A STUDY OF THE MINIMUM SHIFT KEYING MODULATION SCHEME

Prepared By : Mario Antonio Aldera
B.Sc. Electrical and
Electronic Engineering, Cape Town.

Prepared For : The Department of Electrical
and Electronic Engineering at
the University of Cape Town.

30th September, 1988.

Thesis submitted in partial fulfilment of the requirements
for the degree of Master of Science in Engineering.

DECLARATION

I declare that this thesis is my own unaided work, and is being submitted to the University of Cape Town in partial fulfilment of the requirements of a Masters Degree in Engineering. It has not been submitted before to any other university for any degree or examination.

Signed by candidate

M. A. Aldera 30th September, 1988.

ACKNOWLEDGEMENTS

The author wishes to thank the following people for the assistance rendered during the preparation of this thesis:

Dr. R.M. Braun (Thesis Supervisor);

Mr. M.G. Aldera for the loan of the personal computer on which this thesis was prepared;

Dr. L.A. Aldera for the photography which made the thesis complete.

ABSTRACT

This thesis concerns itself with the study of the Minimum Shift Keying (MSK) modulation scheme. The aspects considered are its operation under non-linear conditions as well as an investigation into the hardware implementation of both coherent and non-coherent MSK modems.

The literature on digital data transmission and MSK in particular is surveyed, and a comprehensive theoretical description of MSK is given. In addition, papers on the operation of MSK under non-linear conditions were studied, and their major findings are presented. Due to the lack of theory on the effects of incorrect modulation index on the error performance of MSK, an investigation into this avenue was performed. The design of a correction mechanism for maintaining the modulation index at its correct value is described, and aspects of its implementation are considered. Using the available literature, various modules of which a coherent MSK modem is comprised were developed, and their design is discussed. The design of a non-coherent MSK demodulator is also described.

It is shown that MSK is closely related to both Staggered QPSK, as well as CP-FSK. The maximum ideal spectral efficiency of MSK was found to be 2 Bits/s/Hz, and its ideal error probability was shown to be the same as that for antipodal signaling. It was also found that both MSK and SQPSK benefit in their spectral occupancy under conditions of filtering and non-linear amplification, due to the time offset between their I- and Q-Channels. The effect of incorrect modulation index on MSK was shown to be that the

modulation index tolerance sets the minimum attainable error probability, independent of the signal-to-noise ratio. The constructed MSK modulator was found to offer satisfactory performance, and its hardware implementation was simple. The coherent demodulator was more complex in its implementation, and several modules in its structure were designed, constructed, and their performance evaluated. The non-coherent MSK demodulator was found to be simple to construct, but it was theoretically shown that it is more than 3 dB worse off in noise performance than the more complex coherent demodulator.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

NOMENCLATURE

Antipodal	Two pulses are termed antipodal when one is the additive inverse of the other.
AWGN	Additive White Gaussian Noise. The spectrum of AWGN is approximated as being flat and is denoted as having a two-sided noise density of $N_o/2$ W/Hz. It is termed as being additive as it corrupts a signal in a linear (additive) manner, as opposed to multiplicative noise.
Bessel Spectrum	This is a Frequency Modulation (FM) spectrum, and it is so-called because the amplitudes of the spectral components are given by the values of the Bessel coefficients (which are determined by the modulation index).
Class-C	This is a mode of transistor operation whereby the transistor operates in saturated mode. i.e. it either conducts, or appears as an open circuit, depending on the base voltage.
DSB/SC	Double Sideband/Suppressed Carrier. In this method of modulation, the modulating wave directly multiplies the carrier wave, without the addition of a DC term (which is the case with AM). This effectively suppresses the carrier at the modulator output.

ISI	Intersymbol Interference. ISI is the interference of adjacent pulses with the amplitude of the present pulse.
NRZ	Non Return to Zero. With this pulse format, the pulse duration is equal to the bit duration. i.e. full-width pulses.
NRZ-L	NRZ-Level. This is the case where the pulses correspond directly with the source pulses. i.e. no form of coding is applied.
NRZ-M	NRZ-Mark. This is NRZ, but it only changes state with the appearance of a MARK ("1") in the source pulses. In effect, it is a form of differential encoding.
Orthogonal Signaling	This is the case when the time integral over the pulse duration of the product of two pulses is identically zero.
PD	Phase Detector.
PFD	Phase-Frequency Detector. This type of phase detector only produces an error signal when the edges of the two input waveforms are not aligned. When they are aligned, the output is tri-stated.
RZ	Return to Zero. This form of signaling has its pulse width less than the bit duration. i.e. the pulse returns to zero before the end of the bit interval.

$$Q(y) = \int_{y}^{\infty} e^{-x^2/2}\, dx$$

$$si\ (f) = \frac{\sin\ (f)}{f}$$

$$sinc\ (f) = \frac{\sin\ (\pi f)}{\pi f}$$

CHAPTER 1

INTRODUCTION

A problem facing the Digital Radio Communications industry is the efficient use of the available bandwidth for the transmission of data.

Over the years, many different digital modulation schemes have been developed, each with their own advantages and disadvantages, to meet the requirements of good performance (good error performance in the presence of noise, low bandwidth occupancy) and low cost of implementation. Due to the volume of development work, much literature is available to the reader in this field. The modulation scheme studied in this thesis, Minimum Shift Keying (MSK), is no exception. However, it was felt that a survey of literature on MSK, in conjunction with further research, would prove to be beneficial for any future development work on an MSK data link.

The objectives of this thesis are:

1. To survey the available literature on Digital Data Transmission, and MSK in particular.

2. To investigate the effects that filtering followed by non-linear amplification has on the error performance and spectrum of MSK. The effects on error performance due to incorrect modulation index are also investigated.

3. To investigate the hardware implementation of a

coherent MSK modem.

4. To design and construct a modulation index correction mechanism which maintains the modulator modulation index at its correct value.

5. To design a non-coherent MSK/FSK demodulator.

Due to the unavailability of a coherent MSK modem, it was not possible to obtain experimental data on the performance of MSK with incorrect modulation index. The investigation in this aspect is therefore theoretical.

The thesis begins with a theoretical background to digital data transmission. This culminates in the mathematical description of MSK and the derivation of its ideal performance. The performance of MSK under non-linear conditions as well as with incorrect modulation index is then studied. This is followed by the the design and construction of an MSK modulator and coherent demodulator. The development of a servo control circuit to maintain the MSK modulation index at its correct value is then described. This is followed by the description of the design and construction of a non-coherent MSK demodulator. The thesis ends with conclusions which are drawn from its major findings.

CHAPTER 2

THEORETICAL BACKGROUND TO DIGITAL DATA TRANSMISSION

2.1 SIMPLE BASEBAND SYSTEMS

In order to understand the more complex M-ary bandpass signaling schemes presented later on in this chapter, it is first necessary to study the theory of baseband pulse transmission.

2.1.1 Bandwidth Efficiency

A constraint in Communications Engineering is the amount of bandwidth available for the transmission of data. Ideally, we would like to transmit the maximum amount of information in the shortest possible time, whilst using the minimum amount of bandwidth. It is also desirable to transmit the information in such a way as to minimize detection errors at the receiver. A valid performance indicator of a communication link's efficiency is the Bits per second per Hertz (B/s/Hz) ratio (η). The higher the ratio, the more information we can transmit in a given bandwidth.

Consider a simple baseband data link, such as the one shown in Fig.2.1.

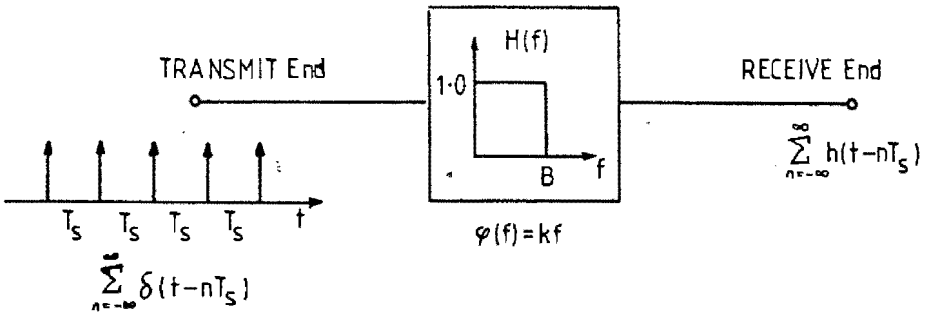


Fig.2.1 A Simple Baseband Data Link

The impulses are transmitted synchronously and are spaced T_s seconds apart. In this simple example, they are assigned to have a positive polarity, but this does not detract from the generality of our analysis. It is desirable, recalling the discussion above, to transmit the impulses at the maximum rate, through the minimum bandwidth, whilst minimizing the error rate at the receive end. In order to bandlimit the impulse train, we pass the waveform through a low-pass, linear-phase filter of bandwidth B Hertz [Hz]. Because of its sharp cut-off, it is commonly referred to as a "Brick-wall" filter. In order to calculate the maximum rate at which the impulses may be transmitted, we must first determine the shape of the impulses at the receive end (to ensure that the pulses do not interfere with eachother).

The response of the filter to a single impulse $\delta(t)$, termed the filter impulse response, is given by;

$$h(t) = F^{-1}\{H(f)\} = \int_{-\infty}^{\infty} H(f).e^{j2\pi ft}\,df$$

$$= \int_{-B}^{B} 1.e^{j2\pi ft}\,df$$

$$= 2B \sin(2\pi Bt) / 2\pi Bt$$

$$= 2B \text{ si } (2\pi Bt)$$

This function is plotted in Fig.2.2.

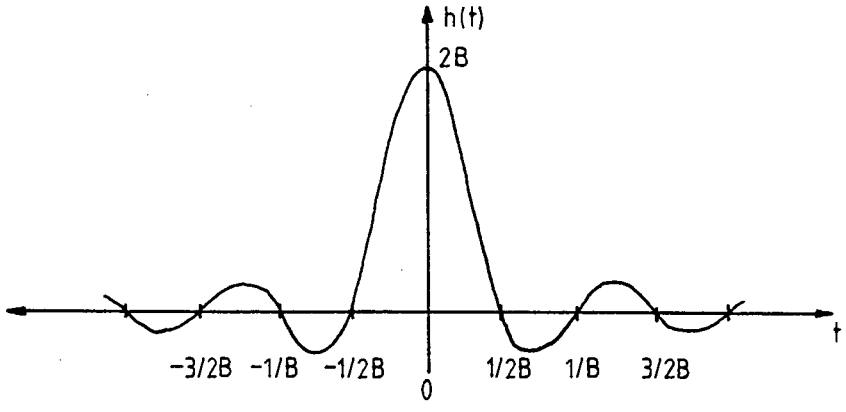


Fig.2.2 Brick-wall Filter Impulse Response

The following characteristics of the impulse response should be noted;

(a) A single impulse has resulted in a response extending from $-\infty$ to $+\infty$ in time (i.e. the filter is non-causal);

(b) The response passes through zero at the times $t = \pm 1/B, \pm 2/B, \ldots, \pm n/B$. Considering the case where $B=1/2T_s$ (recalling that T_s is the impulse spacing at the transmit end), we have;

$$h(t) = 1/T_s \text{ si } (\pi t/T_s).$$

Note that this function passes through zero at integer multiples of T_s, thus it is possible to transmit the impulses spaced T_s seconds apart through a bandwidth of $B = 1/2T_s$ [Hz]. This is due to the fact that at the peak

of the received pulse h(t), the responses due to all the other input impulses pass through zero, thus avoiding intersymbol interference (ISI) which can result in detection errors at the receive end.

In terms of the spectral performance indicator Ω, we are transmitting pulses at at a rate of $f_s = 1/T_s$ through a bandwidth of $1/2T_s$ [Hz], yielding,

$$\begin{aligned}\Omega &= f_s/B = 1/T_s \cdot 2T_s \\ &= 2 \text{ B/s/Hz}.\end{aligned}$$

In our derivation of h(t), the phase response $\phi(f)$ of the filter was ignored. We assume the phase characteristic to be linear. i.e.

$$\begin{aligned}\phi(f) &= k.f \\ \Rightarrow \phi'(f) &= k\end{aligned}$$

This result implies that the group delay of the filter is flat across its passband, and the received pulse is delayed k seconds relative to the input pulse.

The above derivation was first performed in the 1920's by Nyquist and is in fact his first (of three) theorems. In the real world, the brick-wall filtering (even without linear phase) is unachievable as it would require an infinite number of filter sections to realize such a large cut-off slope. Another drawback is that the pulses must be spaced precisely T_s seconds apart as any deviation from this will result in intersymbol interference (the pulse peaks will no longer occur at the zero crossings of the adjacent pulses). In order to circumvent these problems, Nyquist introduced a filtering function having Vestigial Symmetry [2.1].

A skew-symmetric transmittance function, $Y_1(f)$, is added to the brick-wall amplitude characteristic of the low-pass filter, with $Y_1(f)$ defined by the relation;

$$Y_1(f-x) = -Y_1(f+x), \qquad \text{where } 0 < x < f_N$$

and we have defined f_N = Nyquist Frequency = $1/2T_s$.

An infinite number of functions satisfy the skew-symmetric constraint, but one is of particular interest; The Raised-Cosine function. Before Raised Cosine filtering is discussed, we must first introduce the concept of Aperture Equalization.

In the simple baseband model, the input data consisted of a sequence of impulses. In practice, however, full length rectangular pulses of length T_s seconds are used. The Fourier transform of such a pulse has a si (f) shape, while that of an impulse is flat from $-\infty$ to $+\infty$ (i.e. a White Noise spectrum). We must therefore filter the full length rectangular pulses before applying them to the filter. This is commonly referred to as Aperture Equalizing or Pre-Whitening. The filter characteristic A(w) is the inverse of the rectangular pulse's spectrum, thus the equalizer has a 1/si (f) transfer function. For pulses of width T_s, we have $A(w) = 1/\text{si}\ (wT_s/2)$, where $w = 2\pi f$.

We can now define the composite filtering characteristic (which includes both the aperture equalization and Raised Cosine filtering) as;

$$H(jw) = \begin{cases} 1/si\ (wT_s/2) & 0 \le w \le \pi(1-\alpha)/T_s \\ 1/si\ (wT_s/2).\cos^2\{T_s[(w-\pi(1-\alpha))/T_s]/4\alpha\} & \text{for } \pi(1-\alpha)/T_s \le w \le \pi(1+\alpha)/T_s \\ 0 & w > \pi(1+\alpha)/T_s \end{cases}$$

$$\phi(jw) = kw \qquad 0 \le w \le \pi(1+\alpha)/T_s$$

where α is known as the roll-off factor of the filter and lies between $0 \le \alpha \le 1$. For the special case where $\alpha = 0$, we have brick-wall filtering as before, and as α is increased, the filter exhibits a more gradual roll-off in its characteristic (while increasing its bandwidth). A logarithmic plot of H(jw) for various values of α is shown in Fig.2.3.

For the case when $\alpha = 1$, the filtering is referred to as Full-Cosine rolloff filtering, and the bandwidth occupancy of the data spectrum extends to $2f_N$ ($=f_s$). The advantage of having $\alpha = 1$ is that the impulse response of the filter exhibits less oscillation and the impulse tails decay more rapidly (thus reducing the ISI). The more rapid decay of the pulses also reduces the the sensitivity of the system to timing errors in the bit rate frequency. Before, with $\alpha = 0$, the pulses decayed more slowly, so timing errors caused a significant amount of ISI to occur if f_s was not exactly correct. However, the spectral efficiency of the system has reduced from $\Omega = 2$ B/s/Hz to $\Omega = f_s/[(1+\alpha).f_s/2] = 2/(1+1) = 1$ B/s/Hz. The factor α thus represents a tradeoff between bandwidth usage and sensitivity of the system to timing errors.

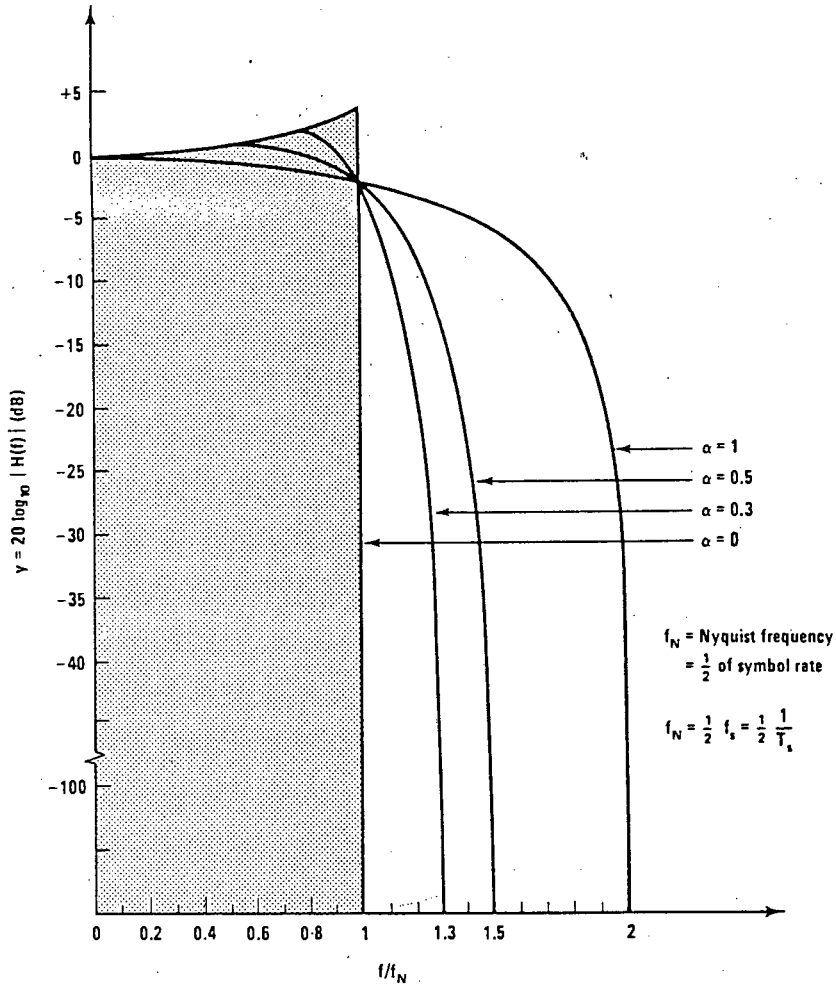


Fig.2.3 Raised Cosine Filtering for Various Values of α [2.1]

A qualitative assesment of the performance of the data link may be made by observing the Eye Diagram of the received pulses. The eye diagram gives one information as to the amount of ISI and jitter present at the receive end of the link. It is, in fact, the superposition of all possible responses of the channel filtering to the transmitted

pulses. The most important feature of the eye diagram is the eye opening, and it is desirable to have a fully-open eye (100 % eye opening) as this means that no ISI is present. Fig.2.4 illustrates an eye diagram for a system with a filter $\alpha = 0.3$ [2.1].

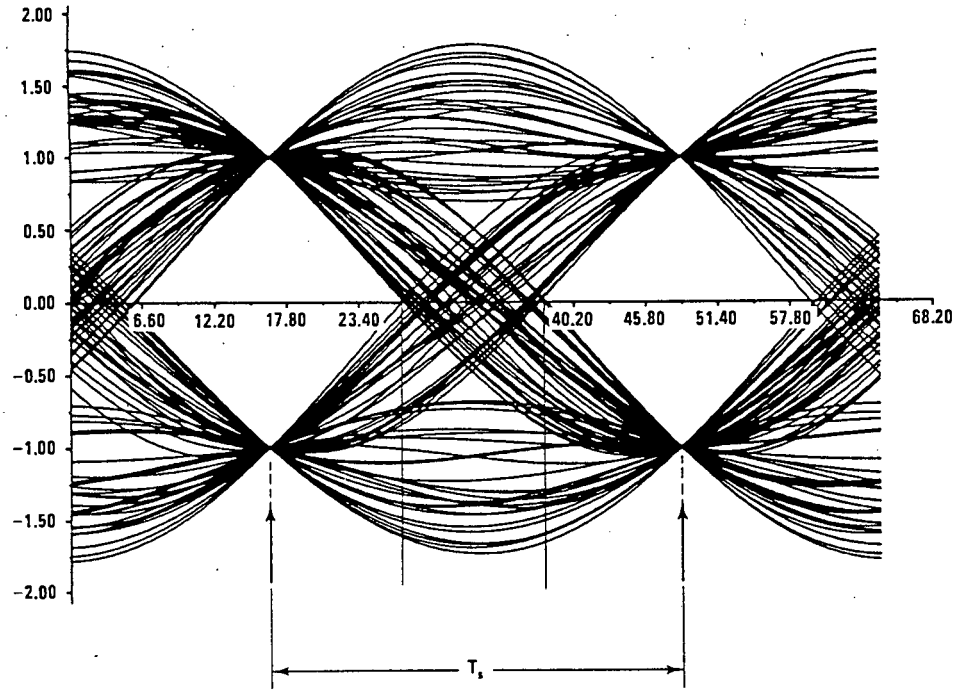


Fig.2.4 Eye Diagram for Filter $\alpha = 0.3$ [2.1]

2.1.2 Calculation of Error Probability

The discussion in the previous section restricted the source of detection errors at the receiver to intersymbol interference only. In practice, one has to take not only ISI into account, but thermal Additive White Gaussian Noise (AWGN) as well. In this section, we address ourselves to deriving the probability of error (P_e) as a function of signal to noise ratio (SNR).

Recalling the well known result of the ideal matched filter [2.2], we know that the optimum receiver filter for an input pulse shape p(t), corrupted by AWGN of two-sided spectral density $S_n(w) = N_o/2$ [W/Hz] is given by;

$$H(w) = k'P(-w)e^{-jwT_o} \qquad \ldots(2.1)$$

where,

k' = arbitrary constant
$P(w)$ = Fourier Transform of the baseband pulse p(t)
T_o = pulse width of p(t)

The error probability of such a system is given by the expression;

$$P_e = Q([2E_p/N_o]^{-\frac{1}{2}}) \qquad \ldots(2.2)$$

where E_p = energy of p(t)

$$= 1/2\pi \int_{-\infty}^{\infty} |P(w)|^2 \, dw \qquad \ldots(2.3)$$

The implementation of a simple baseband link utilizing a matched filter (matched to p(t)) and using p(t) to transmit a "1", and -p(t) to transmit a "0" equiprobably is shown in Fig.2.5.

The operation of the system is as follows; The transmitted pulse (equiprobably p(t) or -p(t)) is corrupted by AWGN present in the channel. The received pulse $\pm p(t)$ is multiplied by the stored replica of p(t) at the receiver, and the product integrated over pulse duration $(0,T_o)$ to yield y(t). At time $t = T_o$, y(t) is sampled and a decision

made by the decision threshold. If $y(T_o)$ is positive, the data bit is assumed to be a "1", and if it is negative, it is assumed to be a "0". When this is completed, the integrator is reset to zero, and the process repeats. Because of its operation, such a receiver is often referred to as an Integrate, Sample and Dump (IS&D) receiver.

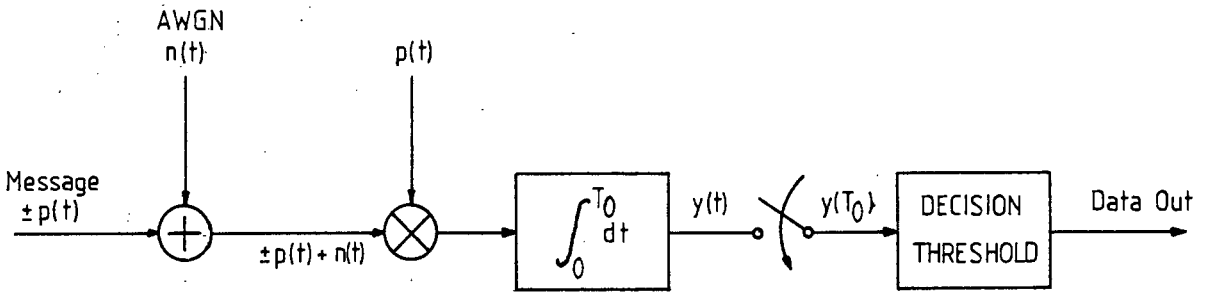


Fig.2.5 Matched Filter Receiver

It is often the case that instead of using a single pulse p(t) to convey the information, two different pulses, e.g. p(t) and q(t) are used. i.e. One transmits p(t) to convey a "1" and q(t) to convey a "0". Using the results of the Optimum Binary Receiver [2.2], it can be shown that if the energy of p(t) is given by E_p (as calculated in equation 2.3), and the energy of q(t) is given by E_q, then the probability of error (for the case when p(t) and q(t) occur with equal probability) is given by;

$$P_e = Q\left[\left[(E_p+E_q-2E_{pq})/2N_o\right]^{\frac{1}{2}}\right] \qquad \ldots(2.4)$$

where $E_{pq} = \int_0^{T_o} p(t).q(t)\ dt$

and the decision threshold $a_o = \frac{1}{2}[E_p - E_q]$. ...(2.5)

Note that the error probability depends only on the energy and not the shapes of the pulses. It is therefore possible for different signaling schemes (with different baseband pulse shapes) to have identical error performances. A few special cases are of interest;

(a) $q(t) = -p(t)$. This is our original signaling scheme (as shown in Fig.2.5), and it is easily shown (from equation 2.4) that the probability of error $P_e = Q(\sqrt{2\mu})$ where we define $\mu = E_p/N_o$ (the Energy per bit to double sided Noise Density ratio). This scheme is referred to as an Antipodal signaling scheme (because $q(t) = -p(t)$), and is also sometimes termed Polar signaling.

(b) $E_p = E_q = E$, and $E_{pq} = 0$. In this case , the cross correlation of the pulses is zero, and the error probability $P_e = Q(\sqrt{\mu})$ ...(2.6). This scheme is known as Orthogonal signaling and it is clear that it is 3dB inferior in noise performance relative to Antipodal signaling.

(c) There is a case where P_e is minimized subject to the minimization of E_{pq}, and this will be discussed in section 2.2.3.

2.2 BANDPASS (CARRIER) SYSTEMS

In this section four digital bandpass systems, namely BPSK, SQPSK, CP-FSK, and MSK are studied. The three systems preceding MSK are discussed as the are closely linked with MSK, and the understanding of their operation aids one in understanding the operation of MSK.

2.2.1 Bi-Phase Shift Keying (BPSK)

2.2.1(a) Theory of Operation. The block diagram, Fig.2.6, shows the basic BPSK modem configuration;

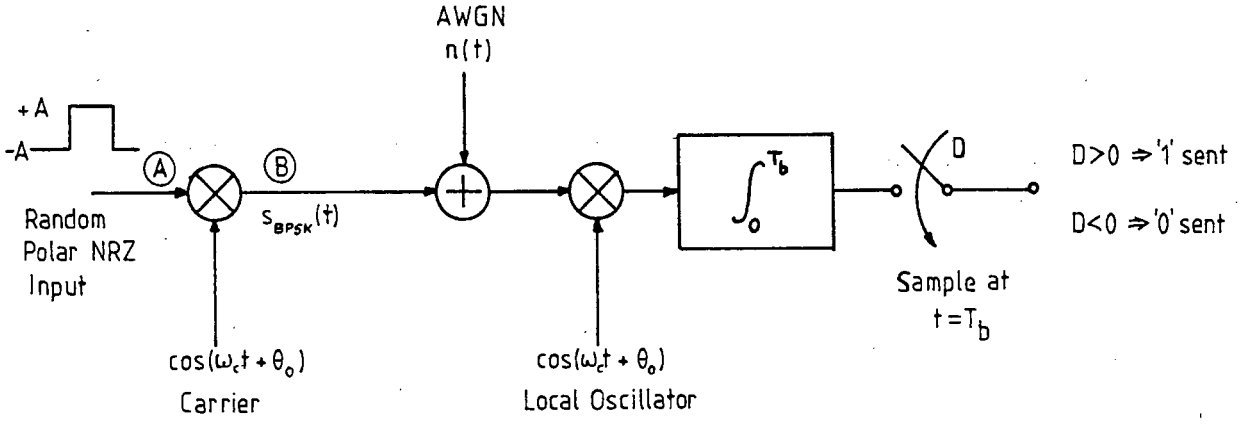


Fig.2.6 BPSK Modem Configuration [2.4]

The incoming random data stream at bit rate $f_b = 1/T_b$ and in polar NRZ format DSB/SC modulates the carrier (at frequency w_c) to produce the BPSK signal $s_{BPSK}(t)$. In effect, the data stream is phase modulating the carrier as when the data value is +A volts, the output of the multiplier is $A\cos(w_c t + \theta_o)$, and when it is -A volts, the output is $-A\cos(w_c t + \theta_o)$ which is equivalent to $A\cos(w_c t + \theta_o + \pi)$. The carrier phase thus abruptly

changes by ±180°, depending on the input data value. The transmitted signal is then corrupted by AWGN in the channel before it is received at the receive end. It is then multiplied by the local oscillator of the receiver (which is precisely phase-locked to the carrier at the transmitter), before being processed by the IS&D circuitry as described in section 2.1.2.

It is clear that BPSK is an Antipodal signaling scheme, as the baseband pulse shapes are identical in shape (rectangular), and opposite in polarity. The error probability is hence given by;

$$P_e(\text{BPSK}) = Q(\sqrt{2\mu}) \qquad \ldots(2.7)$$

2.2.1(b) Spectral Efficiency. As the baseband pulse shape of BPSK is rectangular, the Power Spectral Density (PSD) is expected to be of $si^2\ (f-f_c)$ form. The PSD is shown in Fig.2.7.

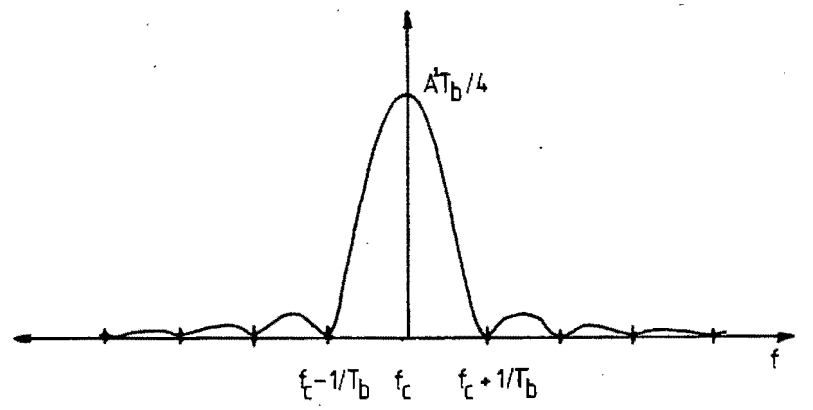


Fig.2.7 BPSK Power Spectral Density

The PSD is seen to be unconstrained, and in order to increase the spectral efficiency (Ω), one needs to apply

Nyquist filtering to the signal. It was shown previously that $\alpha = 0$ filtering allows for the transmission of pulses at baseband at a rate f_b through a bandwidth $f_b/2$. Were such a low-pass filter to be placed at point (A) in Fig.2.6, then the output spectrum of the modulator would extend from $f_c - f_b/2$ to $f_c + f_b/2$. We are therefore transmitting f_b bits per second through a bandwidth of f_b Hz, yielding a maximum spectral efficiency for BPSK of 1 B/s/Hz. It is also possible to perform the filtering at point (B) in Fig.2.6. The filter would then be a bandpass filter of bandwidth f_b Hz, which is commonly referred to as the Double-Sided Nyquist bandwidth. In practice, the spectral efficiency achieved with BPSK is typically 0.8 B/s/Hz [2.3].

A practical feature of the BPSK modems (and any PSK-type modems in general) which is not shown in Fig.2.6 is the differential coder/decoder used in the modulator and demodulator respectively. The coding is necessary in order to remove the ±180° phase ambiguity due to the carrier-recovery circuitry (squaring loops, as will be described in section 5.2) present in the demodulator.

The Constellation Diagram for BPSK is shown in Fig.2.8.

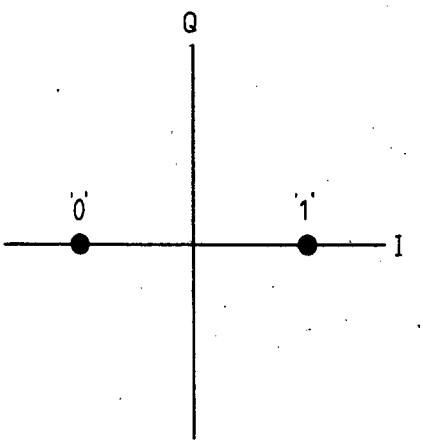


Fig.2.8 BPSK Constellation Diagram

The two dots indicate all possible phases (0°,180°) of the carrier at the decision instants. It should be clear that the margin for error of each of the constellation points is ±90°. Were the carrier phase to lie in the wrong half of the plane at the decision instant, a bit error would result.

2.2.2 Staggered Quadrature Phase Shift Keying (SQPSK)

2.2.2(a) Theory of Operation. To explain the operation of a SQPSK modem, we refer to the block diagram in Fig.2.9 [2.4].

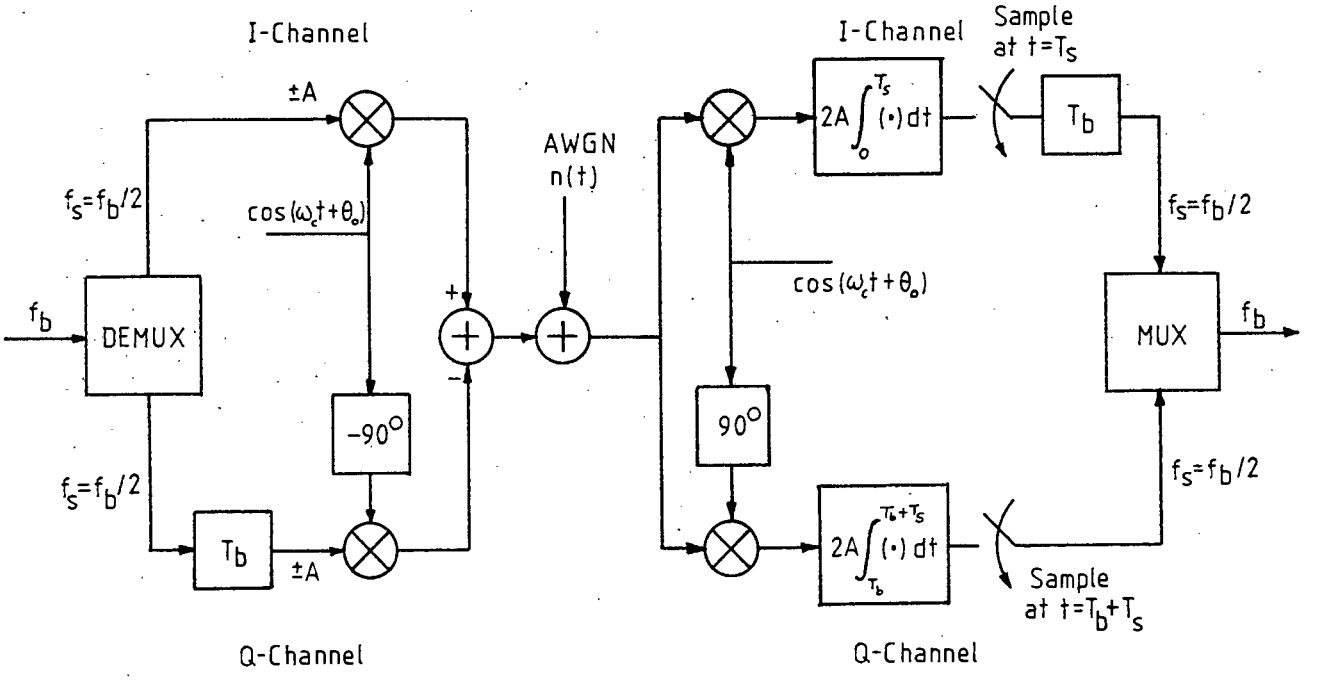


Fig.2.9 SQPSK Modem Configuration

The incoming data stream in polar NRZ format and at bit rate $f_b = 1/T_b$ is divided by the parallel-to-serial converter into two streams, each of symbol rate $f_s = f_b/2$. The symbols are also in polar NRZ format, and have a peak-to-peak amplitude of 2A volts. The lower channel, denoted the Q-Channel, symbols are delayed by one-half a symbol period (i.e. a bit period) with respect to the upper channel (the I-Channel). Fig 2.10 shows typical I- and Q-Channel waveforms.

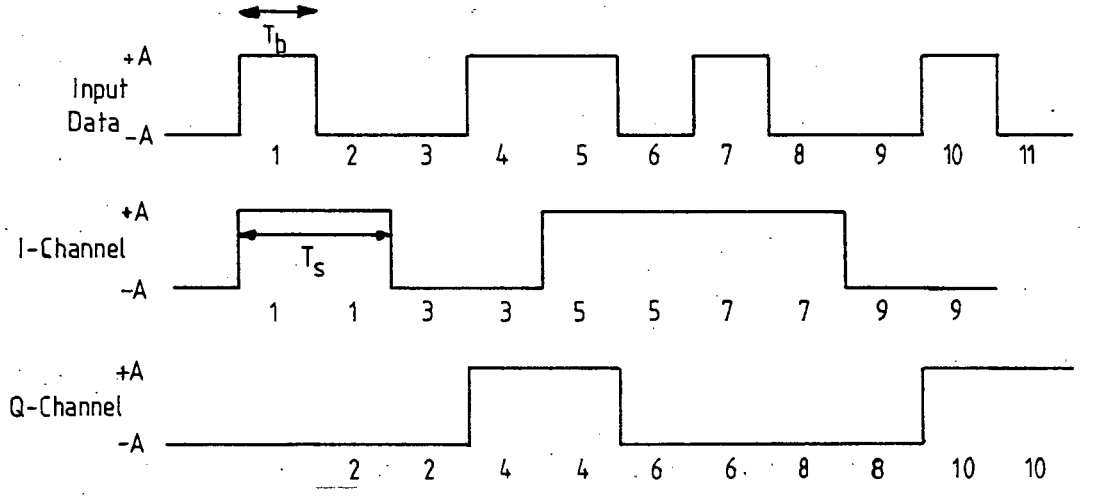


Fig.2.10 Typical I- and Q-Channel Waveforms for SQPSK

It is clear that the symbol duration is twice that of the input bit duration and that they are offset ("Staggered") by a bit period. The symbols DSB/SC modulate the I- and Q-channel carriers and note that the Q-Channel carrier is delayed 90° relative to that of the I-Channel, hence the term (Staggered) Quadrature Phase Shift Keying. The two channels are then subtracted to yield the waveform $s_{SQPSK}(t)$.

The transmitted waveform is corrupted by AWGN in the channel. At the receiver, the signal is powersplitted into two channels; the upper channel is multiplied by a locally generated carrier which is phase-locked to the transmitter carrier w_c, while the lower channel is multiplied by the same carrier phase shifted by 90°. The upper channel is now denoted the I-Channel, while the lower channel is denoted the Q-Channel. The I-Channel IS&D circuit operates over the interval $(0,T_s)$, while that of the Q-Channel operates over (T_b,T_s+T_b). Note that the receiver channels are also staggered, and that the symbols are observed over a $2T_b$ (T_s) interval before a decision is made.

The T_b offset between the channels is compensated for by inserting a T_b offset delay in the I-Channel. The channels

are combined by means of a parallel-to-serial converter which yields the output data stream at a bit rate of f_b Bits/s.

From the topology of the SQPSK modem, it is apparent that it consists of two parallel BPSK modems operating in phase quadrature. From this we may conclude that the probability of error for each channel is identical and the total probability of error is given by [2.4];

$$P_e \text{ (SQPSK)} = Q(\sqrt{2\mu}) \qquad \ldots(2.8)$$

2.2.2(b) Spectral Efficiency. As was the case with BPSK, the baseband pulse shape of SQPSK is rectangular. We therefore expect SQPSK to have a $si^2(f-f_c)$ spectrum. However, a major difference between BPSK and SQPSK is that the symbol length of SQPSK is twice the bit period. This means that the main lobe of the PSD extends from f_c-f_s to f_c+f_s. i.e. the main lobe width is $2f_s = f_b$ wide, as compared to $2f_b$ wide for BPSK. The PSD is shown in Fig.2.11.

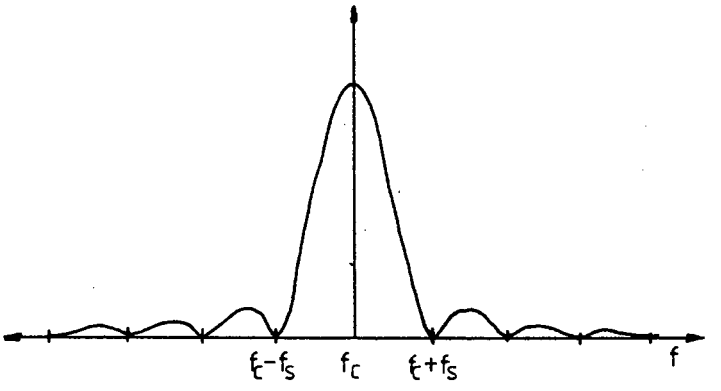


Fig.2.11 SQPSK Power Spectral Density

For α = 0 Nyquist filtering, we could either filter the

baseband symbols with a low-pass filter of cut-off frequency $f_s/2$ [Hz], or with a band-pass filter at RF with a bandwidth of f_s [Hz]. This enables us to transmit at a rate of f_b B/s through a bandwidth of $f_s = f_b/2$ [Hz]. We can therefore deduce that the ideal spectral efficiency of SQPSK is;

$$\eta(\text{SQPSK}) = f_b/(f_b/2) = 2 \text{ B/s/Hz} \qquad \ldots(2.9)$$

In practice, spectral efficiencies of 1.9 B/s/Hz can be achieved [2.3].

The constellation diagram, as should be expected, consists of two BPSK constellation diagrams in phase quadrature. This is shown in Fig.2.12.

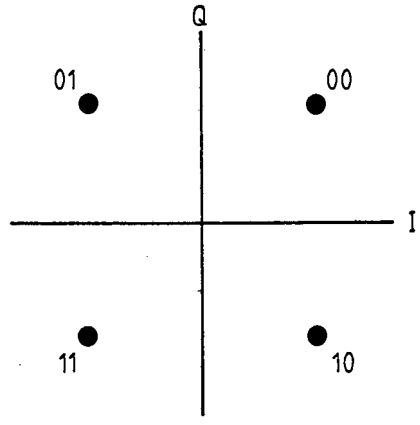


Fig.2.12 SQPSK Constellation Diagram

To conclude the discussion on SQPSK, mention must be made of its related signaling scheme, namely, QPSK. SQPSK is a derivative of QPSK, and it was developed in order to overcome the spectral deficiencies of QPSK under non-linear amplification at the transmitter. The timing diagram for QPSK is similar to that of SQPSK except that the two channels change state at the same instants (they

are not "Staggered"). Considering the constellation diagram in Fig.2.12, it is apparent that as the two channels can both change at the same time, it is possible for the modulator to change from the state "00" to "11", or from "01" to "10". This results in the carrier passing through the origin of the diagram. i.e. the amplitude of the carrier is reduced to zero at certain times. This results in the QPSK waveform having 100% AM modulation present.

In practical transmitters, filtering is performed at I.F. prior to up-conversion and RF amplification as this enables lower Q filters to be used (by definition of Q). The effect of filtering on the waveform is to distort its envelope (introduce AM) and this is exacerbated by the AM already present on the QPSK waveform. The RF amplifiers in most transmitters (especially on satellite links) are often operated in their non-linear region to achieve maximum power efficiency. This non-linear amplification further distorts the envelope by re-introducing the sharp transitions originally removed by the I.F. filters (filtering removes the high frequency components of the waveform, therefore removing the sharp transitions). The net effect is that the filtered sidelobes of the QPSK spectrum are regenerated after amplification, thereby increasing the spectral occupancy.

In SQPSK, the 100% AM is removed by ensuring that only one channel at a time can change state. Referring to Fig.2.12, one can see that the carrier can only change by ±90° per bit time, and not ±180° per symbol time as was the case for QPSK. The envelope of the SQPSK waveform therefore has less envelope fluctuation, and filtering followed by non-linear amplification leads to much lower sidelobe

regeneration [2.4]. The error probabilities and Spectra for QPSK and SQPSK are identical under linear conditions.

2.2.3 Continuous Phase-Frequency Shift Keying (CP-FSK)

2.2.3(a) Theory of Operation. The generalized block diagram for a CP-FSK modulator is shown in Fig.2.13.

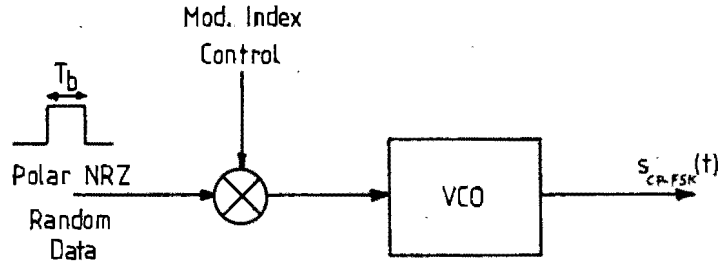


Fig.2.13 Generalized CP-FSK Modulator

The incoming bit stream at bit rate f_b and bit duration T_b is in polar NRZ format. It is processed by the modulation index control circuitry to yield a square wave which alternates between two adjustable voltage levels. This waveform is applied to the Voltage Controlled Oscillator which varies its output frequency between the two frequencies set by the control voltage levels. The modulation index of the modulator is given by the expression;

$$h = \frac{\Delta f}{f_b} \qquad \ldots(2.10)$$

where;

h = FM modulation index

Δf = peak-to peak frequency deviation of VCO

f_b = bit rate of modulating waveform

Fig.2.14 shows the frequency map of the VCO. Note that this is <u>not</u> the spectrum of the generated CP-FSK.

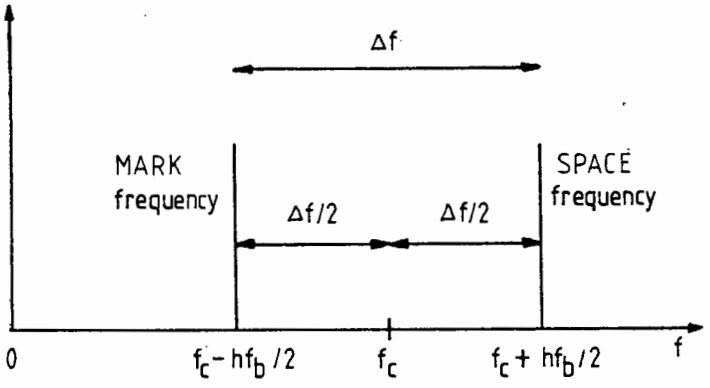


Fig.2.14 Frequency Map of CP-FSK Modulator VCO

The lower frequency, termed the MARK frequency, is selected when the data bit is a "1", and the higher frequency, the SPACE frequency, is selected when the data bit is a "0". It is important to note that since the output waveform is generated by one oscillator (the VCO), phase continuity is guaranteed at the bit transition instants (hence the nomenclature <u>Continuous Phase</u>-Frequency Shift Keying). The time-waveform for CP-FSK is given by the expression;

$$s_{CP-FSK}(t) = \sqrt{2}A.\cos\ (w_c t + u_k.(\pi h/T_b)t + \theta_k)$$

$$\text{for } kT_b \leq t \leq (k+1)T_b \qquad \ldots(2.11)$$

where;

A = RMS amplitude of carrier

w_c = radian frequency of carrier = $(w_m+w_s)/2$, where w_m is the MARK frequency and w_s is the SPACE frequency.

u_k = data bit value in polar NRZ format = ± 1

θ_k = recursive phase constant to ensure phase-continuity at the bit transitions.

Note that the term $u_k(\pi h/T_b)t$ varies the instantaneous frequency between;

$$\pm\ (1/2\pi).(\pi h/T_b) = \pm\ h/(2T_b),$$

therefore we can say;

$$\Delta f = h/T_b = hf_b \qquad \ldots(2.12)$$

Fig.2.14 illustrates this relation clearly.

There are three values of h which are of particular interest;

(i) h = 1. This is the FSK scheme proposed by Sunde [2.5]. It has not found widespread acceptance as its spectrum has two undesired properties, namely;

- There are two discrete spectral lines in its spectrum, which indicate that some power is wasted transmitting two carriers which convey no data;

- The spectrum is wide (the width of the main lobe is $3f_b$ as compared to f_b for SQPSK).

(ii) $h = 0.715$. To understand why this particular modulation index is of interest, we recall equation (2.4) and set;

$p(t) = \sqrt{2}A.\cos(w_c - \Delta w/2)t$ i.e. the MARK pulse

and

$q(t) = \sqrt{2}A.\cos(w_c + \Delta w/2)t$ i.e. the SPACE pulse,

remembering that $\Delta w = 2\pi.h/T_b$ (equation 2.12).

To minimize the probability of error P_e, we must maximize the abscissa of the Q function in equation 2.4. This implies that the E_{pq} term in the numerator must be maximized subject to h. We have;

$$E_{pq} = \int_0^{T_b} p(t).q(t)\,dt$$

$$= A^2T_b[si\,(\Delta w)T_b + si\,(2w_cT_b)]$$

...(2.13)

The intermediate steps of the calculation may be found in [2.2]. In practice, the carrier

frequency is much larger than the bit rate hence the right-hand term in 2.13 may be ignored. Equation 2.13 then reduces to;

$$E_{pq} = A^2 T_b . \text{sinc}\ (2\Delta f T_b) \qquad \ldots(2.14)$$

This function is plotted in Fig.2.15.

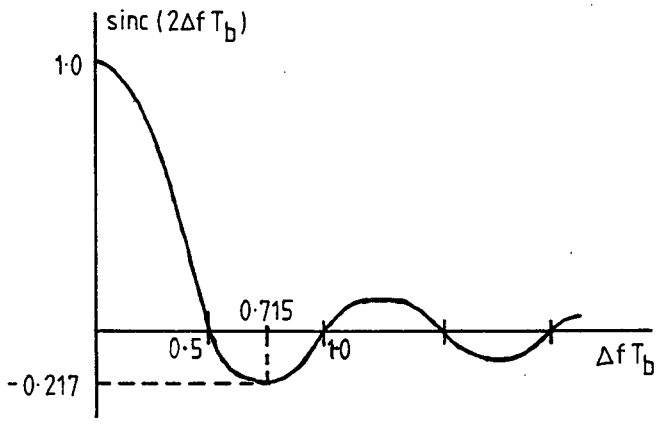


Fig.2.15 E_{pq} vs Modulation Index

The minimum value of E_{pq} is -0.217 and this occurs when $\Delta f T_b = h = 0.715$. We also have that $E_p = E_q = E$ since the pulse amplitudes are the same. This results in an error probability of

$$P_e(\text{CP-FSK, } h=0.715) = Q(\sqrt{1.217\mu}) \qquad \ldots(2.15)$$

which is a 0.85 dB improvement over orthogonal FSK [2.4].

(iii) h = 0.5. This is MSK by definition and will be discussed in section 2.2.4.

In the preceding discussion, no mention was made of the

demodulator structure for CP-FSK. This was deliberate as it is not generally possible to construct coherent receivers for arbitrary h values, only for h being one half, integers, and integers plus one-half [2.6]. It is far simpler to demodulate CP-FSK by means of discriminator detection [2.7], but then the phase information contained in the carrier is lost (with the subsequent degradation in error performance). The results presented in [2.7] suggest that the optimum value for h on wideband channels lies between 0.7 and 0.8.

2.2.3(b) Spectral Efficiency. The Power Spectral Densities for the cases h = 0.5, 0.7, 1.3, and 1.6 are plotted in Fig.2.16 [2.4].

Note that the bandwidth occupancy increases with increasing h, hence it is desirable to choose the smallest h value which meets our requirements. Another important feature of the spectra is their $(f-f_c)^{-4}$ rolloff for large offsets from the carrier, as compared to the $(f-f_c)^{-2}$ rolloff of the BPSK and SQPSK spectra. The reason for the more rapid rolloff is inherent in the continuous phase nature of the CP-FSK waveform. The BPSK and SQPSK schemes allowed for abrupt changes in carrier phase, whereas the CP-FSK scheme gradually changes the carrier phase over a bit time while preserving phase continuity at the bit transition boundaries. This aspect of CP-FSK will be more fully discussed in section 2.2.4.

The results presented in Oetting's paper [2.3], indicate that CP-FSK with a modulation index of 0.7 has an achievable spectral efficiency of 1 B/s/Hz, while a modulation index of 1.0 (Sunde's FSK) has a typical efficiency of only 0.8 B/s/Hz.

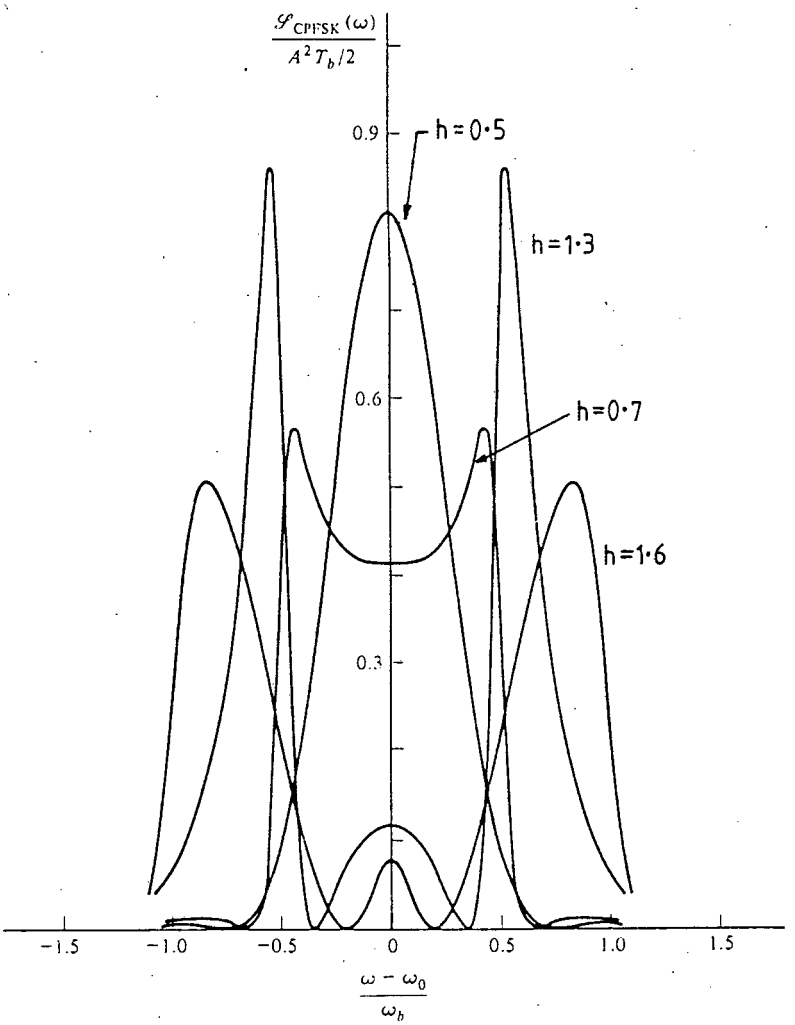


Fig.2.16 Power Spectral Densities for CP-FSK With Different Modulation Indices [2.4]

2.2.4 Minimum Shift Keying (MSK)

2.2.4(a) Mathematical Definition of MSK. We begin by recalling equation (2.10) and setting h = 0.5. This

immediately tells us that the frequency map for MSK consists of two frequencies spaced by;

$$\Delta f = 0.5 f_b \qquad \ldots(2.16)$$

i.e. the MARK and SPACE frequencies are spaced by one-half the bit rate. Following the constraint set by Sullivan [2.8], we choose the apparent carrier frequency (the arithmetic mean of the MARK and SPACE frequencies) to be an odd multiple of the shift frequency (the shift frequency is the peak deviation of the carrier from its centre frequency i.e. $\Delta f/2$).

$$\Rightarrow \quad f_c = N \frac{\Delta f}{2} \quad \text{where N is an odd integer} \qquad \ldots(2.17)$$

For an IF frequency of 70.144 MHz and a bit rate of 2.048 Mbps, we have N = 137, and the MARK frequency is given by;

$$\begin{aligned} f_m &= f_c - \Delta f/2 \\ &= 70.144 - (0.5)(0.5)(2.048) \quad [\text{MHz}] \\ &= 69.632 \text{ MHz}. \end{aligned}$$

The SPACE frequency is given by;

$$\begin{aligned} f_s &= f_c + \Delta f/2 \\ &= 70.656 \text{ MHz}. \end{aligned}$$

Now setting h = 0.5 in equation (2.11), we obtain;

$$s_{MSK}(t) = \cos (w_c t + u_k \pi t/2T_b + \theta_k) \qquad kT_b \le t \le (k+1)T_b \qquad \ldots(2.18)$$

where θ_k, the recursive phase constant, is calculated from;

$$\theta_k = \theta_{k-1} + (u_{k-1} - u_k)\pi k/2 \qquad \ldots(2.19)$$

To calculate the phase shift that occurs over a single bit period, we first substitute t = 0, then t = T_b in the argument of the cosine term in equation (2.18) and obtain their difference. We set u_k = 1, and this yields;

$$\theta(t=0) = \theta_k$$

and

$$\theta(t=T_b) = w_c T_b + \pi/2 + \theta_k.$$

Subtracting, we obtain $\theta(t=T_b) - \theta(t=0) = \Delta\theta$

and $\Delta\theta = w_c T_b + \pi/2$.

We can now see that u_k = 1 has imparted an additional phase of +90° on the carrier, and it is just as easily shown that u_k = -1 retards the carrier by -90° over a bit period. Using this result and equation (2.19), we can sketch the additional phase trellis diagram for MSK (by additional it is meant that the term $w_c T_b$ due to the carrier centre frequency is neglected, as it is invariant to the bit stream). Fig.2.17 shows such a phase trellis diagram for the data sequence 1,-1,-1,1,1,1.

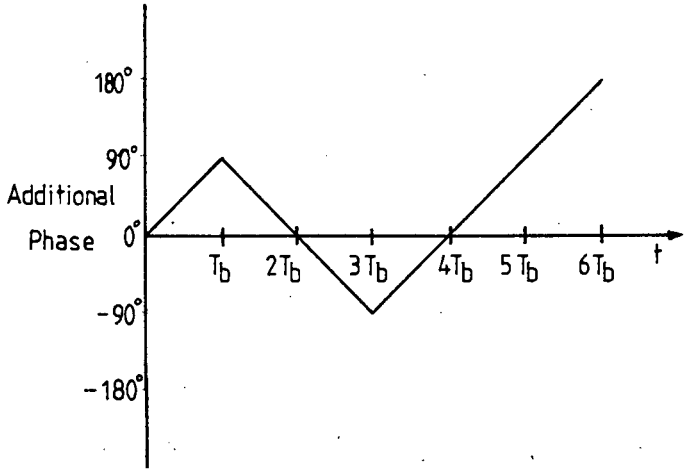


Fig.2.17 MSK Phase Trellis Diagram

2.2.4(b) Calculation of MSK Error Probability. MSK, by definition, is an orthogonal signaling scheme. That is, $E_p = E_q = E$, and $E_{pq} = 0$ as was discussed in section 2.1.2. Sullivan's constraint (equation 2.17) guarantees this, and this can be shown to be true as follows;

For two sinusoids to be orthogonal over a period T_o, one must have a frequency mf_o and the other nf_o, where m<>n, and they are both integers [2.2]. If we let f_o equal the shift frequency ($\Delta f/2$), then for the MARK and SPACE frequencies we have chosen we have;

$$f_m = (136)f_o$$

and

$$f_s = (138)f_o.$$

We can therefore conclude that the MARK and SPACE frequencies chosen guarantee that the MSK generated will be orthogonal.

To calculate the error probability of MSK, we use equation (2.4) with $E_p = E_q = E$, and $E_{pq} = 0$, giving;

$$P_e = Q(\sqrt{[2E/2N_o]}) = Q(\sqrt{\mu}) \qquad \ldots(2.20)$$

It would thus appear that MSK is at a 3 dB disadvantage in error performance compared to antipodal signaling (i.e. we would need to double the average energy per bit-to-double sided noise density ratio μ in order to obtain the same P_e as BPSK or SQPSK).

However, in the calculation of P_e for MSK, it was implicitly assumed that symbols are observed over a single bit period T_b before a decision is made. This is not the maximum time of observation available to us, however. By virtue of the fact that MSK is a continuous phase scheme, the phase in the succeeding bit interval is dependent on the prior bit (this is seen in the phase trellis diagram in Fig.2.17). It will be shown in section 2.2.4(c) that the MSK waveform may be decomposed into two quadrature channels, analogous to the SQPSK scheme. We recall that the symbols in the I- and Q-Channels in the SQPSK receiver were observed over a symbol period $T_s = 2T_b$. It is therefore clear that we may observe the MSK symbols over twice the bit period, thus doubling the observed energy. Were MSK to be demodulated in this manner, the probability of error becomes;

$$\begin{aligned} P_e(\text{MSK}) &= Q(\sqrt{[4E/2N_o]}) \\ &= Q(\sqrt{2\mu}) \end{aligned} \qquad \ldots(2.21)$$

This performance is identical to the antipodal schemes such as BPSK and SQPSK on linear, infinite bandwidth channels. It is not, however, the optimum performance achievable with coherent CP-FSK schemes. Interesting correspondence [2.10,2.11] suggests that $h = 0.67$ and a $4T_b$ observation interval offers a 0.8 dB advantage over MSK. Note, however, that coherent receivers for CP-FSK

with modulation indices unequal to multiples of 0.5 are difficult, if not impossible to construct. MSK is therefore a compromise between excellent noise performance and receiver complexity.

2.2.4(c) Reduction of MSK to SQPSK Format. It will be shown here that MSK is very similar to SQPSK, and this enables us to use the knowledge gained in the study of SQPSK to further understand the operation of MSK.

It can be shown [2.4,2.12] that the expression for MSK in equation (2.18) can be decomposed into the form;

$$s_{MSK}(t) = y_I(t).\cos(\pi t/2T_b).\cos(w_c t) - y_Q(t).\sin(\pi t/2T_b).\sin(w_c t) \quad \text{for } kT_b \le t \le (k+1)T_b \qquad ...(2.22)$$

where;

$y_I(t)$ = I-Channel symbol polarity (±1, NRZ format)
$y_Q(t)$ = Q-Channel symbol polarity
$\cos(\pi t/2T_b)$ = I-Channel symbol weighting
$\sin(\pi t/2T_b)$ = Q-Channel symbol weighting
w_c = carrier frequency as described previously.

It is clear that the MSK waveform has been decomposed into two quadrature channels, one staggered relative to the other by a bit period (this is seen from the fact that the I-Channel symbol is cosinusoidal, while that of the Q-Channel is sinusoidal). Note that the symbol duration is $T_s = 2T_b$ as was mentioned in the previous section. The one major difference between MSK and SQPSK is the shape of the symbols: SQPSK has rectangular symbol weightings. This has a major impact on the difference between the two schemes'

power spectra.

2.2.4(d) Generation of MSK. The form of equation (2.22) suggests what is termed the Parallel generation of MSK (there are two parallel channels operating in phase quadrature). Such a modulator implementation is shown in Fig.2.18.

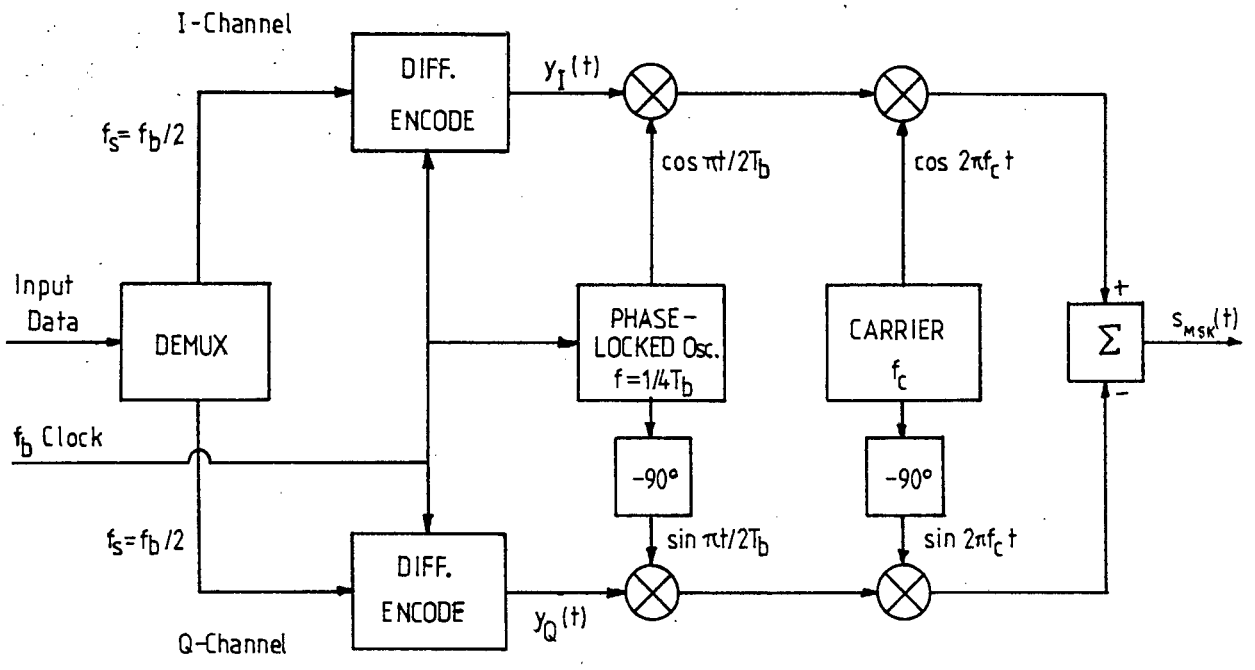


Fig.2.18 Parallel MSK Modulator [2.13]

The incoming bit stream at bit rate $f_b = 1/T_b$ and in polar NRZ format is split into two parallel streams, each at symbol rate $f_s = 1/T_s = f_b/2$. The upper stream is denoted the I-Channel, and the lower stream is denoted the Q-Channel. Both channels are differentially encoded (to remove the 180° phase ambiguity at the receiver) prior to being weighted by the sin- and cosinusoidal symbols. The resultant symbols are then upconverted to the carrier frequency and then subtracted to yield the MSK waveform.

In practice, this may prove to be a complex method of MSK generation, and a simpler implementation is to use the CP-FSK visualization of MSK to generate it. This will be further discussed in Chapter 4.

2.2.4(e) Coherent MSK Demodulation. For coherent, matched filter detection of MSK (achieving the noise performance of antipodal signaling), it is necessary to generate replicas of the MSK cosinusoidal weighting functions at the receiver. The block diagram for de Buda's [2.9] clocked matched filter receiver is shown below.

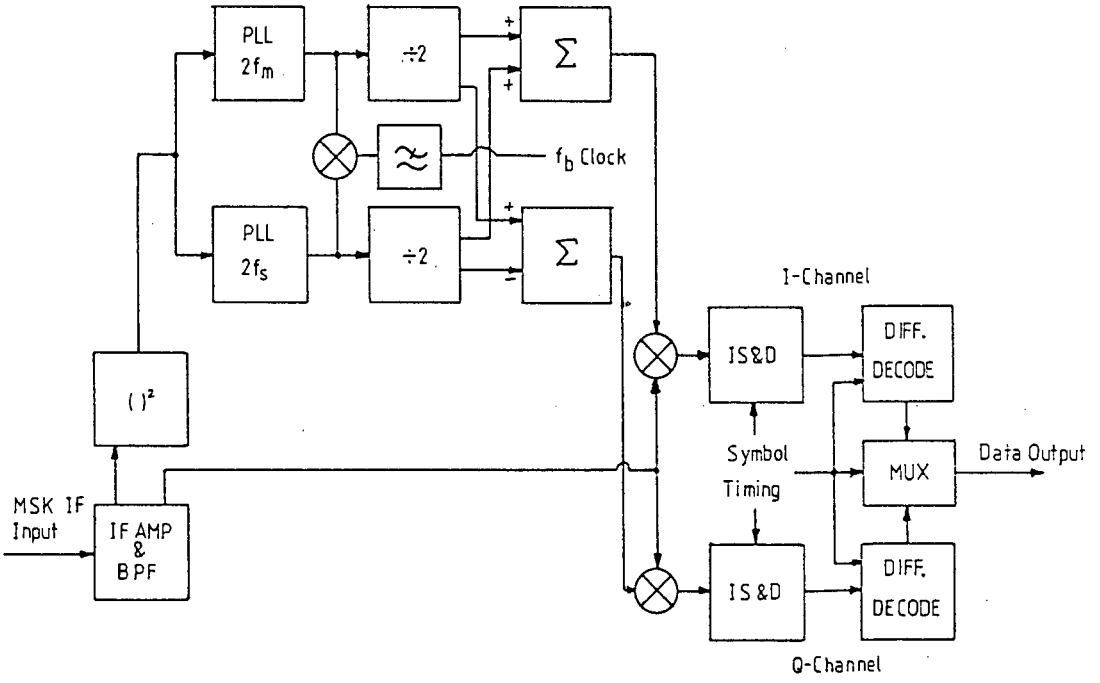


Fig.2.19 Coherent MSK Demodulator [2.13]

In order to produce the weighting functions at the receiver, it is necessary to recover the MARK and SPACE frequencies from the received spectrum. It will be shown

later that the MSK spectrum contains no discrete spectral components (making it an efficient signaling scheme), and so a non-linear operation (squaring) is necessary to regenerate these frequencies. In fact, by squaring the MSK signal, the two spectral lines $2f_m$ and $2f_s$ appear out of the continuous spectrum. APPENDIX A shows mathematically how this arises. The squared spectrum is actually Sunde's FSK (with h = 1) as the frequency doubling effectively doubles the MSK modulation index to 1.0. A plot of the squared MSK spectrum is to be found in Fig.5.3.

The two spectral lines at $2f_m$ and $2f_s$ are extracted by means of Phase-Locked Loops (PLL's), and because they are squaring loops, they perform a divide-by-two operation on the input frequencies, yielding f_m and f_s. The signals at $2f_m$ and $2f_s$ are used to recover the symbol rate clock.

We have the signals $\pm\cos 2\pi f_m t$ and $\pm\cos 2\pi f_s t$ extracted by the PLL's with which the weighted quadrature carriers $\cos(\pi t/2T_b)\cos(w_c t)$ and $\sin(\pi t/2T_b)\sin(w_c t)$ must be generated. Notice the ± sign ambiguity present on the extracted MARK and SPACE frequencies. This is inherent in any system where the frequency undergoes a divide-by-two operation. This is overcome by differentially encoding the data at the modulator and differentially decoding the recovered data at the demodulator.

Adding the MARK and SPACE waveforms yields;

$$(\pm\cos 2\pi f_m t) + (\pm\cos 2\pi f_s t)$$

and recalling that $f_m = f_c - 1/4T_b$, and $f_s = f_c + 1/4T_b$, we get;

$$\pm\cos 2\pi(f_c - 1/4T_b)t \pm \cos 2\pi(f_c + 1/4T_b)t.$$

Using the simple trigonometric relation $\cos(\alpha+\beta)+\cos(\alpha-\beta) = 2.\cos(\alpha).\cos(\beta)$ the expression reduces to;

$$s_I(t) = \pm 2.\cos(\pi t/2T_b).\cos(2\pi f_c t).$$

Similarly, subtracting the MARK and SPACE waveforms yields;

$$s_Q(t) = \pm 2.\sin(\pi t/2T_b).\sin(2\pi f_c t).$$

We recognize these waveforms as the I- and Q-Channel phasors generated at the modulator. A photograph of these phasors is shown in Fig.5.21. In order to downconvert the MSK waveform, we form the products;

$s_{MSK}(t).s_I(t)$ in the I-Channel

and

$s_{MSK}(t).s_Q(t)$ in the Q-Channel.

We are only interested in the low frequency products (they represent the degree of correlation between the received and locally generated symbols) and therefore discard the higher frequency terms of the product ($f \geq f_c$).

The I-Channel multiplication yields;

$$s_{MSK}(t).s_I(t) = \pm 2.y_I(t).\cos^2(2\pi f_c t).\cos^2(\pi t/2T_b)$$

After low pass filtering, the product becomes;

$$s_{MSK}(t).s_I(t) = \pm \tfrac{1}{2} y_I(t)[1 + \cos(\pi t/T_b)] \quad ...(2.23)$$

Similarly, the Q-Channel product becomes;

$$s_{MSK}(t).s_Q(t) = \pm\tfrac{1}{2}y_Q(t)[1 - \cos(\pi t/T_b)] \quad \ldots(2.24)$$

These waveforms are raised cosine waveforms which are in anti-phase with eachother. They are randomly polarized by the $y_I(t)$ and $y_Q(t)$ symbol polarities assigned at the modulator. In the ideal matched filter detection of MSK the waveforms would each be integrated over a $2T_b$ period and a decision made. However, it will be shown in section 5.6 that IS&D circuits are not necessary, and that the low-pass waveforms may be directly sampled at maximum eye opening and a decision made as to the symbol polarity. When this is done, a small degradation in noise performance is incurred, but it greatly simplifies the receiver circuitry.

The waveforms in equations (2.23) and (2.24) are plotted in Fig.2.20.

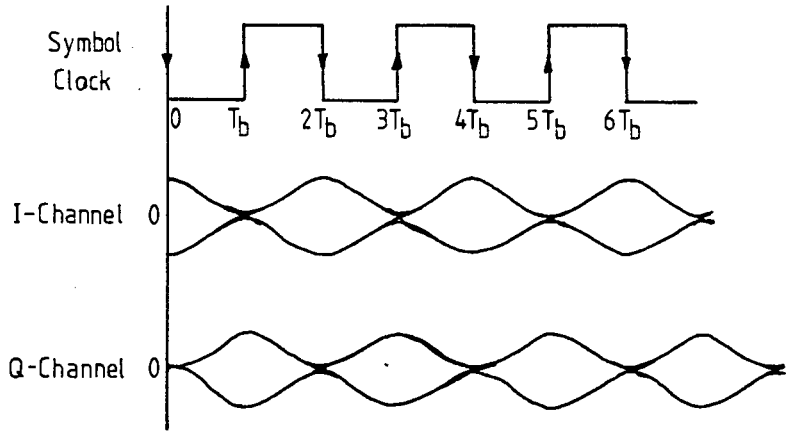


Fig.2.20 I- and Q-Channel Eye Diagrams

It is seen that the I- and Q-Channel eyes are in phase

quadrature and that each channel yields one-half the transmitted bits. The falling edge of the symbol rate clock ($f_s = 1/2T_b$) as shown can be used to sample I-Channel at maximum eye opening, while the Q-Channel could be sampled on the rising edge of the clock.

The bit rate clock is obtained from the $2f_m$ and $2f_s$ frequencies by the relation;

$$f_b = 2\Delta f = 2f_s - 2f_m \qquad \ldots(2.25)$$

We can therefore obtain the bit rate clock by mixing the $2f_m$ and $2f_s$ frequencies and low pass filtering the product. The symbol rate clock is then obtained by dividing the bit rate clock by 2. As mentioned before, the recovered symbol rate clock is used to sample the I and Q eyes, as well as to drive the channel differential decoders and parallel-to-serial converter.

2.2.4(f) Spectral Efficiency. The derivation of MSK's power spectral density is complex and part of its derivation is to be found in [2.12]. The PSD is given by;

$$G_{MSK}(f) = \frac{8P_cT_b(1 + \cos[4\pi fT_b])}{\pi^2(1 - 16T_b^2f^2)^2} \qquad \ldots(2.26)$$

and this is plotted (together with the PSD of SQPSK) in Fig.2.21.

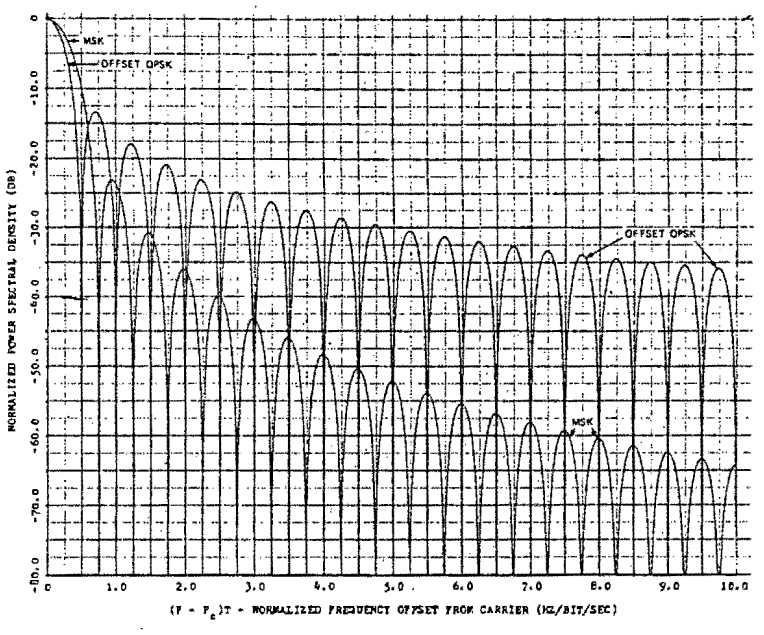


Fig.2.21 MSK and SQPSK Power Spectral Densities [2.12]

The two important differences between the PSD's of MSK and SQPSK are;

- The main lobe width of MSK is $1.5f_b$, which is 50% wider than that of SQPSK (which is $1.0f_b$ wide).
- The PSD of MSK falls off as an inverse fourth power of offset frequency from the carrier, while SQPSK falls off as an inverse square power.

Both differences can be explained in terms of the baseband

pulse shapes of the schemes. MSK uses sinusoidal symbols, while SQPSK uses rectangular symbols. Intuitively, one can see that the sinusoid has more continuous derivatives than the rectangle. This implies that it is smoother than the rectangular waveshape, hence more power will be concentrated in the lower frequency region of the spectrum (hence the rapid spectral rolloff). An interesting correspondence [2.14] provides a simple proof that any baseband pulse shape satisfying the same conditions as MSK does (i.e. symmetric etc.) will always have a wider main lobe than that of a rectangular pulse of the same duration. There clearly exists a trade-off between pulse smoothness (which sets the main lobe width) and rate of spectral rolloff. In closely packed radio channels, MSK may well be at a disadvantage relative to SQPSK which has a narrower main lobe width [2.12].

A plot of out of band power (P_{ob}) vs offset from the carrier for MSK and SQPSK is shown in Fig.2.22. P_{ob} is defined to be;

$$P_{ob} = 1 - \{ \int_{-B}^{B} G_{MSK}(f)\, df \;/\; \int_{-\infty}^{\infty} G_{MSK}(f)\, df \} \quad ...(2.26)$$

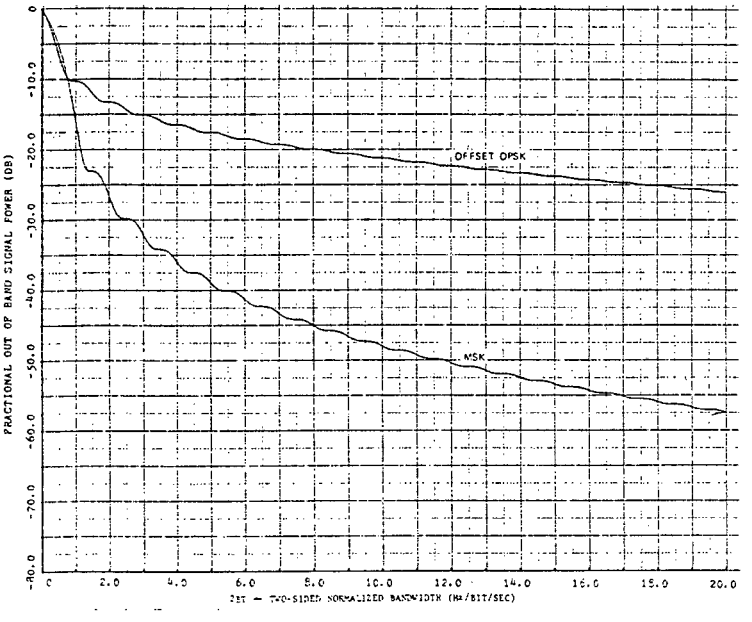
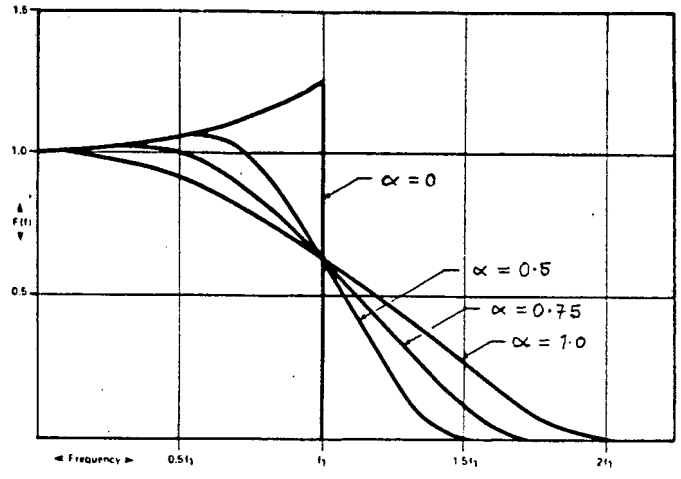


Fig. 2.22 Plot of P_{ob} for MSK and SQPSK [2.12]

MSK contains 99% of its power within a bandwidth of $1.15f_b$ [2.13], thus for a bit rate of 2.048 MBps, the 99% power bandwidth is 2.36 MHz. In order to increase the spectral efficiency of MSK, Nyquist filtering may be applied. Note, however, that the pulses now being filtered are half sinusoids, not rectangular as was the case with SQPSK, so the aperture equalizer transfer function has to be the inverse of the cosinusoid Fourier transform. The transform is derived in [2.15], and the resultant composite Nyquist filtering transfer function is given in [2.16] as;

$$H(f) = \begin{cases} \dfrac{1 - (2f/f_s)^2}{\cos(\pi f/f_s)} & 0 \le f \le (1-\alpha)f_s/2 \\ \tfrac{1}{2}[1 - \sin(\pi/2\alpha)(2f/f_s - 1)] \cdot \dfrac{1 - (2f/f_s)^2}{\cos(\pi f/f_s)} & \text{for } (1-\alpha)f_s/2 \le f \le (1+\alpha)f_s/2 \\ 0 & f > (1+\alpha)f_s/2 \end{cases} \quad \ldots(2.27)$$

The transfer functions of H(f) for various values of α are plotted in Fig.2.23.



$f_1 = f_s/2$

Fig.2.23 Nyquist Filtering for MSK [2.16]

For $\alpha = 0$, the spectrum is constrained to a bandwidth of $f_s/2 + f_s/2 = f_s$ (recall that the RF bandwidth is twice the baseband bandwidth). We can therefore conclude that the maximum spectral efficiency of MSK is;

$$\Omega(\text{MSK}) = f_b/f_s = f_b/(f_b/2) = 2 \text{ Bits/s/Hz} \quad \ldots(2.28)$$

The filtering need not be Nyquist filtering, and Chebyshev filters, for example, may be used [2.17]. One must expect a certain amount of performance degradation, but this may be minor when compared to the simplicity of the non-Nyquist filtering.

To conclude the discussion on the theory of MSK, a sketch of its constellation diagram is shown in Fig.2.24.

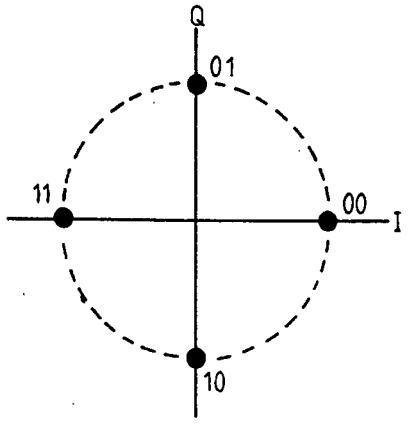


Fig.2.24 MSK Constellation Diagram

As one should expect, the constellation diagram is similar to that of SQPSK, but note that the points lie on the four quadrant axes (this is to be expected if one observes the phase trellis diagram in Fig.2.17). The points are Gray coded (each point corresponds to a symbol i.e. 2 bits), and this is in agreement with the fact that the carrier can only shift by ±90° during a bit period. As a result of its CP-FSK properties, MSK has a constant envelope, and this results in the locus of the carrier in the constellation diagram being circular. This constant envelope property is desirable in situations where non-linear amplification of the carrier is

performed. Note that as a consequence of the Gray coding, a symbol error only results in a single bit error (as opposed to 2 bit errors were the coding not Gray). The probability of a bit error is thus one-half the probability of a symbol error. This fact should be borne in mind when viewing MSK as an M-Ary PSK scheme (where one always calculates in terms of symbols, not bits).

REFERENCES

[2.1] Feher, K. and Engineers of Hewlett-Packard Ltd., "Telecommunications measurements, analysis, and instrumentation", 1st ed., Prentice-Hall, New Jersey (1987).

[2.2] Lathi, B. P., "Modern digital and analog communication systems", 1st ed., Holt-Saunders (1983).

[2.3] Oetting, J. D., "A comparison of modulation techniques for digital radio", IEEE Trans. on Comm., Vol. COM-27, Dec. 1979, pp.1752-1762.

[2.4] Peebles, P. Z. Jr., "Digital communication systems", 1st ed., Prentice-Hall, New Jersey (1987).

[2.5] Sunde, E. D., "Ideal binary pulse transmission by AM and FM", Bell Syst. Tech. J., Vol. 38, Nov. 1959, pp.1357-1426.

[2.6] Osborne, W. P., and Luntz, M. B., "Coherent and non-coherent detection of CP-FSK", IEEE Trans. on Comm., Vol. COM-22, Aug. 1974, pp.1023-1036.

[2.7] Chen, C. H., "Discriminator detection of wide-band PCM/FM", IEEE Trans. on Aerospace and Elect. Systems, Vol. AES-5, Jan. 1969, pp.126-127.

[2.8] Sullivan, W. A., "High capacity microwave system for digital data transmission", IEEE Trans. on Comm.,

Vol. COM-20, June 1972, pp.466-470.

[2.9] de Buda, R., "Coherent demodulation of frequency shift keying with low deviation ratio", IEEE Trans. on Comm., Vol. COM-20, June 1972, pp.429-435.

[2.10] Tou, F. and Simpson, R. S., "Optimum deviation ratio and observation interval for continuous-phase binary frequency-shift keying", IEEE Trans. on Comm., Vol. COM-21, Sept. 1973, pp.1067-1069.

[2.11] de Buda, R., "About optimal properties of fast frequency shift keying", IEEE Trans. on Comm., Vol. COM-22, Oct. 1974, pp.1726-1727.

[2.12] Gronemeyer, S. A. and McBride, A. L., "MSK and offset-QPSK modulation", IEEE Trans. on Comm., Vol. COM-24, Aug. 1976, pp.809-820.

[2.13] Mathwich, H. R., Balcewicz J. F. and Hecht, M., "The effect of tandem band and amplitude limiting on the E_b/N_o performance of minimum (frequency) shift keying (MSK)", IEEE Trans. on Comm., Vol. COM-22, Oct. 1974, pp.1525-1540.

[2.14] Boutin, N. and Morissette, S., "Do all MSK-type signaling waveforms have wider spectra than those for PSK?", IEEE Trans. on Comm., Vol. COM-29, July 1981, pp.1071-1072.

[2.15] Bennett, W. and Davey, J., "Data transmission", McGraw-Hill, New York (1969).

[2.16] Feher, K. and Morais, D. H., "Bandwidth efficiency

and probability of error performance of MSK and offset QPSK systems", IEEE Trans. on Comm., Vol. COM-27, Dec. 1979, pp.1794-1801.

[2.17] Austin, M. C. et al, "QPSK, Staggered QPSK, and MSK- a comparative evaluation", IEEE Trans. on Comm., Vol. COM-31, Feb. 1983, pp.171-182.

CHAPTER 3

THE PERFORMANCE OF MSK UNDER NON-LINEAR CONDITIONS

3.1 INTRODUCTION

All the properties of MSK derived in the previous chapter were for an ideal MSK modem operating under linear filtering (where applied), and with no distorting elements between the modulator and demodulator.

We know that the real world is different from these idealized conditions, and so it is necessary to study the effects that typical real world imperfections have on the performance of MSK. The imperfections to be studied are;

(a) The effects of bandlimiting, then hard-limiting on the spectrum and error performance of MSK. Such a situation arises when MSK is band constricted then amplified in a non-linear manner (for example, by a Travelling Wave Tube or ILA operating at only a few or zero dB backoff). The reason for operating the amplifiers in their non-linear region is that they perform at their highest efficiency here, and High Power Amplifier (HPA) linearizers are costly, and consequently, are sometimes not used.

(b) The effects of modem imperfections on the error performance of MSK. A major factor in determining the quality of the modulator is the degree to which it approximates a true MSK generator. A measure of this is how closely the modulation index is kept to its nominal value of

0.5. A deviation from this value will result in performance degradation and an analytical investigation into the effects of incorrect modulation index will be presented.

3.2 THE EFFECTS OF BAND- AND HARD-LIMITING ON THE MSK SPECTRUM

In practice, it is desirable to perform bandlimiting on the modulated spectrum at IF, rather than at RF frequencies. The reason is that as the IF frequency is lower than that of RF, the centre frequency-to-bandwidth ratio (i.e. the Q of the filter) is smaller for the IF filter than for the RF filter. It is therefore an asset of a modulation scheme to be as unaffected as possible after bandlimiting then non-linear amplification at RF. Fig.3.1 shows a simple baseband equivalent model of a digital link which includes band- and hard-limiting.

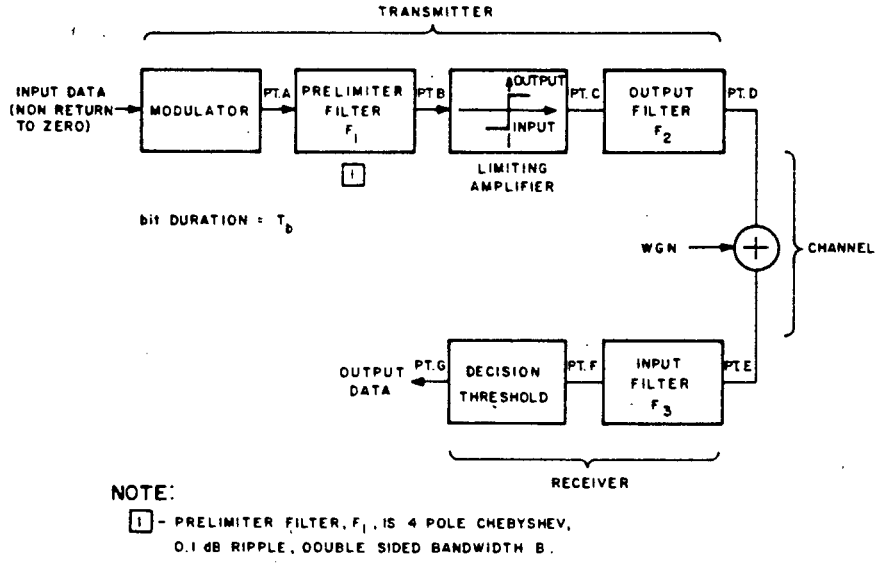


Fig.3.1 A Baseband Model of a Non-Linear MSK System [3.1]

In this model [3.1], the pre-limiter filter F_1 (which performs the main bandlimiting at IF) is a 4-pole 0.1dB ripple Chebyshev filter of 3dB bandwidth B Hz. The limiting amplifier produces a constant output envelope with no phase shift from input to output. This is an accurate model of ILA's used in terrestrial microwave links. The output RF filter (F_2) is modeled as an ideal brick-wall filter of linear phase and having 50% excess bandwidth (the sharp cut-off can only be approximated in practice). The input filter of the receiver, F_3, is such that it presents a Raised Cosine spectrum of 50% excess bandwidth at the decision threshold input (i.e. $\alpha = 0.5$). This model was implemented on a computer and the results were obtained by simulation.

Plots of the spectrum of MSK for the following cases are shown in Fig.3.2.

(a) The case with no filtering and no hard-limiting.
(b) Filtering with $BT_b = 1$ (i.e. the prelimiter filter with B equal to the bit rate).
(c) Filtering with $BT_b = 1$, then hard-limiting.
(d) Filtering with $BT_b = 0.75$
(e) Filtering with $BT_b = 0.75$, then hard-limiting.

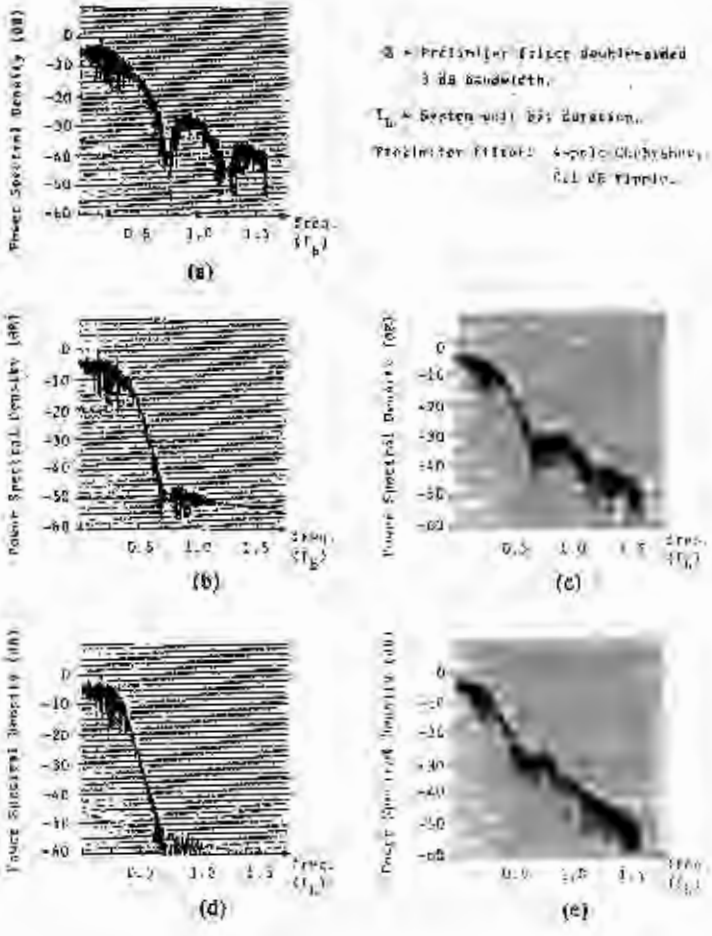


Fig.3.2 Plots of MSK Spectra [3.1]

The following features should be noted;

(a) In the unfiltered, non hard-limited case, the first sidelobe of MSK is approximately 22dB down from the peak of the main lobe.

(b) After filtering with $BT_b = 1$, the first sidelobe is attenuated to 43dB below the main lobe, but it is partially restored after hard-limiting to 27dB below the main lobe.

(c) After filtering with BT_b = 0.75, the first sidelobe is attenuated to 51dB below the main lobe, but is partially restored to 24dB below after the hard-limiting.

(d) In all the cases above, the main lobe of MSK is nearly fully restored after hard-limiting.

One can therefore conclude that hard-limiting partially restores the sidelobes after the signal has been bandlimited. The authors of [3.1] also give results for QPSK and SQPSK, and these show that the sidelobes of QPSK are nearly fully restored to their unfiltered levels after the hard-limiting (the reasons for this were discussed in section 2.2.2(b)). SQPSK is shown to perform very similarly to MSK (i.e. much better than QPSK), and this is due to both schemes' bit period offset between the I- and Q-channels. The fact that the main lobe width of MSK is 50% wider than that of SQPSK may count against it in situations where RF filtering is difficult to implement and the RF emission mask is narrow (recall that the main lobe of MSK is nearly fully restored after hard-limiting). In such situations, SQPSK may be the most suitable scheme [3.1].

An interesting insight into MSK's behaviour under the conditions described above (band- and hard-limiting) is given in [3.1]. Fig.3.3 shows the I- and Q-Channel symbols (note that they are half-sinusoids) for the data sequence 1,-1,-1,1,1,-1,1,1,-1,1.

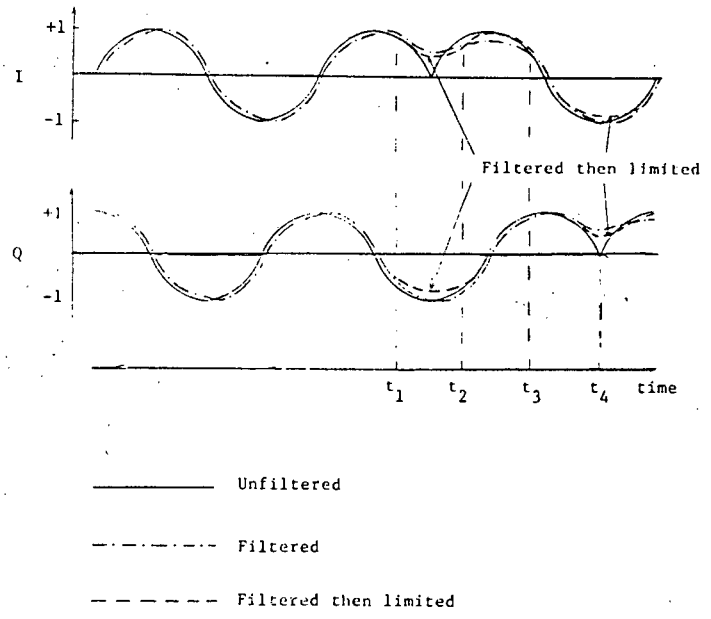


Fig.3.3 MSK I- and Q-Channel Diagram

One can see that the effect of filtering the MSK waveform is to smoothe out the hard transitions, such as those in the time interval (t_1,t_2) in the I-Channel. The reason for the smoothing is that the filtering removes the high frequency components of the waveform, thereby eliminating the sharp transitions. This, in effect, reduces the spectral occupancy of the waveforms which is the reason for the filtering function.

Recalling that the hard-limiter produces a constant output envelope, we can deduce that the vector sum of of the output I- and Q-Channel waveshapes must always be constant. This criterion is always satisfied when MSK is unfiltered (by definition of MSK, it is a constant envelope scheme), thus without pre-filtering, hard-limiting does not affect its

spectral qualities. However, it was shown above that pre-filtering modified the symbol shapes (it smoothed them), so the vector sum of of the I-and Q-Channel symbols at the limiter input is no longer constant. One must therefore expect the hard-limiter to affect the MSK spectrum.

Considering the midpoint of the time interval (t_1,t_2), we see that the filtered I-Channel should be at zero but it is not (due to the smoothing function of the filtering). At this instant, the Q-Channel symbol is at its greatest negative excursion. This yields a vector sum which is greater than that permitted at the hard-limiter output. The limiter therefore has to distort the output symbols (by reducing their amplitudes). Observing the waveforms of the hard-limiter output, one can see that some sharpness in the symbol transitions has been re-introduced. This results in the symbols having a greater high frequency content, causing the partial restoration of the sidelobes.

It should be intuitively clear that were the symbols not offset from eachother (such as in the case of QPSK), then the situation is far worse as then both symbols could pass through zero at the same instant, resulting in severe symbol distortion at the hard-limiter output.

3.3 THE EFFECT OF BAND- AND HARD-LIMITING ON MSK ERROR PERFORMANCE

The result of band- and hard-limiting of the MSK waveform is to distort the symbol waveshapes. This effect was discussed in the previous section. It is clear that since the symbols undergo distortion, then there is a mismatch between the received symbols and their stored replicas at the receiver. This, coupled with the fact that filtering removes some of

the signaling energy results in degradation of the modem error performance.

Fig.3.4 shows the results obtained in [3.1], for the system illustrated in Fig.3.1.

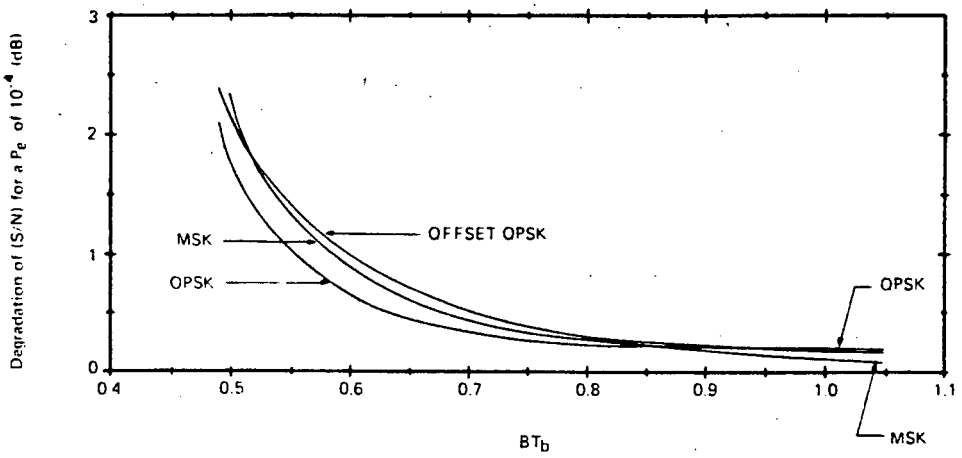


Fig.3.4 Degradation Due to Filtering on MSK and SQPSK [3.1]

For BT_b products greater than 0.8, the degradation due to the filtering and hard-limiting is minor (<0.3 dB), but it then increases rapidly, and for BT_b = 0.5 (bandwidth equal to the double sided Nyquist bandwidth) the degradation is 2dB for MSK. Note that for greater bandlimiting than this, the performance of SQPSK is better than MSK. The reason is that since the main lobe of MSK is wider than that of SQPSK, it is more sensitive to bandlimiting than SQPSK, and at some point (when $BT_b \approx 0.5$) the performance of MSK degrades rapidly because a significant amount of energy is being filtered away.

The curves of Fig.3.4 may prove to be somewhat optimistic in

practice, as the filtering used guarantees that an $\alpha = 0.5$ Raised Cosine spectrum is present at the decision threshold input. Some authors realized their channel filtering by means of 7 Pole, 0.1dB ripple Chebyshev filters [3.2,3.3]. The results presented in [3.2] are shown below in Fig.3.5.

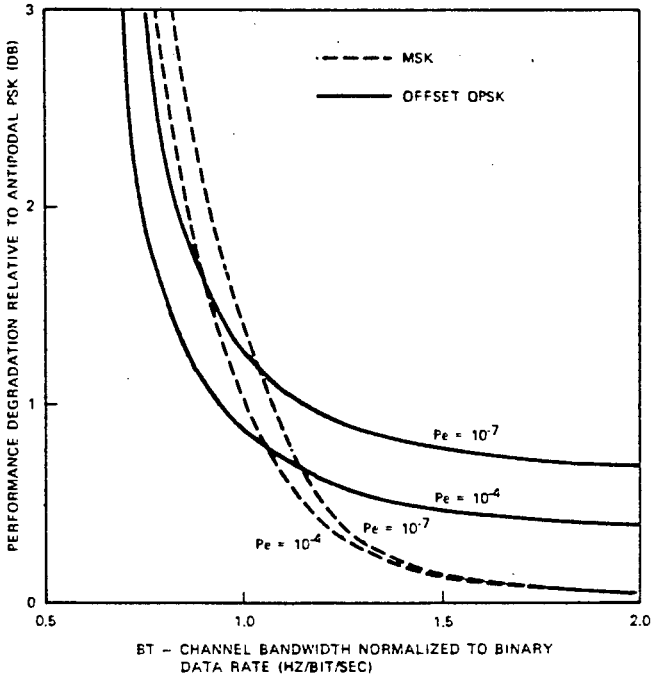


Fig.3.5 Noise Performance Degradation of MSK and SQPSK After Chebyshev Filtering and Hard-Limiting [3.2].

These results show that for $P_e = 10^{-4}$, a point is reached ($BT_b = 1.1$) where the MSK and SQPSK curves intersect. Note that the degradations shown in these curves is more severe than those in Fig.3.4, and this is to be expected as the filtering is non-Nyquist, and neither are the filters phase-equalized.

3.4 THE EFFECTS OF INCORRECT MODULATION INDEX ON THE MSK ERROR PERFORMANCE

An exact theoretical analysis of these effects is difficult, and simplifying assumptions have to be made. The analysis is presented in APPENDIX B.

It is shown in the analysis that an incorrect modulation index has the effect of an accumulation of phase errors which may eventually result in a symbol error. This error probability is fixed, and independent of signal-to-noise ratio. It therefore represents an upper-bound in performance of the modem, and this must be taken into account when specifying the modulation index tolerance.

REFERENCES

[3.1] Feher, K. and Morais, D.H., "The effects of filtering and limiting on the performance of QPSK, Offset QPSK and MSK systems", IEEE Trans. on Comm., Vol. COM-28, Dec. 1980, pp.1999-2009.

[3.2] Gronemeyer, S. A. and McBride, A. L., "MSK and offset QPSK modulation", IEEE Trans. on Comm., Vol. COM-24, Aug. 1976, pp.809-820.

[3.3] Austin, M. C. et al, "QPSK, staggered QPSK, and MSK- a comparative evaluation", IEEE Trans. on Comm., Vol. COM-31, Feb. 1983, pp.171-182.

CHAPTER 4

HARDWARE IMPLEMENTATION OF AN MSK MODULATOR

The basic block diagram for a coherent MSK modem is well known and is given in several papers [4.1,4.2].

However, the actual hardware implementation is not well documented, and the hardware development described in this chapter and the following one (Chapter 5) is aimed at bridging this gap. The modules to be described operate at an IF frequency of 70.144 MHz, and at a data rate of 2.048 Mbps. The modules were designed to be versatile and are therefore conducive to further development work if required. The design of a coherent MSK demodulator is described in Chapter 5.

The modulator consists of four modules;

(i) A Pseudo Random Bit Sequence (PRBS) Generator
(ii) Differential Encoder module
(iii) Level Translation Circuitry (LTC)
(iv) 70 MHz Voltage Controlled Oscillator (VCO)

The modulator block diagram is shown in Fig.4.1.

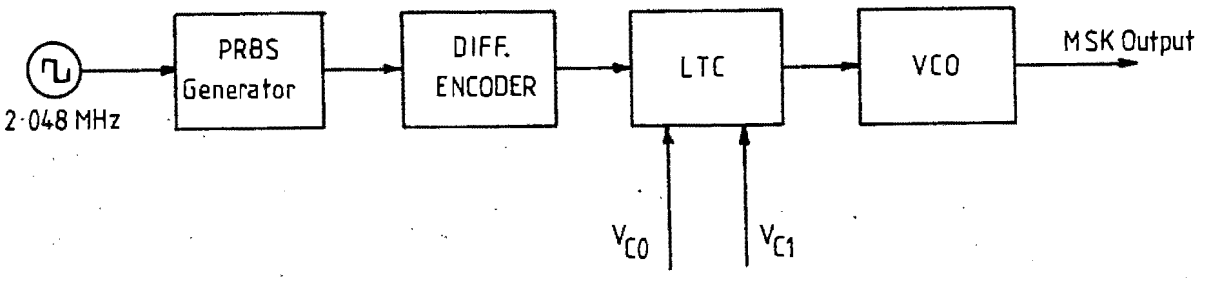


Fig.4.1 MSK Modulator Block Diagram

This particular implementation is very similar to that of the generalized CP-FSK modulator shown in Fig.2.13. It was decided to follow this implementation as it is simpler than the Parallel-type modulator as shown in Fig.2.18. Note that the PRBS generator is not strictly part of the MSK modulator, but it has been included for convenience. For transmission of actual data, the PRBS generator may be disconnected, and the data stream applied to the Differential encoder.

4.1 DESIGN AND CONSTRUCTION OF MSK MODULATOR MODULES

4.1.1 Pseudo Random Bit Sequence (PRBS) Generator

The PRBS generator is necessary in order to simulate live data traffic for the modulator. It was decided to use a 16 stage register configuration, and this will later be shown to be perfectly adequate for the chosen data rate.

The generalized configuration of a PRBS generator is shown in Fig.4.2.

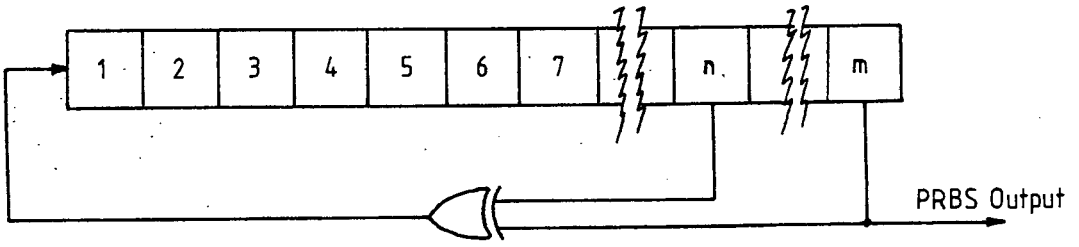


Fig.4.2 Generalized PRBS Generator

The PRBS generator consists of m registers connected in series and clocked at the data rate f_b (in this instance, m = 16 and f_b = 2.048 Mbps). The input to the generator is provided by the output of an EXOR gate which performs a modulo-2 addition of selected taps (m and n) in the register chain.

The maximum possible sequence length of an m-bit register is $K = 2^m-1$, and not 2^m as the state of all zeroes is not permitted (this state cannot be exited). A sequence of length K is termed a Maximal Length Sequence, or an m-sequence. The criterion for a sequence being of maximal length is that the polynomial $1+x^n+x^m$ be irreducible and prime over the Galois field [4.3]. After K steps in the sequence, the pattern repeats, so the sequence is not truly random. For the given value of m, it is desirable to make the sequence be of maximal length. This is guaranteed if we take the feedback taps off at n = 4,13,15 and m = 16 [4.3]. The sequence length is then;

$$K = 2^{16} - 1$$
$$= 65\ 535$$

Fig.4.3 shows the circuit diagram for the PRBS generator.

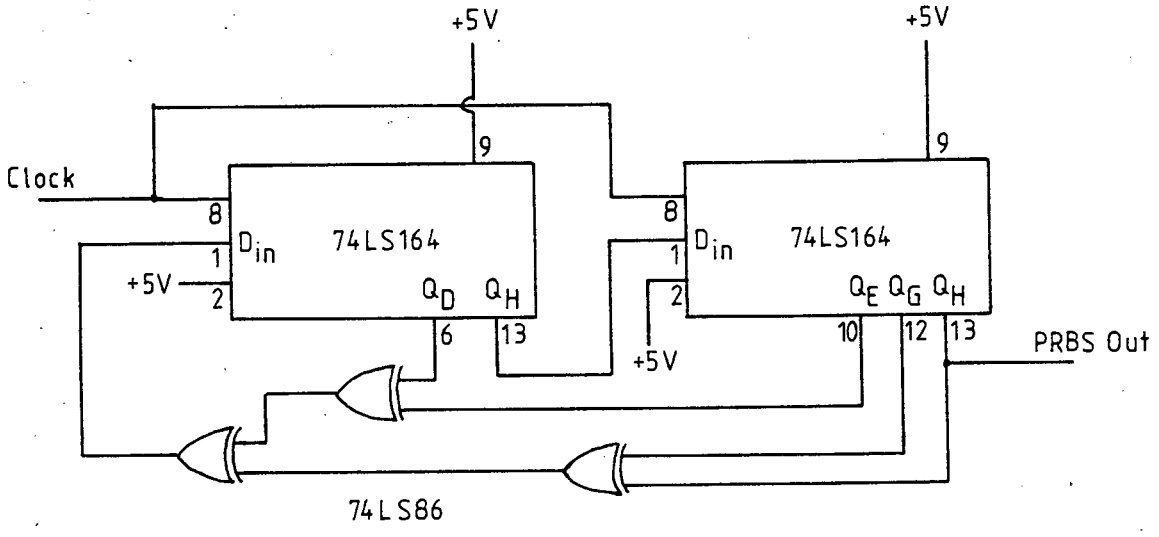


Fig.4.3 Circuit Diagram of PRBS Generator

The Power Spectral Density of a PRBS is obtained from the Fourier transform of its Autocorrelation function (ACF). The ACF (given that the output of the generator is unipolar, NRZ) is triangular in shape, with width $2T_b$. It has a repetition frequency of 1/K. The PSD is given in [4.4] as;

$$P_{PRBS}(f) = \frac{K+1}{K^2}\left[\frac{\sin(\pi f T_b)}{\pi f T_b}\right]^2 \sum_{\substack{n=-\infty \\ n<>0}}^{\infty} \delta(f-n/KT_b) + (1/K^2)\delta(f) \qquad \ldots(4.1)$$

where (for m = 16),

$$K = 65\ 535$$

and

$$T_b = (2.048\ E6)^{-1} = 488.28\ ns.$$

From this expression, one can see that the spectrum consists of a delta function at DC, and a series of delta functions spaced by $1/KT_b$ Hz symmetrically about DC, and weighted by the si^2 (f) envelope. It is clear from the expression that as $K \rightarrow \infty$ (i.e. the sequence length becomes infinite), the discrete spectral lines merge into a continuous spectrum. In our case, the lines are spaced 31.25 Hz apart, which is perfectly adequate for our needs [4.4]. The measured PSD of the sequence generated by the circuit in Fig.4.3 is shown in Fig.4.4.

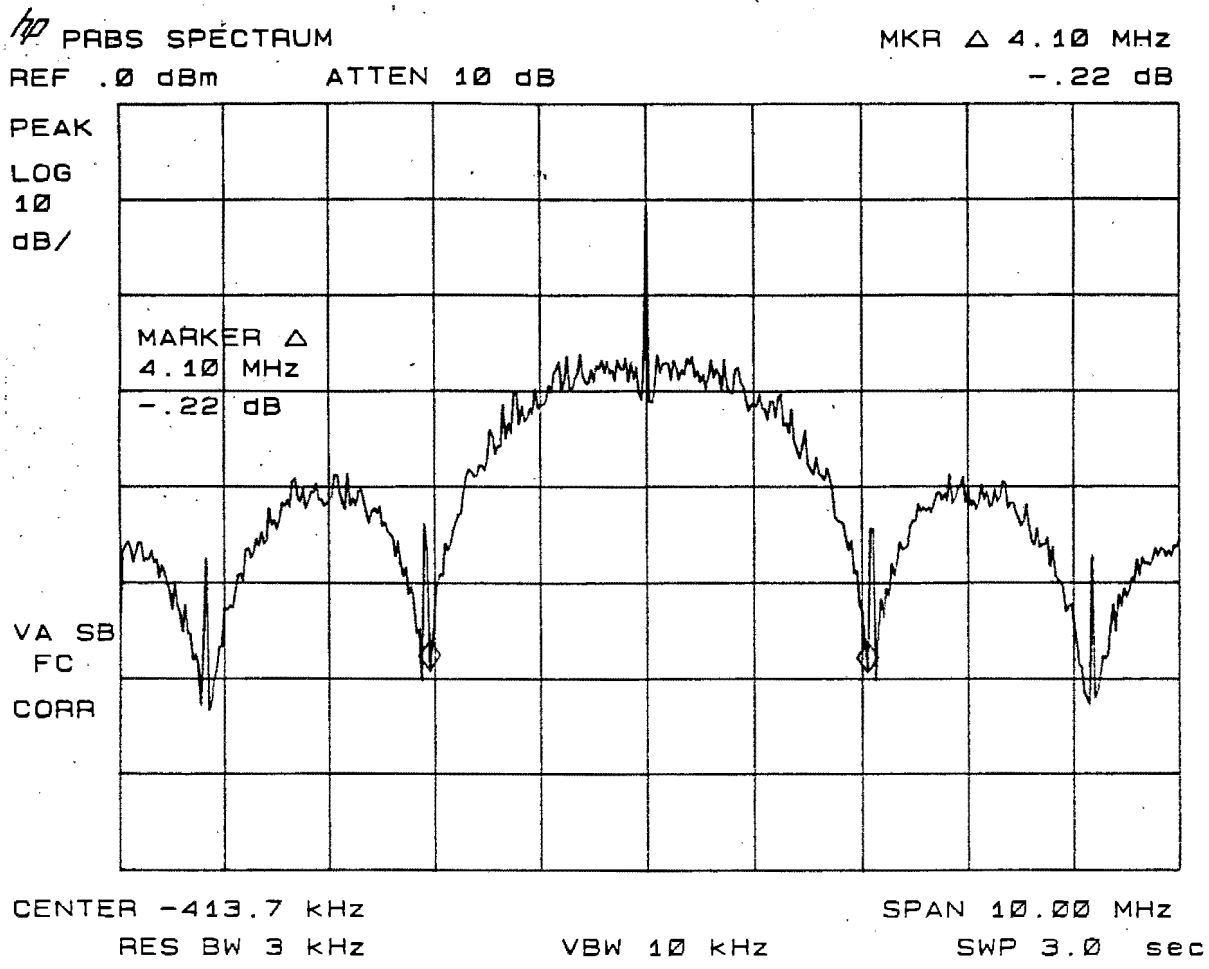


Fig.4.4 PRBS Spectrum (m = 16)

Note that as the PRBS generator output is in unipolar NRZ format, no significant spectral lines are present at the clock frequency or multiples thereof [4.5].

4.1.2 Differential Encoder Module

In order to remove the 180° phase ambiguities at the demodulator (as was discussed in section 2.2.4(e)), it is necessary to implement differential encoding at the modulator, and differential decoding at the demodulator.

The differential encoding rule is defined in [4.2] to be;

$$y_E(n) = y_E(n-1) \oplus y_L(n)$$

where $y_E(n)$ is the present (encoded) bit, $y_E(n-1)$ is the previous encoded bit and $y_L(n)$ is the present data bit to be encoded. The symbol $\oplus$ represents modulo-2 addition (equivalently the EXOR function). Fig.4.5 shows the TTL realization of the differential encoder.

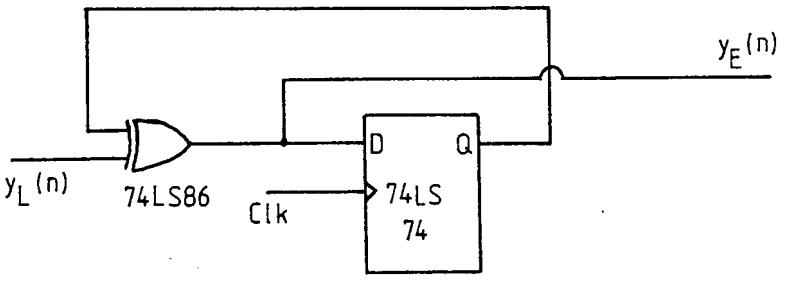


Fig.4.5 Differential Encoder Module

Note that the coder converts the NRZ-L data bits into NRZ-M format, and this does not affect the spectral properties of the generated MSK waveform.

4.1.3 Level Translation Circuitry (LTC)

As the modulator VCO will require control voltage levels in the continuum (0V,15V), and the Differential Encoder output is at TTL levels, the Level Translation Circuitry (LTC) is required to interface the two modules. Design specifications of the LTC include;

- Good high speed switching performance (at 2MHz)
- Fast settling time (minimal oscillation)
- Low output impedance (to achieve a high slew rate)
- Output voltage variable between 1V and 14V
- High input impedance voltage control inputs
- TTL compatible input.

Fig.4.6 shows the level translation function performed by the LTC.

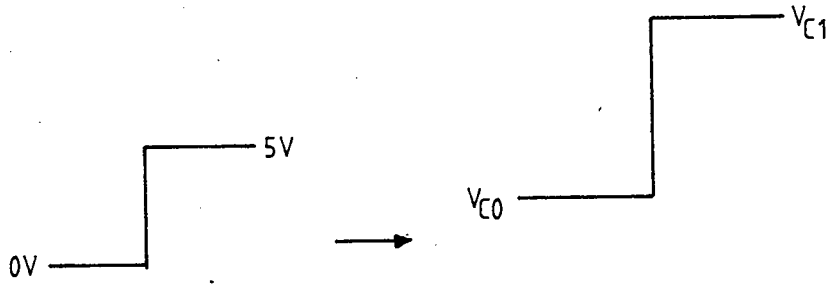


Fig.4.6 Level Translation Function

The input is in TTL format, while the output is at level V_{co} for 0V input, and at level V_{c1} for 5V input. One can therefore adjust the MARK and SPACE frequencies (and hence the modulation index h of the modulator) by varying the high impedance control voltage inputs V_{co} and V_{c1}. The circuitry was designed with the specifications outlined above in mind, and the circuit diagram of the LTC is shown in Fig.4.7.

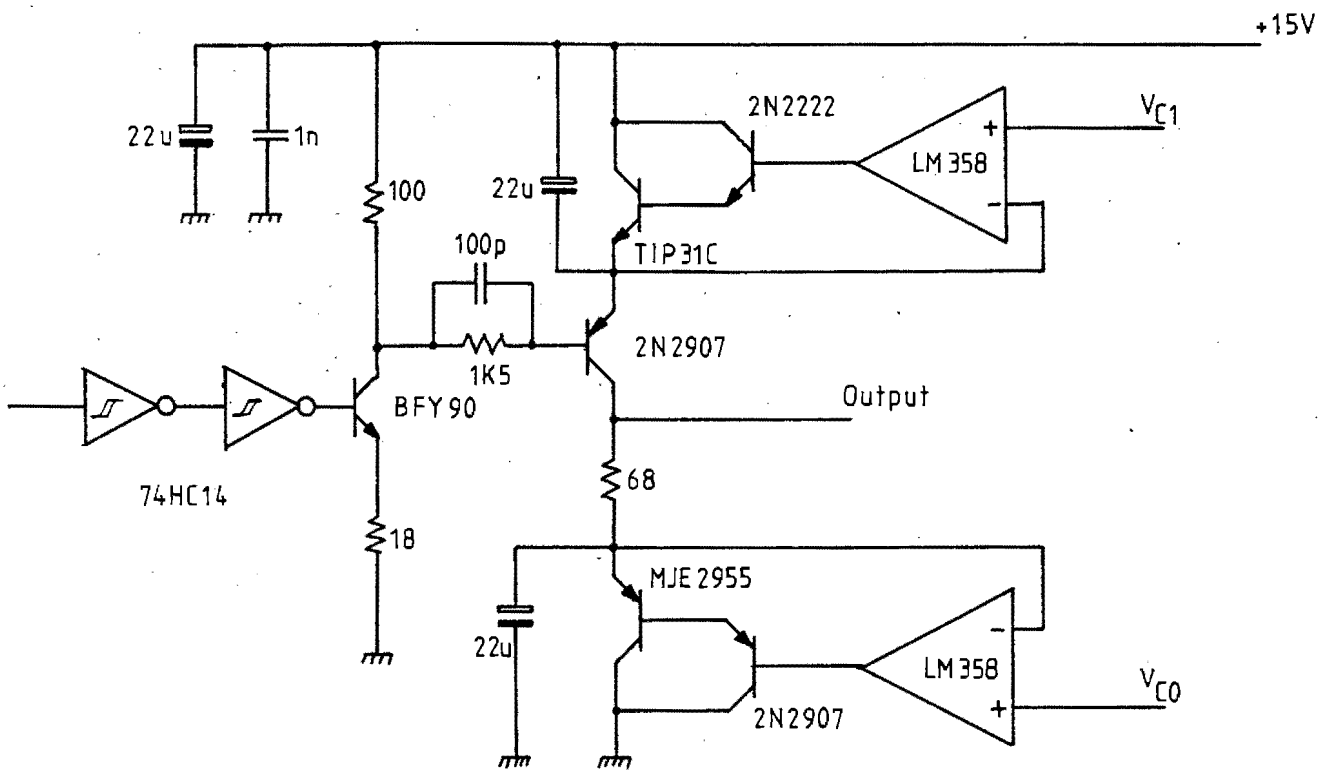


Fig.4.7 Level Translation Circuitry

The input TTL waveform is squared and buffered by the 74HC14 Schmitt trigger buffer. A BFY 90 high frequency transistor inverts the waveform, and because it operates in Class-C mode, the collector swings between ≈1V and ≈15V. This stage is necessary in order to ensure that the second switching transistor will always be operating in Class-C mode, irrespective of its collector and emitter voltage levels (which are both variable between 1V and 14V).

The collector of the BFY 90 is coupled via a parallel resistor-capacitor network to the base of the second switching transistor. The capacitance acts as a speed-up

capacitor and aids in the faster switching of the second transistor [4.3]. Note, in general, the low resistance values used in the the circuitry: this is to reduce the RC time constants in the circuit and hence increase the switching speed.

The control voltages V_{c0} and V_{c1} are first buffered in the two Darlington Emitter-Feedback circuits before being applied to the collector and emitter (respectively) of the second switching transistor. The voltages set by the control inputs are guaranteed to be the same applied to the switching transistor, independent of the V_{BE} drops of the Darlingtons or their temperature dependence ($\approx$ -4mV/°C). The high impedance control inputs V_{c0} and V_{c1} are obtained from the wipers of two 10-turn potentiometers connected between ground and +15V. These inputs may be used in a feedback system which maintains the modulation index at 0.5. This will be discussed in detail in Chapter 6.

When the input to the LTC is low (0V), the first transistor is OFF, and so is the second. The output is then provided by the 68 Ω collector resistor of the second transistor, which is supplied by a voltage level V_{c0} (note that the resistor is low in value in order to decrease the output impedance of the module). When the input to the LTC is high (5V), the first transistor is ON, and so is the second. The second transistor is saturated, hence the output of the circuit is $\approx V_{c1} - 0.2V$. Note that due to the asymmetry of the the output stage, one expects the rise-time to be shorter than the fall-time (the output is active pull-up, and passive pull-down).

The shunt 22 μF capacitors at the outputs of the V_{c0} and V_{c1} buffer circuits provide decoupling for extra short-term

stability, while the +15V Vcc supply is also well decoupled.

The rise-time (with the VCO connected), was measured to be 20 ns, while the fall-time was 36 ns. (rise time is defined to be the time elapsed between the waveform's transition from 10% to 90% of its final value). These values are ≈ 10% of the clock period which is considered to be adequate. The rising edge slew rate was found to be +333V/μs, and the falling edge slew rate was -219V/μs.

4.1.4 Voltage Controlled Oscillator (VCO)

The Voltage Controlled Oscillator is required to meet the following specifications;

- Good phase noise performance (<-70 dBc/Hz at 10 kHz offset).
- Fast response to switching of control input i.e. being capable of square wave modulation at 2 MHz.
- Tunable, with a design centre frequency of 70 MHz
- 50 Ω output impedance.
- < 5% AM on output envelope.
- Stable output frequency.
- Linear transfer function.

Fig.4.8 shows the circuit diagram of the VCO used in the modulator.

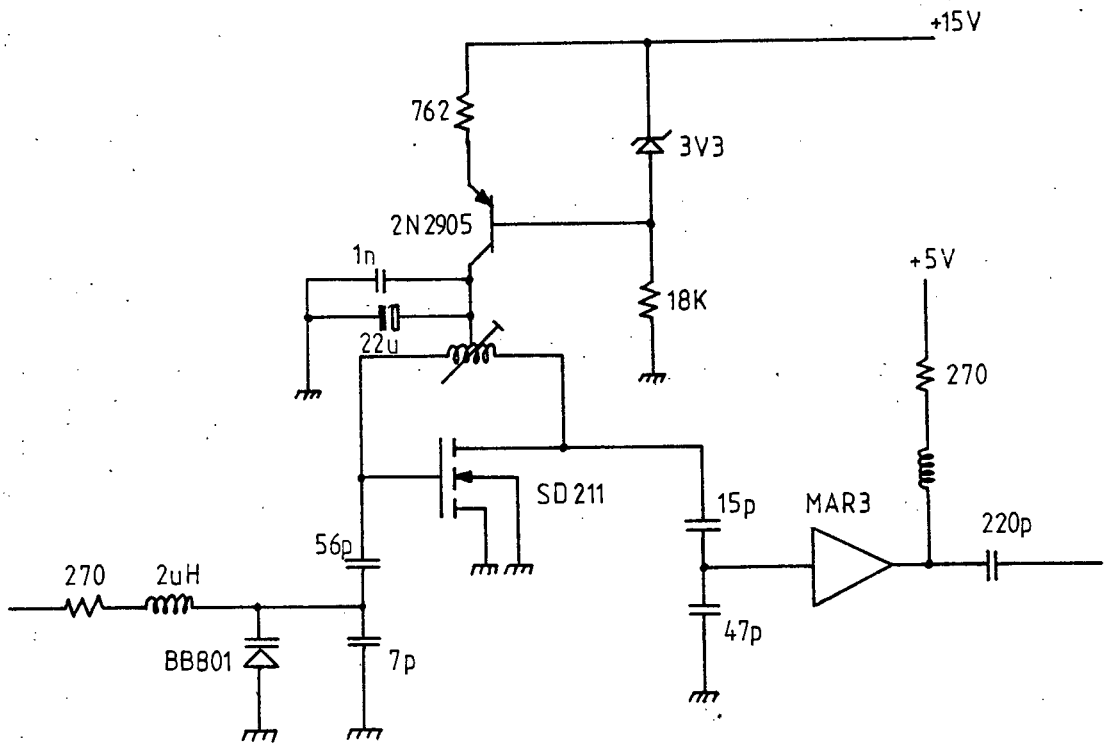


Fig.4.8 VCO Module

As the output impedance of the Level Translation Circuitry is low, its output cannot be directly applied to the varactor as this will result in severe degradation of the Q of the oscillator's tank circuit. The degradation of the Q-factor manifests itself in the extinguishing of the oscillator output, especially at the edges of the modulating square wave (this results in 100% AM being present on the output envelope). This problem is circumvented by using an inductance (2 μH in this circuit) to isolate the 70 MHz oscillations across the varactor from the low impedance of the LTC output. The inductance presents an impedance of ≈ 880 Ω to the 70 MHz oscillations, and this provides sufficient RF isolation between the modules. The combination of the inductance and the varactor capacitance is, in effect, an (undesirable) LC resonant circuit, and were it

not damped, would introduce ringing on the control voltage waveform (thus generating spurious FM modulation). A series damping resistance of 270 Ω is included, and this provides the correct amount of damping (without degrading the slew rate).

Note that a back-biased (free-wheel) diode is not connected across the inductance (it would be there in order to absorb any negative-going voltage spikes generated by the edges of the modulating square wave). It was found that its inclusion resulted in the extinguishing of the VCO oscillations on the falling edges of the VCO waveform. The reason for this is simple: consider the situation where V_{co} = 2V and V_{c1} = 6V. With V_{c1} being presented to the varactor, its steady state voltage is 6V. When the input falls to V_{co} (2V), the diode becomes forward biased and the impedance across the varactor becomes greatly reduced (the inductance is effectively shorted). The resultant degradation in Q introduces large amounts of AM into the VCO output. In practice, it was found that the damping resistance removed nearly all of the negative spikes, thus making the free-wheel diode superfluous.

It is considered good practice to dampen out the maximum possible variation in capacitance achievable by the varactor (this is provided that the full variation is not required). Doing this reduces the VCO's susceptibility to induced noise. The BB 801 varactor used has a C_{max}/C_{min} ratio of 9, and this is reduced to 2 by the 7 pF shunt capacitance.

The centre-tap of the tank circuit inductance provides for the Vcc supply point. This is normally applied from a resistor connected to the positive supply rail (the resistance acts as a crude current source). This was found

to be inadequate as the sharp transitions on the modulating waveform cause current surges in the circuit and affect the MOSFET's operating point (causing unwanted AM in the VCO output). The MOSFET quiescent current was optimized by modulating the VCO with a 2 MHz square wave and observing its output on a spectrum analyzer. Were no AM present, the resultant Bessel Power Spectrum would be symmetrical about the carrier frequency. With AM present, however, destructive interference occurs on the odd lower sideband spectral components as the AM spectrum is even, and the Bessel spectrum has the property;

$$J_{n\ odd}(\beta) = - J_{-n\ odd}(\beta)$$

where $J_n(\beta)$ is the n^{th} order Bessel coefficient for a particular modulation index β. We can therefore gauge the quality of the VCO output by measuring the balance between the first order sidebands. The supply current of 3.5 mA was found to give a sideband imbalance of < 0.5 dB, and the AM was measured on an oscilloscope to be less than 3.6%. The current is supplied to the VCO by means of a simple one-transistor current source.

The output impedance of the VCO is set to be approximately 50 Ω by the Thévenin equivalent reactance of the output capacitors. This was done in order to maximize the power transfer from the VCO to the amplifier stage.

The VCO output frequency vs control voltage input curve is plotted in Fig.4.9.

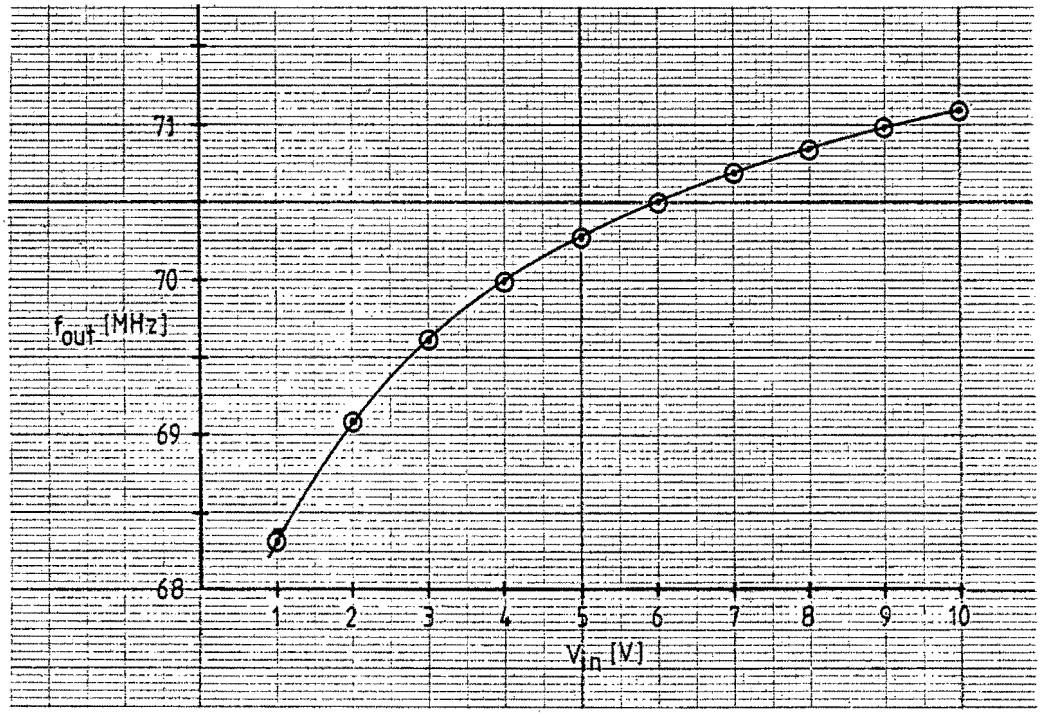


Fig.4.9 Modulator VCO Transfer Function

Note the non-linear characteristic exhibited by the VCO. This is not considered to be serious, provided that the slew rate of the modulating waveform is high enough (if this is done, the non-linear characteristic is of no real consequence as it is traversed for only a very small portion of a bit period). There are methods of linearizing the curve if it is so desired, but this is beyond the scope of this discussion (all that will be said is that mathematics will show that the curve closely approximates a logarithmic function, and this can be compensated for by pre-distorting the control voltage input).

One of the design specifications of the VCO was that it had to have good phase noise performance (i.e. good short-term stability). The measured output spectrum of the VCO for a

constant (zero) input voltage is shown in Fig.4.10.

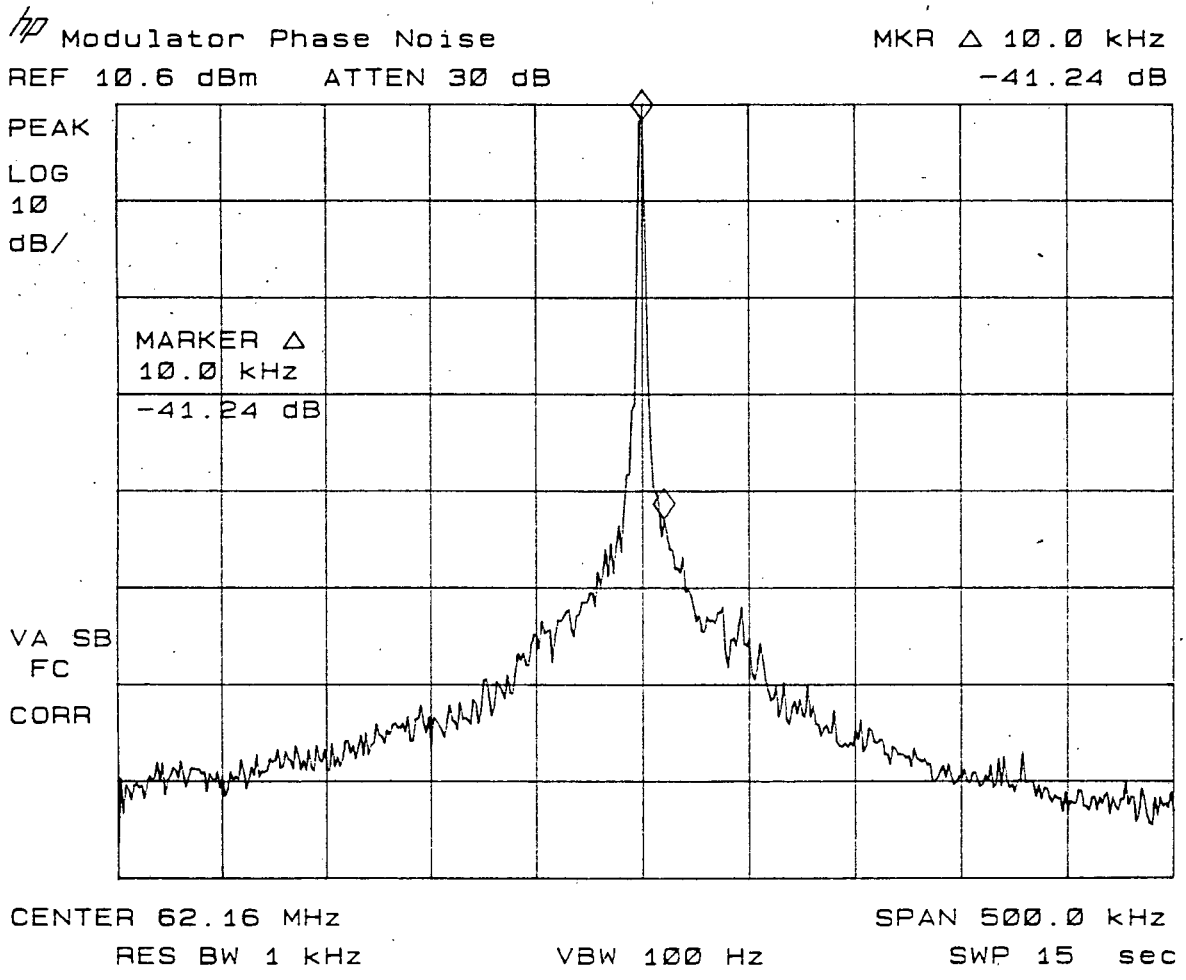


Fig.4.10 Modulator VCO Phase Noise Measurement

The method of making phase noise measurements will be described in section 5.2.3. The phase noise is found to be at a level of -71.24 dBc/Hz at a 10 KHz offset from the carrier. This meets our design specification.

A photograph of the constructed MSK Modulator is shown in Fig.4.11.

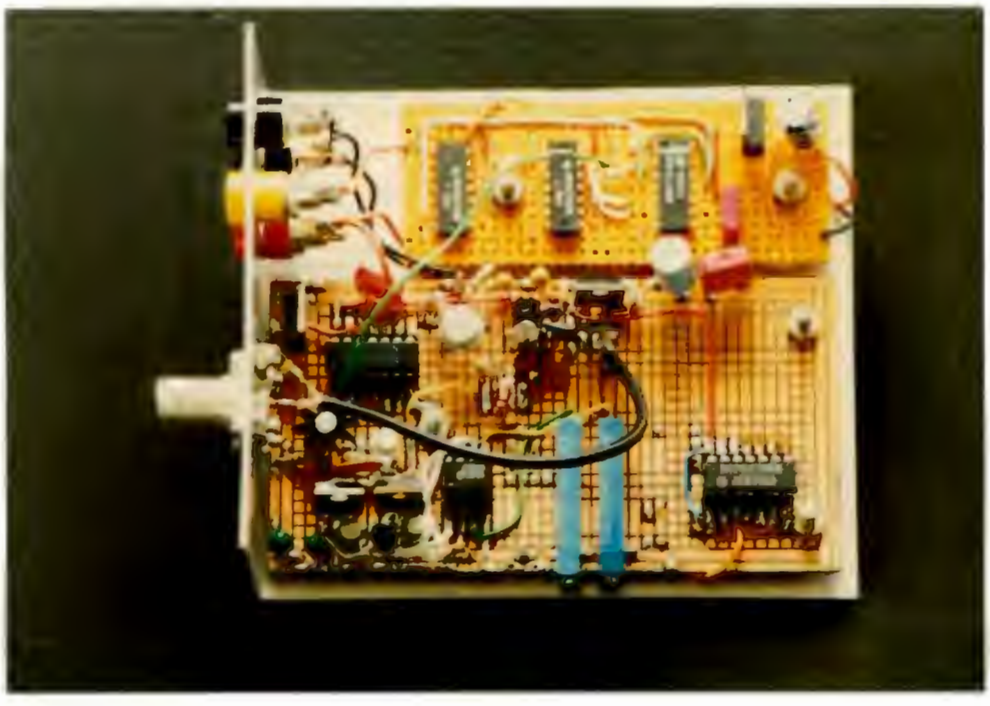


Fig.4.11 Completed MSK Modulator

4.2 MSK MODULATOR PERFORMANCE

The MSK spectrum generated by the modulator is shown in Fig.4.12.

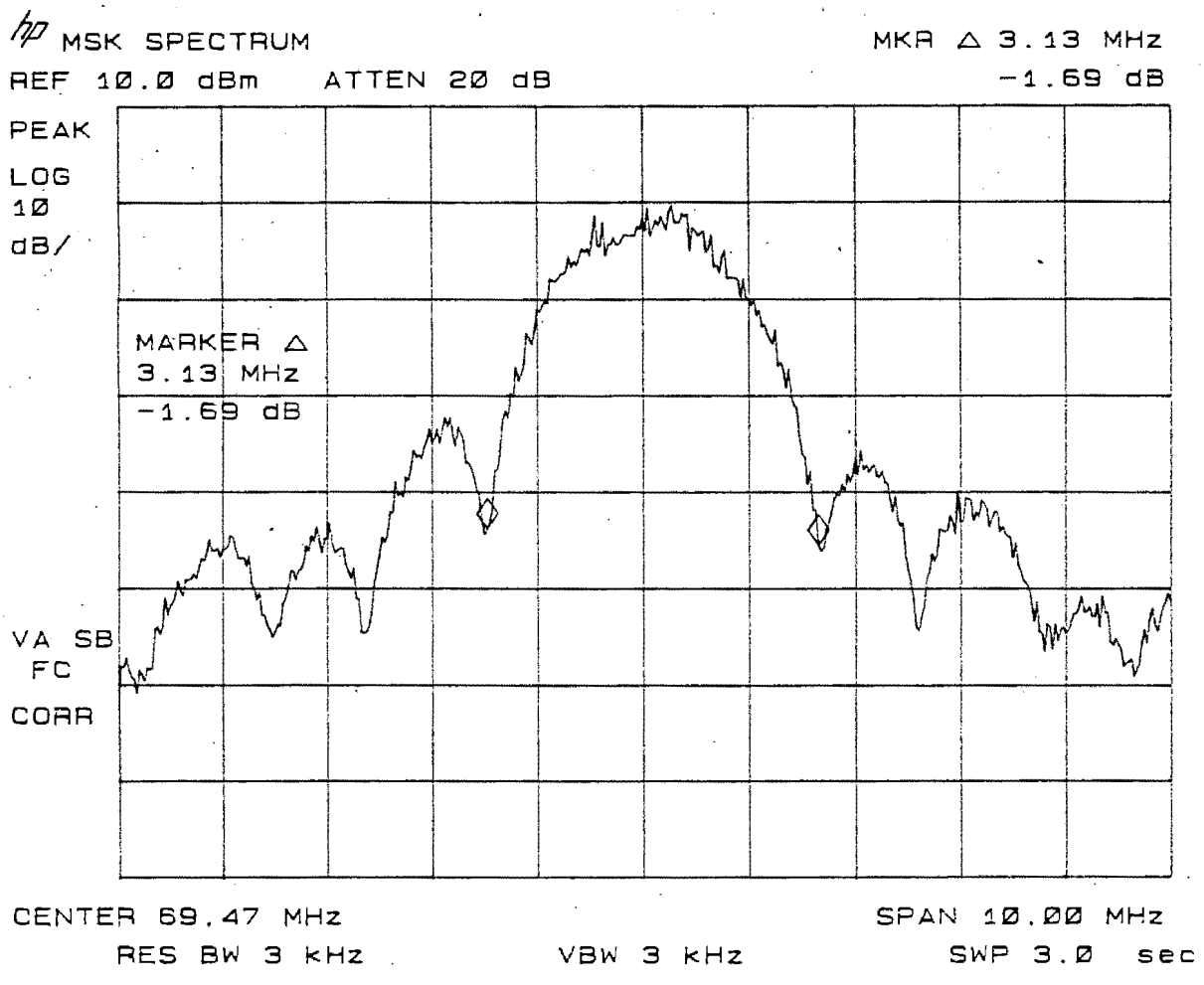


Fig.4.12 Generated MSK Spectrum

Features which should be noted are;

- The main lobe width is measured to be 3.13 MHz. This agrees well with the predicted value of 3.072 MHz, as predicted in section 2.2.4(f).

- The first sidelobes are ≈ 23 dB down from the main

lobe level, in good agreement with Feher's results shown in Fig.3.2.

- The 99% power bandwidth (as measured on the HP 8590A spectrum analyzer) is 2.5 MHz, as compared to the value of 2.4 MHz predicted in section 2.2.4(f). For interest, the 3 dB bandwidth was found to be 1.1 MHz, and the 6 dB bandwidth was 1.7 MHz.

As was mentioned at the outset of this chapter, the circuitry designed is versatile, and the modulator can be used to generate CP-FSK waveforms over a wide range of modulation indices (the modulation index is easily adjusted by varying the V_{c0} and V_{c1} control inputs). These inputs also allow for interfacing with a computer via a DAC, if it is so desired.

REFERENCES

[4.1] de Buda, R., "Coherent demodulation of frequency-shift keying with low deviation ratio", IEEE Trans. on Comm., Vol. COM-20, June 1972, pp.429-435.

[4.2] Mathwich, H. R., Balcewicz, J. F. and Hecht, M., "The effect of tandem band limiting on the E_B/N_o performance of MSK", IEEE Trans. on Comm., Vol. COM-22, Oct. 1974, pp.1525-1540.

[4.3] Horowitz, P. and Hill, W., "The art of electronics", 8th ed., Cambridge University Press, New York (1984).

[4.4] Feher, K. and the Engineers of Hewlett-Packard, "Telecommunications measurements, analysis and instrumentation", 1st ed., Prentice-Hall, New Jersey (1987).

[4.5] Lathi, B. P., "Modern digital and analog communication systems", Holt-Saunders, New York (1983).

CHAPTER 5

HARDWARE IMPLEMENTATION OF A COHERENT MSK DEMODULATOR

The block diagram of a coherent MSK demodulator is shown in Fig.5.1.

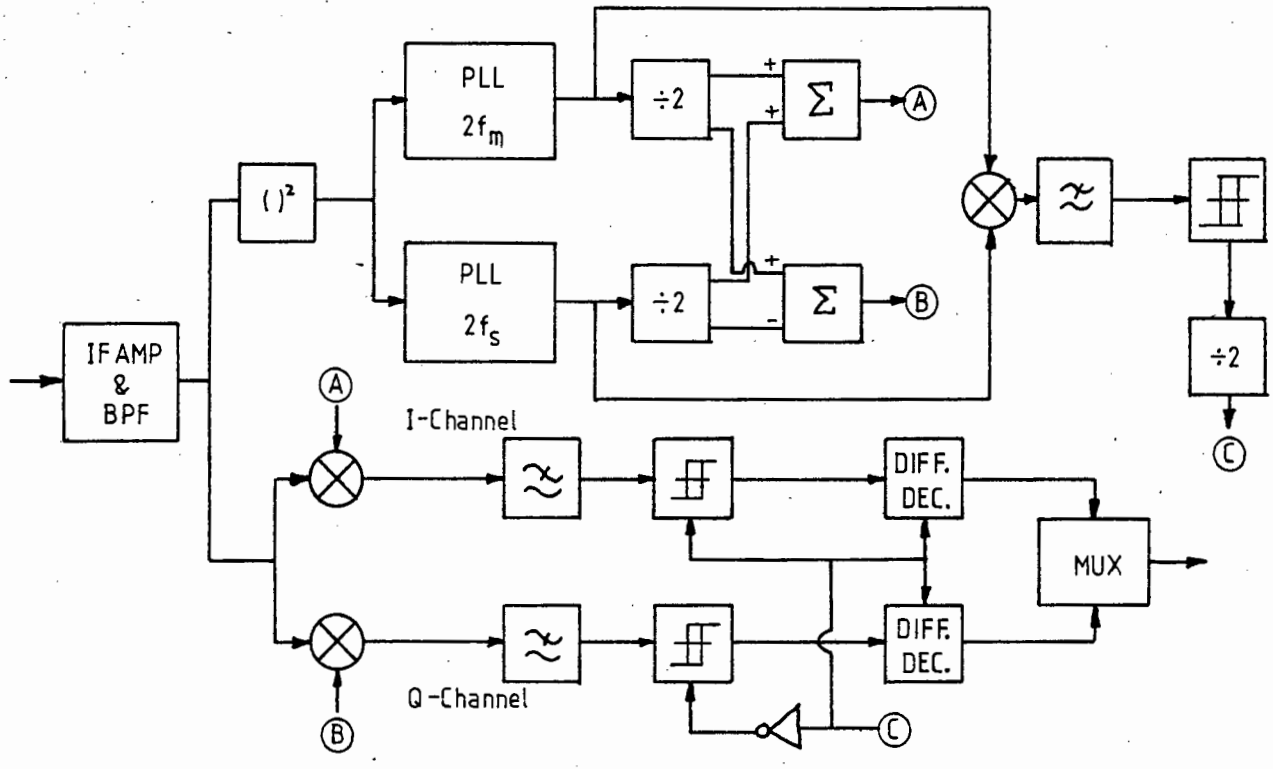


Fig.5.1 Coherent MSK Demodulator Block Diagram [5.1,5.2]

Not all these modules will be constructed as some are trivial in their implementation. The modules to be considered are;

(1) The Frequency Doubler Module;

(2) The Phase-Locked Loop Modules;

(3) Symbol Clock Recovery Module;

(4) I- and Q-Channel Phasor Regeneration Circuitry (Adder/Subtractor Module);

(5) Coherent Down Conversion Module;

(6) Detection Filter Module.

It has been assumed that the received microwave signal has undergone downconversion to IF, followed by IF amplification and bandpass filtering prior to being applied to the circuitry described here. Note that Fig.5.1 is similar to Fig.2.19, except for the difference in circuitry after the down conversion mixers. In Fig.2.19, there are IS&D circuits (which perform matched filter detection on the received symbols), whereas in Fig.5.1, there are lowpass detection filters followed by sampling circuits. This more simple implementation will be more fully discussed in section 5.6.

5.1 FREQUENCY DOUBLER MODULE

As was discussed in section 2.2.4(e), and shown analytically in Appendix A, frequency doubling of the MSK waveform results in two spectral lines being produced (one at twice the MARK frequency $2f_m$, and the other at twice the SPACE frequency $2f_s$). These frequencies are used to recover both the symbol rate clock, and the I- and Q-Channel phasors for coherent down conversion.

The nominal drive level for this module is 0 dBm. The circuit diagram is shown in Fig.5.2.

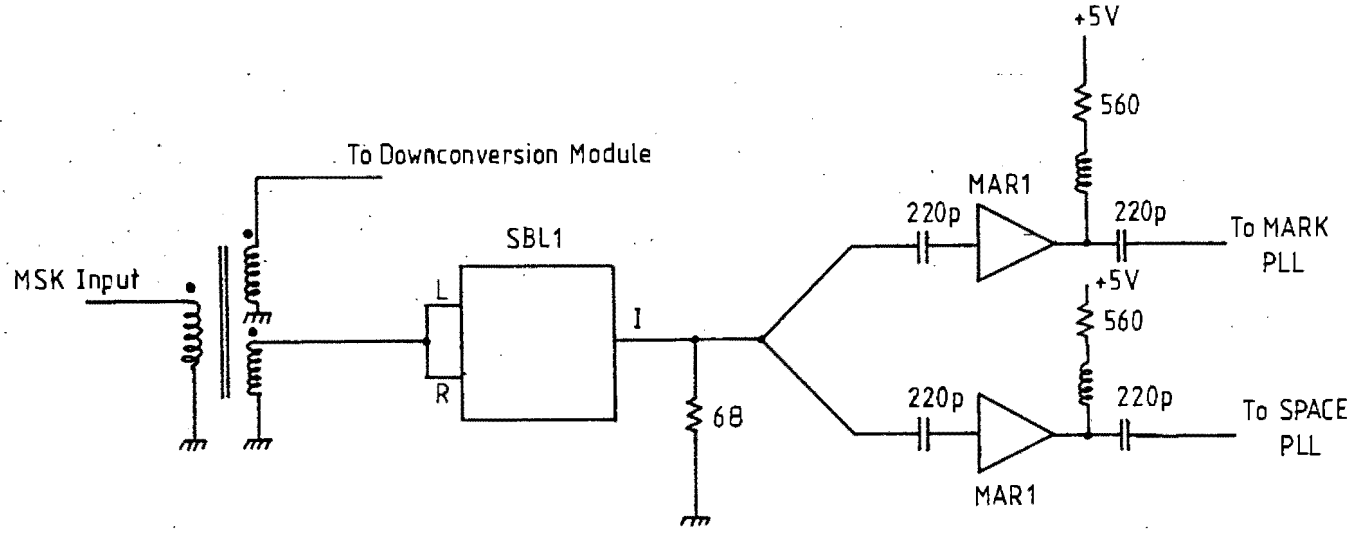


Fig.5.2 Frequency Doubler Module

The incoming MSK waveform is powersplitted into two paths by means of a trifilar-wound transformer; One path is connected to the Coherent Downconversion Module, while the other is used for the frequency doubler. An SBL-1 double balanced mixer (DBM) is used to perform the frequency doubling. This is accomplished by applying the same waveform (the MSK waveform) to the L- and R-ports of the DBM. This effectively squares the waveform, resulting in the strong generation of the first harmonic. The output is taken at the I-port of the mixer, and it is the frequency doubled version of the input. As the DBM exhibits a conversion loss of ≈ 7dB, it is necessary to amplify the frequency doubled signal. For this function, mini-circuits MAR-1 amplifiers are used. There are two amplifiers in the circuit: one feeds the $2f_m$ PLL, and the other feeds the $2f_s$ PLL. Fig.5.3 shows the spectrum at

the output of the module.

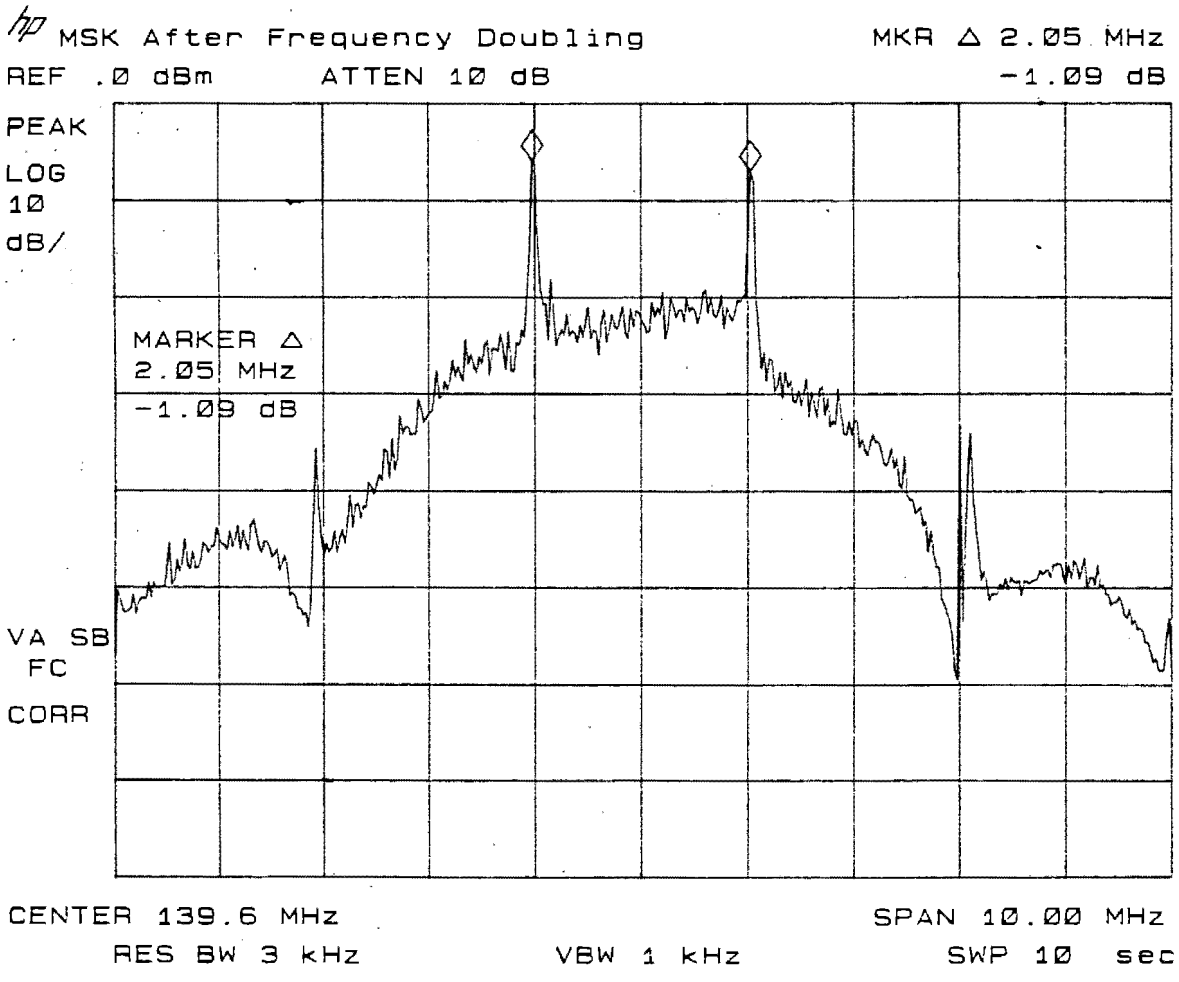


Fig.5.3 Spectrum of Frequency Doubled MSK

The two spectral lines are clearly seen, and they are at the frequencies;

$$2f_m = 2*69.632 \text{ MHz} = 139.264 \text{ MHz}$$

$$2f_s = 2*70.656 \text{ MHz} = 141.312 \text{ MHz}.$$

Note that the lines are spaced 2.05 MHz apart (as read off from the markers on the plot), and this agrees well with the predicted value of 2.048 MHz (equation 2.25). One can also see that the effect of the frequency doubling is to widen the main lobe to 6 MHz, which shows that MSK has a 50% narrower main lobe than Sunde's FSK (h = 1).

5.2 PHASE-LOCKED LOOP MODULES

In order to regenerate the I- and Q-Channel phasors and recover the symbol rate clock, it is necessary to extract the two spectral lines $2f_m$ and $2f_s$ from the spectrum shown in Fig.5.3.

This may be accomplished by using narrow bandpass filters centered at the $2f_m$ and $2f_s$ frequencies, and then amplifying the signals. This is not considered to be a satisfactory approach for the following reasons;

- Were the Bandpass filters to be of 5 kHz bandwidth, the required Q of the filter would be approximately 28000, which clearly excludes the use of LC filters.

- For the chosen bandwidth above, the modulator frequency tolerance would have to be ± 0.0018% which may not be easily achievable.

A more elegant solution is to use Phase-Locked Loops (PLL's) to extract the two frequencies. The advantage of using PLL's is that very narrow noise bandwidths may be realized without sacrificing the receiver tolerance to frequency uncertainties in the received signal (PLL's can be regarded as tracking filters). It is desirable to extract the signals

while allowing as little noise as possible through the bandpass of the PLL's. This ensures that the recovered carriers have minimal phase noise present, allowing for efficient detection. As the theory of PLL's is extensive, only the topics relevant to their use in the coherent MSK demodulation structure will be discussed. The actual hardware implementation will be described, and finally, their performance will be evaluated.

5.2.1 Basic Phase-Locked Loop Theory

5.2.1(a) Derivation of PLL Transfer Function. The block diagram of the PLL used in in the demodulator is shown in Fig.5.4.

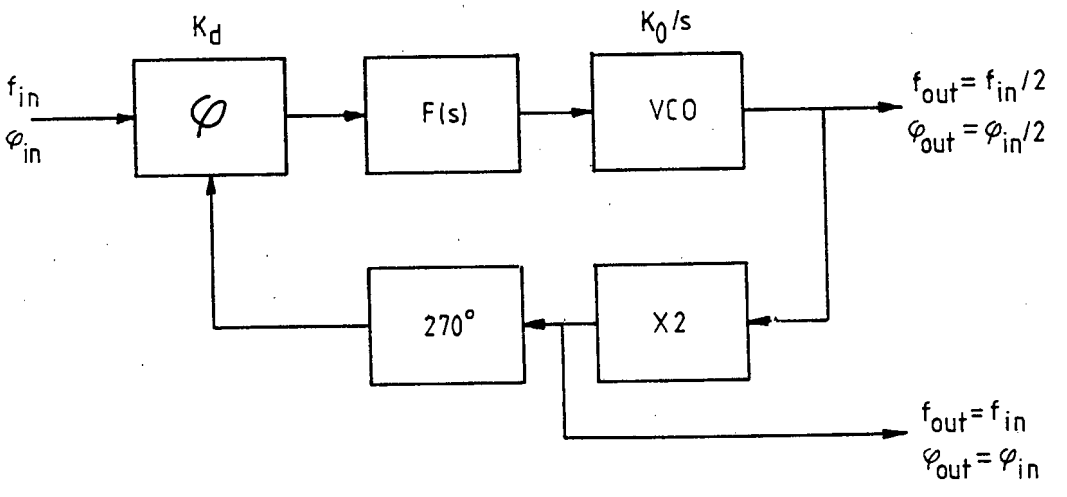


Fig.5.4 Squaring Loop PLL Block Diagram

The various gains in the loop are defined as follows;

K_d = Phase Detector (PD) gain constant [V/rad]
F(s) = Loop Filter Transfer Function [V/V]
K_o/s = VCO Gain constant [rad/s/V]

The X2 block squares (hence frequency doubles) the VCO output frequency, hence explaining the terminology "Square Law Tracking Loop" often found in the literature when describing the PLL shown in Fig.5.4. The 270° phase shifter has an amplitude gain of 1, and we ignore it in the derivation of the loop transfer function (its presence will be explained in section 5.2.2(c)). The open loop transfer function is given by;

$$OLTF = (2K_oK_dF(s))/s. \qquad ...(5.1)$$

The closed loop transfer function is given by;

$$CLTF = OLTF/(1 + OLTF)$$

therefore the closed loop transfer function H(s) for the loop is;

$$H(s) = \frac{2K_oK_dF(s)}{s + 2K_oK_dF(s)} \qquad ...(5.2)$$

The loop filter sets both the Type and Order of the loop. By Type, it is meant the number of perfect integrators in the loop. All PLL's are at least of Type I, as the VCO is a perfect integrator (the 1/s term in its gain constant implies that it is an integrator). Were F(s) to be implemented by a passive RC filter, the loop would still be of Type I, as the passive network (always of low-pass configuration) does not provide a good approximation of a perfect integrator. However, with F(s) realized with an active filter, the Type of the loop would at least of Type II (depending on the number of integrators), as the active filter provides a good approximation of a perfect integrator. The order of the loop is determined by the

degree of the denominator polynomial of the closed loop transfer function. F(s) is often approximated be means of an active filter to be;

$$F(s) = \frac{s\tau_2 + 1}{s\tau_1} \qquad \ldots(5.3)$$

This is the classic lag-lead filter, and it has a pole at DC (provided the open loop gain of the op-amp is high), and a zero at $s = -1/\tau_2$. Substituting (5.3) into (5.2) yields;

$$H(s) = \frac{2K_oK_d(s\tau_2 + 1)/\tau_1}{s^2 + 2K_oK_d(\tau_2/\tau_1)s + 2K_oK_d/\tau_1} \qquad \ldots(5.4)$$

It is clear that this choice of F(s) gives a second order (and Type II) loop transfer function. Second order PLL's are often used in synchronization structures. The reasons are that they are simple to implement, and for step changes in input frequency, show only a small residual phase error (this is important in coherent schemes as the carrier must be accurately extracted). The form of equation (5.4) is the classic second order low pass type, from which the damping factor ζ and natural frequency (w_n) may be inferred to be;

$$2\zeta w_n = (2K_oK_d\tau_2)/\tau_1 \qquad \ldots(5.5)$$

$$w_n^2 = 2K_oK_d/\tau_1 \qquad \ldots(5.6)$$

It is common practice to set the damping factor to 0.707. This reflects a compromise between the transient response speed, and overshoot of the PLL to step changes in phase inputs. Substituting equation (5.6) into (5.5), and using $\zeta = 0.707$, we get;

$$\tau_2 = \sqrt{2}/w_n \qquad \ldots(5.7)$$

and

$$\tau_1 = 2K_oK_d/w_n^2 \qquad \ldots(5.8)$$

By measuring the loop parameters K_o and K_d, and specifying the loop natural frequency, we can determine the two time constants τ_1 and τ_2 in the loop filter. The open loop transfer function is plotted in Fig.5.5.

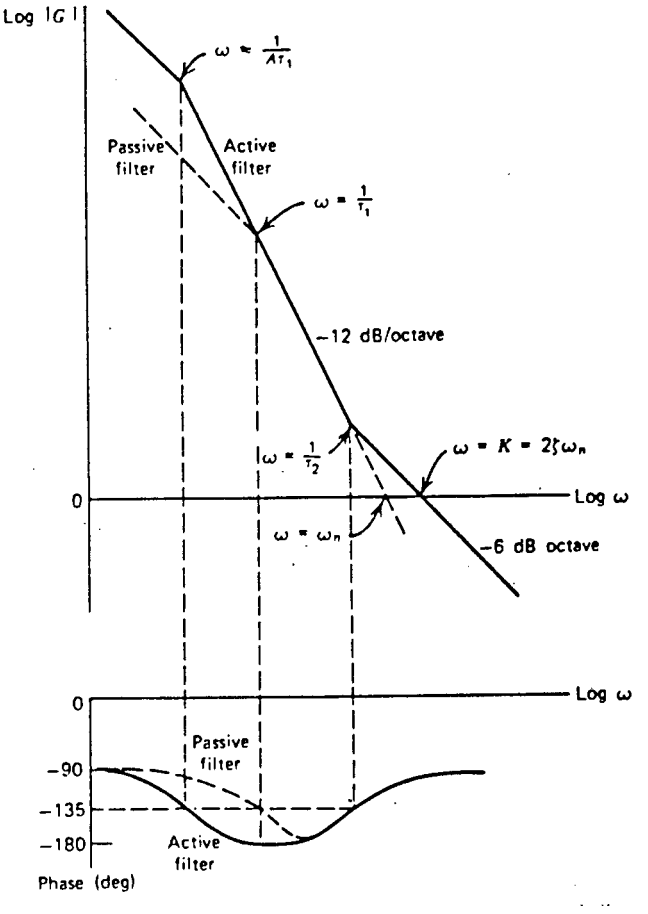


Fig.5.5 PLL Open Loop Transfer Function [5.3]

One can see that the initial -6 dB/octave slope due to the

VCO (recall that the VCO gain constant contains a 1/s term) is increased to -12 dB/octave at a frequency of $1/A\tau_1$ [rad/s], where A is the DC gain of the loop filter (this is typically a very large number, so the breakpoint is close to DC). After this pole, the phase response of the loop approaches -180°. At a radian frequency of $1/\tau_2$, the zero introduced by the loop filter decreases the slope to -6 dB/octave, and the loop phase approaches -90°. To allow for comfortable phase margin, the zero must be placed so that the unity gain point of the open loop transfer function is reached when the phase of the loop is less than 135° (i.e. 45° phase margin). The Amplitude response of the loop filter is shown in Fig.5.6.

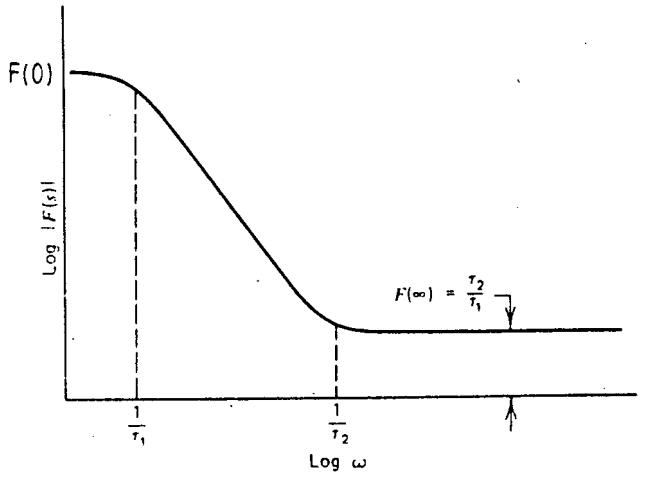


Fig.5.6 Loop Filter Transfer Function [5.3]

The loop filter (because it is active) has a DC gain of F(0). To ensure good tracking performance, F(0) should be made as large as possible. The high frequency gain of the filter can be seen from equation (5.3) to be $F(\infty) = \tau_2/\tau_1$. With these definitions in mind, two important loop gains

are defined;

$$K_v = 2K_oK_dF(0) \qquad \ldots(5.9)$$

and

$$K = 2K_oK_dF(\infty) \qquad \ldots(5.10)$$

We now proceed to discuss four loop parameters which are of interest to the PLL designer.

5.2.1(b) Pull-in Limit (Δw_{pi}). When the PLL is switched on, the free running frequency of the VCO will rarely be in agreement with the input signal frequency. The output of the phase detector will therefore be a beat note, equal in frequency to the frequency difference between the VCO and input frequency. An important property of the beat note is that it contains a DC component. This arises from the fact that the beat note itself frequency modulates the VCO, thus generating a Bessel spectrum with line spacings equal to the modulating frequency (the modulating frequency is the frequency difference between the VCO and input signal frequency). The spectra are shown in Fig.5.7.

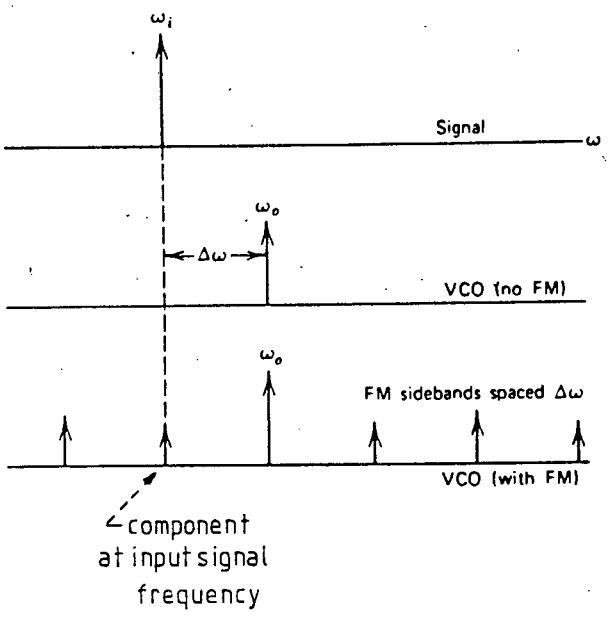


Fig.5.7 Spectra Generated During Pull-in [5.3]

One can see that one of the generated FM spectral components is at the same frequency as that of the input signal. The output of the phase detector therefore has a DC component due to the self-mixing of this frequency. This DC level is integrated by the loop filter, and the resultant voltage is applied to the VCO control input. This pulls the VCO frequency towards the input signal frequency by the phenomenon known as Pull-in. Note that since the loop filter attenuates the beat note to some extent, the maximum initial frequency difference between the VCO and input frequency for which the PLL will pull-in is partly determined by $F(\infty)$. This maximum frequency limit is termed the Pull-in Limit (Δw_{PI}), and is given in [5.3] as;

$$\Delta w_{PI} = \sqrt{2K_vK}$$

$$= K_oK_d\sqrt{8F(0)F(\infty)} \quad \text{[rad/s]} \quad \ldots(5.11)$$

To make Δw_{PI} as large as possible, it is desirable to make the DC gain of the loop F(0) large. The time taken for the loop to pull-in for an initial frequency offset of Δf, and loop noise bandwidth B_L (with $\zeta = 0.707$) is termed the Pull-in time, and is given by;

$$T_P = 4.2(\Delta f)^2/B_L^3 \quad [s] \quad ...(5.12)$$

The loop noise bandwidth B_L is related to the loop natural frequency w_n by the expression;

$$B_L = 0.53w_n \text{ [Hz]}, \text{ with } \zeta = 0.707 \quad ...(5.13)$$

As was mentioned previously, it is desirable to make B_L as small as possible (to limit the amount of phase noise), hence the pull-in time can be quite large in practice. The phenomenon is also unreliable, so it is often the case to construct frequency acquisition aids for PLL's [5.3].

5.2.1(c) Lock-in Range (Δw_L). This is the range of frequencies (which is smaller than the Pull-in Limit), for which the PLL will fall into lock without slipping cycles (the output of the VCO is merely a phase transient). This Lock-in phenomenon is very rapid, and an approximation for its value is given in [5.3] as;

$$\Delta w_L = \pm K$$

$$= \pm 2K_oK_dF(\infty) \quad [rad/s] \quad ...(5.14)$$

It will be shown shortly that for small noise bandwidths, $F(\infty)$ is a small number ($<<1$), so the Lock-in range is much less than the Pull-in Limit.

5.2.1(d) Hold-in Range (Δw_H). Once the PLL has acquired lock, there is a range of frequencies, the Hold-in Range, over which the PLL will track the input signal without losing lock. The limit set on the Hold-in range is determined by which loop components (PD or loop filter) saturates first. It should be borne in mind that as the input frequency varies, the phase detector output has to vary in order to adjust the VCO frequency to agree with the input frequency. The phase detector has a maximum output limit of a few hundred millivolts (for diode ring PD's) and this sets the Hold-in range. Based on this fact, [5.3] gives the Hold-in range as;

$$\Delta w_H = \pm K_v$$

$$= \pm 2K_oK_dF(0) \qquad \text{[rad/s]} \qquad ...(5.15)$$

Since F(0) has already been specified as large (>>1), it is clear that the Hold-in range is large (typically several MHz, but this is dependent on the frequency range of the VCO).

5.2.1(e) Pull-out Frequency (Δw_{PO}). With the PLL in lock, there is a maximum permissible step in input frequency which will not result in the PLL losing lock before re-acquiring it. This maximum frequency step is termed the Pull-out frequency, and is given in [5.3] as;

$$\Delta w_{PO} = 1.8w_n(\zeta + 1)$$

$$= 3.073w_n. \qquad \text{[rad/s]} \qquad ...(5.16)$$

5.2.2 Design and Construction of the PLL Modules

As the two PLL's used in the coherent demodulator lock onto the $2f_m$ and $2f_s$ spectral lines, and it is necessary to use the f_m and f_s frequencies to regenerate the I- and Q-Channel phasors, it is necessary to divide the input frequencies by a factor of two. It is for this reason that the PLL configuration shown in Fig.5.4 (Square Law Tracking Loop) is used. The PLL's not only provide the f_m and f_s frequencies, but the $2f_m$ and $2f_s$ ones as well.

5.2.2(a) Voltage Controlled Oscillator Module. The VCO itself is similar in design to that of the MSK modulator (section 4.1.3). The circuit is shown in Fig.5.8.

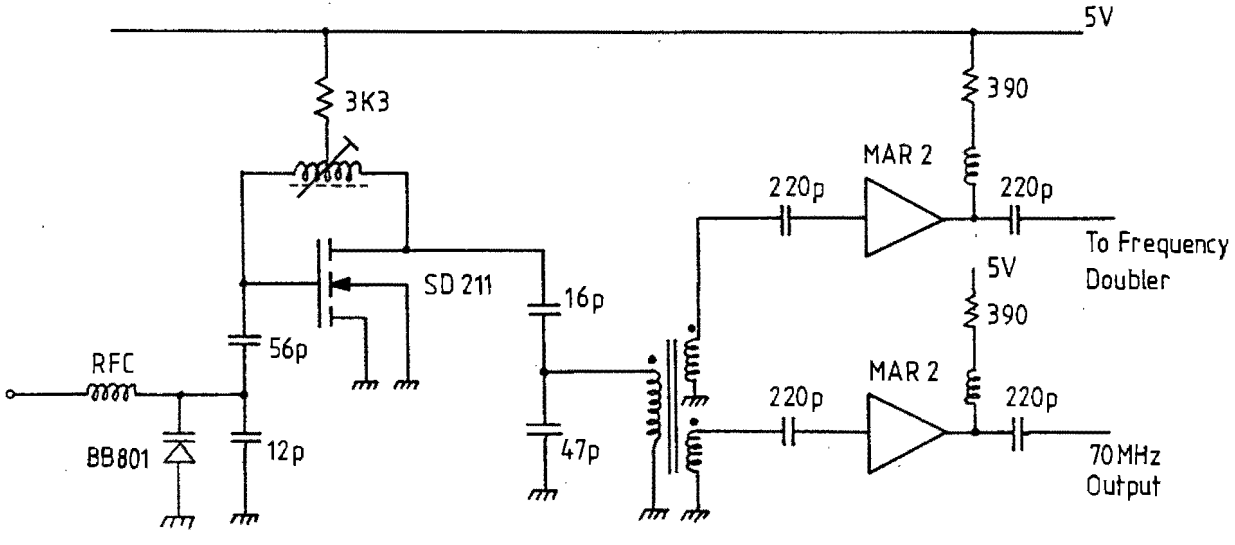


Fig.5.8 VCO Circuit Diagram

The control voltage input (which is variable between 0V and 15V) is coupled to the varactor by means of an RF isolation choke. This is done to prevent impairment of the

VCO tank circuit Q by the low output impedance of the loop filter. In order to restrict the range of frequencies attainable by the VCO, a 12 pF capacitor is connected in shunt with the varactor (this prevents the PLL from acquiring the wrong frequency).

As the modulation requirements of this VCO are not as stringent as those for the modulator, a simple resistance suffices as the current source for the MOSFET. The output of the VCO is powersplitted into two paths by means of a trifilar-wound transformer. One path is amplified to approximately 5 dBm and is used as the PLL 70 MHz output (it provides the f_m or f_s frequency, depending on which PLL it is). The other path is amplified to a similar level and it feeds the frequency doubler module. The transfer function for the $2f_m$ PLL VCO is shown in Fig.5.9.

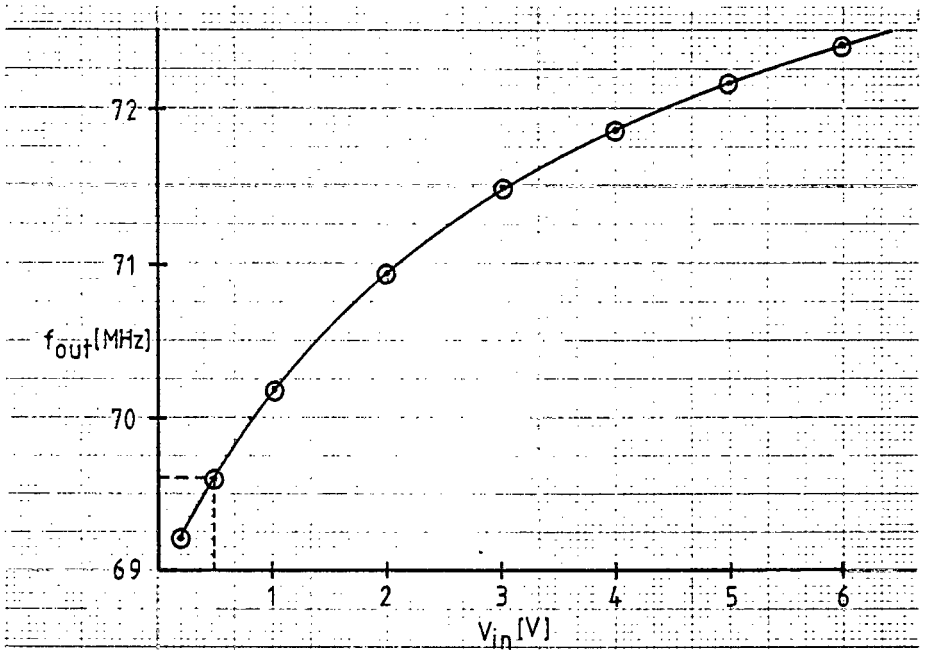


Fig.5.9 MARK PLL VCO Transfer Function

The operating point is f_{out} = 69.632 MHz, and v_{in} = 0.5V. From the curve, the slope at v_{in} = 0.5V is found to be;

$$K_o\Big|_{v_{in}=0.5V} = 8.42\ E6\ rad/s/V.$$

The value of K_o for both the MARK ($2f_m$) and SPACE ($2f_s$) PLL's are similar.

5.2.2(b) Frequency Doubler Module. The circuit diagram for this module is shown in Fig.5.10.

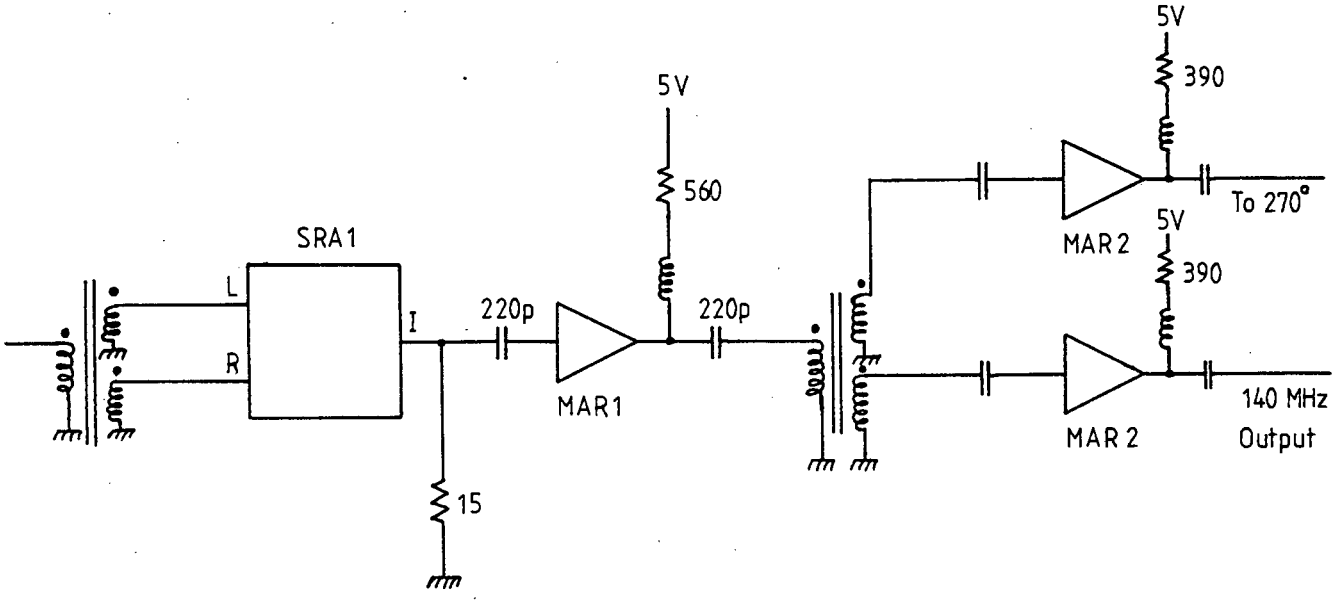


Fig.5.10 Frequency Doubler Module

As was the case with the squaring circuit described in section 5.1, a diode ring double balanced mixer is used for the frequency doubling function. The mixer used, the mini-circuits SRA-1, was found to give best performance when the L- and R-ports were driven by means of a power splitter. The frequency-doubled output is taken at the I-port of the mixer, and is amplified and power splitted into two paths: one serves as the 140 MHz output, while the other is used to drive the phase detector via the 270°

phase shifter. The output levels of this module are ≈2dBm.

5.2.2(c) 270° Phase Delay Module. An inherent feature of the phase detector to be described in the next section, is that it provides zero DC output when the waveform on the L-port lags the waveform on the R-port by 90° i.e. without the 270° phase delay present, the PLL output would be lagging the waveform onto which it is locked by 90°.

The 270° phase delay circumvents this problem by introducing an extra 270° phase delay in the loop, which results in no net phase difference (modulo 360°) between the input to, and output from, the PLL. Ideally, three Hilbert transformers in cascade would be used, as they provide a fixed phase shift independent of frequency. However, it is difficult to implement this and the phase delay is realized by means of a tuned length of co-axial cable. The cable used was RG-174U, which has its EM phase velocity equal to 0.66c, where c is the speed of light (≈ 3 E8 m/s).

$$\text{i.e.} \qquad c/\sqrt{\epsilon_r} = 0.66c$$

$$\Rightarrow \qquad \sqrt{\epsilon_r} = 1.515,$$

where ϵ_r is the relative permittivity of the cable dielectric. For the MARK frequency PLL (operating at 139.264 MHz), the free space wavelength of the EM wave is given by;

$$\lambda = c/f = 3\ E8/139.264\ E6 = 2.1542\ m.$$

In the co-axial cable, this is reduced by a factor $1/\epsilon_r$, and remembering that a 270° phase delay corresponds to 75%

of the free space wavelength, we get;

$$\ell_c = 0.66*0.75*2.1542$$
$$= 1.066 \text{ m},$$

where ℓ_c is the required length of cable for the MARK frequency PLL. The length of cable for the SPACE PLL is 1.051 m.

Note that each of the phase shifters is optimized for only one frequency, but for the range of frequency uncertainty expected, the maximum phase error is less than 1%.

5.2.2(d) Phase Detector Module. It was decided to use a diode ring mixer for this purpose. These mixers have the advantage of easy implementation, and wide bandwidth. As there are no external components, the circuit diagram is not shown here. The R-port is fed by one of the outputs of the frequency doubler module (as described in section 5.1), while the L-port is driven by the Phase Delay module. For optimum performance, the levels at the input to the mixer should be approximately equal (≈ 1 dBm), and constant in amplitude (if this is not the case, the output level of the mixer becomes amplitude as well as phase dependent). The output voltage vs input phase differential for the phase detector is shown in Fig.5.11.

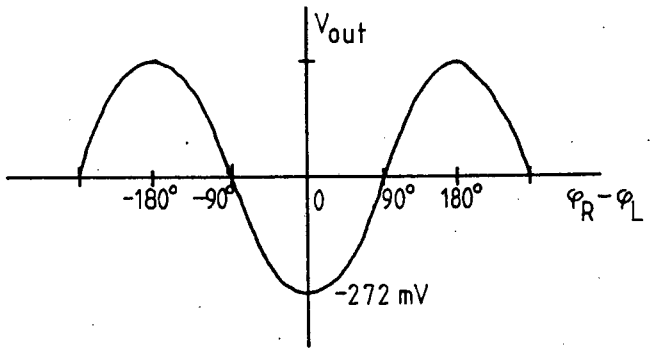


Fig.5.11 Phase Detector Transfer Function

The phase detector produces zero DC output for phase differentials being odd multiples of 90°, and maximum output for even multiples of 90°. Note that the characteristic is cosinusoidal, as is common in diode ring phase detectors. The maximum output level was easily determined by driving the L- and R-ports with the same 140 MHz, 0 dBm signal, and measuring the DC output. The maximum output level was -272mV, hence we can say;

$$V_{out} = -0.272.\cos \phi \quad [V],$$

where ø is the input phase differential.

$$\Rightarrow \quad \frac{dV_{out}}{d\phi} = 0.272.\sin \phi \quad [V/rad]$$

and for ø = 90° (the steady state phase differential when the PLL is locked), we have;

$$K_d = \left.\frac{dV_{out}}{d\phi}\right|_{\phi=90°} = 0.272 \ [V/rad],$$

where K_d is the phase detector gain constant.

It should be noted that as the maximum output level of the phase detector is less than 300 mV, it is necessary to amplify this level to a more usable range for the VCO (i.e. the range 0V to 15V). It is for this reason that an active filter is used to realize the loop filter, as it can have gain (the passive filter cannot).

5.2.2(e) Loop Filter Module. This module is used to implement the F(s) transfer function given in equation (5.3). The op-amp realization of such a transfer function is shown in Fig.5.12.

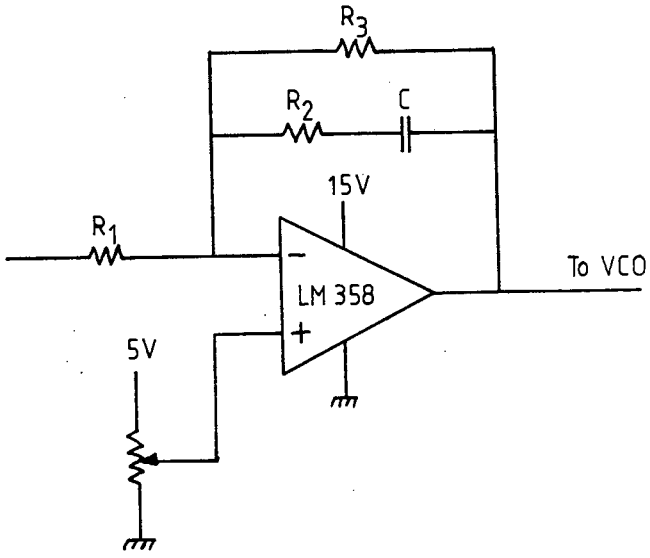


Fig.5.12 Loop Filter Module

The transfer function for this filter is given in [5.3] as;

$$F(s) = \frac{-A(sCR_2 + 1)}{sCR_2 + 1 + (1+A)(sCR_1)}$$

The DC gain A is (without R_3 present) normally a very large number (typically 10^5). The presence of R_3 reduces this gain to F(0), which is still large (typically 100). The reason for including R_3 will be explained later. With A large, F(s) may be reduced to;

$$F(s) \approx - \frac{sCR_2 + 1}{sCR_1}$$

$$= - \frac{s\tau_2 + 1}{s\tau_1}$$

where $\tau_1 = R_1C$,

and $\tau_2 = R_2C$.

We can therefore realize the time constants τ_1 and τ_2 in equations (5.7) and (5.8) by means of the capacitance C and resistances R_1 and R_2. It was mentioned previously that R_3 was included in the op-amp feedback path to reduce the DC gain of the filter. The DC gain is given by;

$$F(0) = R_3/R_1$$

and we desire this to be large ($\approx$100). The reason we restrict the DC gain is that any voltage offsets present in the phase detector output will result in the op-amp reaching saturation levels were the gain too high. We therefore set the DC gain to be;

$$F(0) = 100.$$

In order to calculate τ_1 and τ_2, we need to specify a value for the loop natural frequency w_n. As the loop noise

bandwidth is dependent on w_n, we desire to make w_n small. Experience indicated that $w_n \leq 2\pi.1000$ [rad/s] provided satisfactory results. In the ensuing calculations, we assume $w_n = 2\pi.1000$ [rad/s].

Using the loop gain constants measured or defined in the previous sections, we arrive at the following filter parameters;

$$\tau_1 = 116.0 \text{ ms}$$

and

$$\tau_2 = 225.1 \ \mu\text{s}.$$

With C = 10 μF, we obtain;

$$R_1 = 11.6 \ \text{K}\Omega$$

and

$$R_2 = 22.5 \ \Omega.$$

$R_3 = 100R_1$ therefore $R_3 = 1.16 \ \text{M}\Omega$.

The op-amp used was an LM358 single supply device. Note that it is not possible to use LF351-type devices as these are bipolar supplied, and their output is consequently incompatible with the varactor in the VCO. The potentiometer at the non-inverting input to the op-amp sets the quiescent output voltage, and this is fixed at 0.5V. The measured loop filter response is shown in Fig.5.13.

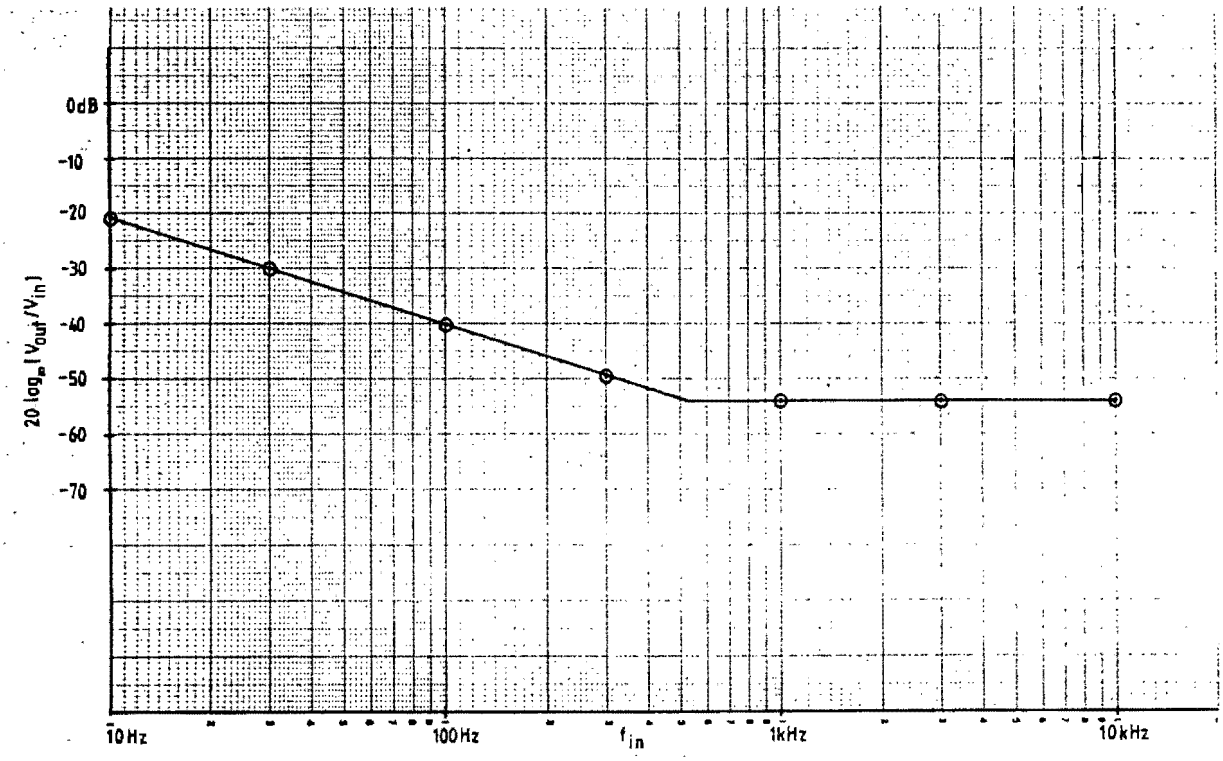


Fig.5.13 Loop Filter Frequency Response

Using equations (5.11), (5.13), (5.14), (5.15), and (5.16), we can calculate the expected performance of the PLL. Fig.5.14 tabulates these results.

Parameter	Calculated Value
Pull-in Limit	454.1 kHz
Hold-in Range	72.9 MHz
Lock-in Range	1.41 kHz
Pull-out Frequency	3.07 kHz
Noise Bandwidth	3.33 kHz

Fig.5.14 Calculated Loop Parameters

Note that the Hold-in range is unrealistically large, and

one must interpret this value as meaning that the PLL can track widely varying input frequencies. The Lock-in range and Pull-out frequencies are seen to be small, and this is a result of the small value of w_n. However, the noise bandwidth is also small because of the chosen value of w_n.

5.2.3 Phase-Locked Loop Performance Evaluation

In order to assess the spectral quality of the PLL output, it is necessary to measure the phase noise content of the recovered carrier.

Any non-linear operation (such as squaring, as described in section 5.1) results in the regeneration of discrete spectral lines, together with a random noise-like spectrum which is caused by the randomness of the modulating data stream. The PLL's act as narrow bandpass filters to extract the spectral lines, but some of the random noise spectrum is admitted through the passband of the PLL's.

The phase noise manifests itself as noise sidebands on the PLL output spectrum. Ideally, the PLL output spectrum consists of a single delta function-like spectral line at the acquired frequency. A comparison between the ideal case and phase noise corrupted case is shown in Fig.5.15.

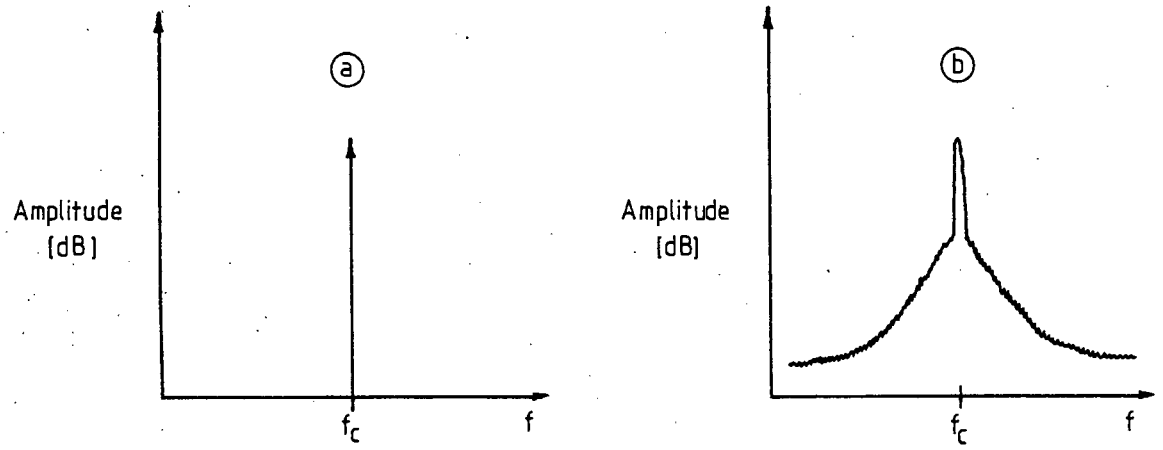


Fig.5.15 (a) Ideal Spectrum
(b) Phase Noise Corrupted Spectrum

Since the noise sidebands are close to the carrier, they can be thought of as low modulation index type frequency modulation. In the time domain, this translates to a waveform which slowly shifts its zero crossings relative to their nominal positions i.e. the waveform is said to have jitter. Fig.5.16 illustrates how phase noise on a carrier may appear in the time domain.

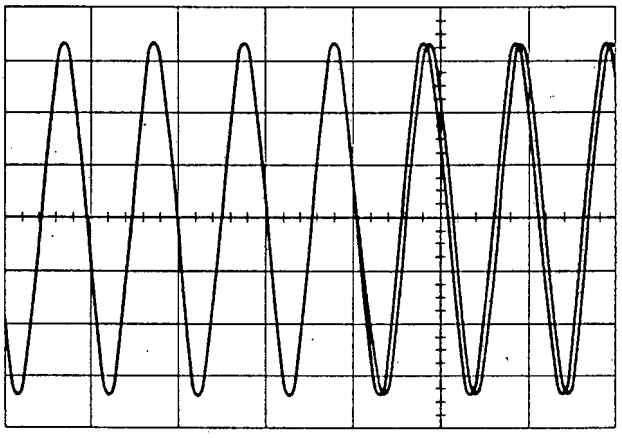


Fig.5.16 Phase Noise Viewed in the Time Domain

There are several methods of quantifying phase noise and these are described in [5.4]. The method chosen here is the Direct Spectrum Measurement method as it is an easy and rapid method of measurement. It suffers the disadvantage that it cannot distinguish between AM and FM noise. This is no disadvantage in our case, as the PLL output is free of AM modulation. A second limitation of this method is that the phase noise can only be measured down to the noise floor of the spectrum analyzer ($\approx$ -70 dBm). Fig.5.17 shows an example of a carrier with phase noise present.

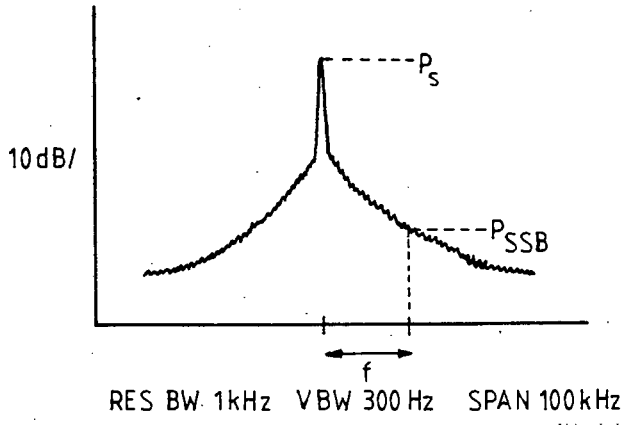


Fig.5.17 Carrier Phase Noise Measurement

The measurement is made as follows: The carrier power P_s is first measured. Then, for a frequency offset f (typically 10 kHz), the noise power (referred to as the Single Sideband Noise Power P_{SSB}) is measured. Because the resolution bandwidth of the spectrum analyzer is, in this case 1 kHz, the power in 1 Hz is given by;

$$P_{SSB}\Big|_{1Hz} = P_{SSB}(dBm) - 10\log_{10}\left[\frac{\text{RES BW [Hz]}}{\text{1 [Hz]}}\right]$$

For a resolution bandwidth of 1 kHz, RES BW = 1 E3, so;

$$P_{SSB}\Big|_{1Hz} = P_{SSB}(dBm) - 30.$$

We can now define the result of the direct spectrum measurement (called the Single Sideband Noise-to-Carrier Power ratio, $\mathcal{L}(f)$) to be;

$$\mathcal{L}(f) = P_{SSB}(dBm) - 30 + C_m - P_s(dBm) \quad [dBc/Hz] \qquad ...(5.17)$$

where C_m is a correction factor (normally 2.5dB) for analogue spectrum analyzers when making noise measurements [5.4].

The measured output spectrum of the MARK frequency PLL is shown in Fig.5.18.

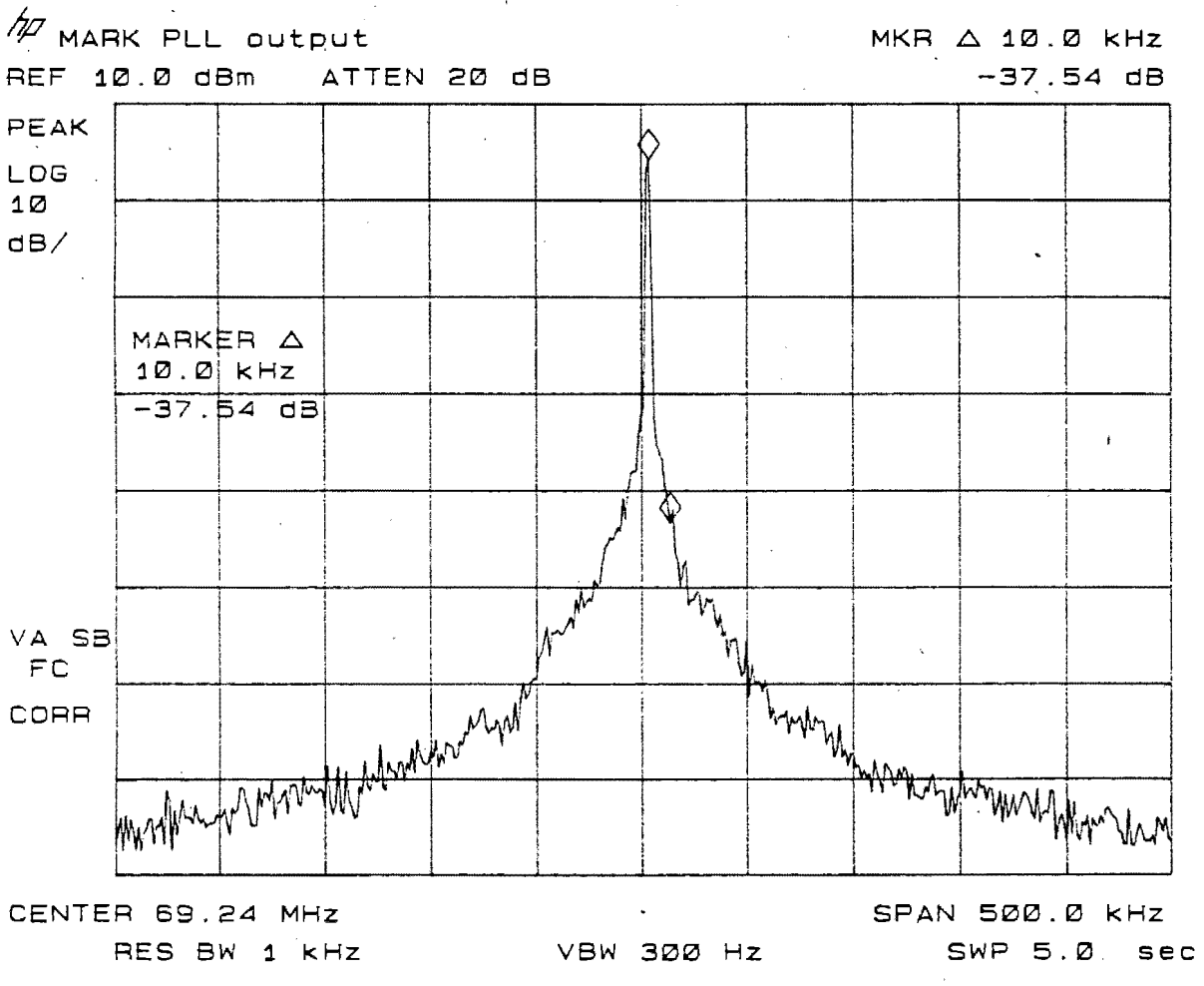


Fig.5.18 MARK Frequency PLL Output Spectrum

Using equation (5.17) and the values measured off the

plot, we get;

$$\mathcal{L}(10\text{kHz}) = -37.54 - 30 + 2.5$$

$$= -65.04 \quad [\text{dBc/Hz}]$$

This performance may be improved by more careful circuit layout, and a narrower loop bandwidth.

5.3 SYMBOL CLOCK RECOVERY MODULE

In order to sample the I- and Q-Channel eyes at their maximum eye openings (as shown in Fig.2.20), it is necessary to recover the symbol rate clock from the received spectrum, and use this to synchronize the sampling circuits.

It has already been shown (equation 2.25) that the bit rate clock is given by the difference frequency of the $2f_m$ and $2f_s$ carriers. The first step then, is to mix the two 140 MHz outputs of the MARK and SPACE frequency PLL's. The difference frequency is then extracted by means of a low-pass filter. The circuit diagram of the module is shown in Fig.5.19.

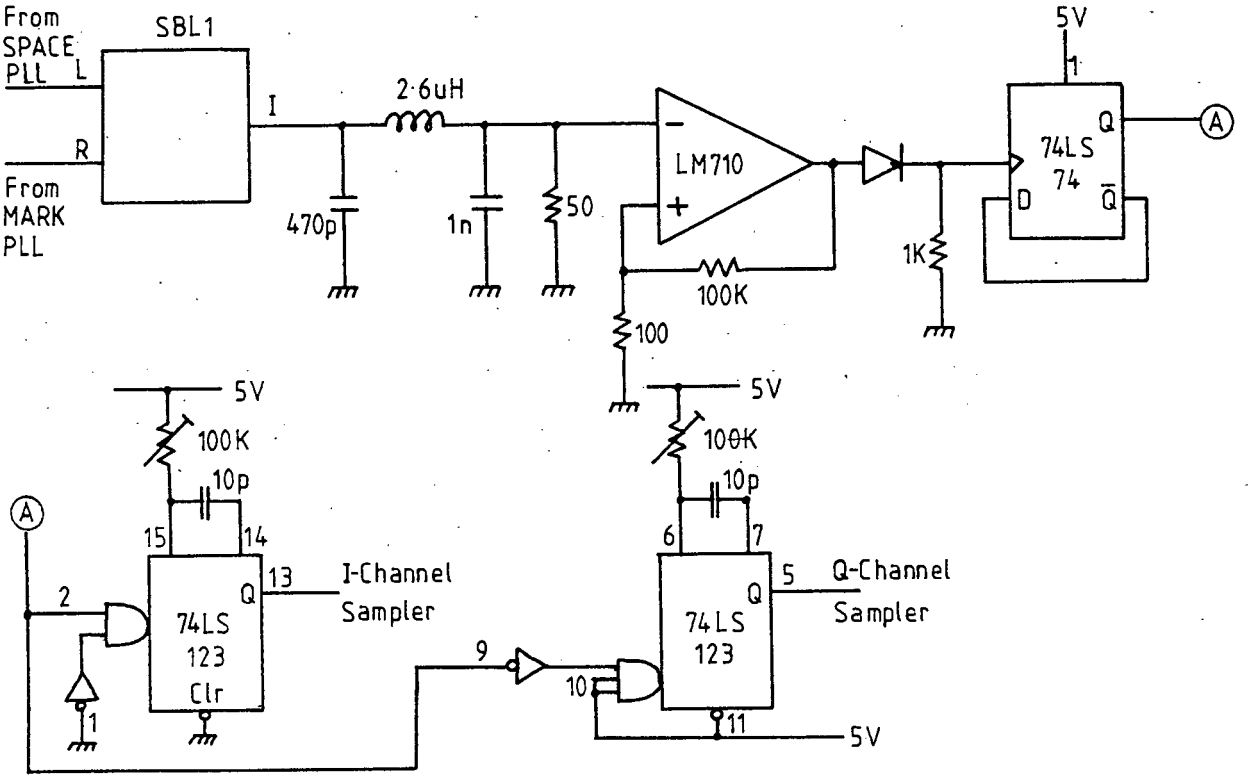


Fig.5.19 Symbol Clock Recovery Module

The mixing of the two PLL outputs is performed by an SBL-1 mixer. The resultant sinusoidal output is then lowpass filtered by means of a 3 Pole, 5 MHz cutoff Butterworth filter. The output of the filter is the bit rate frequency of 2.048 MHz.

An LM710 voltage comparator with 5mV of hysteresis then detects the zero crossings of the waveform and produces a bipolar square wave output. This is made TTL compatible by means of a diode-resistor rectifying network. To obtain the symbol rate clock, it is necessary to divide the bit rate frequency by two, and the 74LS74 D-type flip-flop performs this function.

Observing Fig.2.20, one can see that the I-Channel must be

sampled at the falling edge of the symbol clock, while the Q-Channel is sampled at the rising edge of the clock. As the sampling is optimal at the maximum eye opening, the I- and Q-Channel sampling waveforms (their falling edges) must be capable of being variable relative to the rising edge of the symbol clock (i.e. the sampling instant for each channel must be variable to adjust it for optimum operation).

This requirement is met by the use of a dual monostable (74LS123). One monostable triggers on the falling edge of the symbol clock, and its output pulse width may be adjusted by means of a potentiometer so that its falling edge corresponds to the maximum eye opening of the I-Channel. The other monostable triggers on the rising edge of the symbol clock, and its pulse width is also adjustable for optimum operation.

5.4 ADDER/SUBTRACTOR MODULE

As was shown in section 2.2.4(e), the I- and Q-Channel phasors (the quadrature weighted carriers) may be generated by either adding the extracted f_m and f_s frequencies (this generates the I-Channel phasor), or subtracting them (thus generating the Q-Channel phasor).

The I-Channel quadrature carrier was shown to be;

$$s_I(t) = \pm 2.\cos(\pi t/2T_b).\cos(2\pi f_c t),$$

and that of the Q-Channel to be;

$$s_Q(t) = \pm 2.\sin(\pi t/2T_b).\sin(2\pi f_c t).$$

For addition and subtraction at 70 MHz, RF adders and

subtractors are required. It was decided to use Trifilar-wound broadband transformers [5.5] as the addition/subtraction circuit elements. Fig.5.20 shows the circuit diagram.

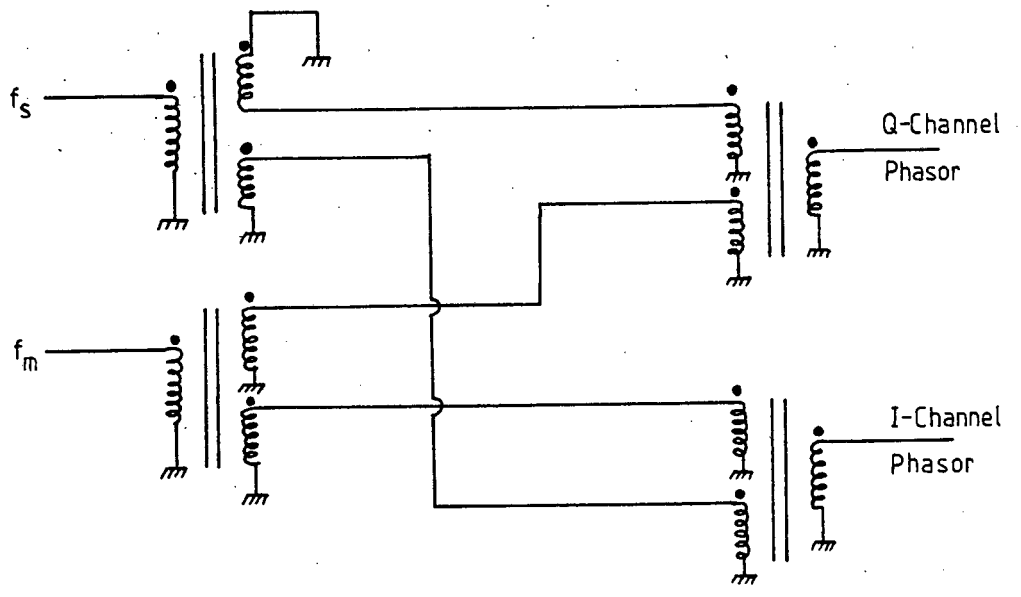


Fig.5.20 Adder/Subtractor Module

The MARK frequency PLL feeds the 0°-0° power splitter, and each output of this splitter feeds an input to the the 0°-0° power combiners. The SPACE frequency PLL feeds the 0°-180° power splitter, and each output of this splitter feeds the other two inputs of the 0°-0° power combiners. The upper power combiner therefore effectively subtracts the PLL outputs, while the lower one adds them. The addition/subtraction process is linear provided that the transformers are not operated in their non-linear (saturated) region. For the powers used in this application, the transformers are operated well out of their non-linear region.

The isolation between the two input ports was measured (at an input power level of 5 dBm) by connecting an RF source to one input port, and measuring the power level at the other input port. The isolation was found to be the same for both ports, and its measured value was 36.5 dB. This is considered adequate, and the good performance is due to the inherent symmetry of the transformer arrangement.

A photograph of the regenerated quadrature phasors is shown in Fig.5.21.

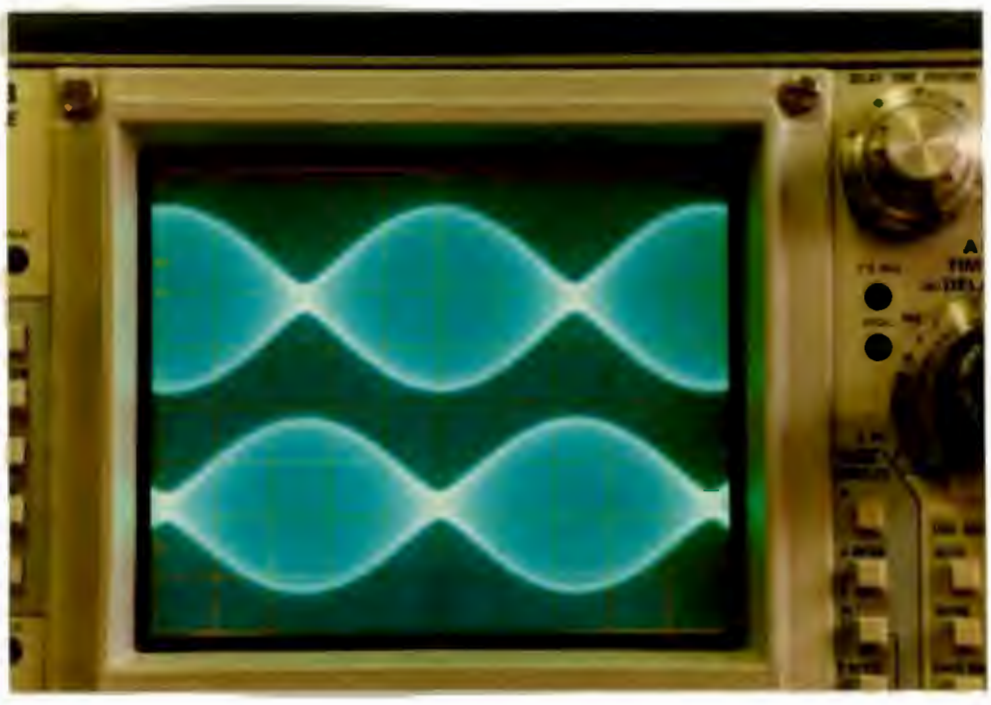


Fig.5.21 Regenerated Quadrature Phasors
(Horizontal Scale: 0.2μs/div)

5.5 COHERENT DOWN-CONVERSION MODULE

The regenerated quadrature phasors are used in the coherent

downconversion to baseband of the received MSK waveform. As was discussed in section 2.2.4(e), the I-Channel is obtained by forming the product;

$$\text{I-Channel} = s_I(t).s_{MSK}(t),$$

where $s_I(t)$ is the I-Channel quadrature phasor, and $s_{MSK}(t)$ is the received MSK waveform at IF.

The Q-Channel is obtained from the product;

$$\text{Q-Channel} = s_Q(t).s_{MSK}(t).$$

In order to perform these multiplications, SBL-1 double-balanced mixers are used, as is shown in Fig.5.22.

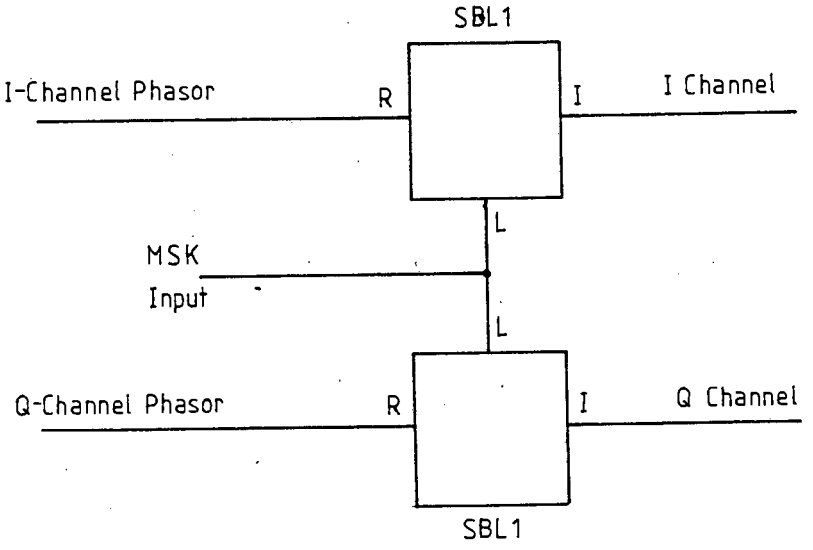


Fig.5.22 Coherent Down Conversion Module

For correct mixer operation, the MSK waveform power level is set at 0dBm, and that of the quadrature phasors at 5 dBm.

5.6 DETECTION FILTER MODULE

Ideally, the outputs of the coherent downconversion module would be integrated over a $2T_b$ period, sampled at the end of the integration, and a decision then made as to the symbol value. This matched filter receiver implementation was discussed in section 2.2.4, and it was mentioned there that a simpler implementation of the decision structure is possible.

It is easier to lowpass filter the I- and Q-Channel outputs of the downconversion module, and then sample the resulting waveforms at their maximum eye openings. In the paper by Austin et al. [5.6], useful design curves for determining the exact parameters (order and bandwidth) of the lowpass filters are presented. Fig.5.23 shows the curves for 2-Pole Butterworth detection lowpass filters.

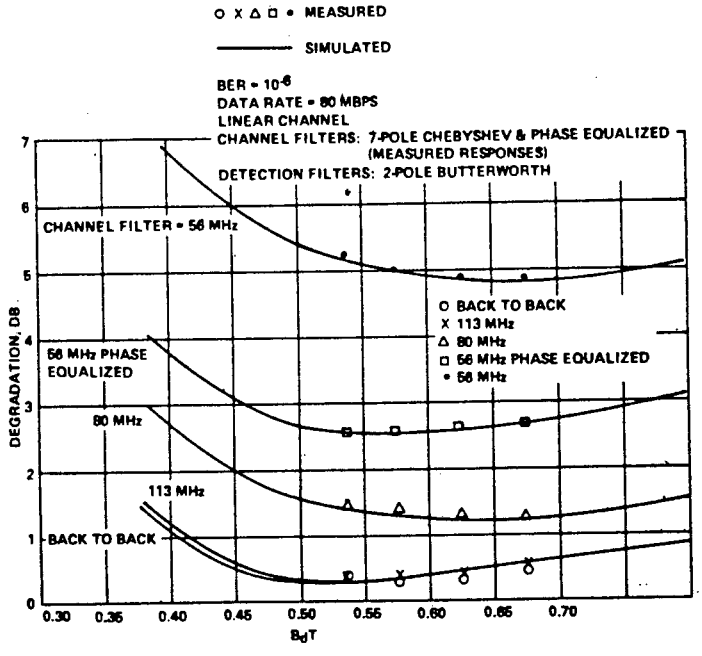


Fig.5.23 Design Curves for Lowpass Detection Filters [5.6]

The horizontal axis shows the 3 dB bandwidth (B_d), symbol duration (T) products. The vertical scale indicates the performance degradation to be expected relative to ideal matched filter detection, for various levels of channel filtering. The lowest curve (for back-to-back modem operation) is the relevant one for our design. For this curve, least degradation occurs at;

$$B_dT = 0.525,$$

and the degradation at this value is 0.33 dB, which is minor. For $T = 2T_b = 976.6$ ns, the 3 dB bandwidth of the detection filter is;

$$B_d = 537.6 \text{ kHz}.$$

The filter is designed for source and load impedances of 50Ω, and its circuit is shown in Fig.5.24.

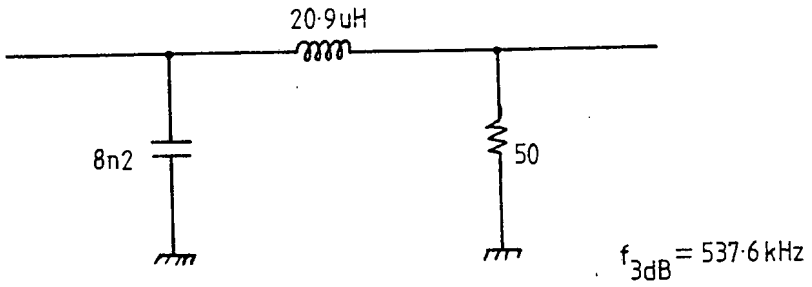


Fig.5.24 Detection Filter Module

The constructed Coherent MSK Receiver is shown in Fig.5.25.

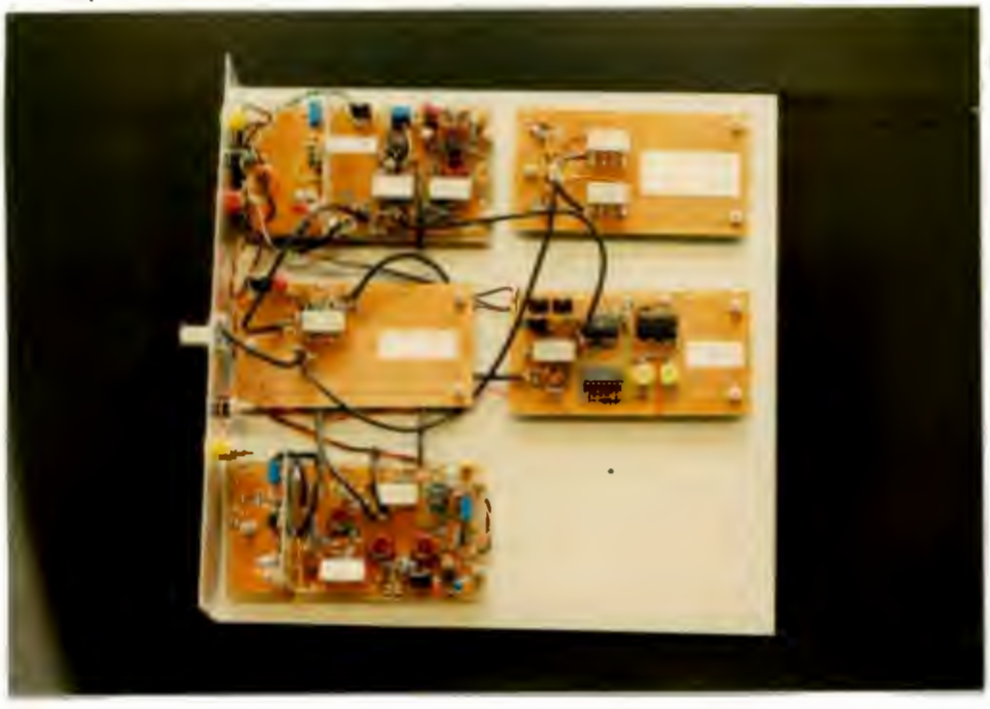


Fig.5.25 Constructed Coherent MSK Receiver

REFERENCES

[5.1] Mathwich, H. R., Balcewicz, J. F. and Hecht, M., "The effect of tandem band and amplitude limiting on the E_b/N_o performance of MSK", IEEE Trans. on Comm., Vol. COM-22, Oct. 1974, pp.1525-1540.

[5.2] de Buda, R., "Coherent demodulation of frequency-shift keying with low deviation ratio", IEEE Trans. on Comm., Vol. COM-20, June 1972, pp.429-435.

[5.3] Gardner, F. M., "Phaselock techniques", 2nd ed., Wiley, New York (1979).

[5.4] Feher, K. and the Engineers of Hewlett-Packard, "Telecommunications measurements, analysis, and instrumentation", 1st ed., Prentice-Hall, New Jersey (1987).

[5.5] Bowick, C., "RF circuit design", Sams, Indiana

[5.6] Austin, M. C. et al, "QPSK, staggered QPSK, and MSK-a comparative evaluation", IEEE Trans. on Comm., Vol. COM-31, Feb. 1983, pp.171-182.

CHAPTER 6

DESIGN AND CONSTRUCTION OF A MODULATION INDEX CORRECTION MECHANISM

The effects of an incorrect modulation index on the error performance of an MSK link were investigated in Chapter 3.

In this chapter, the design and construction of a correction mechanism (to be denoted "Servo Control Circuitry") to control the modulation index is discussed.

6.1 BASIC DESIGN STRATEGY

It was decided that any correction mechanism to be implemented should be done so at the modulator (as opposed to the demodulator), as this results in simpler hardware realization. At the modulator, we have a-priori knowledge of the correct modulation index, and if we were able to compute the actual modulation index of the modulator and compare it with the nominal value, a correction mechanism could be constructed to maintain it within tolerable limits.

The strategy used to compute the modulation index of the modulator was to use the clock recovery circuitry (as developed for the demodulator) as the starting point. Recalling equation (2.25), we had;

$$f_b = 2f_s - 2f_m,$$

where f_b is the bit rate clock and f_s and f_m are the SPACE and MARK frequencies respectively. For the situation where the modulation index h is not 0.5, we have;

$$f_b' = 2(0.5 \pm \delta h)f_b$$
$$= f_b \pm 2\delta h f_b,$$

where δh is a small positive number representing the modulation index error. It is clear that the recovered clock f_b' differs from the data rate clock (which is precisely 2.048 MHz) by the term $2\delta h f_b$. Thus, by comparing the recovered clock with the hardware clock, we may produce an error signal which can be used to adjust the MARK and SPACE frequencies of the modulator so that the modulation index is kept at its nominal value. A block diagram of such a correction mechanism is shown in Fig.6.1.

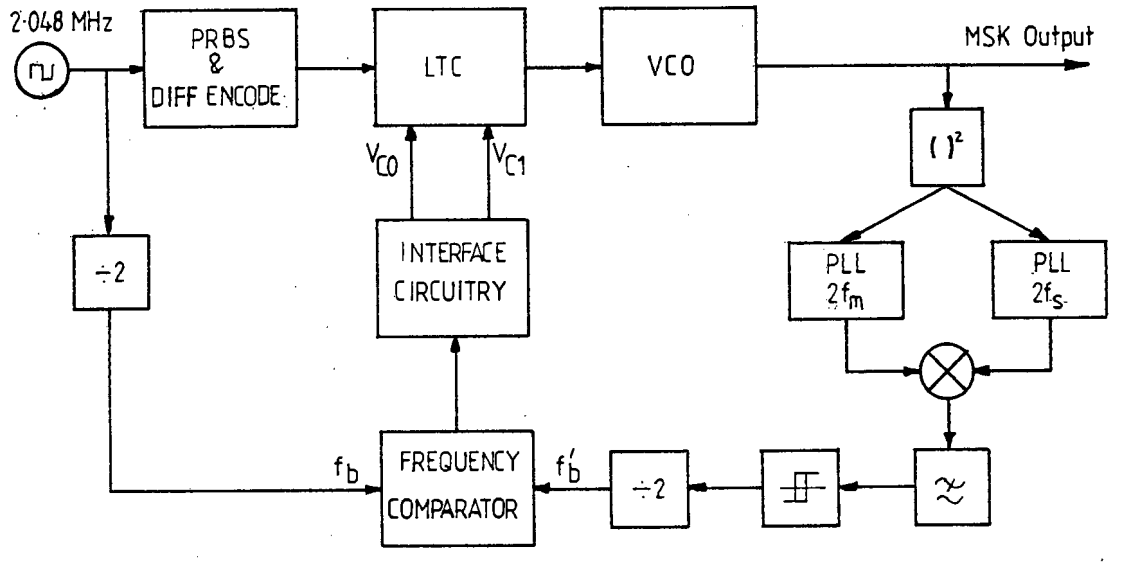


Fig.6.1 Block Diagram of the Modulation Index Servo Control Circuitry

The modulator VCO output is sampled by the input to the symbol clock recovery circuit. The recovered symbol clock is compared with the hardware symbol clock and an error signal

(proportional to the frequency difference between the two inputs) is produced. This is used to drive the interface circuitry which derives, from the error signal, the voltage control inputs to the level translation circuitry necessary to maintain the modulation index at its correct value. Instability of the loop is avoided by making the loop time constant large, thus introducing a form of dominant pole compensation.

6.2 DESIGN OF SERVO CONTROL CIRCUITRY

There are two modules of which the Servo Control Circuitry is comprised. They are;

- The Frequency Comparator
- The Interface Circuitry

Their design and construction is now described.

6.2.1 Frequency Comparator Module

This module is required to produce a bipolar (±15V) output for a maximum input frequency differential of 1 MHz. Because of its simple implementation, a Differential Charge Pump is used to perform the frequency comparison.

The output voltage of a differential charge pump is given by the following expression;

$$V_{out} = (V_1 - 2v_d)(f_1 - f_2)C_1R, \qquad ...(6.1)$$

where;

V_1 = Amplitude of input waveform

v_d = Forward voltage drop of a diode (=0.6V)

f_1, f_2 = The input frequencies to be compared

C_1 = Input capacitor value

R = Output load resistance of the charge pump.

The charge pump operates off a single (+15V) supply, and the necessary level translation to bipolar (±15V) format is done by means of an inverting summer op-amp. Fig.6.2 shows the circuit for the module.

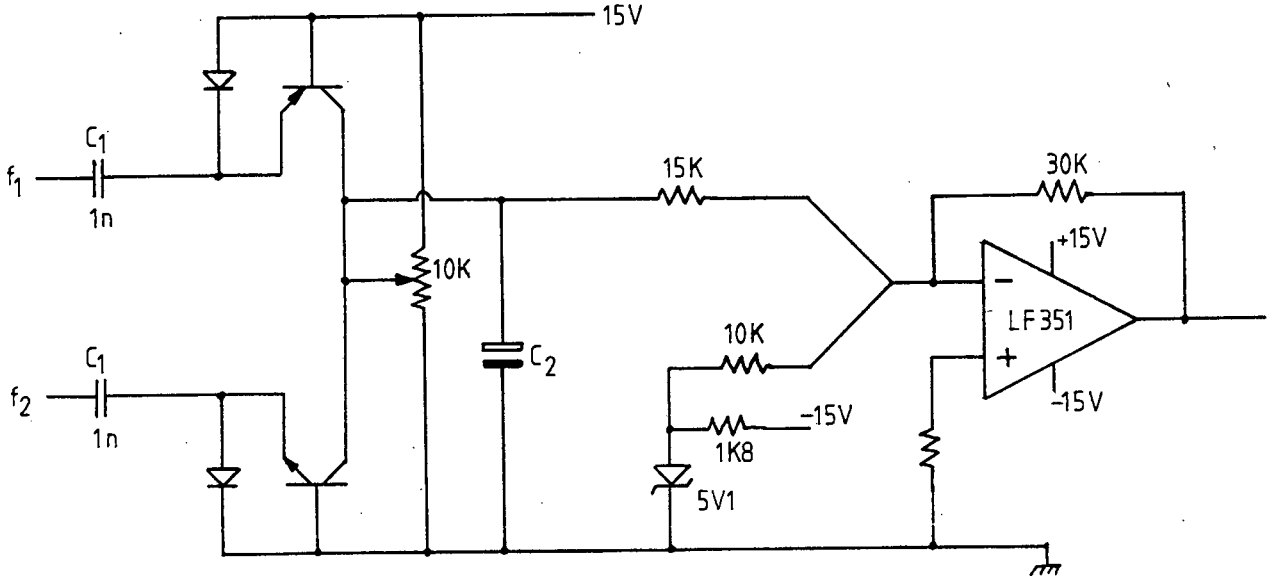


Fig.6.2 Frequency Comparator Module

Since the input to the charge pump is in TTL format, V_1 is approximately 4V, and hence $V_1 - 2v_d = 2.8V$. For a maximum input frequency differential of $(f_1 - f_2) = 1$ MHz, the maximum output voltage swing V_{out} (max) must not exceed 7V in order to ensure correct transistor operation. We thus have;

$$V_{out}(max) = 7 = 2.8 * 1\ E6 * C_1R.$$

For a 10KΩ potentiometer at the output of the charge pump, $R \approx 2.5K\Omega$ (Thévenin equivalent resistance), and hence $C_1 = 1nF$. For $f_1 = f_2$, the output potentiometer is set to give $V_{out} = 7.5V$ (i.e. Vcc/2). The smoothing capacitor is chosen to be large to ensure that the response of the circuit is relatively slow to enhance the overall loop stability.

The output of the charge pump is converted into bipolar (±15V) format by means of the inverting summer circuit. The required mapping function is given by;

$$V_{out} = 15 - 2V_{in}.$$

Hence the op-amp is required to provide a constant 15V output, and subtract from it twice the input voltage. Considering the circuit as a whole, the output voltage is -15V for $f_1 >> f_2$, and +15V for $f_2 >> f_1$. For the two frequencies being equal, we have the output voltage equal to zero. The measured transfer function for the module is plotted in Fig.6.3.

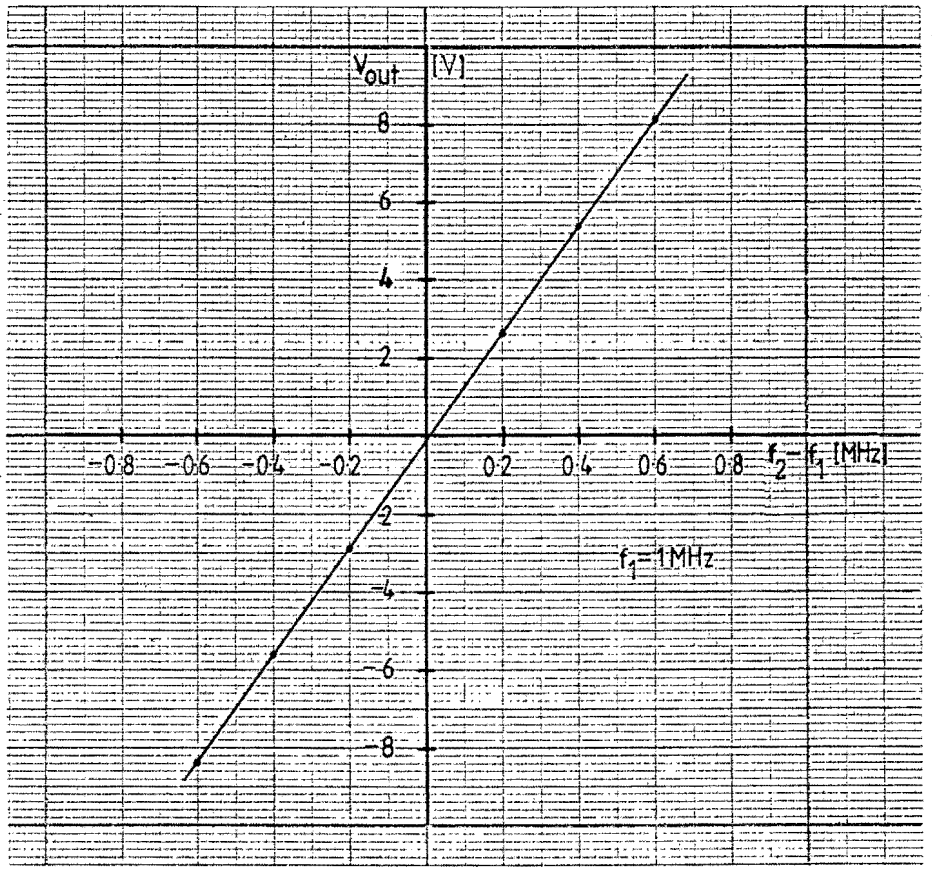


Fig.6.3 Measured Transfer Function of Frequency Comparator Module

Using actual circuit values, the slope of the curve (hence the module gain) is calculated to be +14mV/kHz, and this agrees well with the the actual curve slope of +13.9mV/kHz. The curve exhibits a high degree of linearity with a correlation coefficient of 0.9987.

6.2.2 Interface Circuitry

In order to use the single-ended bipolar output produced by the frequency comparator module, it is necessary to convert it into two unipolar levels which are applied to the level translation circuitry inputs V_{c0} and V_{c1} (see Fig.6.1). The interface circuitry was designed to perform this function.

The operation of this module is best understood by referring

to its circuit diagram shown in Fig.6.4.

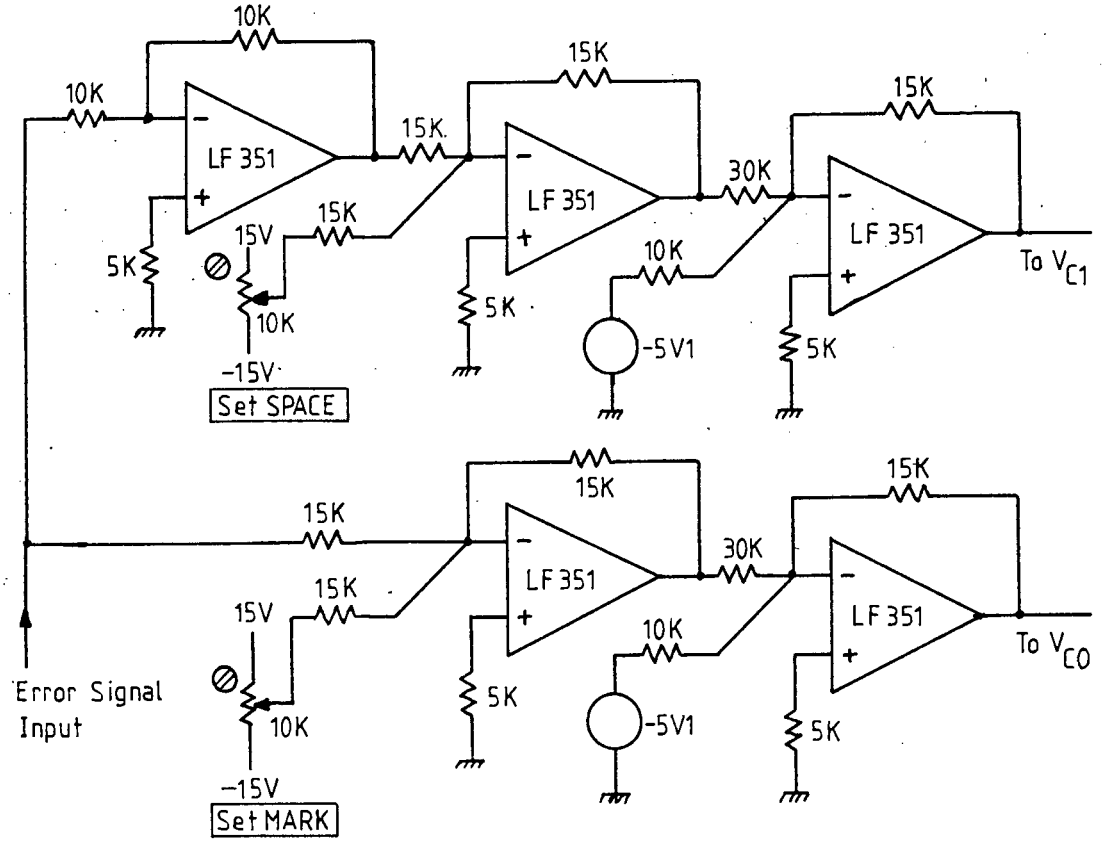


Fig.6.4 Interface Circuitry

If, for example, the modulation index of the modulator were too large, the recovered symbol-rate clock frequency would be greater than the nominal frequency of 1.024 MHz, and this results in the output of the frequency comparator module increasing from its nominal value of 0V. From Fig.6.4, one can see that the input is split into two paths: the upper one controls the SPACE frequency, while the lower one controls the MARK frequency. Considering the lower path, it is clear that the error signal is added to the voltage set by the preset resistor. The function of the preset resistor is to set the nominal MARK frequency (exactly the same

function as the preset resistances that were originally used to set V_{c0} and V_{c1} in the level translation circuitry described in section 4.1.3). Note that the sense of the error signal is inverted because the op-amp is operating in the inverting summer configuration.

Since the output of the op-amp is bipolar, it has to be converted back to unipolar format (0V to 15V) as required by the inputs to the level translation circuitry. An inverting summer is used to do this, and the required mapping function is;

$$V_{out} = 7.5 - 0.5V_{in}.$$

The op-amp thus multiplies the input by -0.5 and adds this to a constant level of 7.5V. The 7.5V source is generated by amplifying the -5V source (set by the zener) by a gain of -1.5. Note that the sense of the error signal is again inverted and this results in the output of the lower path varying in-phase with the error signal. For zero voltage input to the module (i.e. for correct modulation index), the output of the lower path is dependent only on the voltage set by the preset resistance, which sets the nominal MARK frequency.

Looking at the upper path (the SPACE frequency path), it is apparent that it is identical to that of the lower path, except that the error signal is inverted (multiplied by -1) prior to being added to the voltage set by the preset resistance (which sets the nominal SPACE frequency). Considering the upper path as a whole then, we can easily see that its output varies in anti-phase with the error signal. For zero input error signal to the module, the output of the upper path is set by the preset resistance

which sets the nominal SPACE frequency.

It is clear that for a modulation index greater than the nominal value, the outputs of the module tend to narrow their difference, and for modulation indices less than the nominal value, the outputs tend to increase their difference. Were the outputs of the interface circuitry to be connected to the level translation circuitry, they would adjust the MARK and SPACE frequencies appropriately to return the modulation index to its correct value. The operation of the Servo Control Circuitry can be summarized as follows;

h > 0.5: SCC increases the MARK frequency and decreases the SPACE frequency i.e. h is reduced back to 0.5.

h < 0.5: SCC decreases the MARK frequency and increases the SPACE frequency i.e. h is increased back to 0.5.

6.3 THE PROBLEM OF FREQUENCY CREEPAGE

Referring to the circuitry of Fig.6.4, we can see that the gains of the MARK and SPACE frequency control paths were set to be equal. However, if we look at the Modulator VCO curve shown in Fig.4.9, it is clear that the gain of the VCO (in MHz/volt) is greater at the MARK than at the SPACE frequency. Were this not taken into account, one would find that the centre frequency of the VCO drifts away from its nominal position (70.144 MHz), even though the modulation index is maintained correctly by the Servo Control Circuitry.

From the VCO curve (Fig.4.9), the gain of the VCO at the MARK frequency is 0.427 MHz/volt, while at the SPACE frequency it is 0.178 MHz/volt. By a first order approximation then, we can set the gain of the SPACE frequency control path to be 0.427/0.178 = 2.40 times larger than that of the MARK frequency path, thus offsetting the gain differential due to the non-linearity of the VCO.

This can be seen mathematically as follows: considering the Taylor expansion for the VCO frequency about its set point of V volts (the nominal control voltage for either the MARK or SPACE frequencies) and the error voltage ϵ,

$$f(V + \epsilon) = f(V) + f'(V).\epsilon + f''(V).\epsilon^2/2! + \ldots$$

We ignore the higher order terms (containing the ϵ^2 and higher terms). It is clear that the Servo Control Circuitry corrects the the modulation index by means of a first-order approximation (only the first derivative of the VCO characteristic is considered). By making the SPACE frequency control path 2.4 times larger than that of the MARK frequency, we have compensated for the differences in the VCO gain at the two frequencies. The gain of the SPACE frequency control path is adjusted by changing the value of the input resistance to the inverter op-amp.

CHAPTER 7

THE NON-COHERENT DETECTION OF MSK

An advantage of MSK over SQPSK is that it may be coherently detected with the same error performance as BPSK under conditions of low SNR, as well as non-coherently when the SNR is reasonably high.

MSK may be non-coherently detected (by means of discriminator detection) by virtue of the fact that it can be regarded as a special case of CP-FSK with a modulation index of 0.5 (as was discussed in section 2.2.4(a)). In this chapter, we discuss the advantages and disadvantages of the non-coherent detection of MSK, and then proceed to describe the design and construction of such a receiver.

7.1 COMPARISON BETWEEN THE COHERENT AND NON-COHERENT DETECTION OF MSK

It has been previously shown (in section 2.2.4) that for coherent, matched filter MSK demodulation, the ideal error probability is given by;

$$P_e = Q(\sqrt{2\mu})$$

where μ is the average energy per bit-to-double sided noise density ratio. This is identical performance to BPSK and SQPSK on infinite bandwidth, linear channels. The demodulator implementation is, however, relatively complex. If one were to be operating under conditions which do not require the full advantage of MSK's error performance (i.e. under relatively noise-free conditions), discriminator

detection may be implemented, resulting in considerably simpler receiver circuitry. By disregarding the phase information available to us, a degradation in noise performance must be expected. Because MSK is an orthogonal signaling scheme (see section 2.2.4(b)), the probability of error is given by the expression [7.1];

$$P_e = \tfrac{1}{2}e^{-\mu/2}.$$

This is more than 3dB worse in noise performance than with coherent detection [7.2]. Note however, that the spectral advantages of MSK (compact bandwidth, constant envelope etc.) are retained. It is possible to construct non-coherent CP-FSK receivers which outperform antipodal signaling [7.3], but their description is beyond the scope of this discussion.

The block diagram of a non-coherent MSK receiver is shown in Fig.7.1.

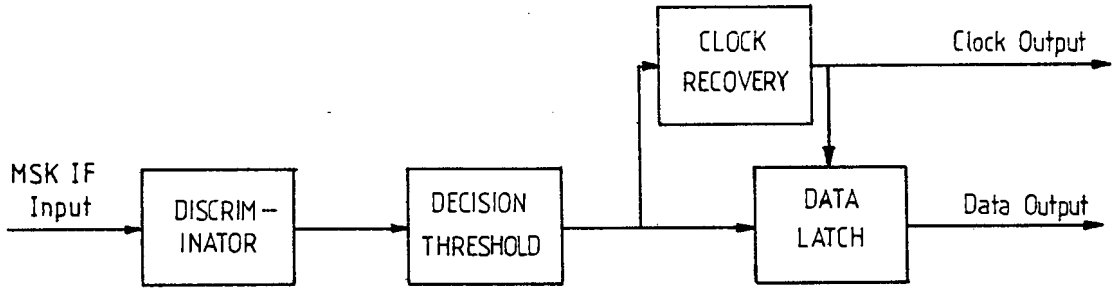


Fig.7.1 Block Diagram of a Non-Coherent MSK Receiver

Note that only the IF modules after downconversion, amplification etc. are shown. The MSK input (as generated by the modulator described in Chapter 4, but without differential encoding) is detected by a frequency discriminator which produces a negative polarity unipolar NRZ output waveform. This waveform is then passed through a decision threshold device which converts it into TTL, NRZ format. From the resultant waveform, the bit rate clock is recovered, and this is used to clock a latch which samples the waveform at the decision threshold output at maximum eye-opening. The data latch output is the recovered data stream. The design of the non-coherent MSK demodulator is now described.

7.2 FREQUENCY DISCRIMINATOR MODULE

It is necessary for the discriminator to have a rapid response time (i.e. be capable of detecting wide bandwidth modulation) as it is required to detect the 2 MHz modulation present on the received signal.

The discriminator used is the Balanced Quadrature Detector [7.4]. It is capable of rapid response times, and its performance for 2 MBps data rates was found to be satisfactory. The circuit diagram is shown in Fig.7.2.

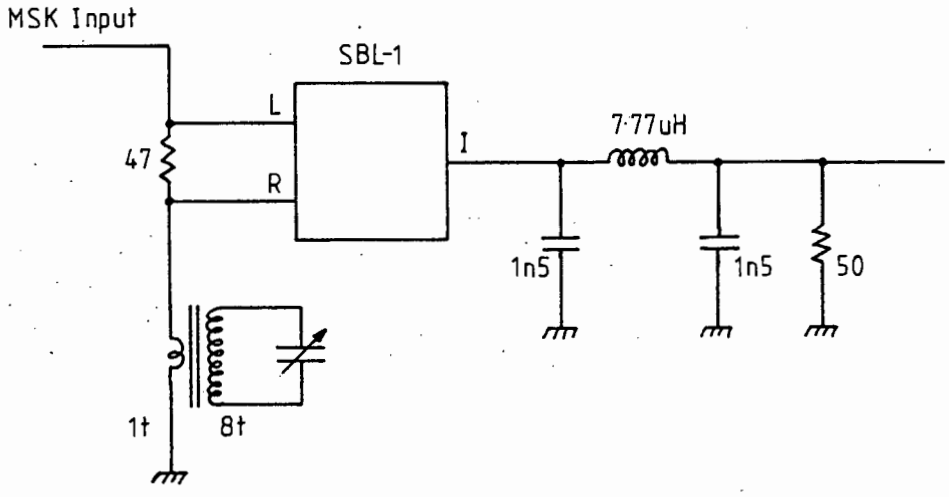


Fig.7.2 Frequency Discriminator Module

The MSK input is connected directly to the L-port of an SBL-1 double balanced mixer. It is also coupled to the R-port of the mixer via a 47Ω resistance. Connected between the R-port and ground is a parallel LC tuned circuit which has a centre frequency of 70.144 MHz. As the input frequency varies (in sympathy with the modulating data stream), the currents in the L- and R-ports of the mixer differ in phase, due to the presence of the LC circuit at the R-port. These phase differentials are detected by the mixer (which acts a phase detector) and the I-port of the mixer is the frequency detected data stream.

Because the LC circuit is connected to the R-port of the mixer (which presents a load impedance of 50Ω), an impedance transformer is necessary to prevent Q degradation (this would result in the loss of sensitivity of the discriminator, as the gain slope is reduced by a smaller Q).

An 8:1 transformer is used, and the resultant load presented to the tuned circuit is;

$$R_L = (8)^2.50$$

$$\approx 3.2\ K\Omega.$$

This high value of load impedance ensures that the Q of the circuit is not too low. The variable capacitor is used to tune the centre frequency of the discriminator to be 70.144 MHz.

The phase detector output, taken at the I-port of the mixer is lowpass filtered to reduce the noise bandwidth of the discriminator, as well as to remove the 70 MHz feedthrough from the mixer L-port. Chen [7.5] produces interesting results on choosing such filtering. It is shown in his paper that using a lowpass filter with a bandwidth equal to the bit rate can significantly improve the noise performance of the demodulator. The results were obtained for a modulation index of 0.71 (not directly applicable here), and for wide IF bandwidths (which is the case here).

Using a 3-pole Butterworth lowpass filter of 3 dB bandwidth 2.048 MHz, the following element values are obtained;

$$C_1 = C_2 = 1.55\ nF$$
$$L_1 = 7.77\ \mu H$$
$$R_{source} = R_{load} = 50\Omega.$$

The measured transfer function of the Discriminator module is shown in Fig.7.3.

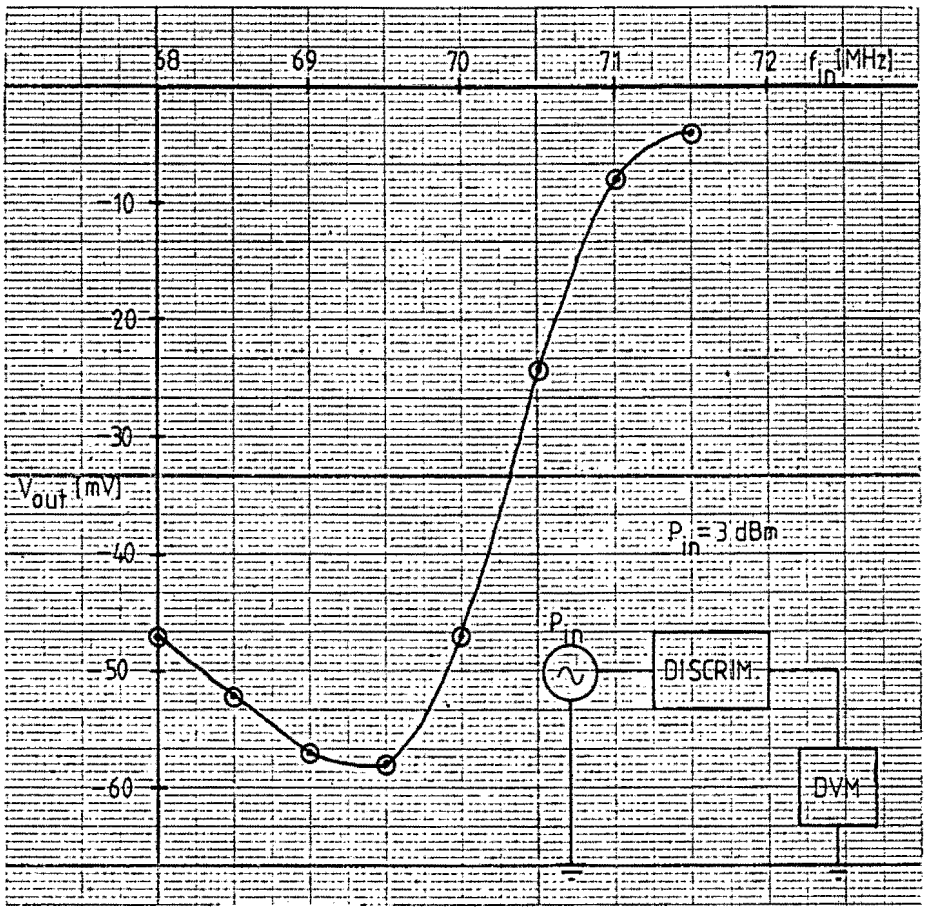


Fig.7.3 Frequency Discriminator Transfer Function

Note that the curve is approximately linear about its centre frequency and that the bandwidth is approximately 1 MHz, which is exactly what is required as the MARK and SPACE frequencies are spaced 1.024 MHz apart. If the measured bandwidth had been too wide, it could have been reduced by increasing the turns ratio of the LC circuit coupling transformer, thereby increasing the tank circuit Q. If the bandwidth had been too narrow, it could have been widened by reducing the tank circuit Q (by connecting a shunt resistance across it).

As can be seen from Fig.7.3, the output voltage of the discriminator is negative in polarity and varies from ≈-55mV to ≈-5mV. Fig.7.4 shows the eye diagram output of the discriminator module.

Fig.7.4 Eye Diagram of Discriminator Module Output
(Horizontal scale: 0.2μs/div)

7.3 DECISION THRESHOLD MODULE

The decision threshold module is used to convert the small-amplitude waveform at the Discriminator Module output into unipolar NRZ (TTL) format. Fig.7.5 shows the circuit diagram.

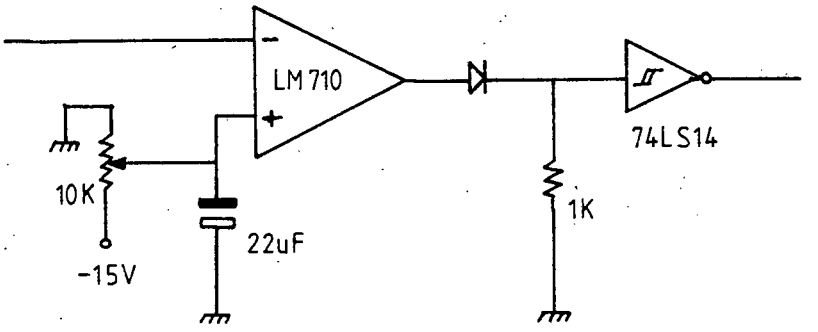


Fig.7.5 Decision Threshold Module

It is not possible to capacitively couple the output of the discriminator module and then use the voltage comparator as a zero-crossing detector (which may be tempting to do as one would like to eliminate any DC offsets present on the detected waveform). This is due to the fact that the input waveform to the comparator is in unipolar NRZ format and thus contains spectral components at, and near DC. By AC coupling the two modules, significant amounts of signaling power will be lost, resulting in degraded performance. The 10KΩ preset potentiometer is used to adjust the decision threshold level of the comparator, which optimally lies at the half-amplitude level of the input waveform (for equiprobable MARKs and SPACE's).

The output of the LM710 voltage comparator varies between ≈5V and ≈-1V, which is not TTL compatible. The diode-resistor network is used to rectify this waveform and make

it TTL compatible. The 74LS14 Schmitt trigger is used to buffer the module output for the circuitry that follows. The signal at this point in the circuitry is used for two functions;

- It is used by the clock recovery circuitry to regenerate the bit-rate clock;

- It is sampled at maximum eye-opening by means of a data latch (which is triggered by the recovered clock), and the resultant waveform is the recovered data stream.

The spectrum of the data waveform after the decision threshold is shown in Fig.7.6.

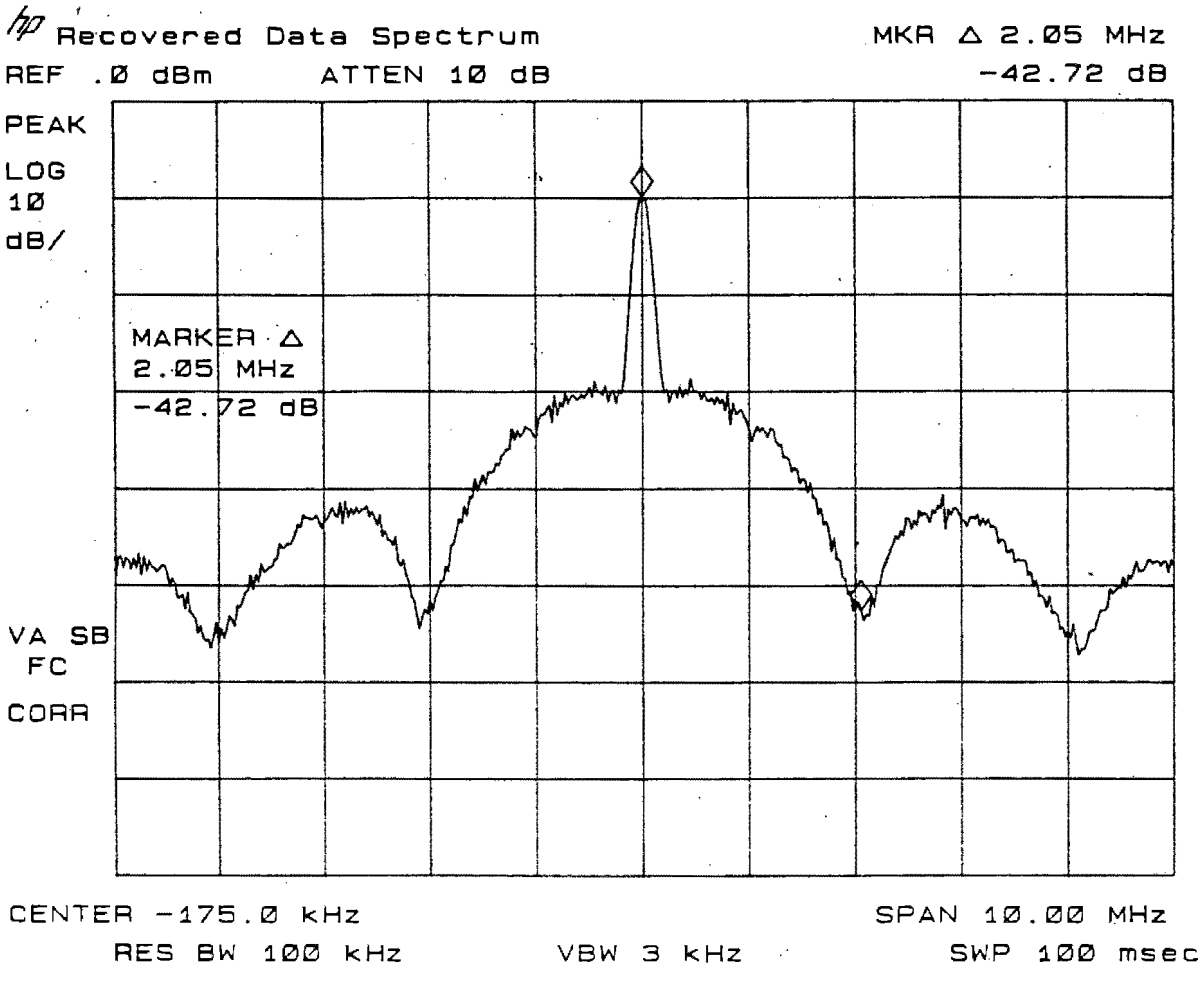


Fig.7.6 Power Spectrum of Decision Threshold Output

7.4 CLOCK RECOVERY MODULE

Observing Fig.7.6, it is clear that no discrete spectral components are present in the spectrum available for clock recovery. This is to be expected, as the spectrum shown is that of a unipolar NRZ random data stream, which contains no deterministic spectral components (the nulls of the spectrum

lie on the clock frequency and multiples thereof). In order to regenerate a deterministic component, a non-linear signal processing operation has to be applied to the waveform.

A widely known method [7.6] of regenerating the clock line is shown in the block diagram in Fig.7.7 below.

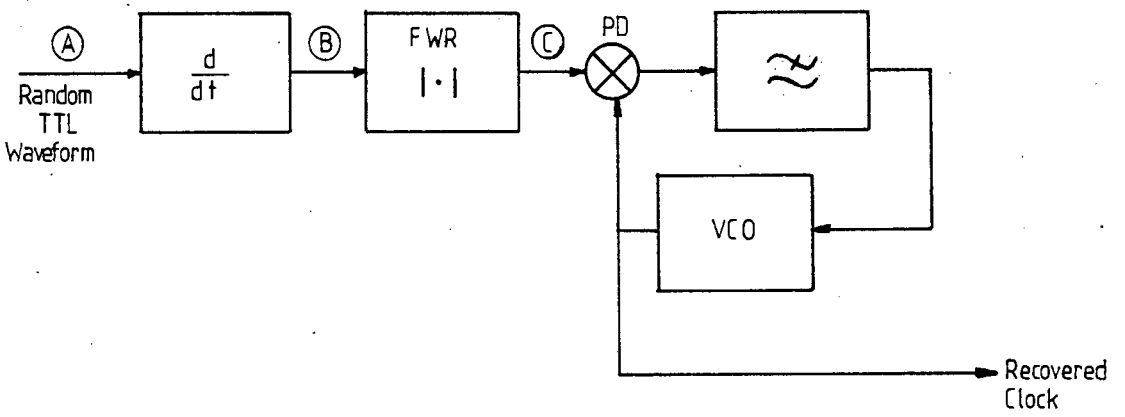


Fig.7.7 Clock Recovery Module Block Diagram

The time domain waveforms for points A, B, and C are shown in Fig.7.8.

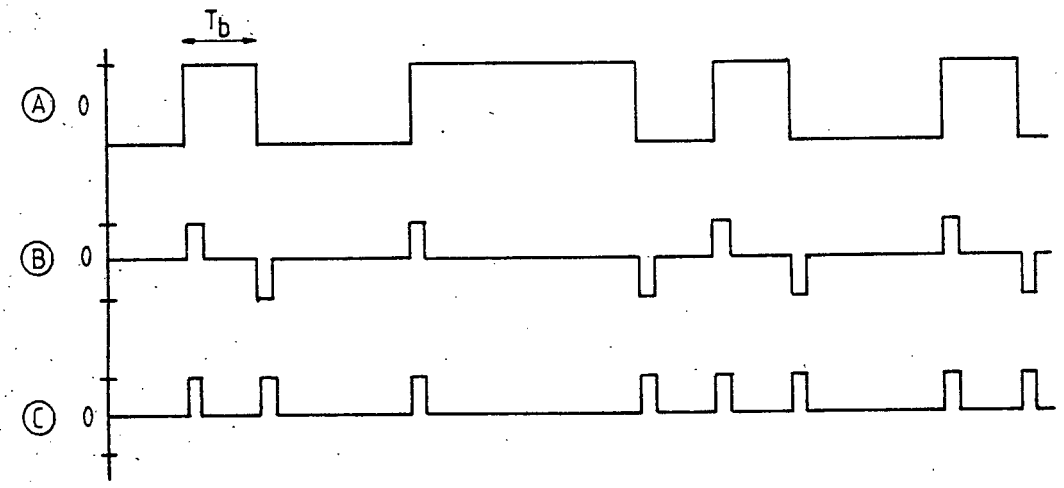


Fig.7.8 Waveforms at Points A, B, and C in Fig.7.7

Fig.7.8A shows a segment of a random TTL waveform. After differentiation, it appears as shown in Fig.7.8B. Full wave rectification produces a unipolar waveform as shown in Fig.7.8C. Note that the waveform in C may be heuristically shown to be the sum of the two waveforms shown in Fig.7.9.

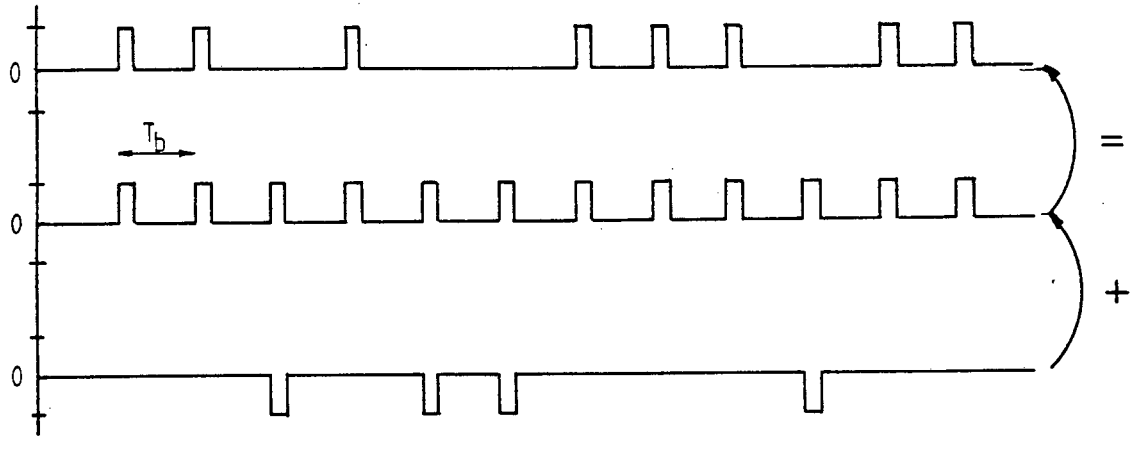


Fig.7.9 Decomposition of a Random Waveform into a Random and Deterministic one

This clearly shows how the random RZ waveform may be decomposed into the sum of a deterministic waveform (the clock waveform), and a random waveform. It is the deterministic component which regenerates the clock line. The waveform at point C in Fig.7.7 is then fed to the Phase-Locked Loop (PLL) which acts as a narrow bandpass filter to extract the clock line at a high carrier to noise ratio.

Each module of which the Clock Recovery Module is comprised is now described.

7.4.1 Differentiator Network

The differentiator network and full-wave rectifier may be implemented by means of a passive RC highpass filter followed by a diode rectifier circuit. This implementation has the problem of the diode drops being present, as well as the low output signal level of the differentiator network (the output of the differentiator may not be enough to overcome the diode forward voltage drops). A more elegant implementation is the time domain differentiator network. This is easily constructed using standard TTL gates, and has the advantage of high output signal levels (5V), which will be necessary in later circuits. The circuit diagram is shown in Fig.7.10.

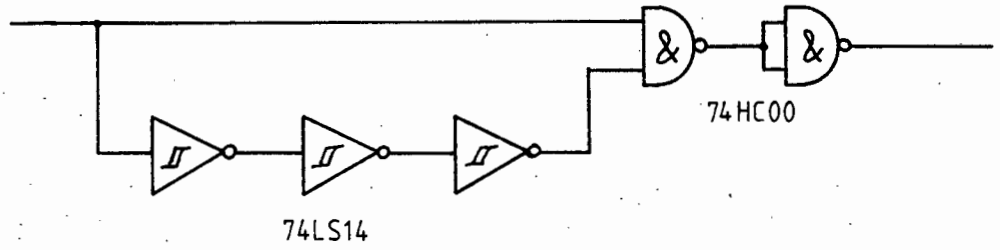


Fig.7.10 Differentiator Network

The circuit is, in fact, a rising edge detector as it produces a narrow positive pulse only on the rising edges of the input waveform. Note that on average, half the clocking information available has been discarded (no pulses are produced for the falling edges of the input waveform). In practice, this was found to be of no disadvantage, and a rising- and falling-edge detector is easily implemented by means of two inverters and an EXOR gate if so desired [7.4]. Wider output pulses are achievable by using slower gates (such as 74C04 inverters) in the delay chain. Wider pulse widths may be desired if greater output amplitude of the clock lines is desired (the pulse widths determine the spectral amplitudes of the clock lines). Fig.7.11 shows the output power spectrum of the differentiator network.

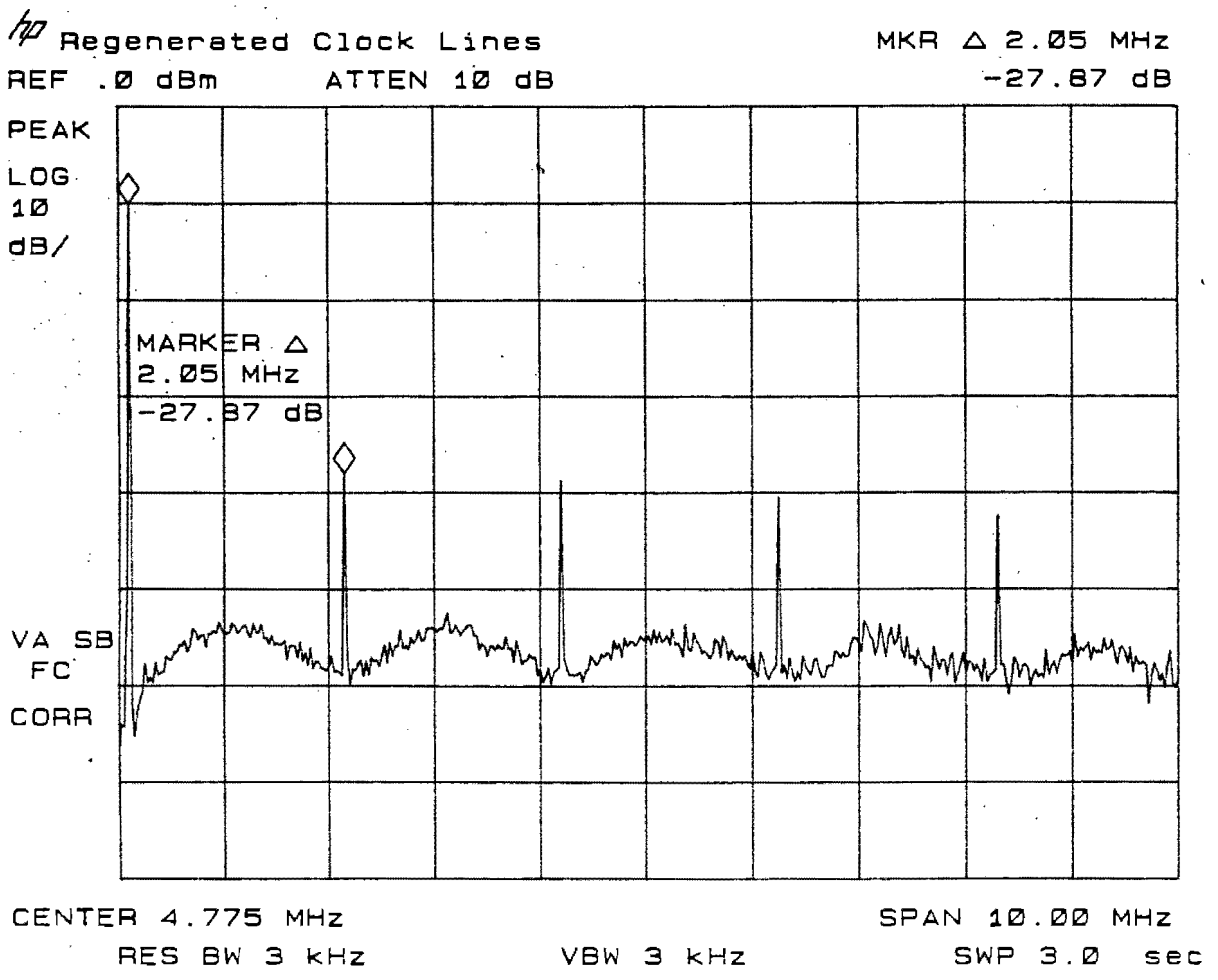


Fig.7.11 Regenerated Clock Lines

The discrete spectral lines at multiples of the clock frequency are clearly seen.

7.4.2 Bandpass Filter And Amplifier

Observing Fig.7.11, it may be tempting to apply the PLL directly to the output of the differentiator network, as the

spectrum implies that a clock frequency is present in the waveform, and may be extracted by a PLL.

However, the PLL to be used (to be described in section 7.4.3) has a Phase-Frequency Detector (PFD) at its input. This particular detector was chosen because, amongst other reasons, of its immunity to the exact duty cycle of the input waveform. The operating principle of the PFD is that it detects the phase difference between the edges of the input and VCO waveforms to produce an error signal [7.4]. The PFD is, as a result, sensitive to missing edges in the input waveform. If the waveform has too many missing edges (which occurs often in random waveforms), the PLL will not lock as the output of the PFD is not usable.

This dilemma is resolved by applying the random pulses to a high-Q tank circuit (in effect, a bandpass filter) prior to applying them to the PLL. The effect of the tank circuit (which is centered at the bit rate frequency of 2.048 MHz) is to ring at its natural frequency after being excited by an input pulse. The Q of the circuit is chosen such that the ringing oscillations persist for the longest expected constant data block length of the data stream. The oscillations add extra zero crossings to the waveform, which enables the PLL to lock reliably. The bandpass filter is shown in Fig.7.12.

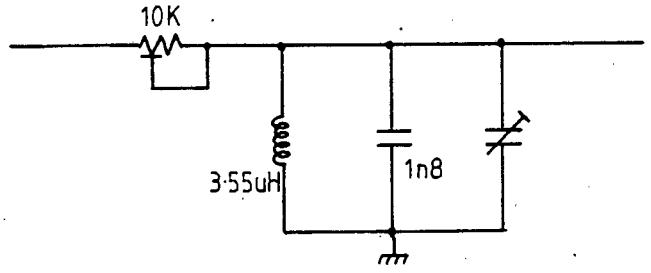


Fig.7.12 Bandpass Filter Circuit

The low output impedance of the TTL circuitry is coupled to the filter via a 10KΩ potentiometer. By varying the resistance, the Q of the circuit as well as the output amplitude may be varied. For a resonant frequency of 2.048 MHz, the inductance is 3.55 μH, and the capacitance is 1.8nF. A trimmer capacitor is used to centre the filter at the correct frequency.

The output amplitude of the bandpass filter is small, and some amplification is necessary in order to present a high enough signal level to the PLL. Measurements made indicated that a gain of approximately 30 was necessary. The amplification is provided by a simple common emitter amplifier, as is shown in Fig.7.13.

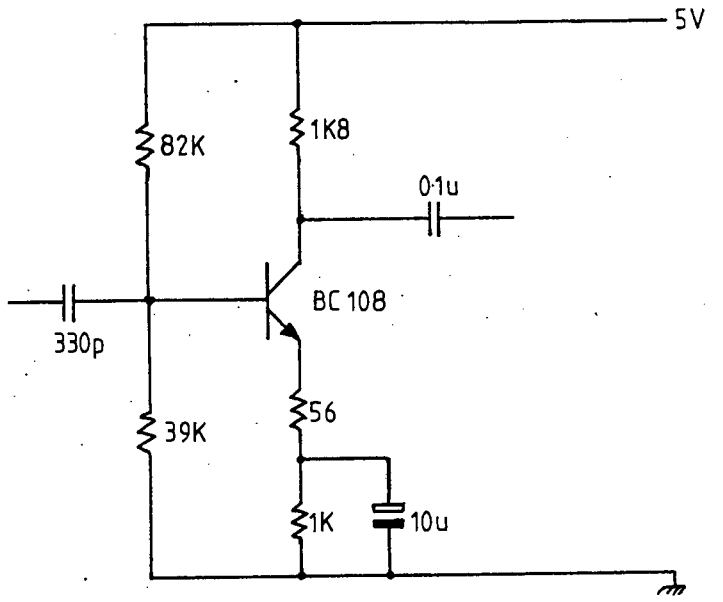


Fig.7.13 Common Emitter Amplifier Circuit

The output of the amplifier has an amplitude of approximately 300mV and this is sufficient to drive the input of the PLL. Fig.7.14 shows clearly the action of the bandpass filtering and amplification on the spectrum of Fig.7.11.

Note how the 2.048 MHz clock line has been amplified and the clock harmonics suppressed. It is onto this spectral line that the PLL locks.

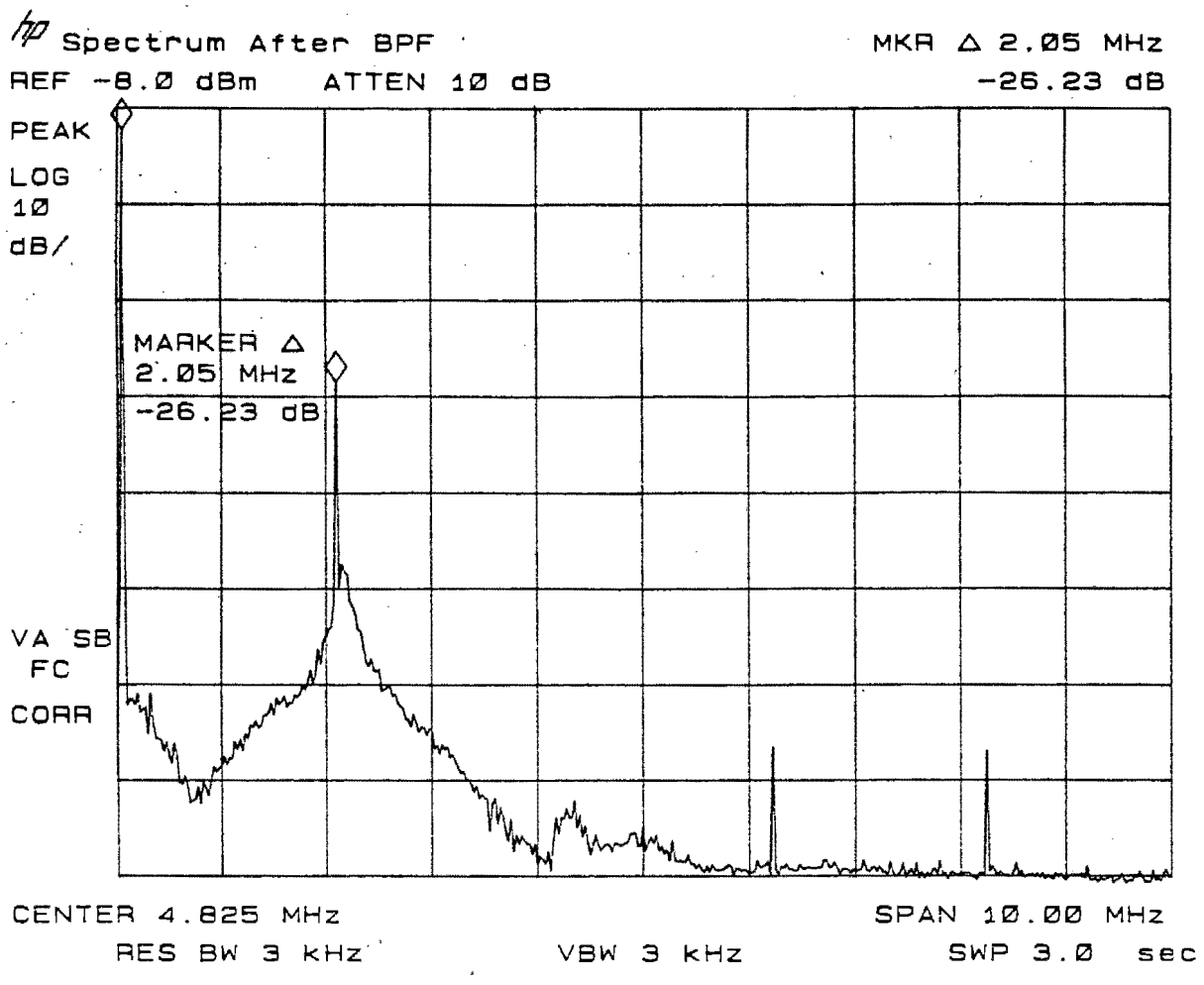


Fig.7.14 Power Spectrum After Bandpass Filtering and Amplification

7.4.3 Phase-Locked Loop

As the frequency at which the PLL must operate (2 MHz) is beyond that of conventional CMOS PLL circuits (CD4046), a high speed CMOS PLL, the 74HC4046, is used. The upper

frequency limit of its operation is 20 MHz.

7.4.3(a) Input Interfacing. The input to the PLL phase detector may either be of TTL levels (these are directly coupled), or small signal levels (≈ 200mV) which are AC coupled to the input. The small signal levels are amplified and hard-limited by an onboard self-biased amplifier. The hard-limiting of the input waveform is a desirable feature as this eliminates any AM present. It is shown in [7.7] that the AM noise is predominantly in the low frequency portion of the spectrum, and is thus passed relatively unattenuated by the loop filter. This may result in excess phase noise in the PLL output.

7.4.3(b) Choice of Phase Detector. The 74HC4046 has three onboard phase detectors from which to choose, and as was mentioned previously, the Phase-Frequency Detector is used as it is insensitive to the duty cycle of the input waveform. Another advantage is that it has a wider phase detection range than the other two PD's (the EXOR, and the RS flip-flop PD's). The use of this detector also results in less phase noise on the PLL output as it only produces an output when the edges of the input signal and the VCO are not aligned (there is consequently less frequency modulation of the VCO resulting in lower phase noise). The other PD's, by nature of their operation, continually perturb the VCO frequency even when the PLL is locked.

As the PFD operates linearly over the phase differential range of $(0 , 2\pi)$, and the maximum output voltage is +5V, the Phase Detector Gain Constant K_d is calculated to be;

$$K_d = 5/2\pi = 0.8 \quad [V/rad].$$

7.4.3(c) Configuration of the VCO. To avoid false locking of the PLL, the frequency excursion of the VCO must be limited to only the frequencies of interest. The lower frequency limit is set at 1.6 MHz, and the upper limit at 2.4 MHz. Using the design curves for the PLL, we obtain;

$$R_2 = 10\ K\Omega$$
$$C_1 = 150\ pF.$$

We next compute $f_{max}/f_{min} = 2.4/1.6 = 1.5$. Hence from the design curves,

$$R_2/R_1 = 0.7$$

$$\Rightarrow \quad R_1 = 14.3\ K\Omega.$$

Fig.7.15 shows the VCO circuit diagram.

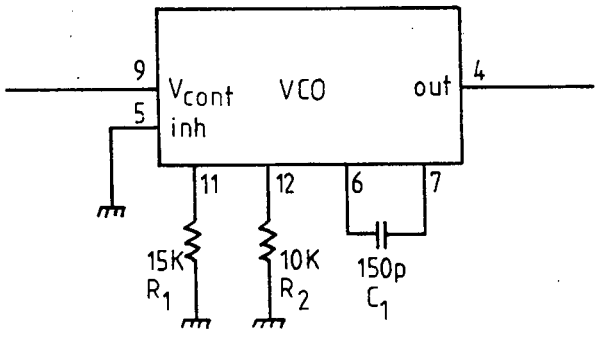


Fig.7.15 VCO Circuit Diagram

The VCO gain constant K_o is calculated from;

$$K_o = 2\pi(f_{max} - f_{min})/(V_{max} - V_{min}),$$

where V_{max} and V_{min} are the maximum (5V) and minimum (0V) control voltages applied to the VCO.

$$\text{i.e.} \quad K_o = 1.0\ E6 \quad [rad/s/V]$$

7.4.3(d) Design of Loop Filter. The loop filter performs the following functions;

- It restricts the noise bandwidth of the PLL (we would like to have a small loop bandwidth to limit the amount of phase noise on the recovered clock).

- It determines the stability of the PLL.

The loop can be made stable by ensuring that the PLL open loop transfer function approaches the 0 dB gain point at a rate of -6 dB/octave or less. The open loop transfer function is shown in Fig.7.16.

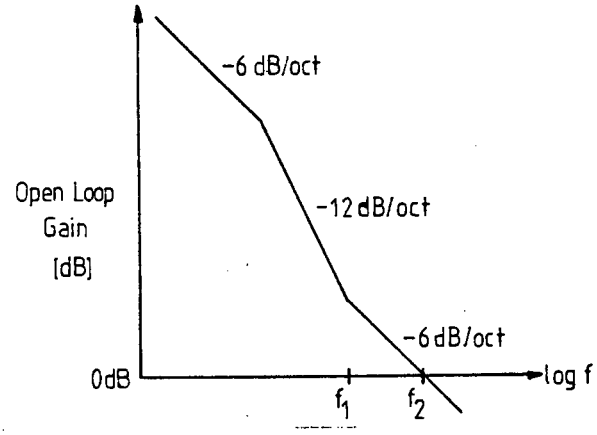


Fig.7.16 PLL Open Loop Transfer Function

The initial -6 dB/octave rolloff is due to the presence of the VCO in the loop (it is an integrator, as was discussed in section 5.2.1(a)). A pole (inserted by the loop filter) then increases the slope to -12 dB/octave. This is reduced back to -6 dB/octave by the loop filter zero at a frequency f_1 [Hz]. The gain slope continues to descend at this rate and the unity gain frequency is reached at a frequency f_2. Provided that f_2 is chosen to be 3 to 5 times f_1, sufficient phase margin will be provided, and the loop will be stable [7.4]. The loop filter that is used is the Lag-Lead filter, as is shown in Fig.7.17.

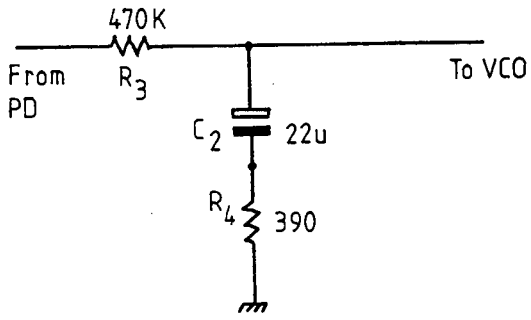


Fig.7.17 Loop Filter Circuit

The pole occurs at a frequency $1/[2\pi(R_3 + R_4)C_2]$, and the zero at a frequency $f_1 = 1/[2\pi R_4 C_2]$. Before we can calculate the zero frequency, the unity gain frequency f_2 must be specified. A value of 100 Hz is chosen, as this results in good phase noise performance. The zero frequency is thus chosen to be one-fifth this i.e. f_1 = 20Hz. The choice of R_3 is critical, as it determines the loop gain at the frequency f_2 (the loop gain must be unity at the frequency f_2).

Choosing the time constant R_3C_2 = 10s, and using the calculated values for K_o and K_d, the loop gain is calculated to be 1.0 at a frequency of 100 Hz i.e. the chosen value of R_3C_2 ensures loop stability. We can now specify C_2 = 22μF, and hence;

$$R_3 = 470\ K\Omega$$

and

$$R_4 = 390\ \Omega.$$

The actual constructed PLL using these component values was found to be stable, and frequency acquisition was rapid (< 5s) and reliable. The recovered clock spectrum is shown in Fig.7.18.

Using the method of phase noise measurement described in section 5.2.3, the Single Sideband Noise-to-Carrier Power ratio is calculated for a 10 kHz offset to be;

$$\mathcal{L}(10\ kHz) = -54.67 - 10.\log[300] + 2.5$$
$$= -76.94\ \ dBc/Hz.$$

This is indicative of a high quality recovered clock.

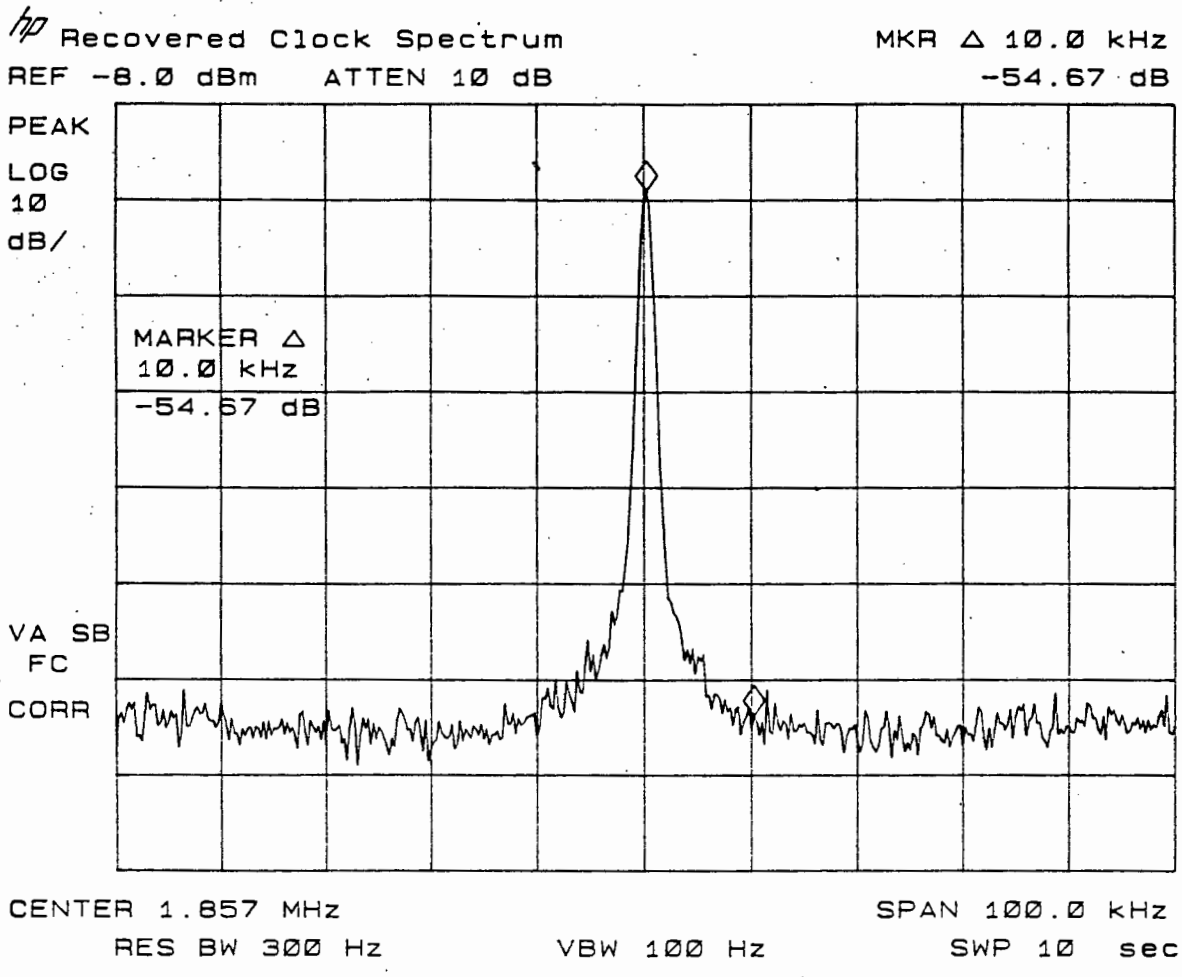


Fig.7.18 Recovered Clock Spectrum

The rising edge of the recovered clock (which will be used to trigger the data latching circuitry) is at a fixed, but arbitrary offset from the centre of the eye-opening of the data waveform at the decision threshold output. For optimum performance, the rising edge must occur at the maximum eye-opening. It is for this reason that a monostable (74LS123) is used to adjust the rising edge of

the recovered clock with respect to the data eye-opening. The circuit to perform this static phase adjustment is shown in Fig.7.19.

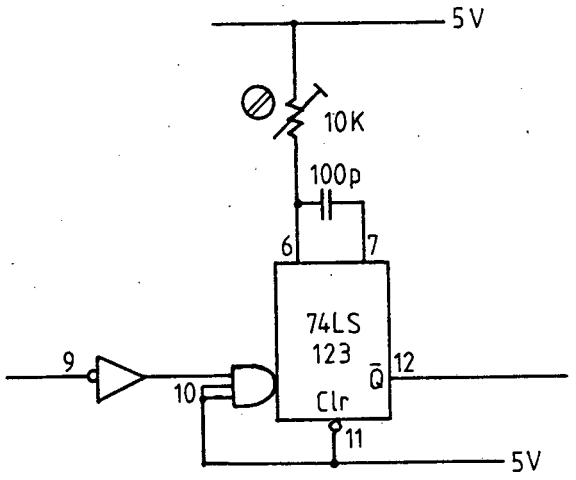


Fig.7.19 Static Phase Adjustment Circuit

The 10 KΩ potentiometer varies the rising edge of the monostable output for optimum performance.

7.5 DATA LATCHING MODULE

The function of this module is to use the recovered clock to latch the random data stream at the point of maximum eye-opening. This function is performed by a 74LS74 D-type flip-flop. Fig.7.20 shows the circuit.

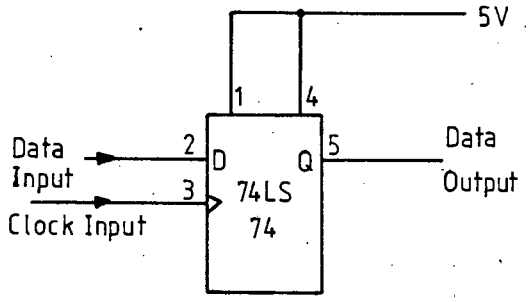


Fig.7.20 Data Latch Module

The operation of this circuit is as follows: The random data stream is applied to the D input of the flip-flop, and it is latched at maximum eye-opening by the rising edge of the recovered clock. The output of this module, in conjunction with the recovered clock, may be used to reconstruct the bit stream back into parallel format etc.

A photograph of the constructed non-coherent MSK demodulator is shown in Fig.7.21.

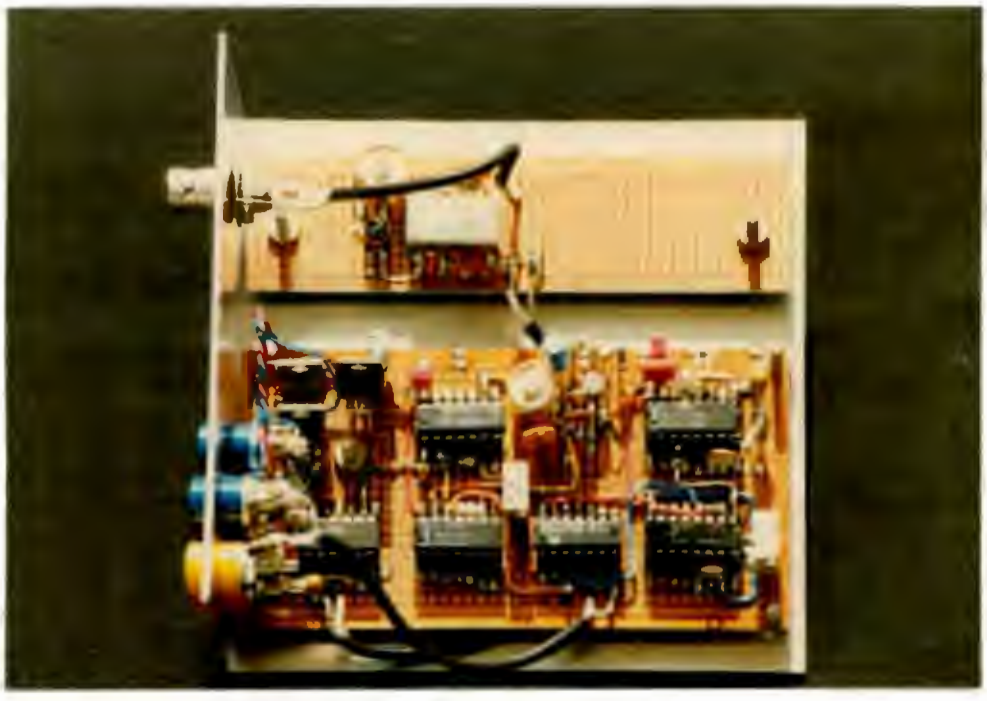


Fig.7.21 Constructed Non-Coherent MSK Demodulator

REFERENCES

[7.1] Lathi, B. P., "Modern digital and analog communication systems", Holt-Saunders, New York (1983).

[7.2] de Buda, R., "Coherent demodulation of frequency-shift keying with low deviation ratio", IEEE Trans. on Comm., Vol. COM-20, June 1972, pp.429-435.

[7.3] Osborne, W. P. and Luntz, M. B., "Coherent and non-coherent detection of CP-FSK", IEEE Trans. on Comm., Vol. COM-22, Aug. 1974, pp.1023-1036.

[7.4] Horowitz, P. and Hill, W., "The art of electronics", 8th ed., Cambridge University Press, New York (1984).

[7.5] Chen, C. H., "Discriminator detection of wide-band PCM/FM", IEEE Trans. on Aerospace and Elect. Systems, Vol. AES-5, Jan. 1969, pp.126-127.

[7.6] Gardner, F. M., "Phaselock techniques", 2nd ed., Wiley, New York (1979).

[7.7] Gardner, F. M., "Self-noise in synchronizers", IEEE Trans. on Comm., Vol. COM-28, Aug. 1980, pp.1159-1163.

CHAPTER 8

CONCLUSIONS

In view of the findings of this thesis, the following conclusions may be drawn;

1. MSK is a bandwidth efficient modulation scheme (having a maximum spectral efficiency of 2 Bits/s/Hz) and a has an error probability identical to that of antipodal signaling.

2. Due to the time offset between the I- and Q-Channels, MSK exhibits less sidelobe regeneration than QPSK after bandlimiting and hardlimiting.

3. The CP-FSK type MSK modulator implementation is simple to implement and its performance is satisfactory.

4. The coherent MSK demodulator is relatively complex to implement, and requires several complex modules.

5. The effect of an incorrect modulation index on the error probability of MSK is that it sets the minimum attainable error probability, independent of signal-to noise ratio.

6. The non-coherent MSK demodulator is much simpler to implement than the coherent demodulator, but its noise performance is more than 3dB inferior to that of the coherent demodulator.

APPENDIX A

THE EFFECT OF SQUARING ON THE MSK SPECTRUM

The effect of squaring (Frequency Doubling) on the MSK spectrum is derived mathematically here. It will be shown by simple mathematical means how each of the features of the frequency doubled MSK spectrum arise.

In the calculations that ensue, the following trigonometric identities are used;

(i) $\cos^2 \theta = \frac{1}{2}[1 + \cos 2\theta]$

(ii) $\sin^2 \theta = \frac{1}{2}[1 - \cos 2\theta]$

(iii) $\sin \theta . \cos \theta = \frac{1}{2}\sin 2\theta$

(iv) $\cos \alpha . \cos \beta = \frac{1}{2}\cos (\alpha + \beta) + \frac{1}{2}\cos (\alpha - \beta)$

The MSK time-domain expression is;

$$g_{MSK}(t) = y_I(t)\cos (\pi t/2T_b).\cos (2\pi f_c t) - y_Q(t)\sin (\pi t/2T_b).\sin (2\pi f_c t).$$

For brevity, we let $y_I(t) = y_I$, and $y_Q(t) = y_Q$. Also, $\pi/2T_b = 2\pi\Delta f = \Delta w$, and $2\pi f_c = w_c t$.

$$\Rightarrow g_{MSK}(t) = y_I \cos \Delta wt.\cos w_c t - y_Q \sin \Delta wt.\sin w_c t.$$

Squaring the MSK waveform yields;

$$g^2_{MSK}(t) = y^2_I\cos^2 \Delta wt.\cos^2 w_c t + y^2_Q\sin^2 \Delta wt.\sin^2 w_c t$$

$$- 2y_I y_Q \sin \Delta wt.\cos \Delta wt.\sin w_c t.\cos w_c t$$

Using identities (i), (ii), and (iii), we may transform the expression to;

$$g^2_{MSK}(t) = \tfrac{1}{4}.y^2_I.[1 + \cos 2\Delta wt][1 + \cos 2w_c t]$$

$$+ \tfrac{1}{4}.y^2_Q.[1 - \cos 2\Delta wt][1 - \cos 2w_c t]$$

$$- \tfrac{1}{2}.y_I.y_Q.\sin 2\Delta wt.\sin 2w_c t$$

Noting that $y^2_I(t) = y^2_Q(t) = 1$ (as these symbol streams are in polar NRZ format with a normalized amplitude of 1 Volt), we obtain;

$$g^2_{MSK}(t) = \tfrac{1}{4}[2 + 2\cos 2\Delta wt.\cos 2w_c t]$$

$$- \tfrac{1}{2}.y_I.y_Q.\sin 2\Delta wt.\sin 2w_c t$$

Using identity (iv), this expression reduces to;

$$g^2_{MSK}(t) = \tfrac{1}{2}.[1 + \tfrac{1}{2}.\cos 2(w_c + \Delta w)t + \tfrac{1}{2}.\cos 2(w_c - \Delta w)t]$$

$$- \tfrac{1}{4}.y_I.y_Q.[\cos 2(w_c - \Delta w)t - \cos 2(w_c + \Delta w)t]$$

From this expression, the following features can be noted;

(a) The squared spectrum contains a DC term ($\tfrac{1}{2}$).

(b) The term $\tfrac{1}{2}.\cos 2(w_c + \Delta w)t$ represents a deterministic spectral component at a frequency of twice the SPACE frequency.

(c) The term $\frac{1}{2}.\cos 2(w_c - \Delta w)t$ is a similar spectral component at a frequency of twice the MARK frequency.

(d) The last term is a random term as it contains the random product $y_I(t).y_Q(t)$. It is responsible for the generation of the random (noise-like) spectrum at the output of the frequency doubler. To ascertain its spectral properties, we set $y_I(t).y_Q(t) = r(t)$. This reduces the last term to;

$$\frac{1}{4}.r(t).\cos 2(w_c - \Delta w)t - \frac{1}{4}.r(t).\cos 2(w_c + \Delta w)t.$$

This is immediately recognized as the difference of two BPSK waveforms, each with the same pseudo-bit rate of f_b bits/s. Recall that the I- and Q-Channel symbols are $2T_b$ in length, but staggered by T_b seconds. Their product can therefore change at a rate of $1/T_b = f_b$ times per second (even though each symbol can only change state every $2T_b$ seconds). The BPSK power spectrum has a si^2 (f) envelope, with the first null spaced f_b Hertz from the centre frequency. Note that one BPSK spectrum is centered at twice the MARK frequency, while the other is centered at twice the SPACE frequency.

Once these features are recognized, it is easy to sketch what the squared MSK spectrum looks like. Fig.A.1 shows the expected appearance of the spectrum.

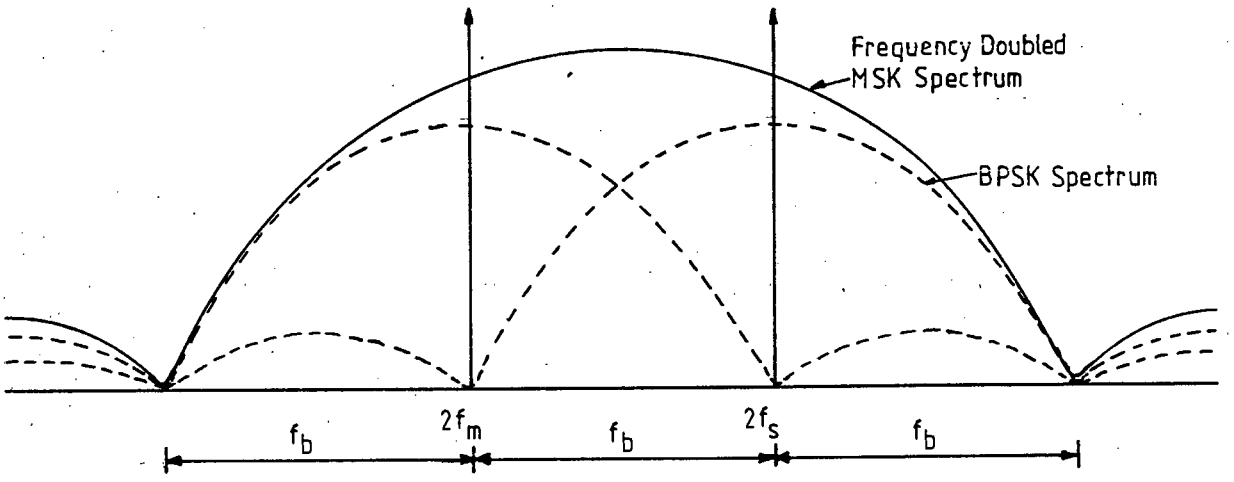


Fig.A.1 Mathematically Derived Spectrum
of Frequency Doubled MSK

Note how this compares very favourably with the measured spectrum shown in Fig.5.3.

APPENDIX B

AN ANALYTICAL INVESTIGATION INTO THE ERROR PERFORMANCE OF MSK WITH INCORRECT MODULATION INDEX

As this is an analytical investigation, some simplifying assumptions regarding the MSK modem have to be made in order to simplify the analysis. The assumption are;

- The recovered carriers (f_m, f_s, $2f_m$, $2f_s$) at the demodulator are assumed to be at the correct frequencies (i.e. they are not recovered from the received carrier). The MARK and SPACE frequencies of the modulator are incorrect and this results in an incorrect modulation index. As the carriers at the receiver are, in practice, extracted from the received MSK signal, the results presented in the analysis are not exact and this must be borne in mind when interpreting them.

- The modulator VCO is assumed to be locally linear about its centre frequency. This ensures that the phase accumulated during a MARK is the same as that accumulated during a SPACE (even though the phase is no longer correct).

- The maximum "0" or "1" block length is assumed to be 6 bits.

Consider the MSK constellation diagram shown in Fig.B.1.

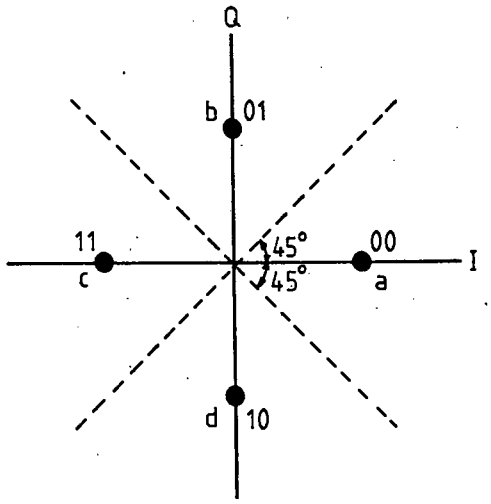


Fig.B.1 MSK Constellation Diagram

The symbols are denoted a, b, c, and d. With the modulation index at its correct value (h = 0.5), the phase changes over the bit intervals T_b are given by;

$$\Delta\theta = \pm\pi h$$

$$= \pm\pi/2.$$

Provided that the constellation points lie within their ±45° quadrants at the decision instants, no symbol error will occur (with no noise present, this will always be the case when h = 0.5).

When the modulation index is incorrect, the phase imparted during a bit time is;

$$\Delta\theta = \pm\pi(0.5 + \delta h)$$

$$= \pm\pi/2 \pm \pi\delta h$$

$$= \pm\pi/2 \pm \Phi_e,$$

where δh is a small positive number and Φ_e is the phase error after a single bit period. It is clear that the phase states of the modulator will no longer be at the points a, b, c, and d at the times T_b, $2T_b$, $3T_b$ etc. Fig.B.2 shows the position of the phase state at $4T_b$ after the bit sequence 0000, and starting at $\theta = 0°$.

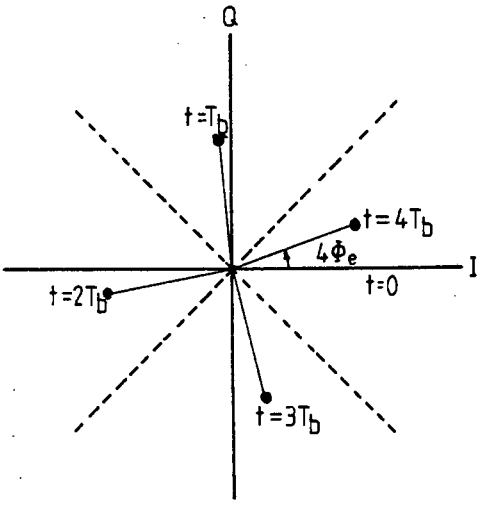


Fig.B.2 Accumulation of Phase Errors
for the Bit Sequence 0000

It is clear that for the chosen value of Φ_e and sequence length, no symbol error will result during, or at the termination of the sequence. This is because the symbols lie within their correct decision boundaries at the decision instants. It should also be apparent that for a given Φ_e there exists a sequence length with calculable probability of occurrence which will eventually result in a symbol

error.

Considering a bit sequence of length four bits, we list all 16 possible sequences;

0000	0001	0010	0011
0100	0101	0110	0111
1000	1001	1010	1011
1100	1101	1110	1111

Fig.B.3 List of All Possible Sequences of Length Four Bits

Because of the assumption that the VCO is at least locally linear about its centre frequency, we may delete all entries of even parity as they result in no net phase change (i.e. the entries 0011, 0101, 0110, 1001, 1010, and 1100).

This assumption also allows us to group the sequences into Equivalent Sequences. For example, the equivalent sequence of 1011 is 11, as the two bits 10 result in no net phase change. For four-bit sequences, there are four sets into which the Equivalent Sequences may be categorized, as shown in Fig.B.4.

	Set A	Set B	Set C	Set D
	0001 0010 0100 1000	0111 1011 1101 1110	1111	0000
Equivalent Sequence	00	11	1111	0000

Fig.B.4 Reduction of Four-Bit Sequences
Into Four Sets of Equivalent Sequences

The probability of Set A occuring (P_A) is four times that of one of its entries occuring (as there are four elements in the set). The probability of any of the entries occuring is 1/16, therefore we can say;

$$P_A = (4).(1/16)$$
$$= 1/4.$$

Similarly, $P_B = 1/4$, $P_C = 1/16$, and $P_D = 1/16$.

For Sets A and B, the maximum tolerable Φ_e is;

$$\Phi_e < 45°/2$$
$$\text{i.e.} \quad < 22.5°,$$

as the equivalent sequence length for Set A is 2 bits and the accumulated phase error must be less than 45° for no symbol error to result. For Sets C and D, the maximum tolerable Φ_e is given by;

$$\Phi_e < 45°/4$$

< 11.25°.

A similar analysis for five- and six-bit sequences was performed, and the results are shown in Fig.B.5 and B.6.

	Set E	Set F	Set G	Set H	Set I	Set J
Eq Seq.	0	1	000	111	00000	11111
Prob.	10/32	10/32	5/32	5/32	1/32	1/32

Fig.B.5 Reduction of Five-Bit Sequences Into Six Sets of Equivalent Sequences

	Set K	Set L	Set M	Set N	Set O	Set P
Eq Seq.	00	11	0000	1111	000000	111111
Prob.	15/64	15/64	6/64	6/64	1/64	1/64

Fig.B.6 Reduction of Six-bit Sequences Into Six Sets of Equivalent Sequences

From these results, one can conclude the following;

I For a phase error of Φ_e = 11.25°, the probability of a symbol error is given by;

$$P_{es} = 2(1/16) + 2(1/32) + 2(6/64) + 2(1/64) = 0.406$$

This expression arises from the fact that the sets C, D, I, J, M, N, O, and P have sequence lengths long enough (i.e. ≥ 4 bits) to result in an accumulated

phase error of ≥ 45° which results in a symbol error.

- For a phase error of $\Phi_e = 9°$, a similar argument leads to;

$$P_{es} = 2(1/32) + 2(1/64) = 0.0938$$

- Finally, for $\Phi_e = 7.5°$, the only sets which allow for a symbol error are Sets O and P. This gives the symbol error probability as;

$$P_{es} = 2(1/64) = 0.03125.$$

These calculations show that even with the simplifying assumptions made at the outset of the analysis, the calculation of the error probability due to incorrect modulation index is tedious to derive, and results are only obtainable for special cases of modulation index errors.