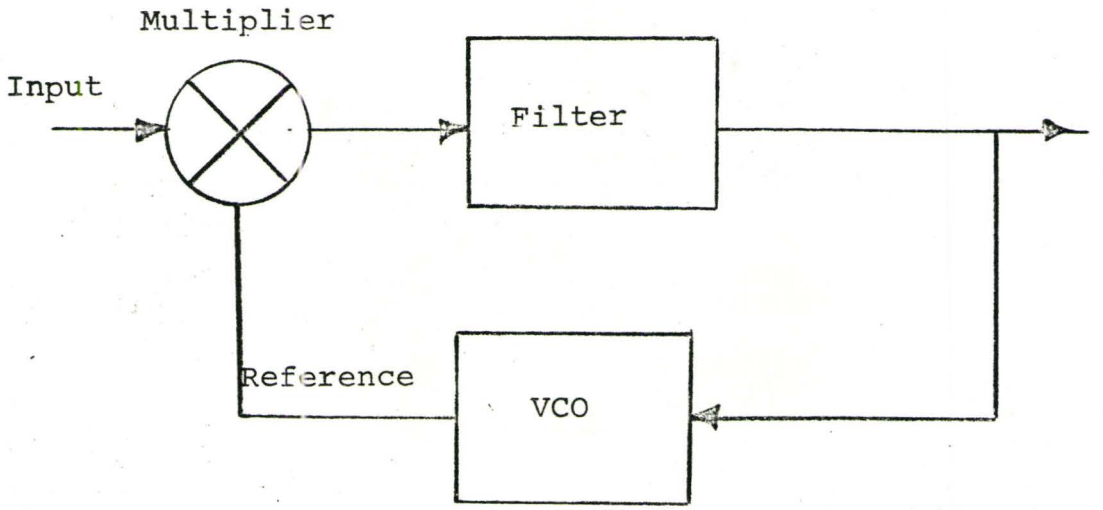
fig.3.8 Probability of Error Curves for Coherent and Non-Coherent Detectors for $\rho = 0$

warrant the diffulties encountered in the system design. In general, if a receiver can be operated coherently, that is, exact replicas of the transmitted signals will be available at the receiver, it is to the designer's advantage to use antipodal signals rather than orthogonal signals because of the possible higher performance of the former.

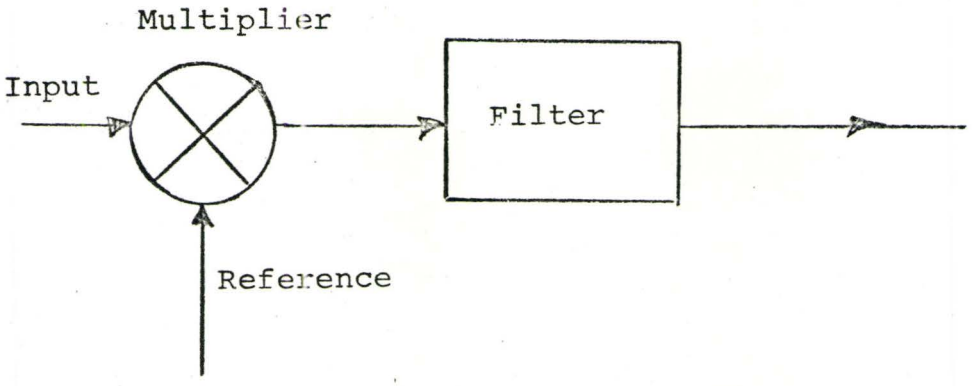
3-9 THE PHASE-LOCK LOOP DETECTOR

The block diagram of a phase-lock loop and a correlator are illustrated in figure (3.9). If we compare the two block diagrams, it appears that the phase-lock loop is just a correlator with feedback provided through a voltage-controlled oscillator (VCO). That is, both systems involve the multiplication of an incoming signal with a reference signal and subsequent low pass filtering or integration of the product. The only difference between the two systems is the source of reference signal. For the correlator, the reference is derived from a second system which is completely independent of the operation of the correlator. However, as described earlier, the reference for a phase-lock loop is derived within the loop itself.

For the coherent correlation detectors, as previously discussed, no mention was made of the source of



a) Phase-Lock Loop



b) Correlator

fig.3.9 Comparison of Phase lock Loop and Correlators

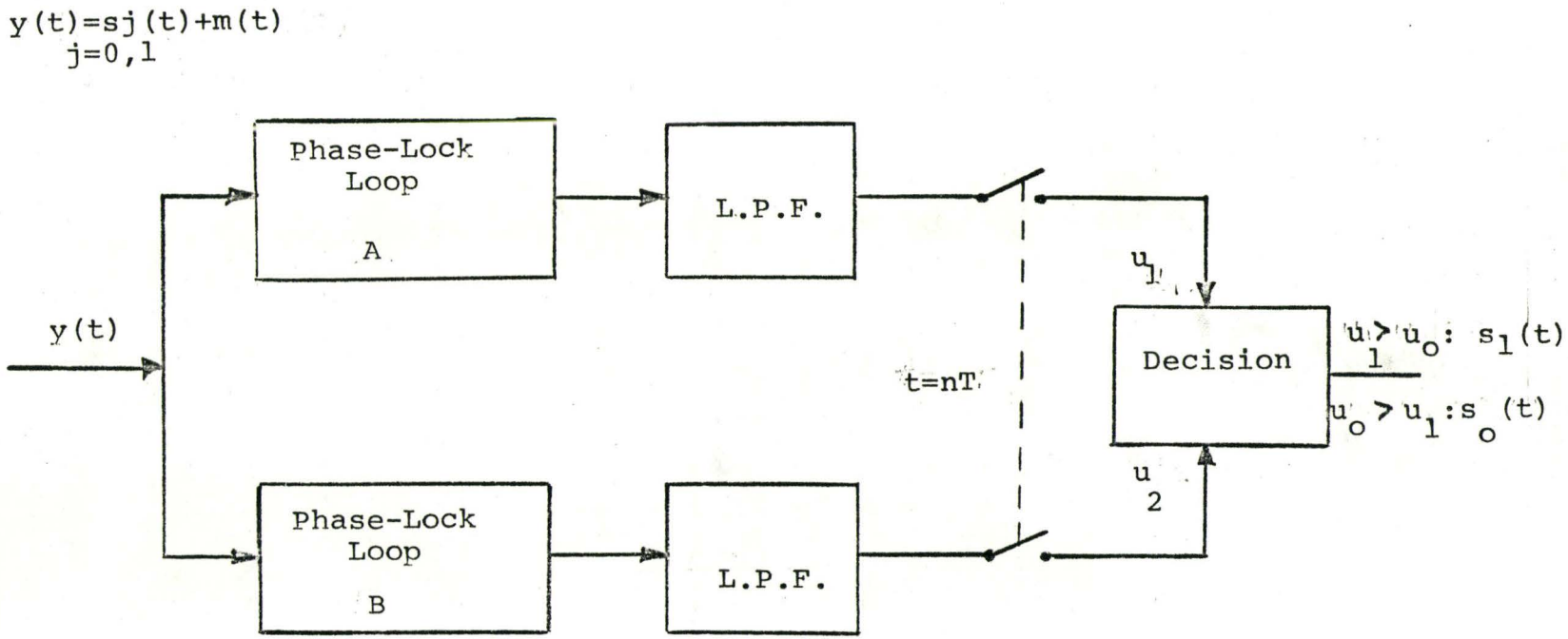


fig.3.10 Phase-Lock Loop Detector

reference signals. It was assumed that the references were available and more important, that they were in phase and frequency synchronism with the possible transmitted signals. For a normal communications system in which transmitter and receiver are located great distances apart, this necessary synchronization leads to complex design problems. For the less frequent, but important, case in which there is relative motion between the transmitter and receiver the problem becomes virtually impossible, unless a method of tracking the transmitted signals is used. Of course, the phase-lock loop is just such a tracking system.

From what has been said above, it seems possible that a phase-lock loop might be used for both the generation of appropriate reference signal and subsequent correlation. This being so, then the correlators in the detectors of the previous section may be replaced by phase-lock loops and the problem of providing suitable references is thereby eliminated. The detector would then be as illustrated in figure (3.10) and we shall refer to it as the phase-lock loop detector.

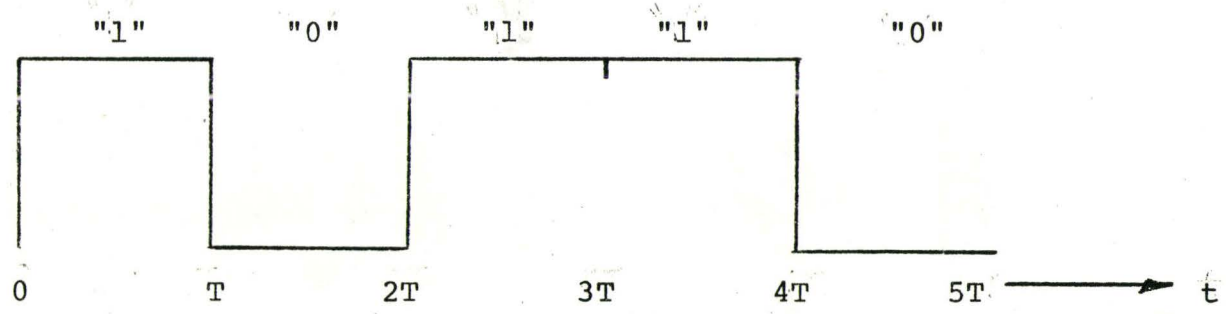
In figure (3.10) we have assumed that loop A is able to respond only to $s_1(t)$; that is, the loop cannot lock onto $s_o(t)$. Similarly, loop B can lock only onto $s_o(t)$. This means that for the system to be able to determine which

of the two signals has been transmitted, the frequencies of the two signals must be different and therefore the detector is restricted to frequency-shift-keyed signals only.

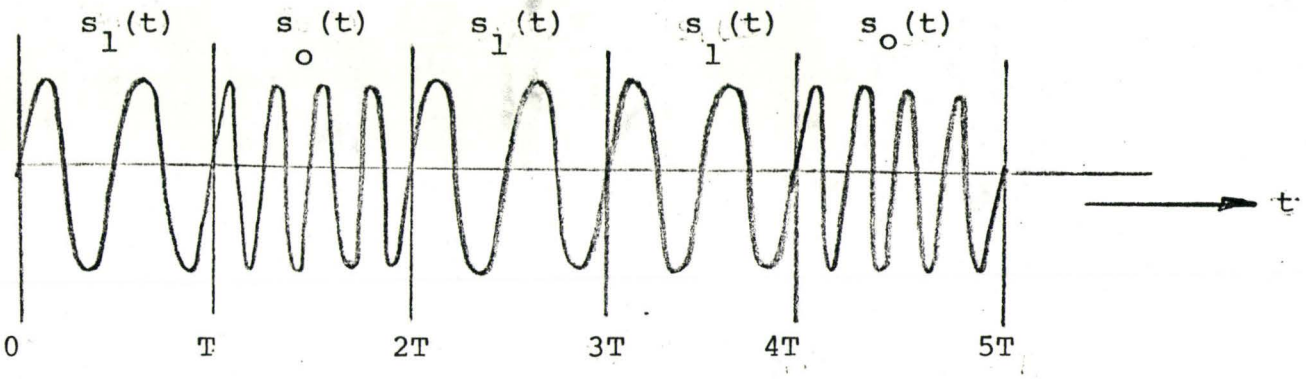
The operation of the detector can best be explained by referring to figure (3.11) where it is assumed that the binary sequence 10110 is to be transmitted by the waveform illustrated in part (b) of the figure. At time zero, $s_1(t)$ is applied at the input of the detector. Since the signal lies within the lock range of loop A, a reference is generated and correlation takes place. If we assume that loop A is able to lock instantaneously and that the frequency of the signal approaches the upper limit of the lock range, then the loop output is

$$AK_1 \sin(\phi) = AK_1 \cos(\phi - \pi/2)$$

where A is the RMS value of $s_1(t)$ and K_1 is the RMS value of the VCO output. We have further assumed that any high frequency terms are eliminated by the low pass filter. By appropriate choice of the low pass filter cut-off frequency, the output of the filter at time T will be $AK_1T \cos(\phi - 90^\circ)$. For $\phi = 90^\circ$ this is exactly the output of a correlator under similar circumstances; that is, AK_1T.



a) Binary Signal



b) Frequency-Shift Keyed Signals

fig.3.11 Waveforms for FSK

Loop B, on the other hand, is unable to lock to the signal and its output would be a beat note which would be removed by the low pass filter. Thus, referring to figure (3.10) at time T^-, u_1 would be greater than u_o and the detector would correctly decide in favor of the signal $s_1(t)$.

At time T^+, a zero is transmitted and $s_o(t)$ appears at the input of the detector. Loop B will respond and lock onto the signal; a reference is generated and correlation takes place. Loop A will go from a locked to unlocked state producing a beat note at its output for the duration of signal $s_o(t)$. Thus at time $2T^-$, $u_o = TAK_1$ is greater than u_1 and the detector correctly decides in favour of $s_{o(t)}$.

Operation of the detector would continue as above. For each transition in the binary sequence, there will be a corresponding transition within the detector, that is, loops changing from a locked to an unlocked state or vice versa.

Based on the discussion above, we can appreciate that the detector using "ideal" phase-lock loops will behave exactly as the correlation detector and its performance in terms of probability of error should be equivalent to that of the correlation detector. However, in actual practice it is impossible to design and construct an ideal loop. Using practical loops, the performance of the detector will be degraded relative to that of the correlation detector for two reasons.

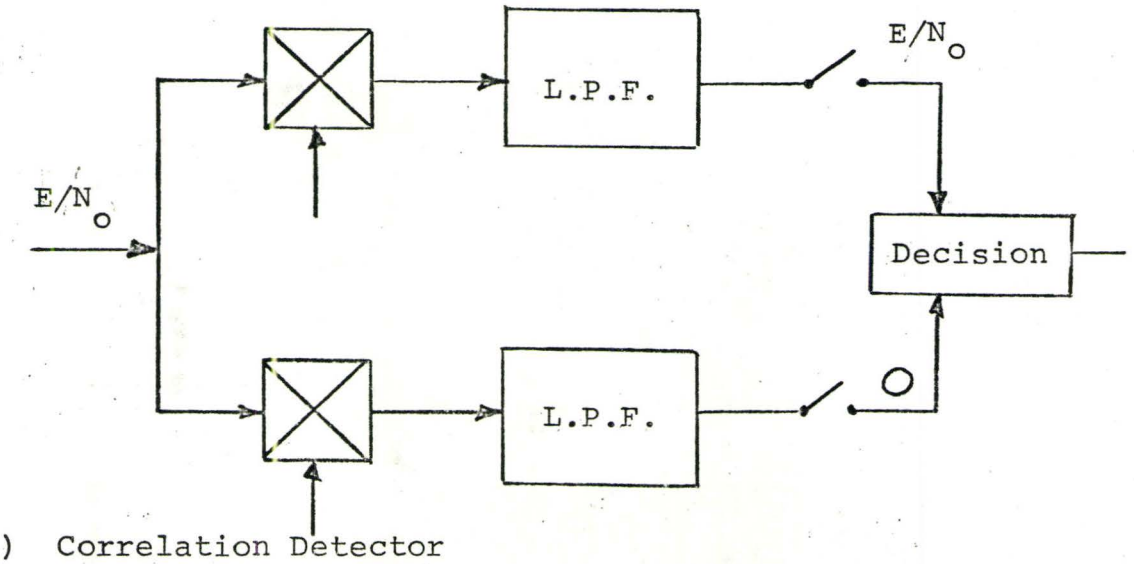
1) One of the assumptions made, was that the loops were able to lock instantaneously; that is, the appropriate loop would be locked for the entire bit duration and therefore the reference signal would be present for the bit duration. Of course, this is an impossibility; a finite time is required for a loop to lock. Thus, after each transition in the binary sequence, the appropriate loop will go through a locking-in process during which correlation does not take place. At the end of the bit, the output of the low pass filter associated with the loop, will therefore not be AK_1T but rather $AK_1(mT)$ where m is a positive fraction. That is, during the first (1-m)T seconds of the bit, the loop is not locked and therefore correlation is not taking place.

2) The second assumption was that the loops were designed such that when locked, the phase error would be 90 degrees and therefore the phase difference between the incoming and reference signal would be zero. This implies that the loop is operating at the upper limit of its lock range and if any disturbance should cause the phase error to exceed 90 degrees the loop will momentarily lose lock. In the first chapter it was pointed out that additive Gaussian noise causes a jitter in the output of the VCO and therefore a jitter in the phase error. Thus, if the phase error has a mean value of 90 degrees, there is a large

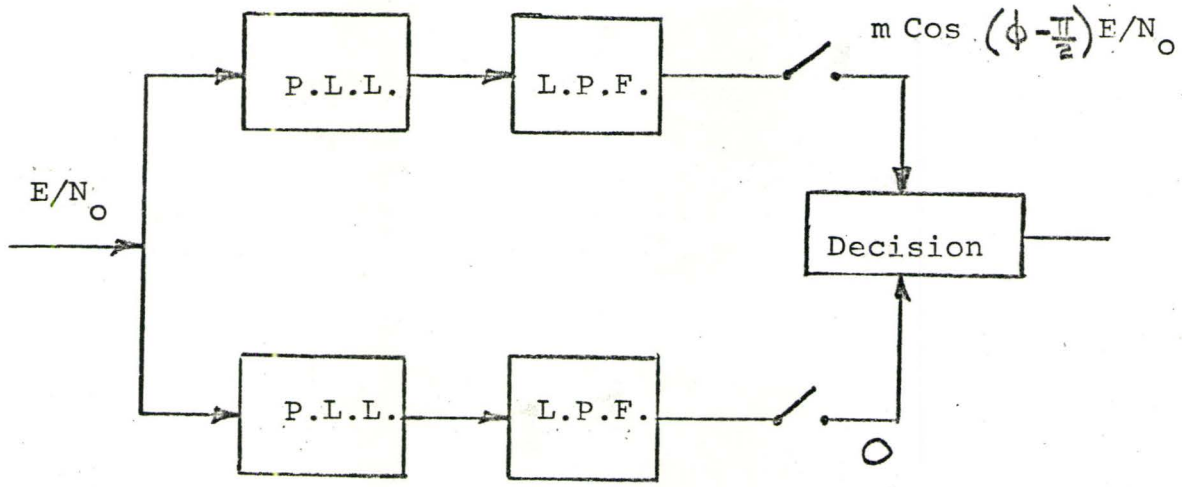
probability that the loop will momentarily lose lock in the presence of the additive noise. To reduce the probability of losing lock it is necessary to make this mean value or phase error of the loop less than 90 degrees. The output of the filter would then be $AK_1(mT)\cos(\phi-90)$.

To further appreciate how the two factors mentioned above degrade the performance of the detector, we can refer for the moment back to the correlation detector. We have already said that at the end of a bit, the output of one channel of the detector is AK_1T. If we make the RMS value of the reference signal equal to that of the input signal, then this output becomes A^2T. But, this is just the energy contained in the input signal during one bit duration of T seconds. Thus for the correlation detector, at the time a decision is made, we have at the output of one channel, the maximum possible signal energy and therefore, for any noise of spectral density N_o the maximum possible signal to noise ratio E/N_o.

For the same situation, we have at the output of the channel of the phase-lock loop detector $A^2T(m)\cos(\phi-90)$ which is something less than maximum signal energy and and therefore at the point of decision, the signal to noise ratio is not maximized. Signal to noise ratios for each detector are shown in figure (3.12).



a) Correlation Detector



b) P.L.L. Detector

fig.3.12 Signal to Noise Ratios for Detectors

Based on the comparison of signal to noise ratios for the two detectors, we might expect that the probability of error curves will be identical in shape but displaced from one another; that is, for identical input signal to noise ratios, the performance curve for the phase-lock loop detector will be displaced to the right. The amount of displacement will depend primarily on the two factors discussed above. We can, of course, reduce this displacement by making the factor m approach unity, which implies designing the loops to lock as quickly as possible. This can be accomplished by increasing the loop bandwidth. We can also improve the performance by making the phase error approach 90 degrees, which means the loops be designed for a minimum phase jitter in the presence of additive noise. A reduction in phase jitter requires a reduction in loop bandwidth. Thus, an improvement in detector performance involves conflicting requirements and a compromise in loop bandwidth is necessary. We might expect then, that the experimental results presented in the next chapter will indicate an optimal loop bandwidth, above and below which the performance of the system decreases.

CHAPTER 4

EXPERIMENTAL RESULTS

4-1 INTRODUCTION

In this chapter we shall give experimental results describing the performance, in the presence of noise, of a detector using digital phase-lock loops and compare this performance with that of a coherent detector using correlators. To make the experimental results meaningful, the measurements on both systems were made under the same conditions.

4-2 THE TEST ARRANGEMENT

A block* diagram of the test arrangement, used to determine the performance curves for the detectors, is illustrated in figure (4.1). The noise generator used for

* The schematic diagrams of all circuits constructed are shown in appendix (A).

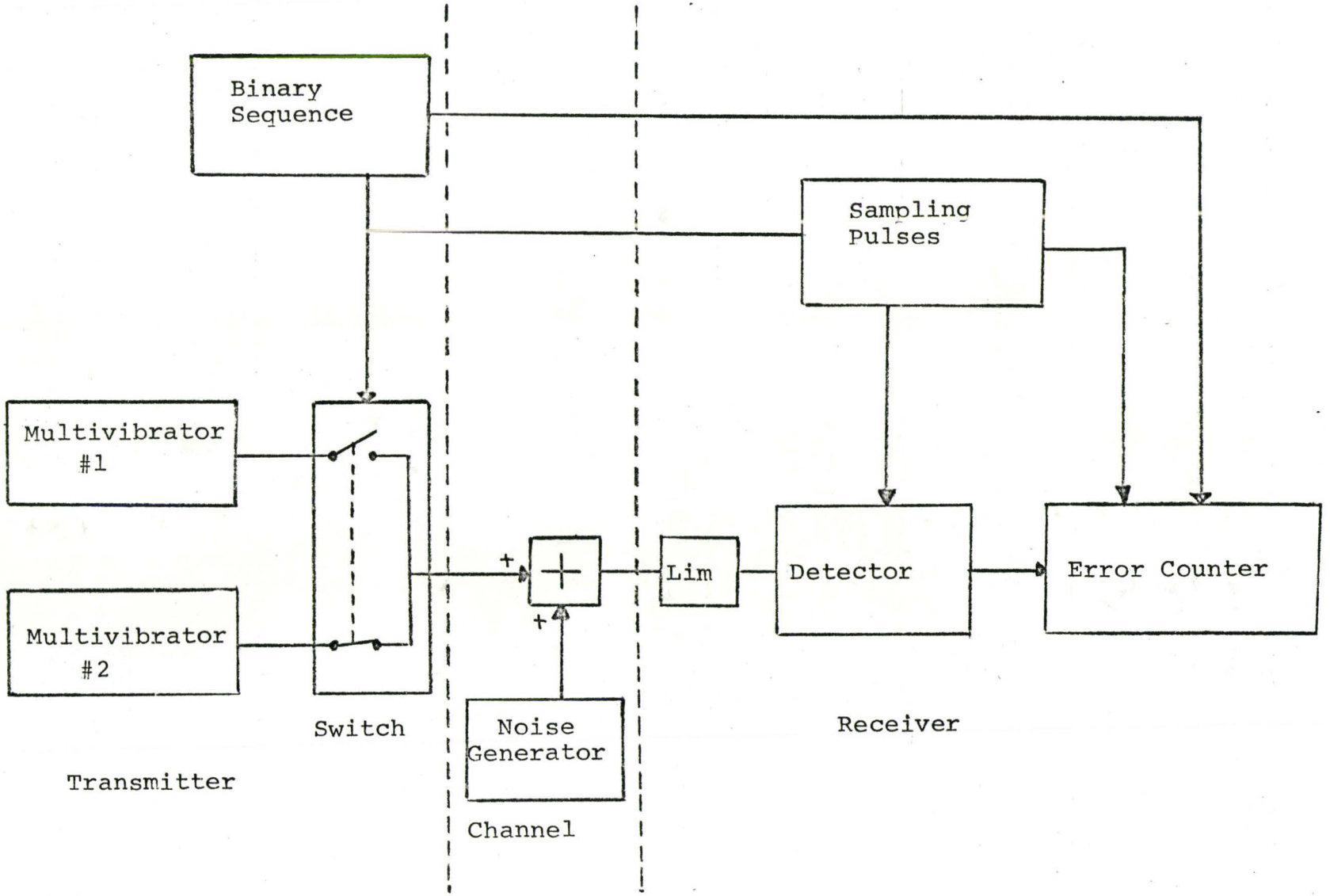


fig.4.1 Test Arrangement

the test had a cut-off frequency in the order of 20KHz. Therefore the system had to be operated with carrier frequencies which would fall in this range. Further, we chose to operate near the low end of this range so that, in the event pre-detection filtering of the carriers was necessary, low pass rather than bandpass filters could be used.

In the implimentation of the so-called "transmitter", two schemes were available. In the first, only one oscillator was used and its output was frequency modulated by the binary sequence to be transmitted. In the second, two oscillators were used and the binary sequence caused a switching action between the oscillators; that is, each time there was a transition in the binary sequence the output of the transmitter electronically switched from one oscillator to the other. The latter arrangement, although requiring more circuitry, has the advantage that one carrier frequency can be varied independent of the other. For this reason, the two-oscillator scheme was chosen.

Since both detector systems were implimented using logic elements, it was necessary to restrict their input signal amplitudes, to the logic levels of the elements used (approximately 0 and 3 volts). That is, the carrier signals applied to the detectors had to be

rectangular waves. To insure that this would be the case, a limiter with an input-output characteristic as in figure (4.2) was included, at the input to the detectors. With such a characteristic, the input signals to the limiter can have any waveshape and the output will be a rectangular wave with the same fundamental frequency. Further, the output will be compatible with the logic elements used in the detectors. The only requirement of the input signal to the limiter is, that it must have only two zero crossings in each cycle. This requirement insures that, the input signal to the detectors makes the transition from the low level to the high level once each cycle. The rectangular wave outputs of astable multivibrators meet this requirement and, for this reason, multivibrators were used as the oscillators in the "transmitter". They were constructed in such a way that their output frequencies could be readily varied.

The summing point at which the signal and noise were added, was constructed using an operational amplifier.

The source of binary data was a commercially built square wave oscillator, which gave us the sequence 101010 10..... The reason for choosing this particular sequence is as follows. In many systems that operate on digital data, it is found that their performance will vary with the binary sequence applied. There is no evidence to

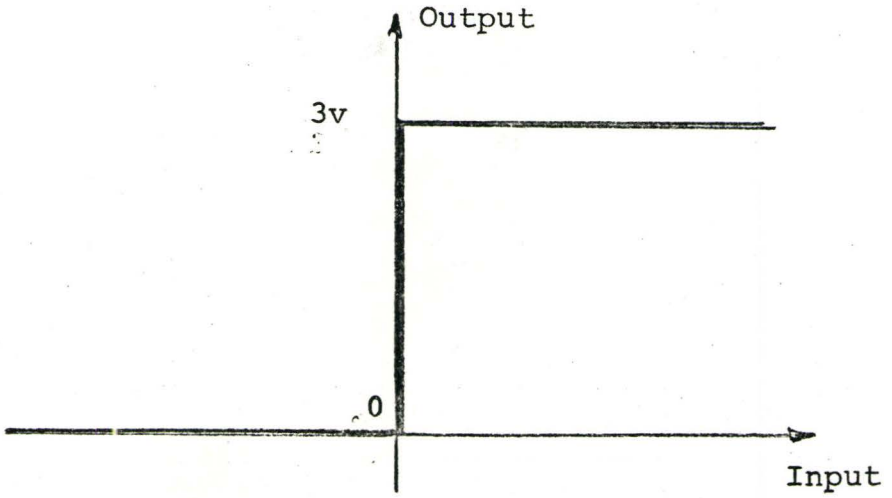


fig.4.2 Limiter Characteristic

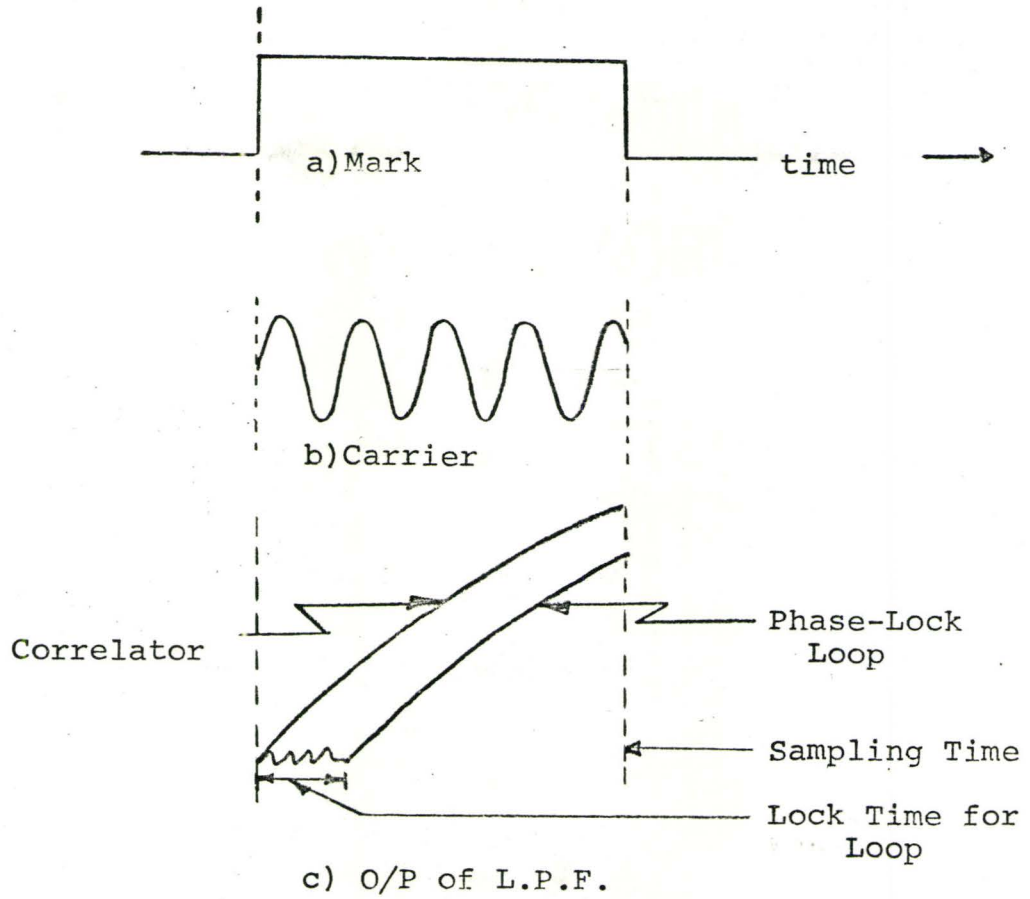


fig.4.3 Output of L.P.F for Both Detectors

indicate that this is true for the coherent correlation detector. However, based on the discussion in the previous chapter we can expect this to be true for the phase-lock loop detector. We have pointed out that because the loops are unable to lock instantaneously the performance of the detector will be degraded. That is, each time a loop goes through a locking-in process, a portion of the signal energy is lost, or rather will not be available at the point at which the decision is made. This is illustrated in fig.(4.3), where we have shown the output of a low pass filter in both the correlation detector and the loop detector. It is apparent, that at the instant of sampling the output in the correlation detector is greater than that in the loop detector and therefore, the probability of making an error is correspondingly higher for the loop detector. Thus we might expect that the performance of the loop detector will depend on the number of transitions in the binary sequence; or equivalently, the number of times the loops must go through a locking-in process. As the number of transitions is increased, in a sequence of fixed length, the probability of making an error will likewise increase. The sequence which we have chosen has the maximum possible number of transitions and therefore we will be able to determine in our tests, the maximum possible probability of error for

any signal to noise ratio. Further, we can expect, that for any other binary sequence, the performance of the detector will be at least as good, if not better.

Also, the choice of this sequence, simplified the circuitry required to derive the sampling pulses necessary for the operation of the detector. The method used is discussed in appendix (A).

In the error counting circuit, both the outputs of the sequence generator and detector were sampled once each bit. The samples were compared and, if different, a pulse was applied to a commercially built counter. In this way, a running count of the number of errors could be made over any length of time.

In figures (4.4) and (4.5) we have illustrated one branch of the phase-lock loop detector and correlation detector, respectively. For both systems, the multiplier is implimented using an "exclusive or," which is sometimes referred to as an anti-coincidence multiplier. That the "exclusive or", performs as a multiplier is evident from fig.(2.12), where we have shown a linear relationship between the phase error of a digital loop and the average value of the "exclusive or" output.

The low pass filter and sampling circuit are identical in both detectors. The outputs of the two branches in either detector were applied to a comparator

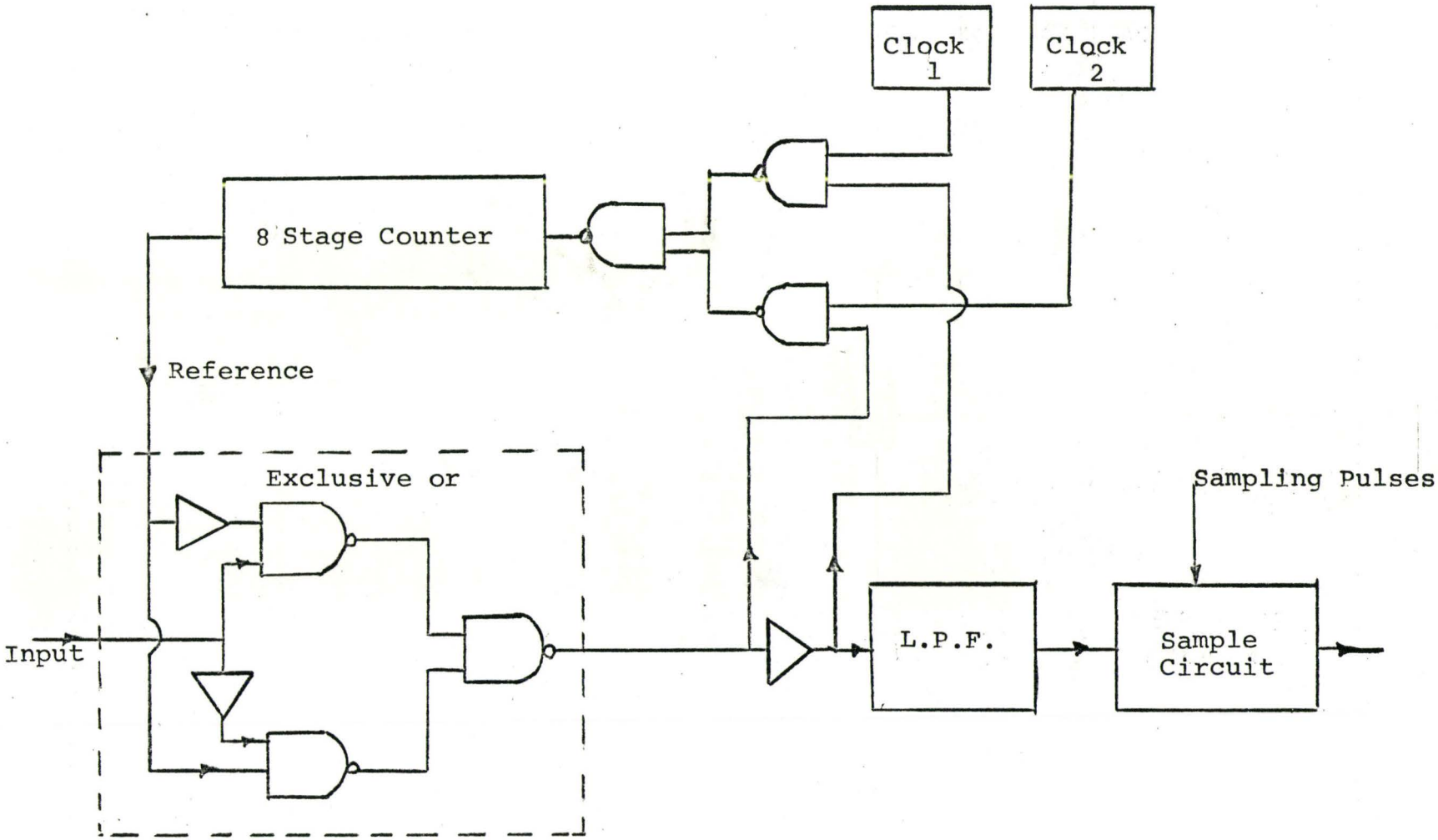


fig.4.4 One Branch of PLL Detector

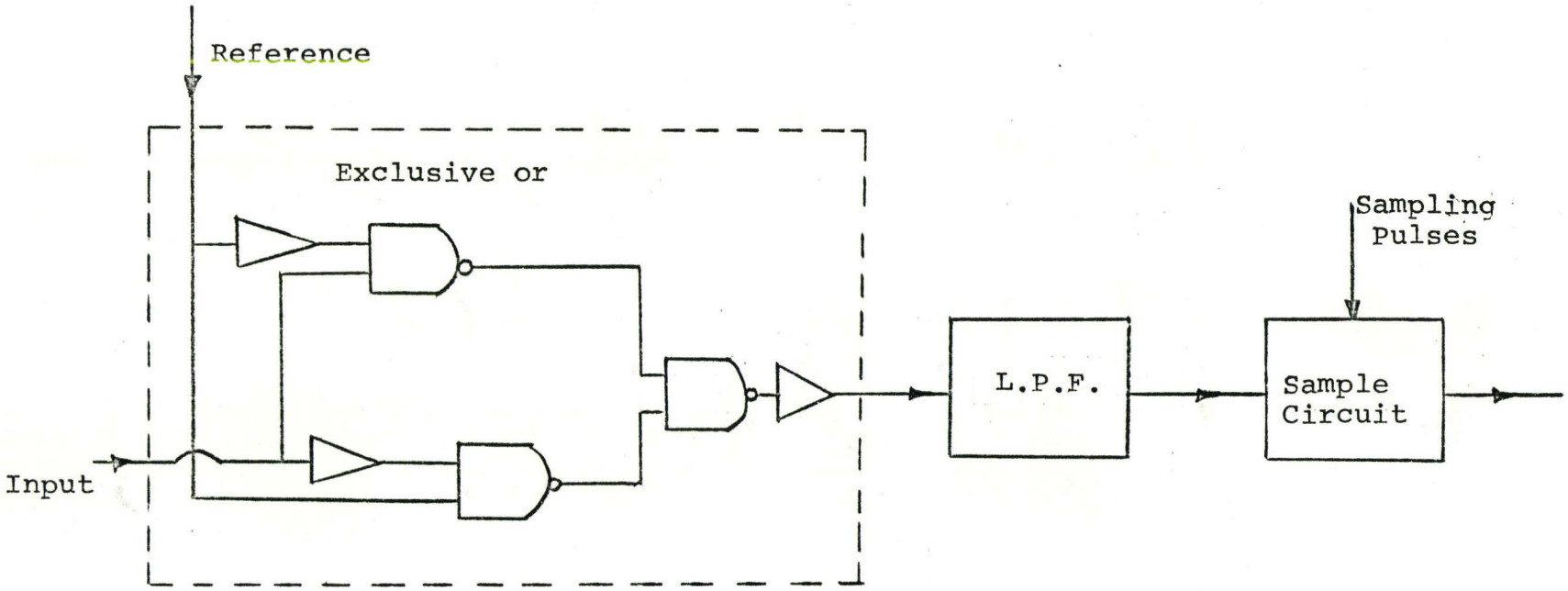


fig.4.5 One Branch of Correlation Detector

circuit which made the decision as to which signal was transmitted and then generated the appropriate bit.

4-3 MEASUREMENTS

In order to obtain the experimental results for both detectors, some simple measurements and calculations on the test arrangement had to be made. An estimate of the probability of error, for a particular signal to noise ratio, was made by using the counter to record the number of errors made over a particular length of time, and dividing this value by the total number of bits transmitted during that time. After a number of measurements had been made, it was found that, for the range of signal to noise ratios we were using, 100,000 transmitted bits were sufficient to give a reliable estimate of the probability of error. For the bit rates used in the tests, this meant that measurements had to be made over lengths of time exceeding ten minutes.

Measurement of the signal to noise ratio (E/N_o) was made at the input to the limiter stage, which was considered to be part of the receiver. Fortunately, a measurement of the RMS value of the noise could be made directly from a meter on the generator. The one-sided spectral density of

the noise could then be calculated by dividing the mean-squared value of the noise by the bandwidth of the noise generator. The RMS value of the square wave carriers was measured using a Hewlitt Packard True RMS Meter. The bit energy could then be calculated by multiplying the mean-squared value, by the bit duration, T seconds. Thus, the signal to noise ratio could be written as

$$E/N_o=(S/N)^2W/B$$

where S is the RMS value of carrier signal,

N is the RMS value of the noise,

W is the noise bandwidth, and

$B=1/T$ is the bit rate or system bandwidth.

As the bit rate increases, the signal to noise ratio decreases, and thus we can expect that for a constant noise level the probability of making an error will increase as the bit rate increases. However, the performance of the detector is described by only one curve, no matter what the bit rate.

We have suggested in a previous chapter, that frequency-shift keyed signals are an example of orthogonal signals; or equivalently, signals whose cross-correlation coefficient is zero. Orthogonality is defined by

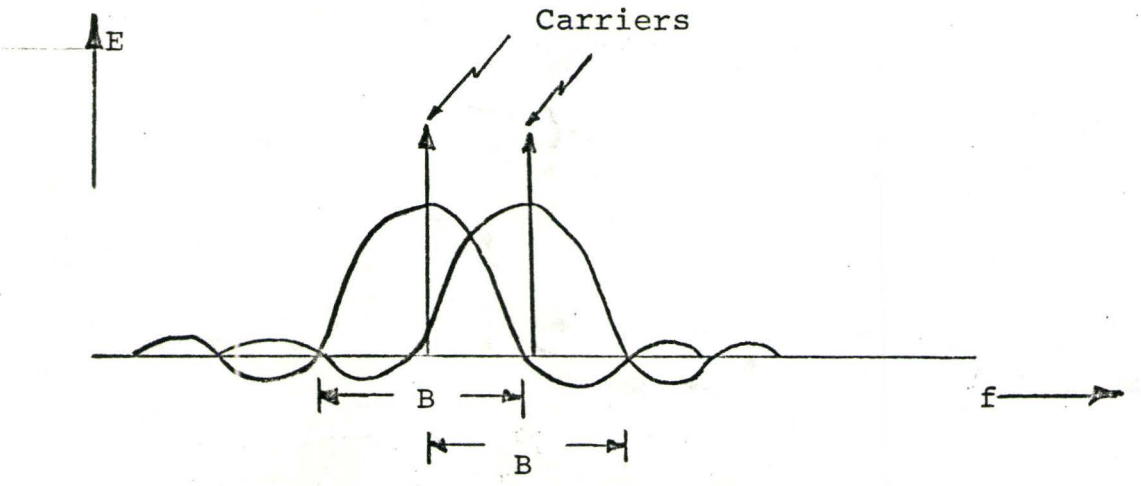
$$\int_o^T s_1(t)s_o(t)\,dt=0$$

Based on this definition, frequency-shift keyed signals are approximately orthogonal only if their spectra do not overlap. In figure (4.6), we have illustrated the spectra of the signals transmitted, in our FSK system. If the carrier frequencies are separated by something less than the system bandwidth B, the cross-correlation coefficient is greater than zero. If the separation of carrier frequencies is greater than the system bandwidth, then the signals are approximately orthogonal and their cross-correlation coefficient is approximately zero. We also showed in the last chapter that the probability of error is a function of the cross-correlation coefficient. To minimize the probability of error, the cross-correlation coefficient must be minimized. In our system we could insure this by maintaining a carrier frequency separation greater than our system bandwidth.

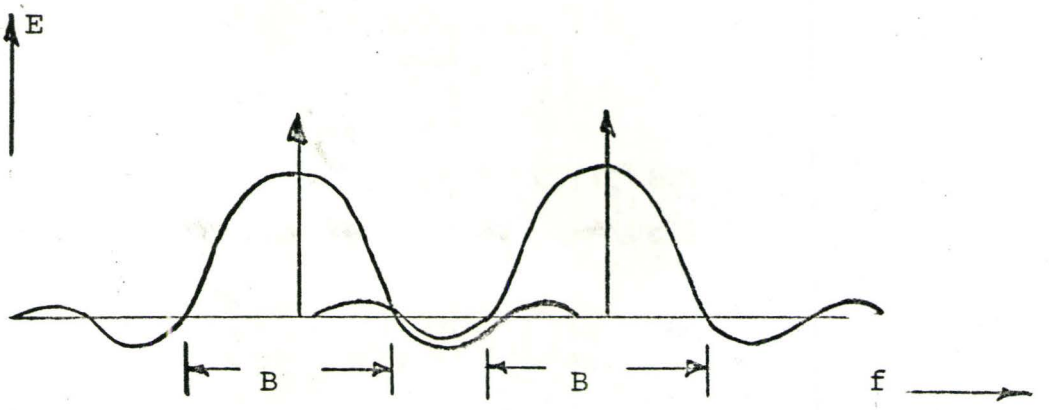
Care also had to be taken in choosing the clock frequencies associated with each loop; for it is these frequencies which determine the lock range and bandwidth of the loop.

4-4 RESULTS

The first and most important observation made on the



a) Spectra Overlap $(0 < \rho < 1)$



b) Spectra Do Not Overlap $(\rho \doteq 0)$

fig.4.6 Spectra of FSK Signals

phase-lock loop detector was, that in the absence of noise, it worked just as expected, reproducing at its output the binary sequence transmitted. The bit rate was arbitrarily chosen at 120 bauds. Variation of loop clock frequencies and carrier frequencies did not have any effect on the output of the system, provided the lock ranges of the two loops did not completely overlap and the carrier frequencies were adjusted to fall within the upper half of the lock ranges.

However, when noise was added to the input signal, performance in terms of probability of error did vary, with the clock and carrier frequencies. A process of trial and error was used to find a combination of frequencies which would give the best performance in terms of probability of error. The combination finally decided upon is shown in figure (4.7), where we have illustrated the static characteristics of the two loops. The clock frequencies necessary to obtain these characteristics can be determined from equation (2.8). The two carrier frequencies are also shown superimposed on the static characteristic.

The performance curve, for the correlation detector operating at the same bit rate and on the same carrier frequencies, was also obtained.

The experimental and theoretical probability of error curves for the coherent correlation detector, are plotted in fig.(4.8). The significant difference between

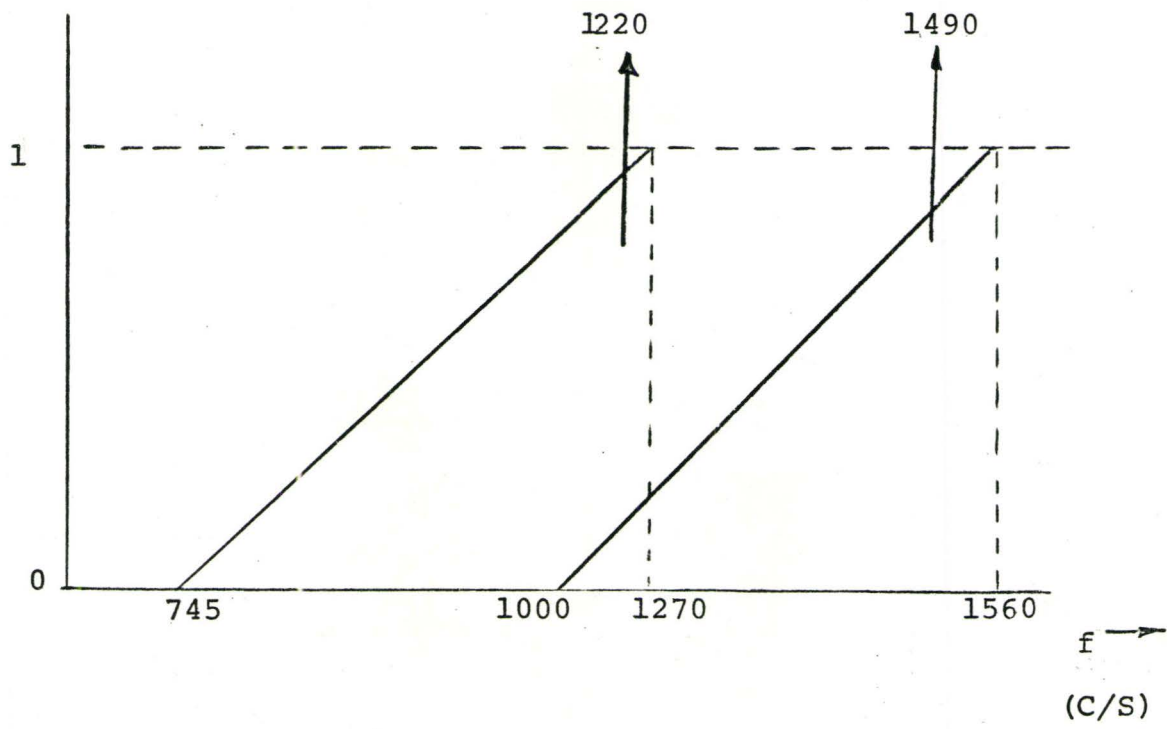


fig.4.7 Static Characteristics of Loops in PLL Detector

the two curves, (approximately 4.9dB at $P_E=0.001$ is due to the non-ideal conditions under which the experimental system was operated. As mentioned earlier, it was necessary to include a limiter stage at the input to the detector; however, the limiter was not considered in the derivation of the theoretical probability of error curve. It is expected that because the limiter is a non-linear device, it will have a degrading effect on the system, particularly at low signal to noise ratios. Further, it was assumed, in the theoretical derivation, that the transmitter was capable of switching instantaneously between the two signals and that the noise process was white and Gaussian. Both assumptions are only approximated, in the experimental test. Finally, although we have maintained the carrier frequency separation greater than the system bandwidth, it is likely that the cross-correlation coefficient of the two signals is slightly positive, thus causing a further degradation in the experimental detector.

It should be emphasized at this point, that the purpose of the experimental tests was to compare a practical correlation detector with a practical loop detector, both operated under the same conditions. It is assumed, that if improvements can be made in the performance of one detector, then, similar improvements can be made in the other.

The experimental probability of error curves for the

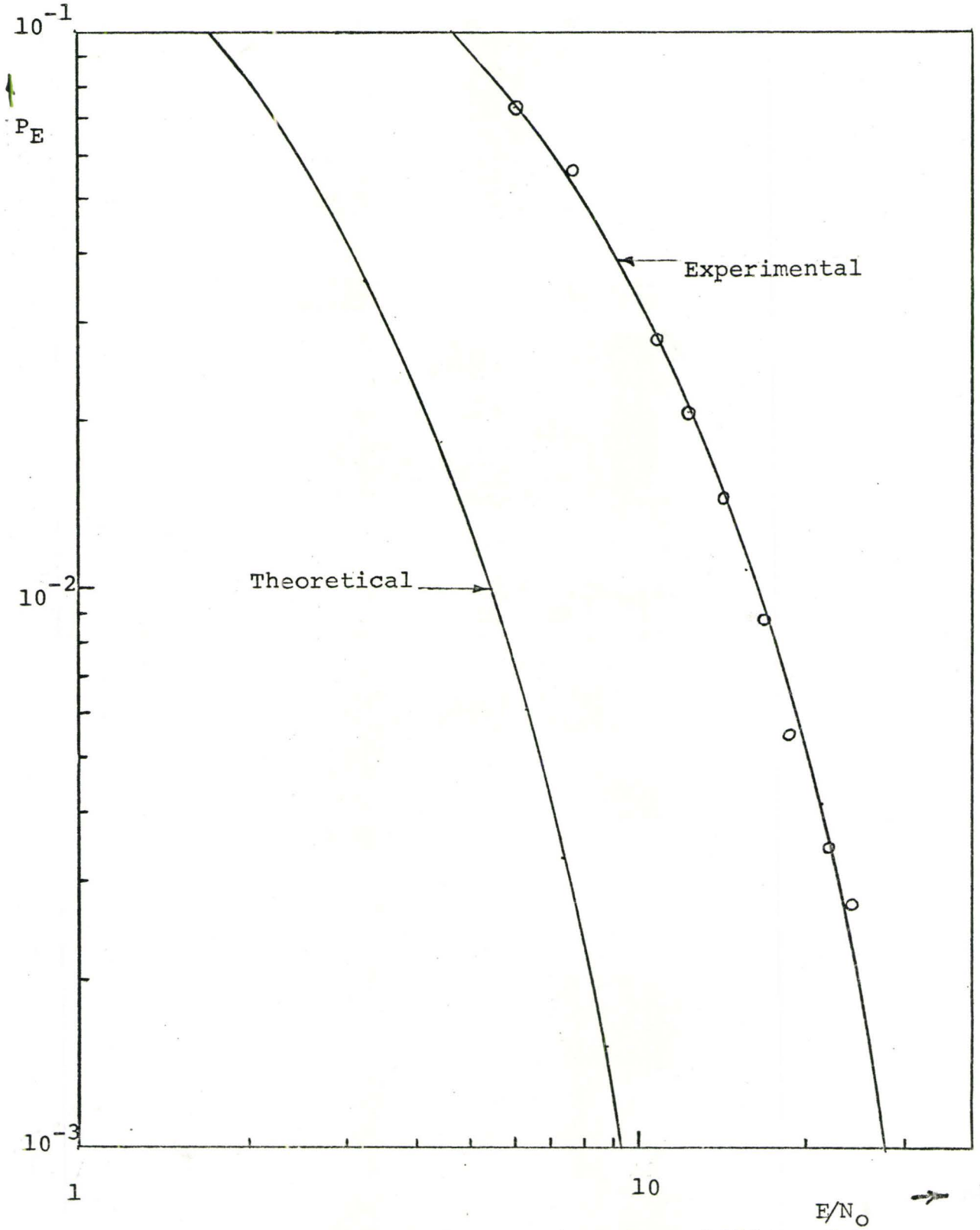


fig.4.8 Experimental and Theoretical Probability of Error Curves for Correlation Detector

two detectors are plotted in figure (4.9). The difference between the two detectors is approximately 1.6dB at a probability of error of 0.001. Based, on the theory of detectors presented in the last chapter, we might expect that an experimental non-coherent detector, implimented using filters and envelope detectors, would have a performance curve lying about 3dB to the right of the coherent detector curve. Thus, we may conclude that the performance of the phase-lock loop detector is somewhere between that of the coherent and non-coherent detectors.

To further illustrate the performance of the phase-lock loop detector, a series of pictures were taken using a storage scope and polaroid camera. In figure (4.10) input and output waveforms were recorded for four different signal to noise ratios. In part (a) of the figure, signal to noise ratio is infinite; that is, no noise is present, and the amplitudes of the two carrier signals were intentionally made different in order that they be more readily distinguishable. It is apparent from these diagrams that as the signal to noise ratio decreases, the probability of error increases correspondingly. In figure (4.11) the output of one branch of the detector is shown for different signal to noise ratios in a manner corresponding to figure (4.10). The degradation in the appearance of the waveforms at low signal to noise ratios as in parts (a) and (b) indicates that the loop is momentarily losing lock.

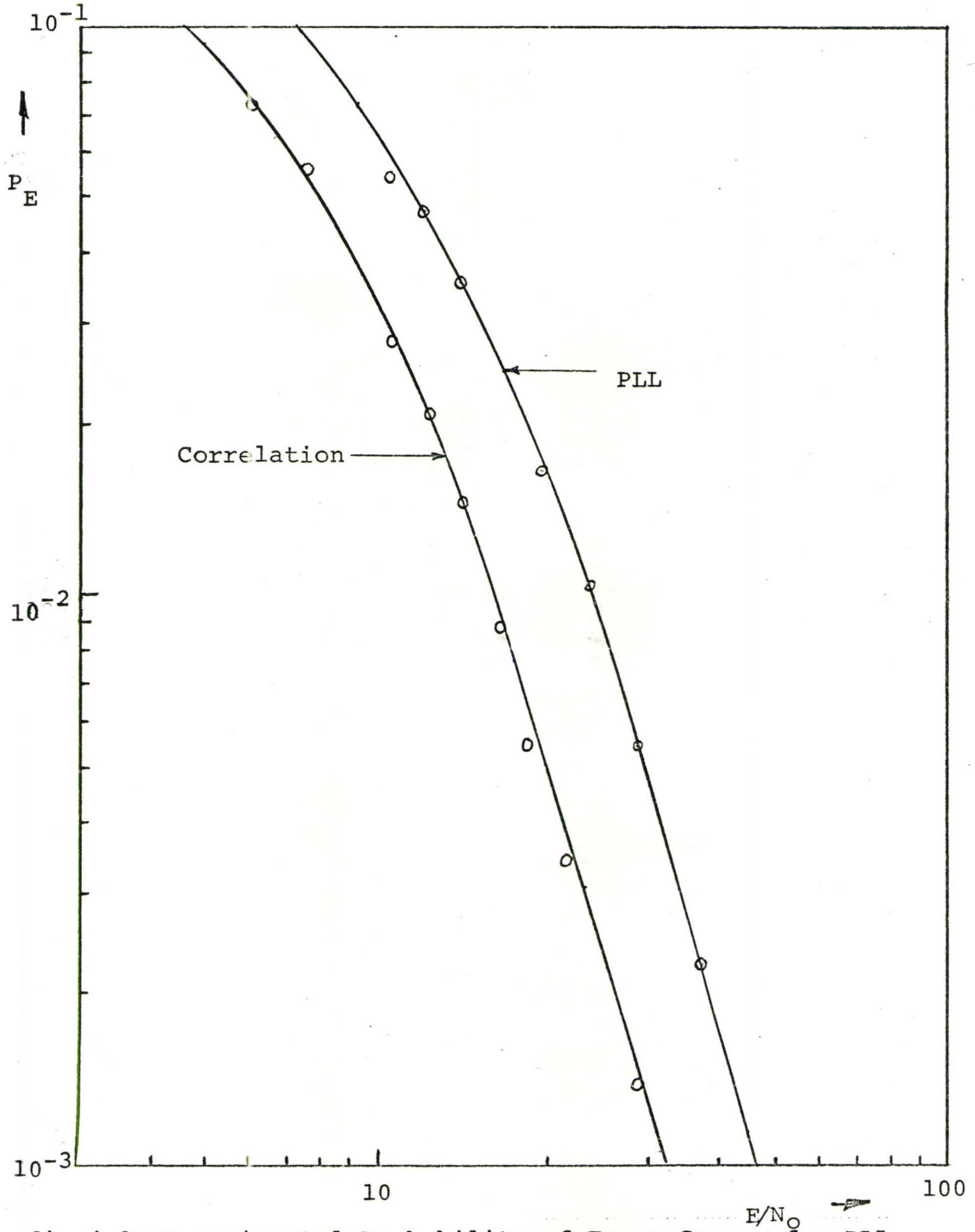
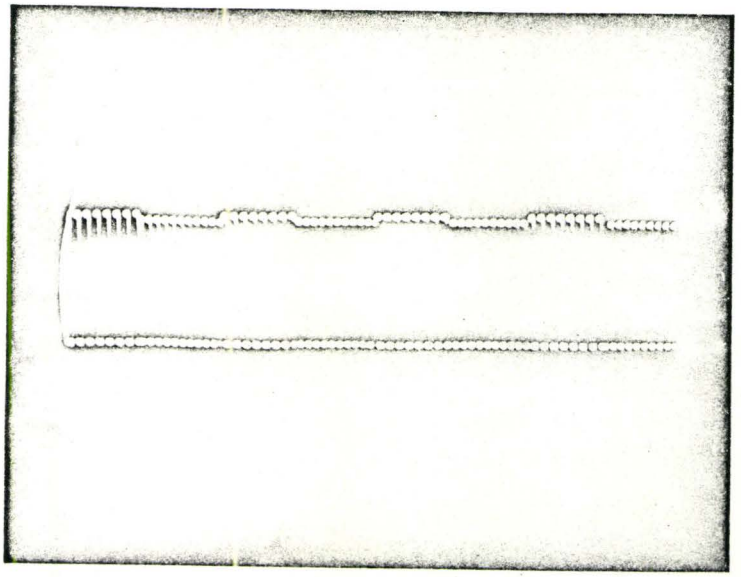


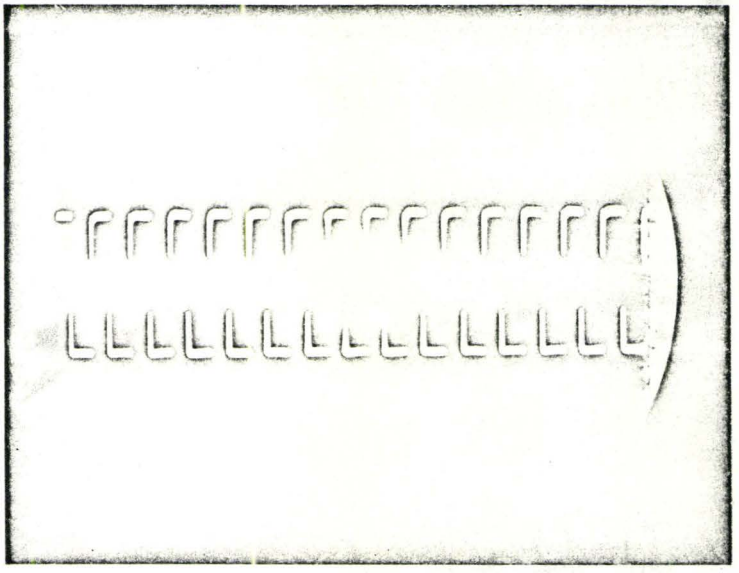
fig.4.9 Experimental Probability of Error Curves for PLL and Correlation Detectors



Horizontal
5m sec./cm

Vertical
.5v/cm

FSK input signal (E/No=∞)

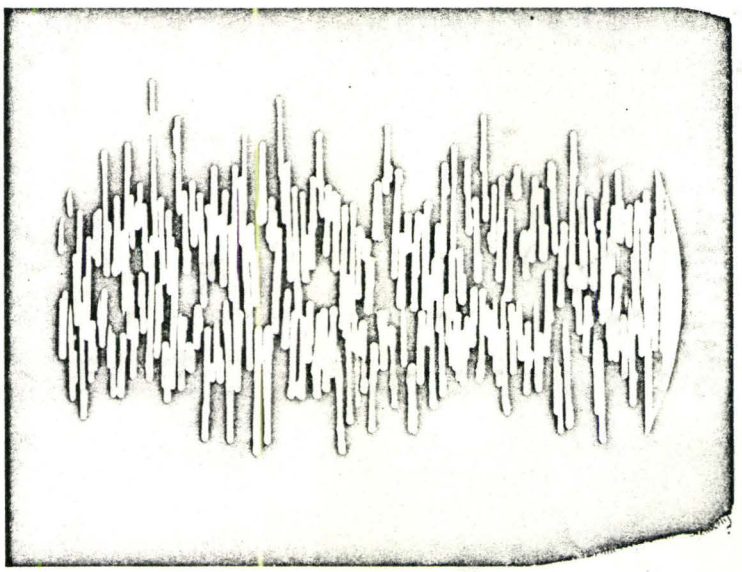


Horizontal
20m sec./cm

Vertical
.2v/cm

Output binary sequence (0 Errors)

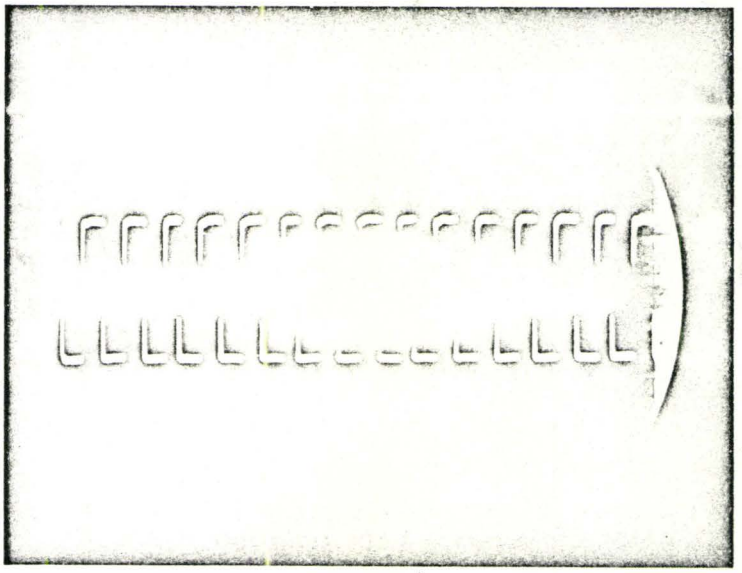
fig.(4.10 a) Input and Output Waveforms for PHASE-LOCK LOOP Detector for Different Signal to noise ratios.



Horizontal
5m sec./cm

Vertical
.5v/cm

FSK input signal (E/N0=24)

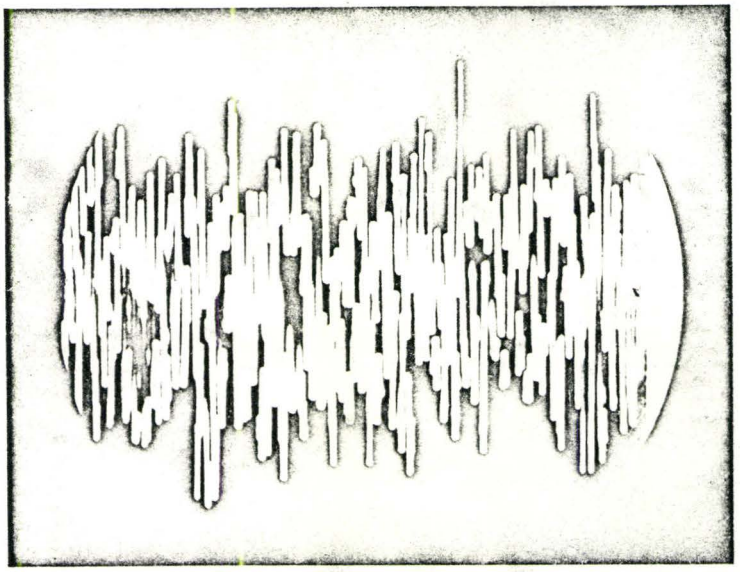


Horizontal
20m sec./cm

Vertical
.2v/cm

Out put binary sequence (0 Errors)

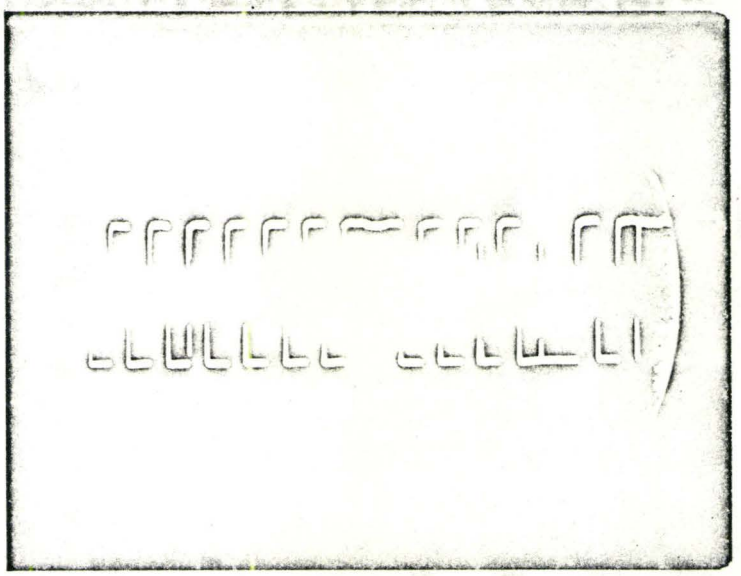
fig. (4.10 b)



Horizontal
5m sec./cm

Vertical
.5v/cm

FSK input signal (E/No= 10.6)

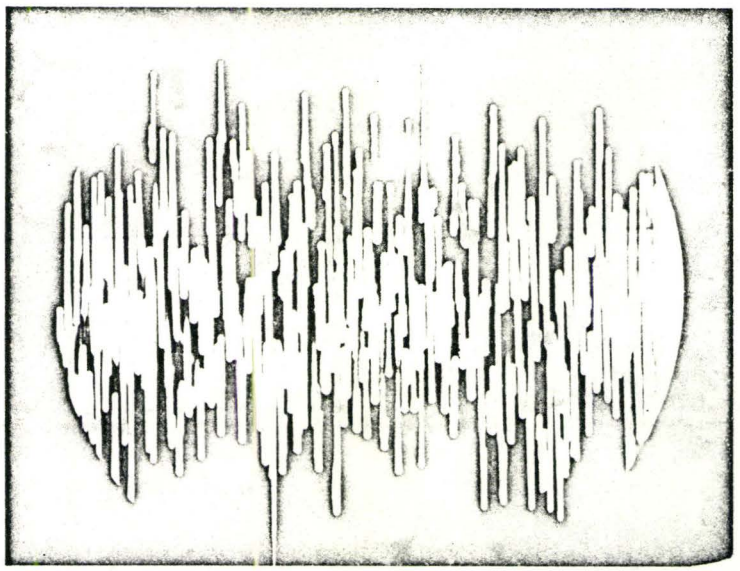


Horizontal
20m sec./cm

Vertical
.2v/cm

Output binary sequence (3 Errors)

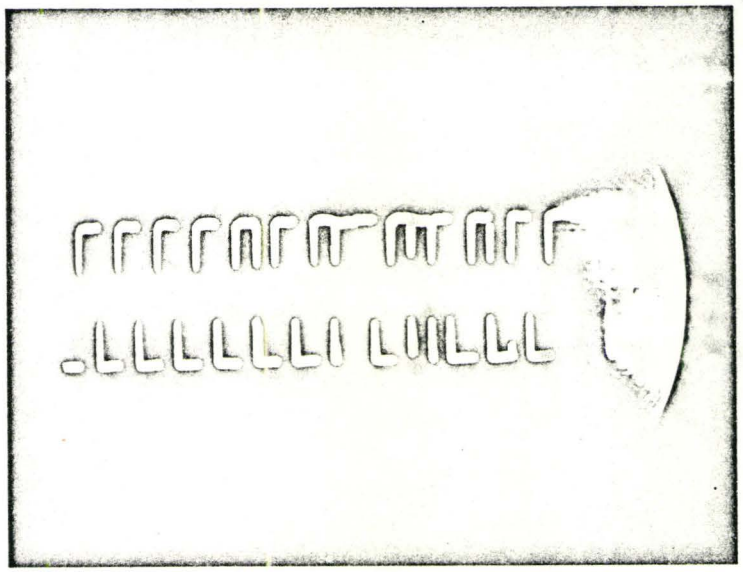
fig. (4.10 c)



Horizontal
5m sec./cm

Vertical
.5v/cm

FSK input signal (E/No=7.4)

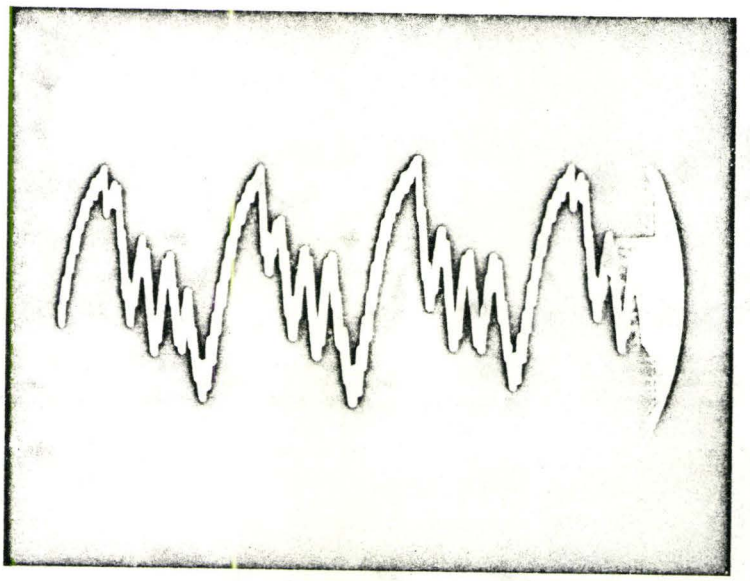


Horizontal
20m sec./cm

Vertical
.2v/cm

Output binary sequence (3 Errors)

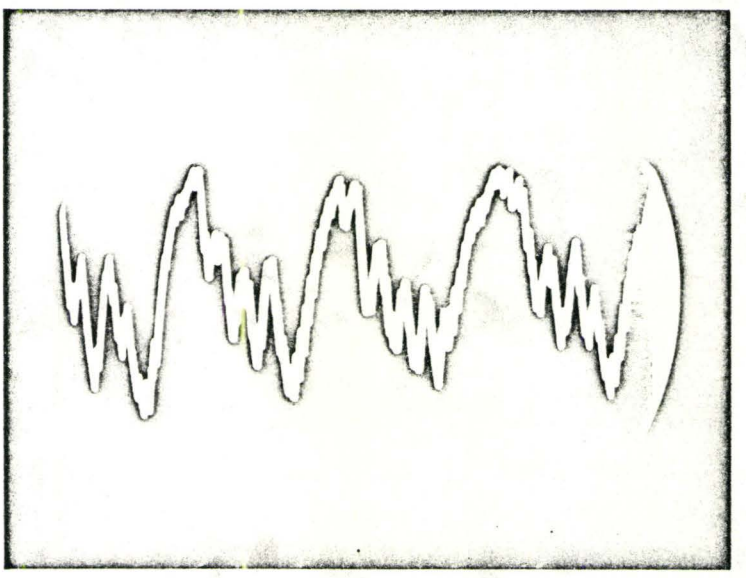
fig. (4.10 d)



Horizontal
5m sec./cm

Vertical
.5v/cm

(a) E/No= ∞

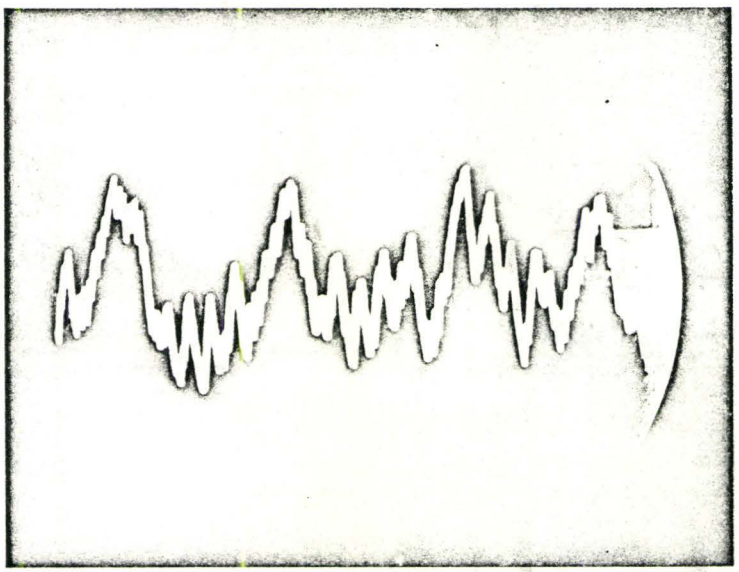


Horizontal
5m sec./cm

Vertical
.5v/cm

(b) E/No=24

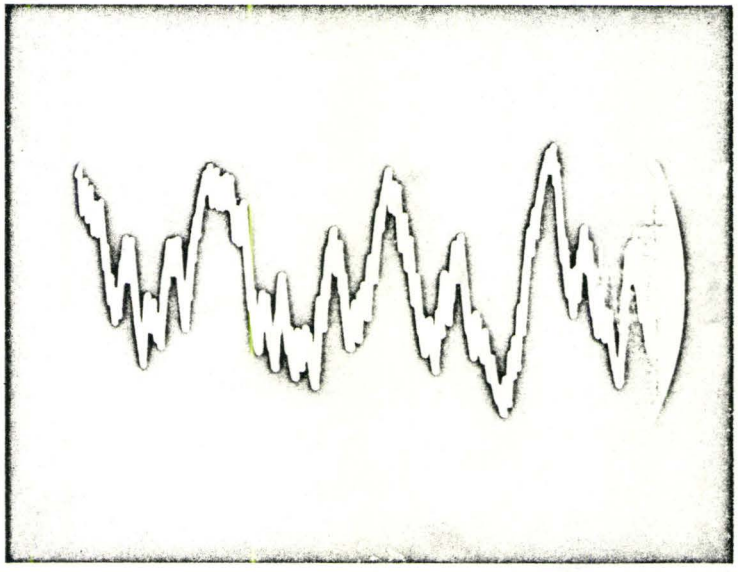
fig.(4.11) Output Waveform of One Branch of PHASE-LOCK LOOP Detector for Different Signal to Noise Ratios.



Horizontal
5m sec./cm

Vertical
.5v/cm

(c) E/No=10.6



Horizontal
5m sec./cm

Vertical
.5v/cm

(d) E/No=7.4

fig.(4.11)

It appears that a useful application of the phase-lock loop detector is for the case in which there is some uncertainty as to the frequencies of the received FSK signals. Such uncertainty would result if the transmitted signals underwent a Doppler shift in frequency. To determine experimentally how the detector would behave under such circumstances, it was decided to obtain performance curves for a number of carrier frequencies. The lock ranges of the two loops were maintained constant and carrier frequencies changed in steps of 10Hz up to a maximum of 40Hz in both the positive and negative direction. For mean carrier frequencies of 1220Hz and 1490Hz, this corresponded to a maximum change of 3% and 2.5%, respectively. The curves for positive changes in frequency are plotted in fig.(4.12); the curves for negative changes lie within the same region and therefore were not plotted. For this range of carrier frequencies, the performance curves vary over only a 1.0dB range of signal to noise ratio.

A similar test was performed on the correlation detector. The frequencies of the reference signals were maintained constant at 1220Hz and 1490Hz, and the incoming carrier frequencies were varied. As one would expect, as soon as the reference and carrier signals become unsynchronized, the coherent correlation detector ceased to operate. This is illustrated in fig.(4.13). This clearly demonstrates that the phase-lock loop detector is superior to a coherent detector when dealing with a FSK signal that has undergone some Doppler shift.

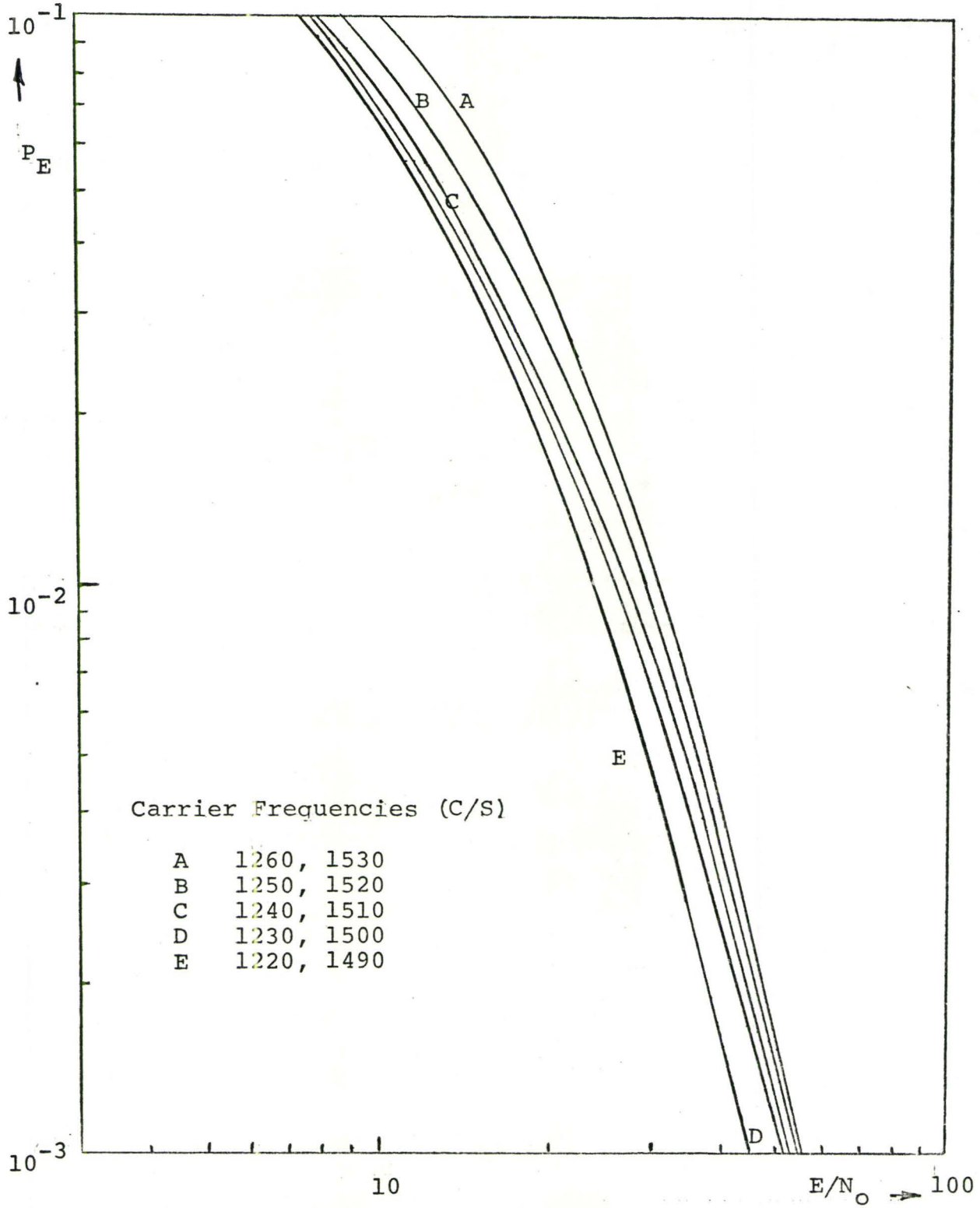


fig.4.12 Experimental Probability of Error Curves for PLL Detector, With Four Different Pairs of Carrier Frequencies

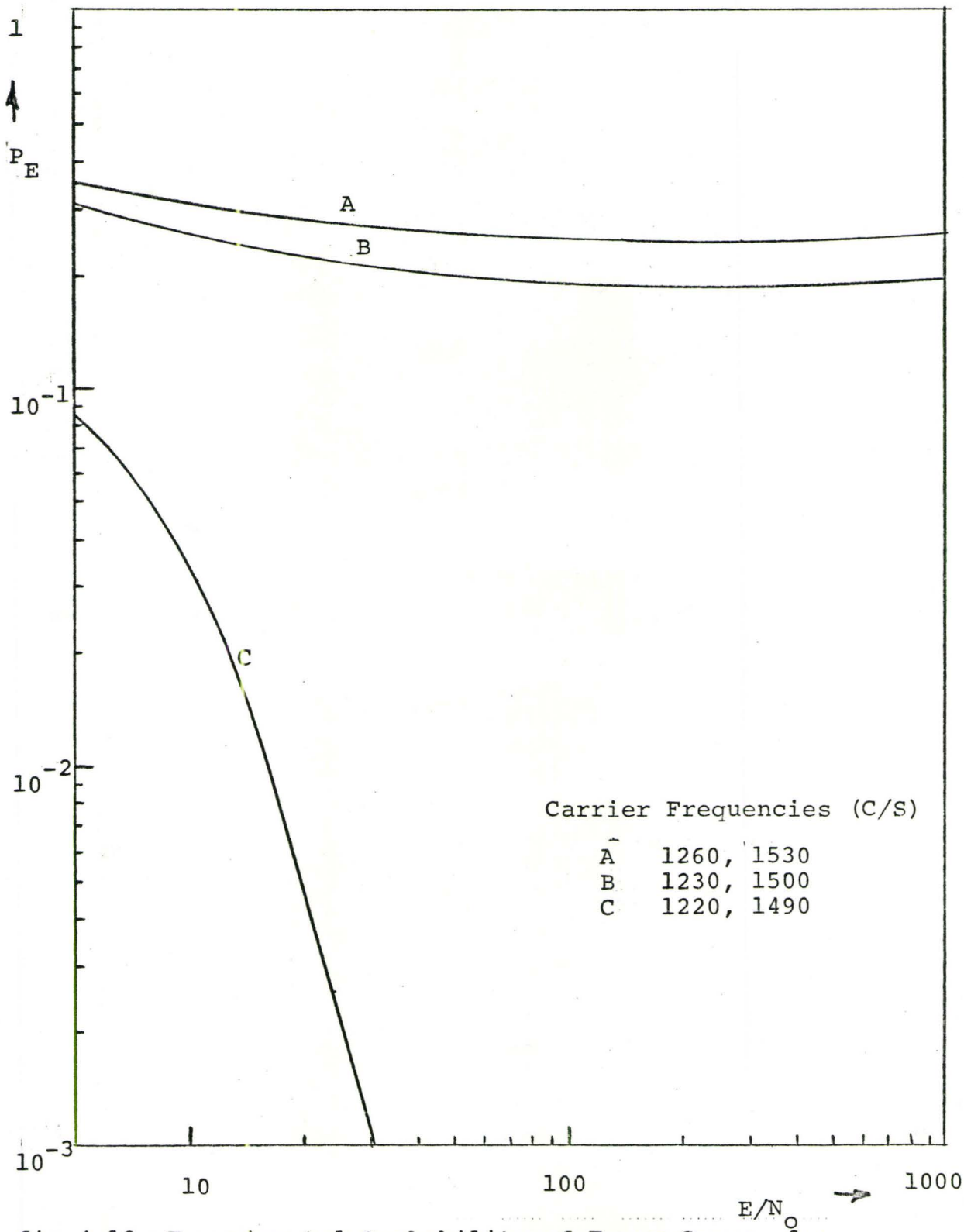


fig.4.13 Experimental Probability of Error Curves for Correlation Detector, With Reference and Input Signals Unsynchronized.

CONCLUSIONS

The usefulness of digital phase-lock loops as part of a binary detection scheme for FSK signals has been successfully demonstrated, From the results of the previous chapter we may conclude that for an ordinary FSK signal, the phase-lock loop detector is inferior to a coherent correlation detector but superior to the optimum noncoherent detector. Since the design and practical implimentation of a digital phase-lock is relatively simple, we may consider a phase-lock loop detector as a practical alternative to a non-coherent detector, for FSK signals.

Furthermore, because a phase-lock loop is, by nature, able to track variations in the frequency of the incoming signal, the performance of a phase-lock loop detector does not suffer a serious degradation when the two frequencies making up the FSK signals, change (due to Doppler shift). In this respect, the phase-lock loop detector has a definite advantage over a coherent correlation detector. As for the noncoherent detector although it will continue to operate with FSK signals that have undergone frequency changes, its performance will be severely degraded because it utilize fixed tuned bandpass filters.

We may thus conclude that, the unique features of the phase-lock loop detector are:

1) It generates its own reference signal.

2) It can operate successfully on FSK signals with changing frequencies.

THE APPENDIX

This appendix contains: Circuit Diagrams
Circuit Explanations

APPENDIX A

CLOCK CIRCUITS

In fig (A.1) we have illustrated three clock circuits using inverters. Referring to (A.1b), the operation can be explained as follows. The device is basically a flip-flop (gates 1 and 3) with a high frequency by-passed inverter in the loop (gate 2). The inverter makes the loop unstable. The capacitor couples gate 1 to gate 3, allowing regeneration to occur for fast rise times, C and R_2 control the frequency, and R_1 varies the symmetry of the output. Since in our application the symmetry of the output was not important the arrangement in (A.1c) was used for the digital PLL. It was found that this arrangement gave the best stability.

"TRANSMITTER"

The multivibrators used are illustrated in figure (A.2). For an explanation of this circuit refer to Millman and Taub[1]. The outputs (points A and B) were applied to what we have called the switching circuit (figure A.3). In fact, this circuit is actually a sampling circuit, where the binary sequence applied at point C serves as the sampling pulses. The diode-potentiometer arrangement at the output, provides a means of varying the amplitude of the FSK signals.

CHANNEL

The summing circuit was implimented using a 741 operational amplifier and the limiter using 710 comparator. Both circuits are illustrated in figure (A.4).

SAMPLING CIRCUITS

The binary sequence was applied to a differentiator (point C, figure (A.5a)) giving positive and negative spikes at the transitions. The negative spikes were inverted using the diode bridge and differential amplifier arrangement illustrated in figure (A.5a). The resulting train of positive spikes were applied to a monostable multivibrator[1] (point E, figure (A.5b)) which generated a positive sampling pulse at the end of each bit.

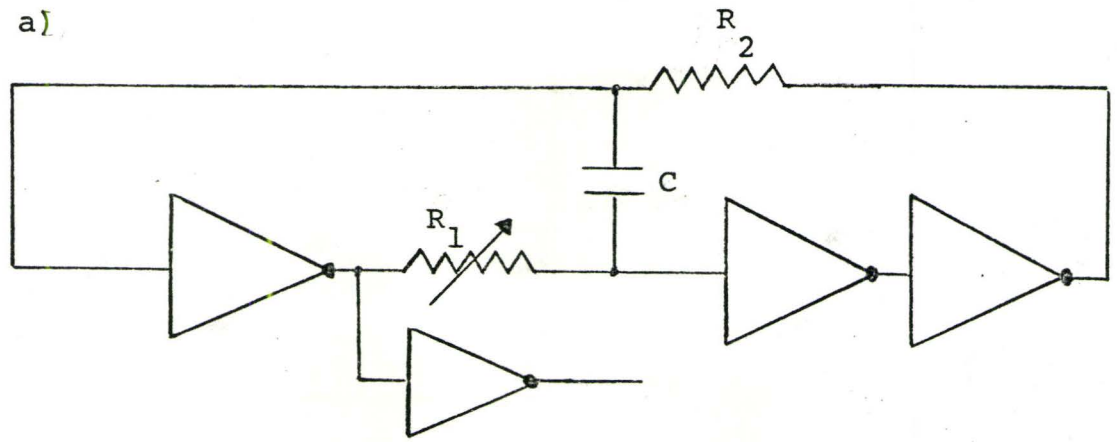
The outputs of the branches of the detector were applied at points F and G where they were sampled at the end of each bit. The resulting samples were compared using a 710 comparator (figure A.6). The output of the comparator was then applied to a hold circuit which generated the appropriate bit at point J.

ERROR COUNTING CIRCUIT

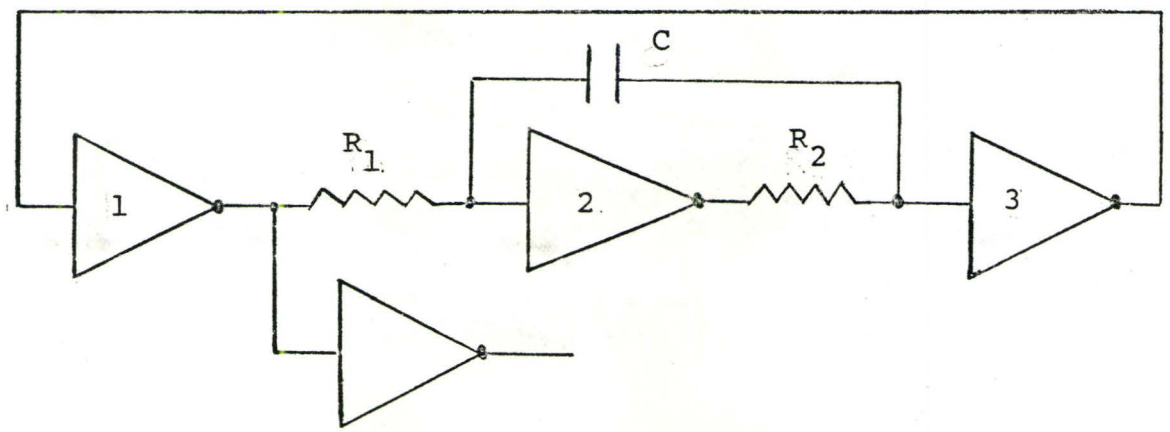
The outputs, of both the detector and source of binary

sequence, were sampled once each bit, as in figure (A.7). The samples were applied to an "exclusive or" for comparison. If the samples were different, the "exclusive or" provided a pulse at its output, which was recorded on a commercially built counter.

a)



b)



c)

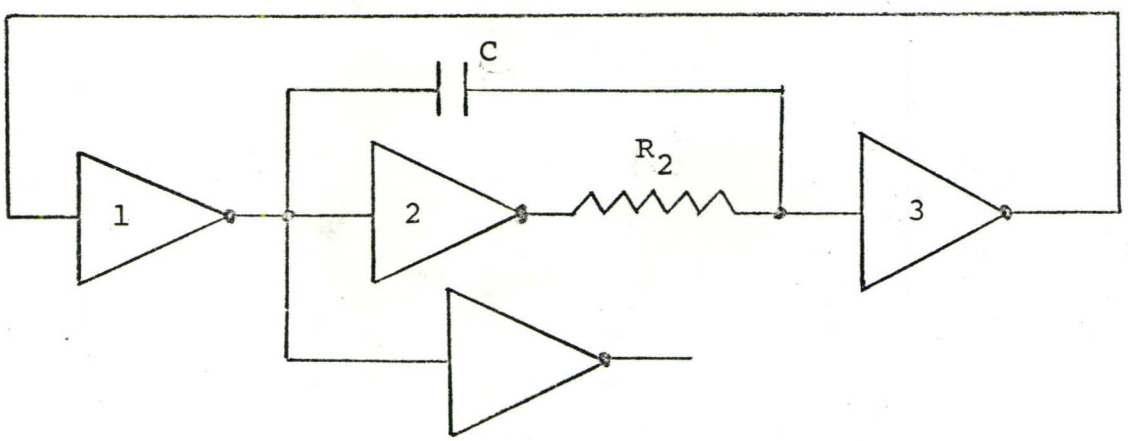


fig.A.1 Clock Circuits

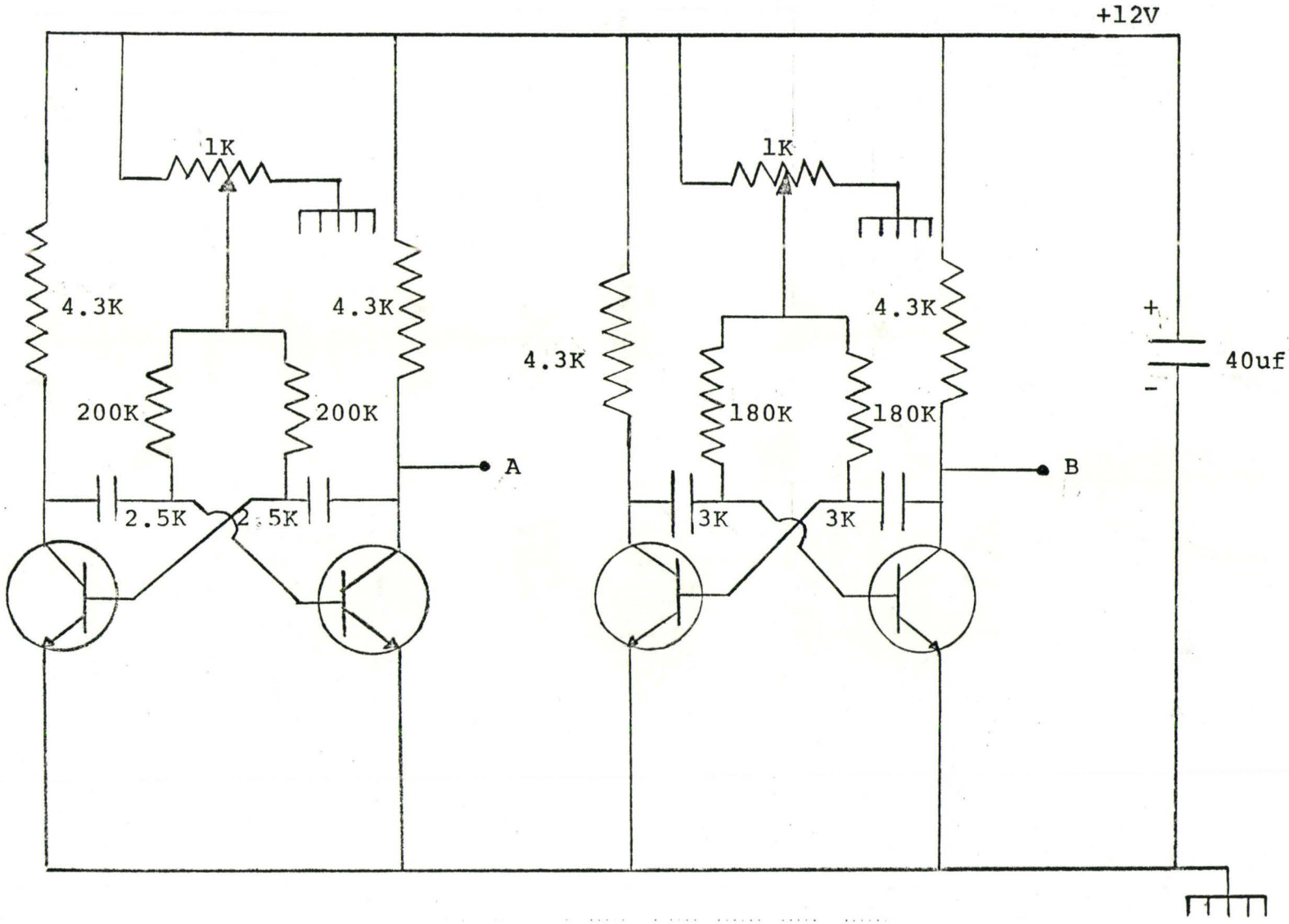


fig.A.2 Multivibrator Circuits

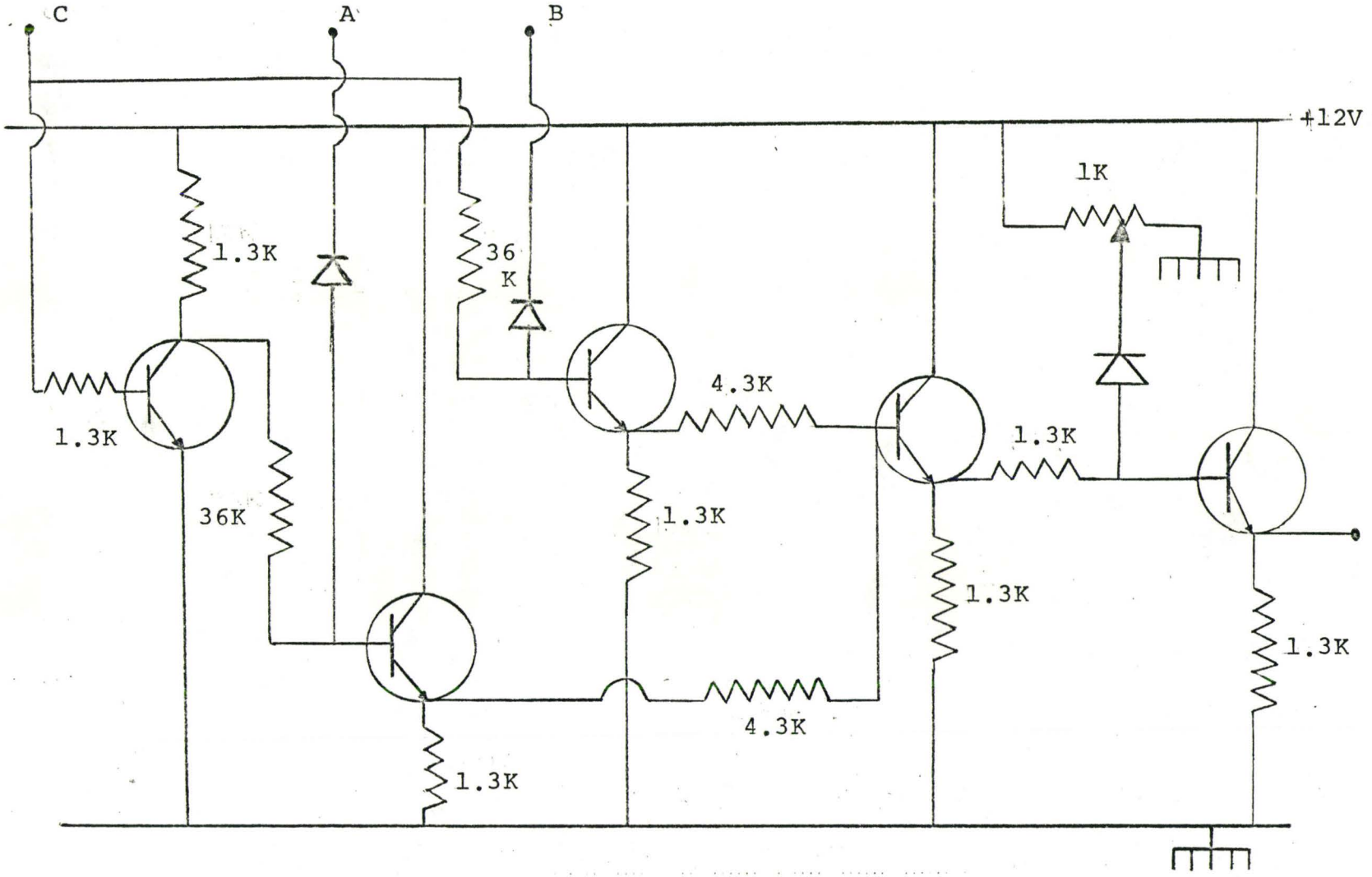


fig.A.3 Switching Circuit

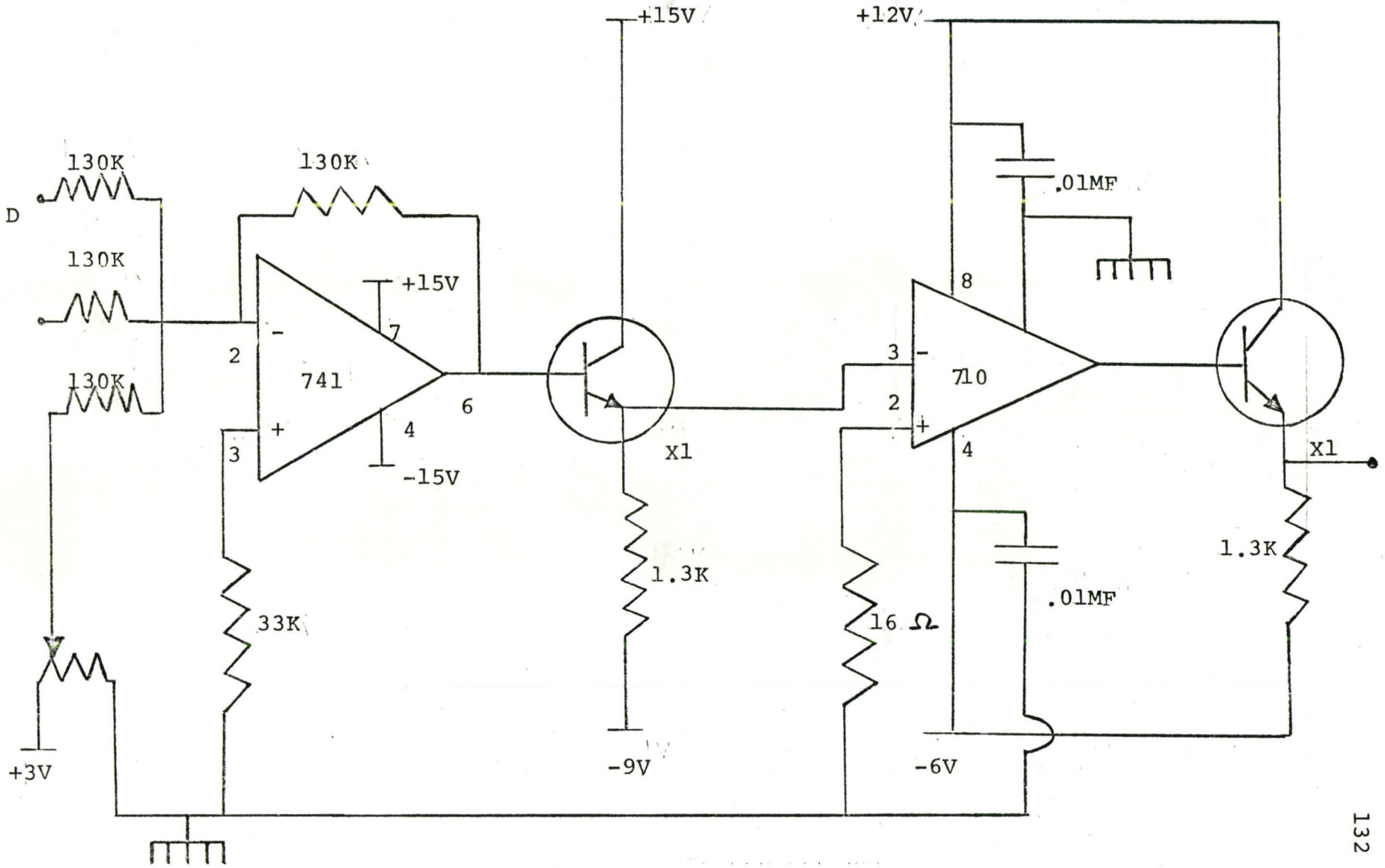


fig.A.4 Summer and Limiter

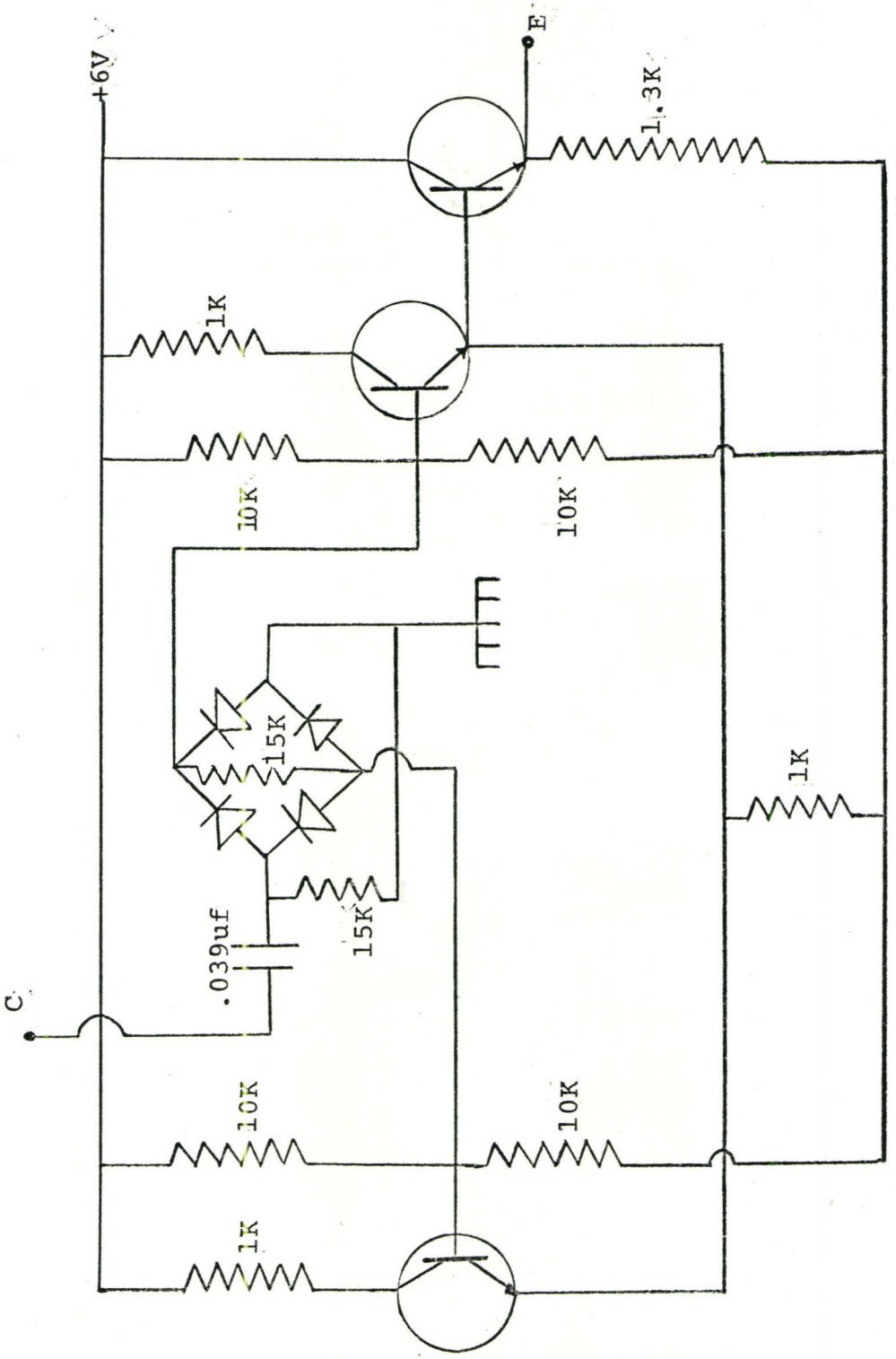


fig.A.5a Sampling Circuits

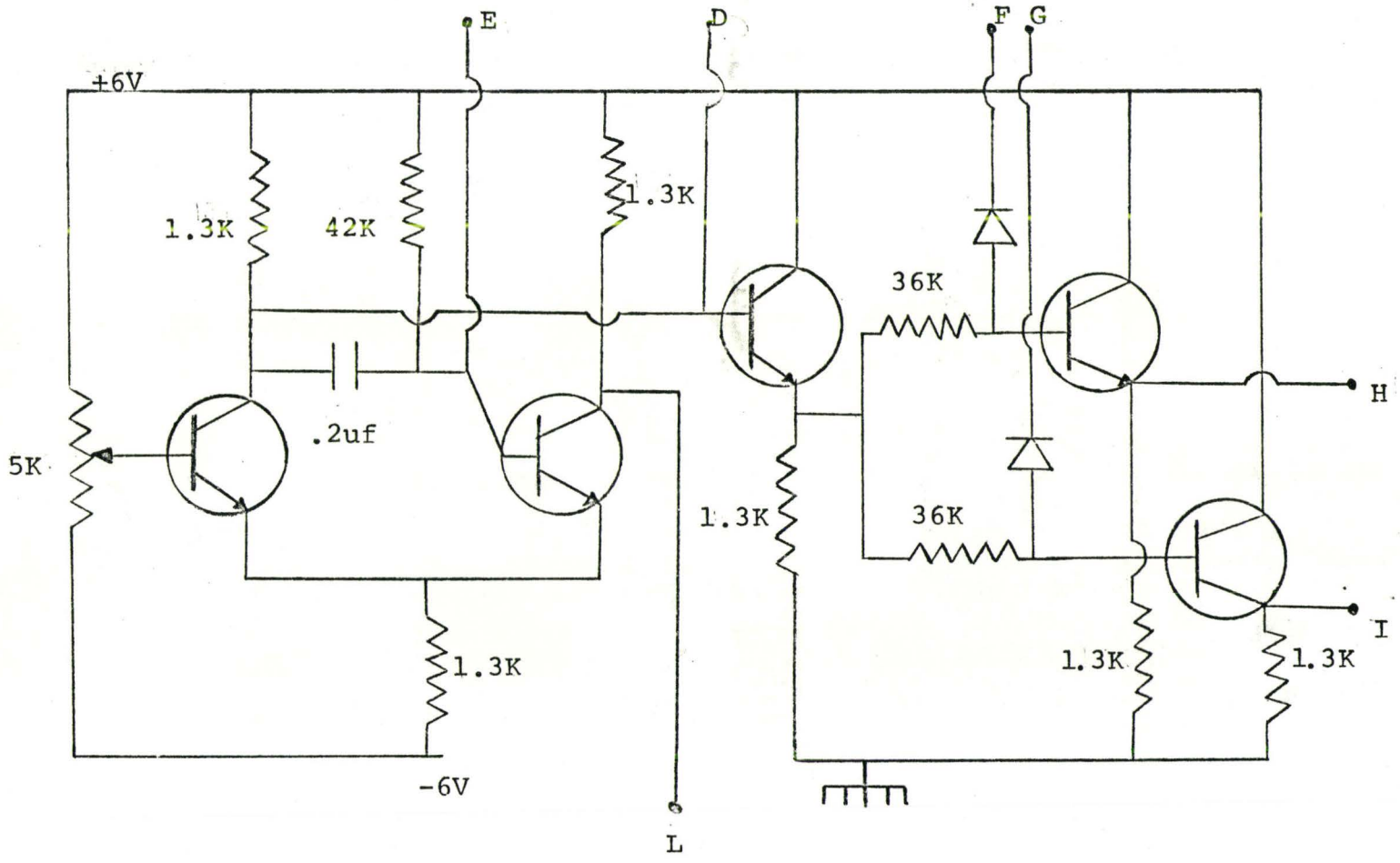


fig.A.5b Sampling Circuits

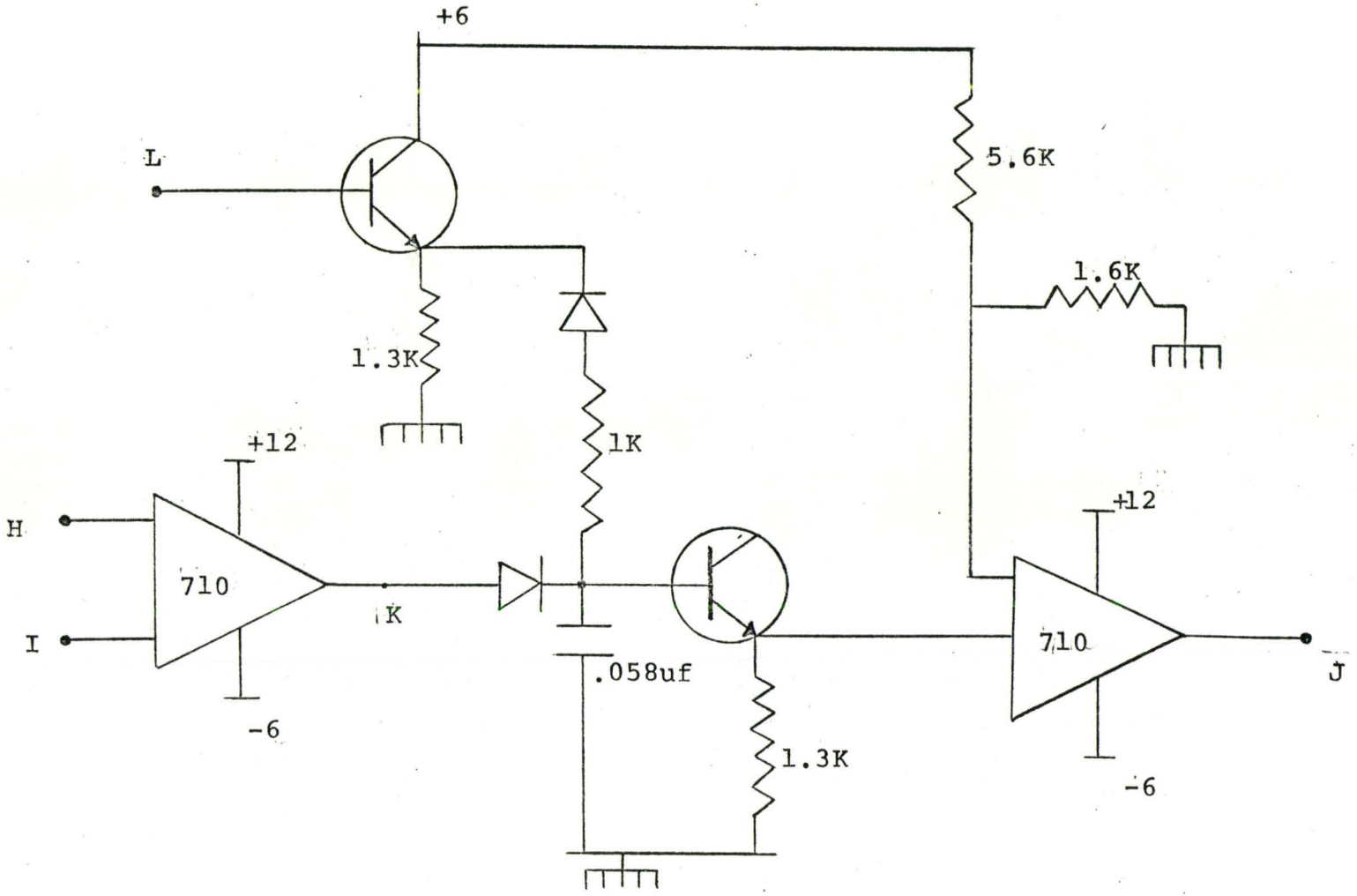


fig.A.6 Comparator and Hold Circuit

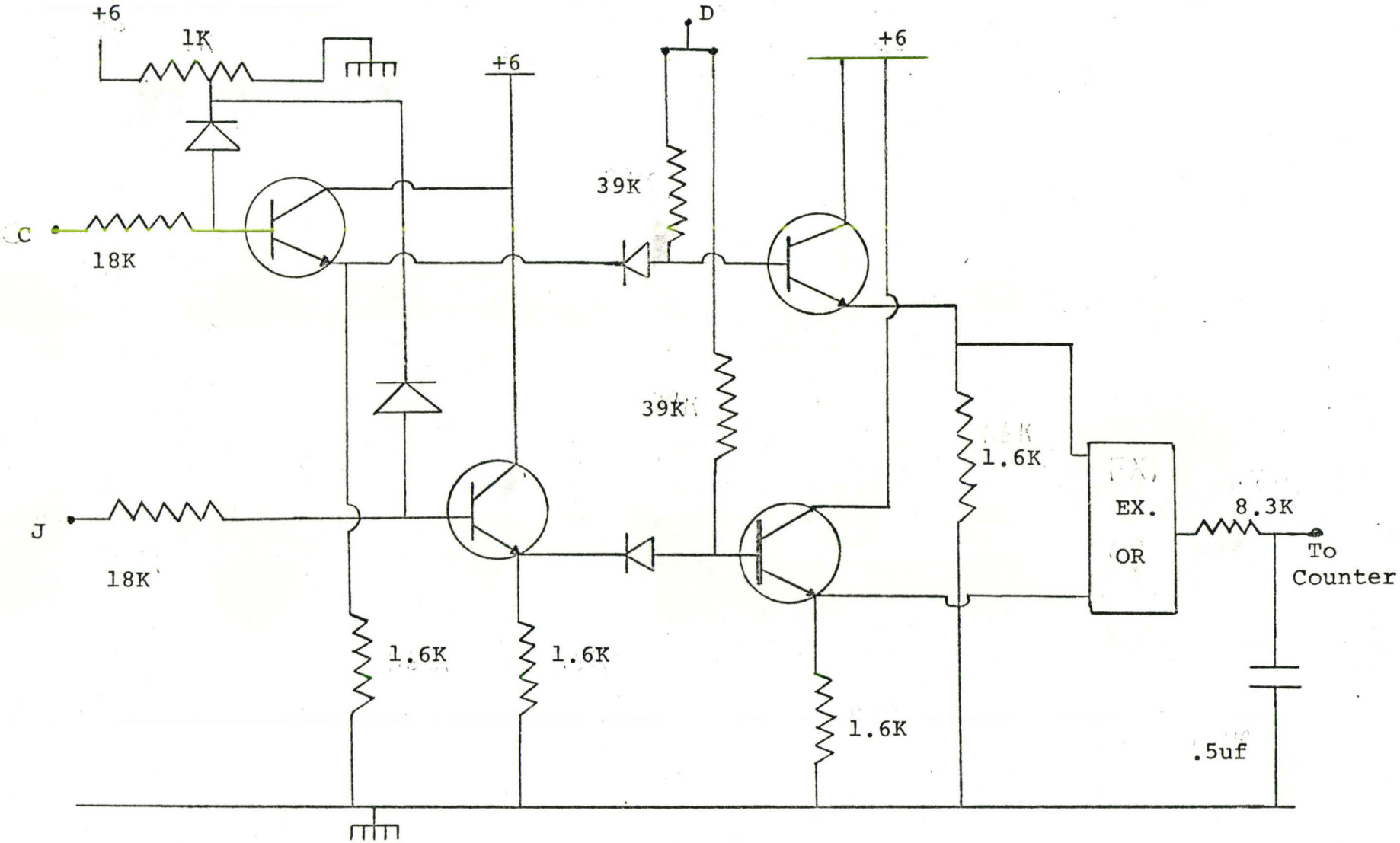


fig.A.7 Error Counting Circuit

REFERENCES

Chapter One:

1 D.O.North, "Analysis of the Factors which Determine Signal/Noise Discrimination in Radar", Report PTR-6C, RCA Laboratory, June 1943.

2 J.Van Vleck and D.Middleton, "A Theoretical Comparison of Visual, Aural, and Meter Reception of Pulsed Signals in the Presence of Noise", Journal of Applied Physics Vol. 17, pp. 940-970, November 1946.

3 "Matched Filter Issue", Trans. PGIT-IRE, Vol. IT-6, No.3, June 1960.

4 Y.W.Lee, T.P.Cheatham, and J.B.Wiesner, "Application of Correlation Analysis to the Detection of Periodic Signals in Noise", PROC.IRE, Vol.38, pp. 1165-1172, October 1950.

5 H.de Belliscize, "La Reception Synchrone", Onde Elect., Vol.11, pp. 230-240, June 1932.

6 T.J.Rey, "Automatic Phase Control: Theory and Design", PROC.IRE, Vol. 48, PP. 1760-1771, October 1960.

7 K.R.Wendt and G.L.Fredendall, "Automatic Frequency and Phase Control of Synchronization in Television Receivers", PROC.IRE, Vol.31, pp. 7-15, January 1943.

8 R.Jaffe and E.Rechtin, "Design and Performance of Phase-Lock Circuits Capable of Near-Optimum Performance", IRE Trans. Information Theory, Vol IT-1, pp. 66-76, March 1955.

9 S.C.Gupta and R.J.Solem, "Optimum Filters for Second and Third Order Phase-Lock Loops", IEEE Trans. Space Electronics and Telemetry, pp. 54-62, June 1965.

10 W.J.Given, "Theory of AFC. Synchronization", PROC. IRE, Vol.41, pp. 1043-1048, August 1953.

11 A.J.Viterbi, "Acquisition and Tracking Behaviour of Phase-Locked Loops", J.P.L. External Publication No.673, July 14, 1959.

12 A.J.Viterbi, "Acquisition and Tracking Behaviour of Phase-Locked Loops", Proc. Symp. Active Networks and Feedback Systems, Polytechnical Institute of Brooklyn, Vol.10, 1960.

13 Davenport and Root, "Introduction to Random Signals and Noise", McGraw Hill Book Co., New York, 1958.

14 A.J.Viterbi, "Phase-Lock Loop Dynamics in the Presence of Noise by Fokker-Plank Techniques", PROC. IEEE, Vol.51, No.3, 1960.

Chapter Two:

1 G.Pasternack and R.Whalin, "Analysis and Synthesis of a Digital Phase-Locked Loop for FM Demodulation", Bell System Technical Journal, December 1968.

2 E.Jury, "Theory and Application of Z Transform Method", John Wiley, 1964.

Chapter Three:

1 K.Ogata, "State Space Analysis of Control Systems", Prentice-Hall, 1967.

2 Hancock and Wintz, "Signal Detection Theory", McGraw-Hill Book Company, 1966.

3 A.J.Viterbi, "Principles of Coherent Communication", McGraw-Hill Book Company, 1966.

Appendix A

1 Millman and Taub, "Pulse, Digital and Switching Waveforms", McGraw-Hill Book Company, 1965