A SURFACE ACOUSTIC WAVE WIDEBAND FM DISCRIMINATOR

by

© T.N.W.M. Somapala Nanayakkara

A Thesis
Submitted to the School of Graduate Studies
in partial Fulfilment of the Requirements
for the Degree
Doctor of Philosophy

McMaster University

October 1979

A SURFACE ACOUSTIC WAVE WIDEBAND FM DISCRIMINATOR

DOCTOR OF PHILOSOPHY (1979)
(Electrical Engineering)

McMASTER UNIVERSITY
Hamilton, Ontario

TITLE: A Surface Acoustic Wave Wideband FM Discriminator

AUTHOR: T.N.W.M. Somapala Nanayakkara

SUPERVISOR: Professor C.K. Campbell

NUMBER OF PAGES: xvii, 269

ABSTRACT

The advantages that wideband frequency modulation (FM) offers over amplitude modulation (AM) are minimum transmission power and better output signal-to-noise ratio. The discriminator in the FM receiver is important since its performance dictates the degree of capability of exploiting these advantages. Conventional FM discriminators are faced with limitations such as bandwidth of operation, linearity of conversion characteristics, dynamic range due to FM threshold, frequency of operation, reproducibility, reliability, and economics. These problems are caused by the basic restrictions found in the principle of operation, and the nature of the devices used to realize these discriminators. The subject of this thesis is to explain alternatives to overcome these problems.

In the class of narrow band FM discriminators, we have suggested two discriminators whose operation is based on Surface Acoustic Wave (SAW) resonators, and SAW differential delay lines. For wide band FM discriminators we have developed a principle of operation based on a modified M-ary correlation receiver principle. The actual discriminator was realized with a system using a SAW chirped matched filter as the central signal processing element, and

a zero-crossing detector technique is employed to minimize the errors in signal detection. The design and fabrication of a SAW-chirp filter is discussed. A simple thin film technique to modify the aperture of an interdigital transducer (IDT) in order to obtain the required frequency response is explained and verified experimentally. To prove the principle of operation, a discriminator operating at 70 MHz with a bandwidth of 18 MHz has been produced. The performance evaluation of the device shows the linearity to be better than 1.8% over the bandwidth. The FM threshold occurs at -3 dB of input carrier-to-noise ratio. The parameters and measurement techniques employed to evaluate the noise performance of the chirp filter are explained, and the capability of the principle of operation for further extension of the FM threshold is explained. Finally, as applications we have proposed a multi-frequency position modulation modem. The principle of operation and implementation of this modem is discussed. In addition, applications to dual modulation schemes, such as pulse amplitude modulation - frequency modulation (PAM-FM), and applications to instrumentation systems, are also discussed briefly.

ACKNOWLEDGEMENTS

I would like to express deep gratitude to my supervisor, Dr. C.K. Campbell, for his consideration, supervision, guidance and approval throughout the duration of this work. The author also acknowledges the constructive criticism and the approval of the other members of the supervisory committee, Dr. S.S. Haykin, Dr. H.D. Barber and Dr. J.B. Anderson. As well, I would like to express my thanks to Dr. D.P. Taylor for the useful discussions which helped in this work.

I would like to thank my colleagues Messrs. P. Edmonson, M.S. Suthers and D.G. Seiler who provided a stimulating work environment. Also, thanks is due to Mr. J.P. Reilly for proof-reading this thesis.

I wish to acknowledge the cooperation of various members of the department of Electrical and Computer Engineering during the course of this work. In particular, I would like to express my thanks to Dr. Haykin for letting me use some of his equipment.

I would also thank Miss P. Dillon of the Word Processing Centre for the typing of this thesis. Finally I would like to thank my wife, Trish, and my daughter, Ivy, who have paid a very heavy price for seeing me through these difficult times.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

LIST OF FIGURES

LIST OF FIGURES (continued)

LIST OF FIGURES (continued)

LIST OF FIGURES (continued)

LIST OF FIGURES (continued)

LIST OF FIGURES (continued)

LIST OF FIGURES (continued)

CHAPTER 1

INTRODUCTION

It seems that man needs some type of turmoil to innovate or discover new phenomena. During World War II, frequency modulation (FM) was used extensively for signalling. This was mainly because of its noise reduction properties. With the conclusion of the War, interest in FM communication remained relatively dormant. However, the field was kept alive by a few intellectuals in research laboratories [1,2]. With the emerging affluence after the War, and the introduction of the long playing record in 1948, the general public in the western world showed an interest in high fidelity sound. With this demand, a somewhat better informed public began to appreciate noise - immune properties of FM. In April 1961 the United States Federal Communication Commission approved the transmission standards for stereophonic multiplex broadcasting.

Communication engineers, however, choose FM communication over amplitude modulation (AM) for certain advantages that it can offer. One such advantage of great importance is that the transmitting power required for FM is much less than AM. In AM the power required for transmission

increases with the degree of modulation index but the efficiency increases. In contrast, FM will not increase the transmission power requirement with increasing modulation index. The reason for this advantage comes about from the fact that the ideal FM-modulated signal is a constant amplitude signal whose amplitude is independent of the modulation index. This means the peak power level is identical to the average power. This obviously implies a constant power level. This certainly is not the case with AM where the peak power varies with the modulation index.

Increasing the modulation index in FM has other consequences. It increases the bandwidth. Theoretically, the FM bandwidth is infinite; however, in practice a major portion of FM waves is limited to a finite number of sidebands. Thus, there is a finite bandwidth required to recover the original information. This bandwidth increases with the increase in modulation index. The other consequence of increasing the bandwidth due to an increase in the degree of modulation or modulation index is the improvement of the output signal to noise ratio $(SNR)_o$. For single-tone modulation, the $(SNR)_o$ is proportional to the square of the modulation index. Thus, the increasing bandwidth will improve the $(SNR)_o$. This is a very important advantage which makes FM preferable to AM.

Needless to say, when communication engineers

considered FM for the purpose of transmission, they wanted to exploit all these advantages, especially, in this era of space communication, when we are looking for power-conserving, compact, and high performance systems. However, it has not been possible to reap the full benefits of these advantages due to the limitation of the devices, and operational principles of the discriminators. One of the most discouraging aspects of FM receivers is the FM threshold which prevents the exploitation of the above-mentioned advantages. As a result, many studies have been conducted to find ways of extending the threshold, and it has been difficult to break through the barriers that the devices confront. However, now the time is ripe to break through these barriers because there are new devices which have better signal processing capabilities than the old ones. A class of devices that comes under this catagory is Surface Acoustic Wave (SAW) devices.

The basic physics of surface waves were first described by Lord Rayleigh [3]. Ash [4] recognized the potential of SAW to realize circuit functions. Since the introduction of the interdigital transducer by White and Voltmar [5], a considerable amount of effort has been spent in the area of SAW devices. This is reflected in the fact that in recent years at the Institute of Electrical and Electronic Engineers (IEEE) Ultrasonic Symposia, SAW topics

have captured over 90% of the papers delivered.

In SAW devices the signals are propagated along the surface of the substrate. As such, they can be easily tapped, modified and guided. These capabilities have been used to produce some sophisticated signal processing functions. Furthermore, the velocity of SAW being 5 orders of magnitude slower than that of electromagnetic (EM) waves; the device size can be reduced by the same order of magnitude. Since coupling structures used in SAW devices are planar structures, they are easily produced using planar technology which has already been developed for integrated circuits. In addition, their frequency range of operation is 10 MHz - 1 GHz. This has normally been a most awkward frequency range for devices. Conventional LC networks are difficult to produce in this range while microwave circuit techniques are not realistic because of the large wavelengths. In addition, they are reproducible, reliable, and if produced on a mass production scale can be very economical.

Since 1965, a large number of SAW signal processing devices have been introduced. Filters, matched filters, correlators, convolvers, tapped delay lines, PN code generators, and chirp filters are a few examples. Design and fabrication techniques have also been developed, yet only a few of these devices have been used in actual civil

communication systems. The reason for this may be that too many devices have been introduced too soon, and the systems engineers are not completely aware of the capabilities of these devices. Indications are that more and more of these devices are finding their places in communication systems. All this indicates is that as communication engineers become more aware of SAW device capabilities, SAW devices are being introduced to new systems. In other words, SAW devices, like other new devices that have been introduced in the past, are taking the natural evolutionary process of being accepted.

There is a great need for conducting systems-oriented research. It is believed that SAW based sub-systems would be the key to the success of SAW devices. This belief makes a great deal of sense, and every indication points towards proving that belief has been correct. Therefore, as an increasing number of communication engineers come to know about SAW devices, and the increase in systems-oriented research, SAW devices will be more widely used in the future.

1.1 SCOPE OF THIS THESIS

As we mentioned earlier, to exploit all the advantages of FM technology we need to find a way to overcome some of the problems associated with the devices in

FM receivers. Also, as previously mentioned, as far as the SAW device field is concerned there is a need for systems-oriented research. Thus, if we can solve some of the FM device problems with a SAW device based sub-system we should be fulfilling the needs of two areas. Therefore, the scope of this thesis is to explain how to build a SAW device based sub-system which helps to exploit the advantages of FM communication.

To carry out this task in some logical manner, we must first identify the problem encountered in the design of FM receivers. The major signal processing device or sub-system in the FM receiver is the FM discriminator. This has been the major cause of the problems that need to be carefully examined. Therefore, in Chapter 2 we will investigate the principles of operation of popular discriminators, their performance and limitations. We will also look at possible ways to use SAW devices in discriminator circuits and speculate on the possible advantages and disadvantages of these new SAW discriminators. After this critical study, we will outline the problems that need to be solved.

In Chapter 3 we will identify the major causes of these problems. Having done this, we will formulate a functional diagram to solve the problems. With the aid of this functional diagram we will develop a principle of

operation of the discriminator, and identify this principle of operation as an equivalent to M-ary receiver. Finally in this Chapter we shall define the general hardware requirements.

Since we claim that SAW devices can meet the hardware requirements, in Chapter 4 we shall review the SAW device principle. We shall deal with the fundamentals of SAW phenomena and introduce the general area of SAW theory pertinent to this work. In doing so, we shall make an attempt to portray the versitility of these devices.

Chapter 5 will deal with the hardware implementation of the discriminator. We shall show how to produce correlators with SAW devices, and discuss their physical mechanism. Also in this Chapter we shall show how to adopt the SAW linear chirp filter as a correlator bank, and discuss how it can be designed. Having done this, we shall explain how to use this device and produce the discriminator. Also we shall explain the other circuit techniques for overcoming some of the inaccuracies in the chirp filter.

Chapter 6 will review the SAW fabrication technology necessary to produce the devices. In this review, technology will be included that is presently used to produce SAW devices operating in the frequency range from 10 MHz - 1 GHz. Also, we shall discuss some of the problem

areas, and what is presently being done to solve them. Furthermore, we shall introduce a simple economical thin film technique to experimentally vary the apodization functions of the SAW transducers.

Chapter 7 will present the experimental results of the new SAW FM discriminator. In doing so, we shall explain the noise performance parameters used to characterize both the SAW device and the discriminator as a whole. Also, we shall explain the measurement techniques used to measure noise performance.

Chapter 8 will briefly discuss the application of this discriminator in the field of communications and instrumentation. In particular, we shall propose a multi-frequency position modem which helps to overcome some of the critical problems found in "frequency hopping" Spread Spectrum systems.

Finally, in the conclusions we will look back at the initial problems of FM discriminators and check the problems that we have solved. Here we shall make an assessment of the progress we have made by comparing our results with the performance results of existing discriminators.

CHAPTER 2

FM DISCRIMINATORS

The central part of any FM receiver system is the FM demodulator. Its main function is to recover the original message signal which is frequently referred to as the baseband signal. In the final analysis, it is this device which dictates the performance characteristic of the receiver. The device in the FM receiver which performs the demodulation is commonly known as the FM discriminator. There are several principles of operation for these devices. These principles will be reviewed in this Chapter. In doing so, we shall make an attempt to bring about some physical insight into the operation of these devices.

In the process of FM modulation, we convert an amplitude variation of the baseband signal into a signal whose frequency varies. In other words, the FM modulator is a voltage-to-frequency converter. Ideally, the FM signal is a constant amplitude signal with a variable frequency. Obviously, in FM demodulation we can translate the variation of frequency into an amplitude variation, and therefore, recover the original baseband signal. That is to say, that the major function of the discriminator is one of frequency-

to-voltage conversion. One is able to realize this conversion by choosing a network whose transfer characteristic H(f) is such that H(f) varies linearly with frequency. Such networks, of course, are limitless in number. However, for the purpose of this review, we shall divide the discriminators into 3 broad categories on the basis of operational principles. These are:

1. Filter Type Discriminators
2. Delay Line Type Discriminators
3. Feedback Type Discriminators

2.1 FILTER TYPE DISCRIMINATORS

Filter type discriminators are perhaps the most widely used discriminators. The principle of operation of these discriminators is depicted in Figure 2.1(a). The first part is the frequency selective network. The transfer characteristic of this network has a linear amplitude variation with frequency. The second part is the envelope detector to detect the variation in amplitude.

As indicated in Figure 2.1(b) a simple discriminator is very easily made. The circuit is tuned at a frequency above that of the signal. The voltage amplitude across the circuit varies as the frequency changes. This variable amplitude signal is fed to a diode detector and converted to a direct voltage. The transfer function of this discrim-

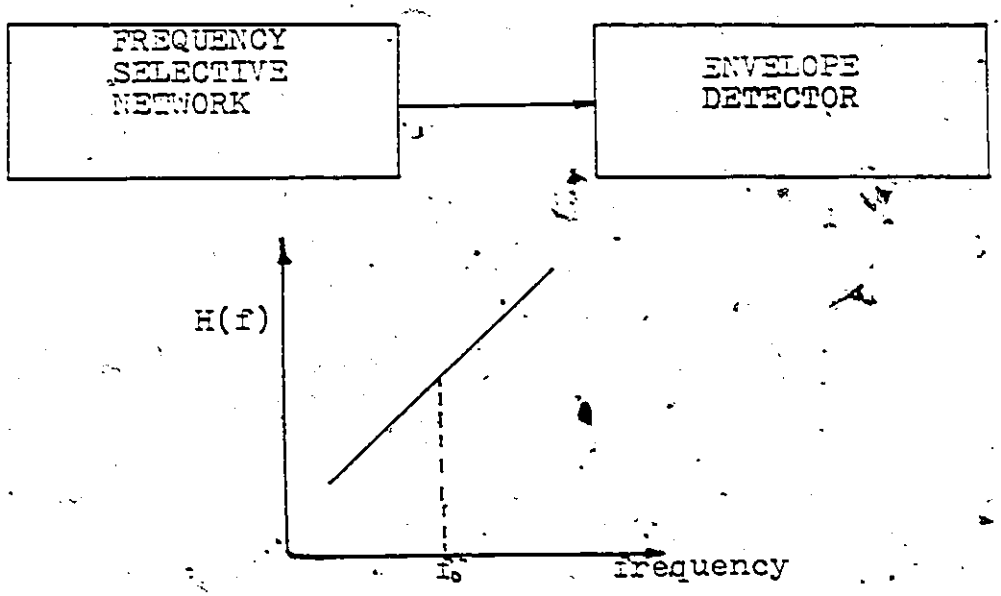


Figure2.1(a) Basic discriminator principle and the ideal transfer characteristics.

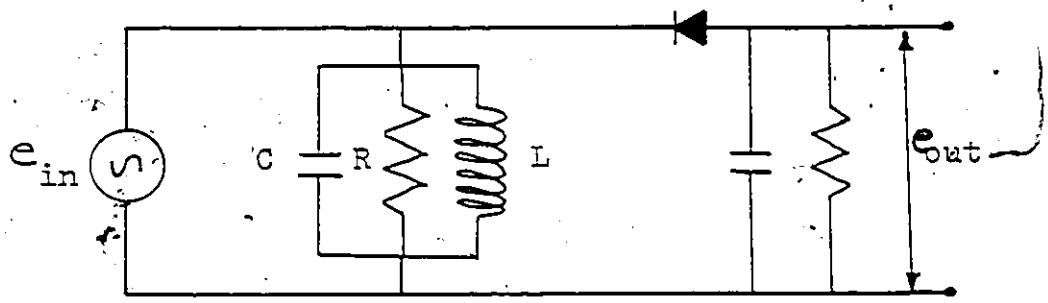


Figure 2.1(b) An elementary FM discriminator.

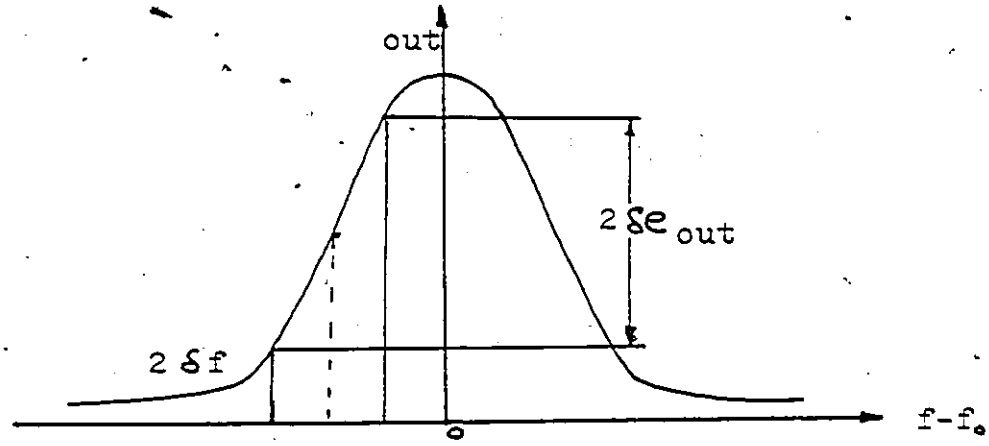


Figure 2.1(c) Transfer characteristics of an elementary discriminator.

inator is shown in Figure 2.1(c).

This elementary discriminator suffers from several drawbacks. Evidently, a signal above the resonant frequency would be discriminated equally well. Thus it has an undesirable selectivity condition. Since the circuit is useful only in the linear region of the transfer function, the useful bandwidth is only a fraction of the device bandwidth. If the limiter prior to the discriminator does not function properly, the amplitude modulation present in the FM signal, due to the noise, will give rise to distortions in the output. These drawbacks do not encourage the use of this type of circuit.

2.1.1 Balanced Discriminator

The basic principle of operation of the so called Balanced Discriminator is not different from the general principle of operation of the Filter Discriminator. However, the realization of this circuit is a little more complicated. An idealized model of a balanced frequency discriminator is shown in Figure 2.2(a). We recognize these as two elementary discriminators staggered in frequency. Here, we have two filters of the same bandwidth 2B and the amplitude responses are identical, but they are different in sign and centre frequency of operation. Furthermore, we idealized these filters to have triangular shape responses,

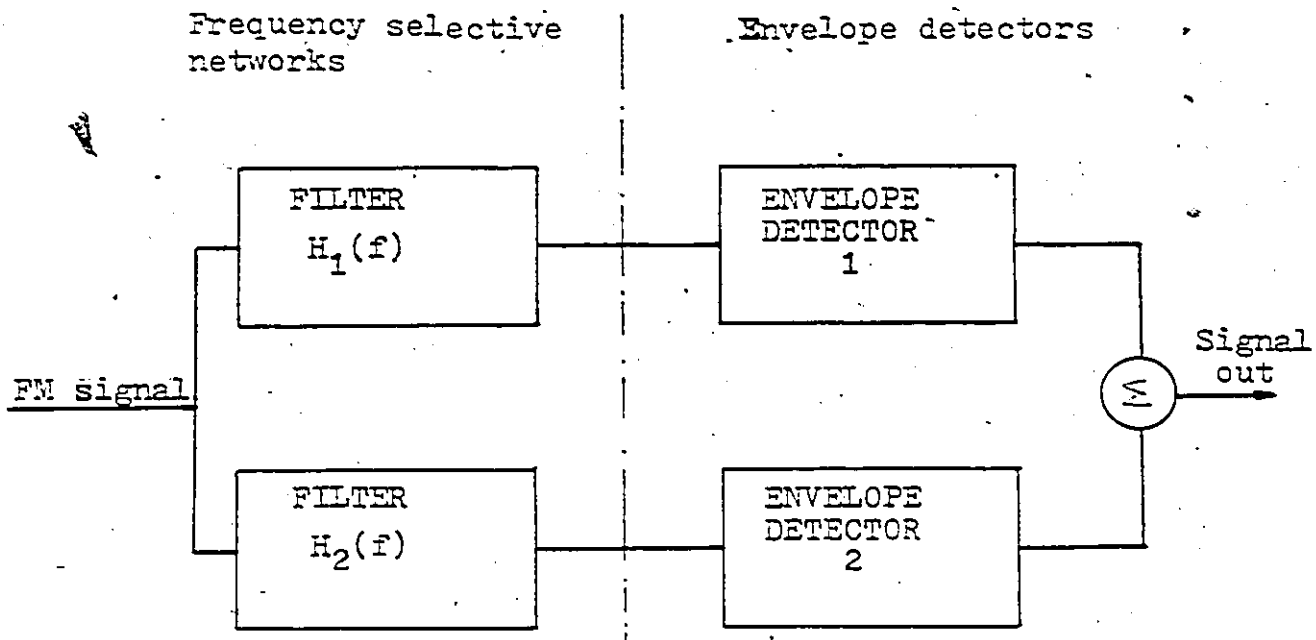


Figure 2.2(a) The model of a balanced frequency discriminator.

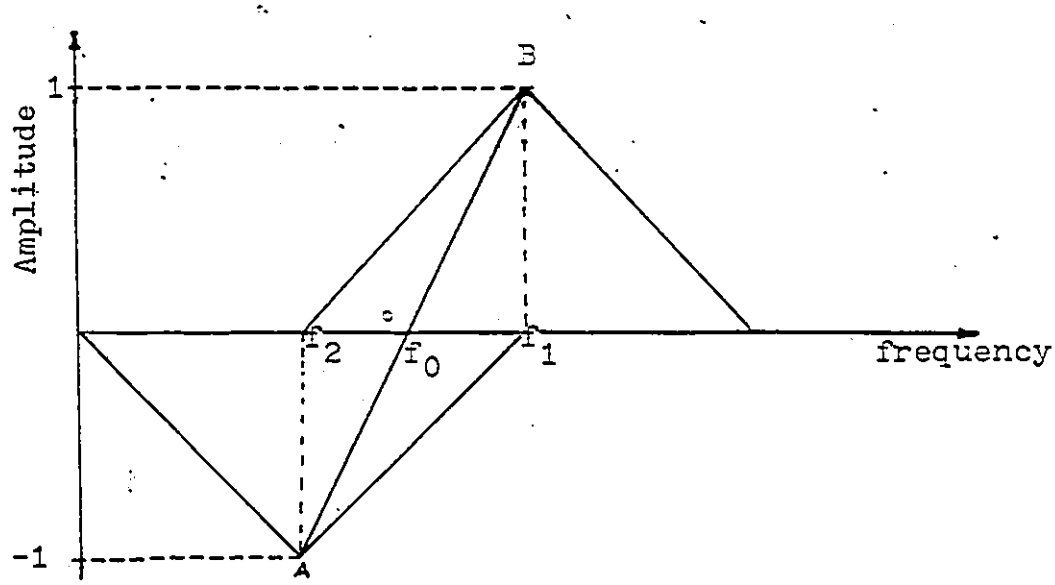


Figure 2.2(b) Ideal response curves for the filters and for the discriminator.

and the relative bandpass frequency characteristics are shown in Figure 2.2(b). The linear operating range of this idealized discriminator is given by AB. The bandwidth of the device is half the bandwidth of a single filter. We also note that if we were to alter the frequency response of the filter or the separation in frequency df_s of the filters, we will alter the discriminator transfer characteristics both in bandwidth and in linearity.

We shall be extremely fortunate indeed if we can find filters exhibiting such a triangular response. A more realistic circuit arrangement is shown in Figure 2.3(a). Here, we have two tuned filters staggered in frequency. The envelope detectors are diode demodulators. The detailed analysis of these circuits has been carried out by Arguimbau [6] and also by Haykin [7]. Here we shall outline the important parameters that dictate the performance of the circuit. The main parameters of the filters, which affect the bandwidth of operation and linearity of the discriminators, are amplitude response of the filters and the frequency separation between the two resonance frequencies. This separation is given by [6]

$$df = a \cdot BW \tag{2.1}$$

where df = separation between resonance peaks

a = is a constant

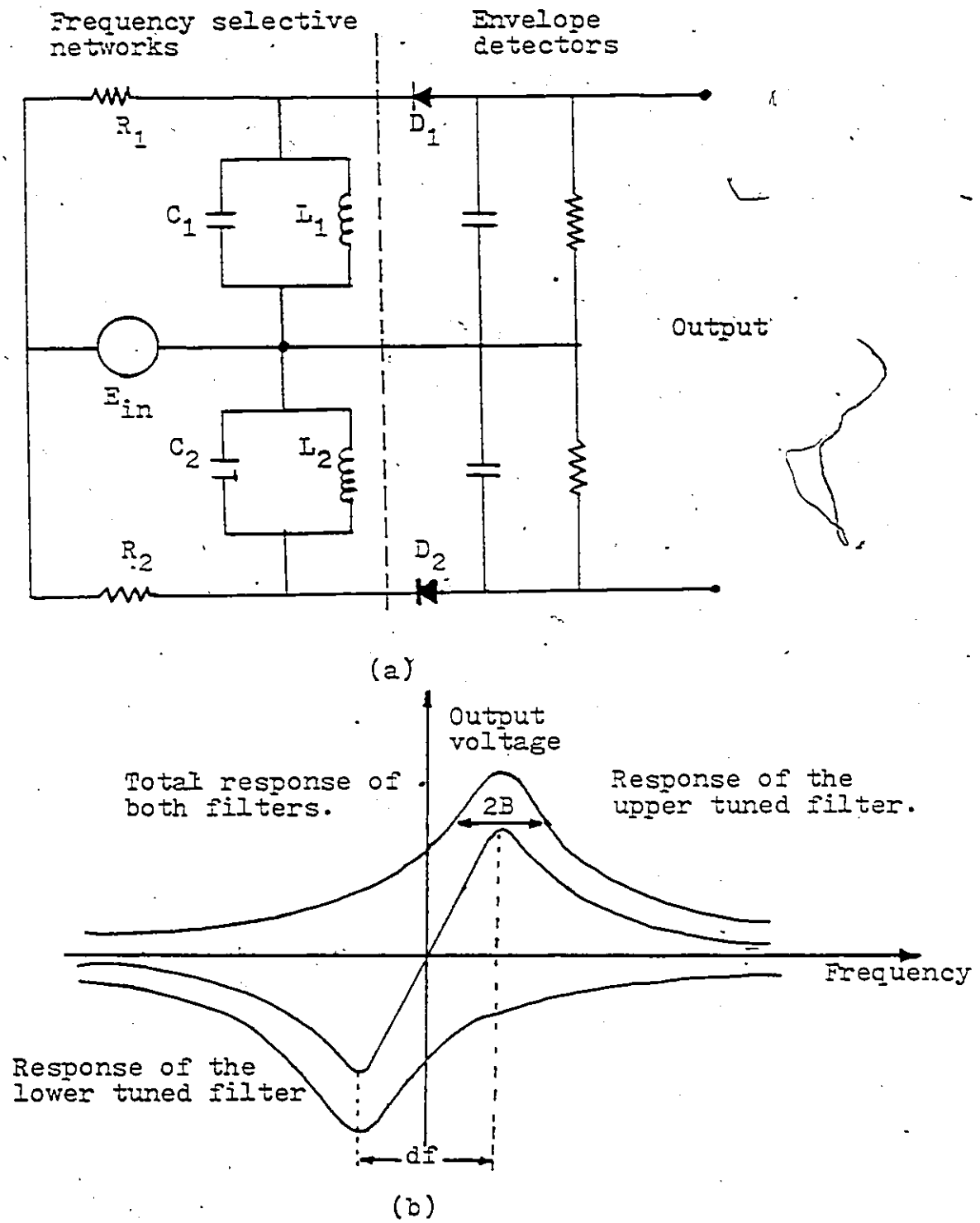


Figure 2.3 Balanced frequency discriminator.
(a) Circuit diagram.
(b) Frequency responses.

BW = Bandwidth of each filter (we assume that both filters have equal bandwidths)

The individual responses of the two filters, along with the total response of the discriminators is illustrated in Figure 2.3(b). It is possible to obtain various shape discriminator curves by varying the frequency separation and of course, varying the response characteristics of individual filters. In general, to obtain good linearity one needs to employ high-Q filters with a smaller value of a. This, obviously, introduces a problem of bandwidth limitation. Another parameter which is affected by the Q of the filters is the sensitivity of the discriminator. The sensitivity of the discriminator is given by

$$S = \frac{dv}{df} \tag{2.2}$$

In other words, S is the slope of the discriminator transfer function. If the Q of the filters is high, then S will be large as well.

As one can easily see from Figure 2.3(b), the tuned filter outputs are not strictly band limited. This introduces some distortion at the output. Another problem which arises from the fact that the filter bandwidths are larger than the discriminator bandwidths is that the device will respond to noise frequencies outside the bandwidth of the device. In fact, there is no protection in this circuit against noise. As such, it can have a very poor FM thres-

hold when used in a FM receiver. Also, another limitation of this device is that it cannot be operated over the whole bandwidth, because the device transfer characteristics are not linear at the band edges of the device.

Quite apart from these problems, we have other problems associated with this type of filter networks. It is very difficult to produce accurate LC components at higher frequencies. Although this type of discriminator has been produced to operate around MHz region, it tends to be a very narrow band device. It is correct to say that in the range of 10 MHz - 1 GHz, it would be extremely difficult to produce wideband high linearity discriminators using this principle.

Difficulties in producing accurate LC circuits generate other problems such as reproducibility, reliability and economic problems. Invariably, most of these devices require correction circuits to combat inaccuracies in the LC circuits. This means that large scale production of these types of circuits is virtually impossible. In fact, these devices are often produced mostly on a one-to-one basis. Even then, it is difficult to reproduce the same performance to a high degree of accuracy. Needless to say, this type of situation can definitely result in poor reliability, and of course, the cost of production increases the price of these devices.

2.1.2 SAW Filter Discriminators

Problems such as reproducibility, reliability and economy can be solved using surface acoustic wave devices [8,9]. It is possible to realize filters based on surface acoustic wave interactions. The principles of operation of surface acoustic wave filters will be dealt with in Chapter 4.

In principle, there is no difference between the LC filter discriminators and SAW filter discriminators except for the fact that the frequency-selective circuits of the discriminator employ SAW filters. In Figure 2.4 we illustrate an arrangement for a SAW discriminator. Notice that both filters are fabricated on the same substrate. The solution to the problems mentioned above comes about as a result of inherent qualities of SAW technology. Namely SAW devices are highly reproducable, extremely reliable and eventually can be economical.

SAW filter discriminators can extend the frequency of operation up to about 1 GHz without having to face the same kind of problems that one is faced with within LC discriminators. There are no tuning circuits in SAW filter circuits such as one finds in LC circuits.

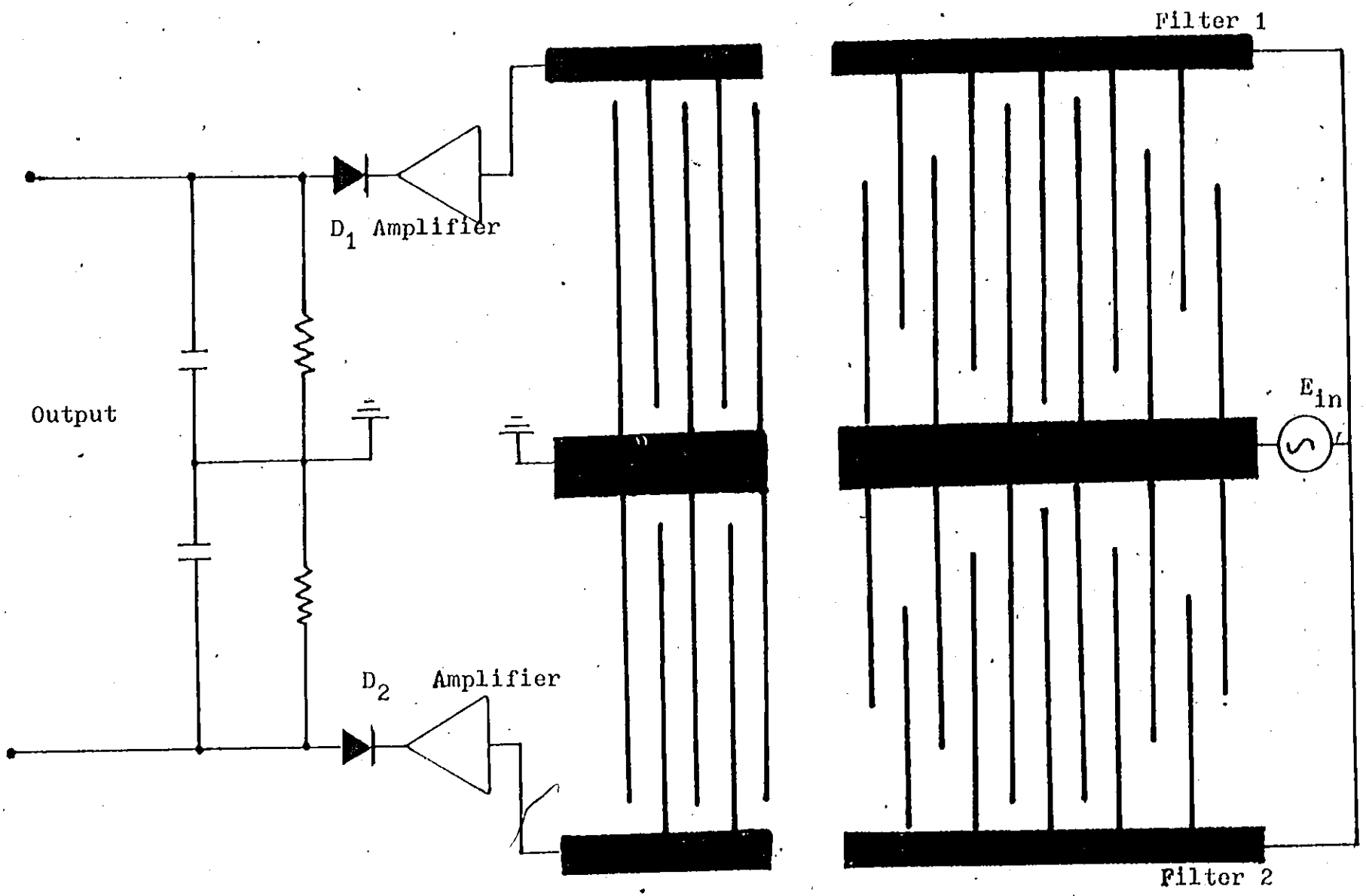


Figure 2.4 A filter type SAW discriminator.

Although SAW discriminators solve the above mentioned problems, they do not solve linearity and bandwidth problems. That is because they are a result of the fundamental limitations in the principle of operation of these discriminators. Quite apart from that, truncation errors in the SAW filter generate ripples in the filter frequency responses. They, in turn, yield poor linearity response in the discriminator transfer characteristics. We shall have more to say about truncation errors in Chapter 4. The frequency responses of the filters are also affected by other higher order effects such as bulk wave coupling, finger reflections, triple transit echoes, edge reflection, etc. However, there are methods to reduce these effects.

At this point one might add, no complete studies have been conducted to evaluate the ultimate capabilities of these devices. Hartemann [10], who recognized the possibility of using SAW filters in discriminators, carried out his investigation with the objective of proving the basic principle. His discriminator was designed to operate at 30 MHz. The linearity of this device was only 2% and the bandwidth was 1.4 MHz. He claims that these devices can operate in the frequency range of 5 MHz - 400 MHz, and also, he indicates that the maximum frequency separation of 100 MHz can be achieved. However, these claims have not been proven and also the claim on the frequency range is no

longer valid since we know it is possible to build filters to operate at higher frequencies with advanced fabrication technology such as electron-beam lithography.

Another area of performance we know little about is the noise performance of these devices. To our knowledge, nobody has conducted any studies in regard to this aspect of the device. There are good reasons to believe their noise performance is comparable to conventional LC discriminators if not better. Dwelling on that subject would be pure speculation rather than fact. Therefore, we shall conclude that it seems that SAW filter discriminators have a lot to offer to the state of the art of narrow band discriminators. However, further research and development has to be carried out before they can be useful.

2.1.3 Bulk Wave Resonator Filter Discriminators

It is also possible to realize frequency selective circuits with bulk wave resonators. However, the frequency of operation of these resonators is limited to below 30 MHz. The reason for this frequency limitation is that it is difficult to produce high frequency bulk wave resonators. The resonance frequency of operation of bulk wave resonators depends on the physical thickness of the resonator material, with the required resonator thickness reducing with increasing frequency. Therefore, at higher frequencies this

thickness becomes hard to achieve. Their reliability of operation at higher frequencies is known to be poor. It may be said that above 10 MHz one could achieve better performance with SAW filter discriminators than with bulk wave discriminators.

2.1.4 SAW Resonator Filter Discriminators

While we are on the topic of the filter discriminators, it is appropriate to point out the possibility of using SAW resonators for very narrow-band high-linearity discriminators. SAW resonator filter discriminators have never been discussed in the open literature to our knowledge. The purpose of mentioning this here is to point out the possibility of using these resonators in narrow-band discriminators to achieve better sensitivity, linearity and noise performance.

Earlier in this Chapter we pointed out that, if we can produce a frequency response of the filters that is perfectly triangular, then we can produce high linearity discriminators. Also we mentioned that it is difficult to produce such filters. With SAW resonators however, it is possible to produce pseudo-triangular shaped responses. In fact, if we are only concerned with extremely narrow-band discriminators of high performance, it is possible to increase the Q of these resonators to obtain almost

triangular responses. Furthermore, in the case of SAW filters, we found that the cause of poor linearity arises from the ripples in the filter responses. In other words, the cleanliness of the slope of the triangular responses curve will determine the linearity, providing that we have the proper frequency separation.

Surface acoustic wave resonators can be designed accurately [11] and prove to have extremely good pseudo-triangular responses. Thus, it is possible to use these devices in frequency selective circuits in discriminators, and thereby achieve good linearity. The sensitivity of these discriminators increases on two accounts: first, due to the fact that the resonators have a high Q; second, from the fact that it is possible to have a high passband to off band attenuation ratio by cascading resonators.

We have seen one mechanism responsible for poor noise performance. This relates to the relatively poor bandpass characteristic of the LC type filters, which respond to the out of band noise as well as to the inband noise, because of the poor band limiting properties of the filter. We find SAW resonators have very good band rejection properties. Consequently, they are less likely to respond to the out-of-band noise.

It was pointed out that bulk wave resonators perform poorly above 10 MHz. On the other hand, surface wave

resonators can be produced in the frequency range of 10 MHz - 1 GHz with good reproducibility, reliability and economy. In Figure 2.5 we illustrate a surface acoustic wave resonator and possible arrangement for surface acoustic wave resonator discriminator.

2.2 SAW DELAY LINE DISCRIMINATORS

It is known [12] that delay lines can be used for frequency discriminators. Conventional delay lines have a number of disadvantages such as delay line distortions, dispersion and aging effects, etc. These disadvantages can be overcome by using acoustic delay lines. In principle these acoustic delay lines can be either bulk acoustic delay [12] lines or surface acoustic wave delay lines. However, surface acoustic wave delay lines are preferred due to their ease of fabrication, ruggedness, and of course, due to other advantages such as reproducibility, reliability, etc. There are quite a few SAW discriminators which use delay line principles [10]. However, we shall explain one possible way of producing a SAW discriminator using a differential delay of two delay lines.

In Figure 2.6 we illustrate the circuit arrangement of the SAW differential delay line discriminator. To explain the operation of the device, let us feed the input FM signal to the input transducer. We can formulate this

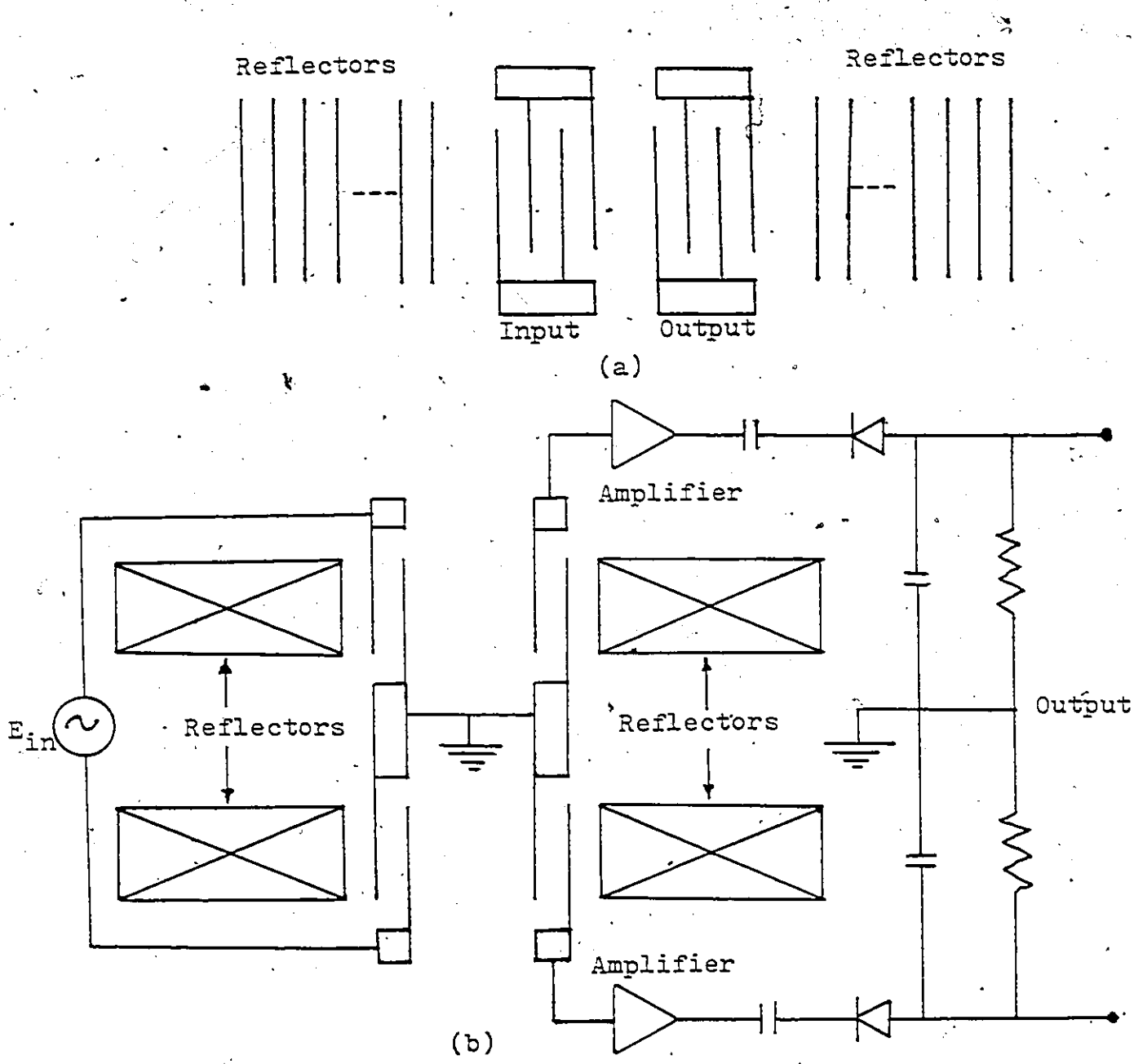


Figure 2.5 SAW resonator filter and a discriminator.
(a) SAW resonator filter.
(b) SAW resonator filter discriminator.

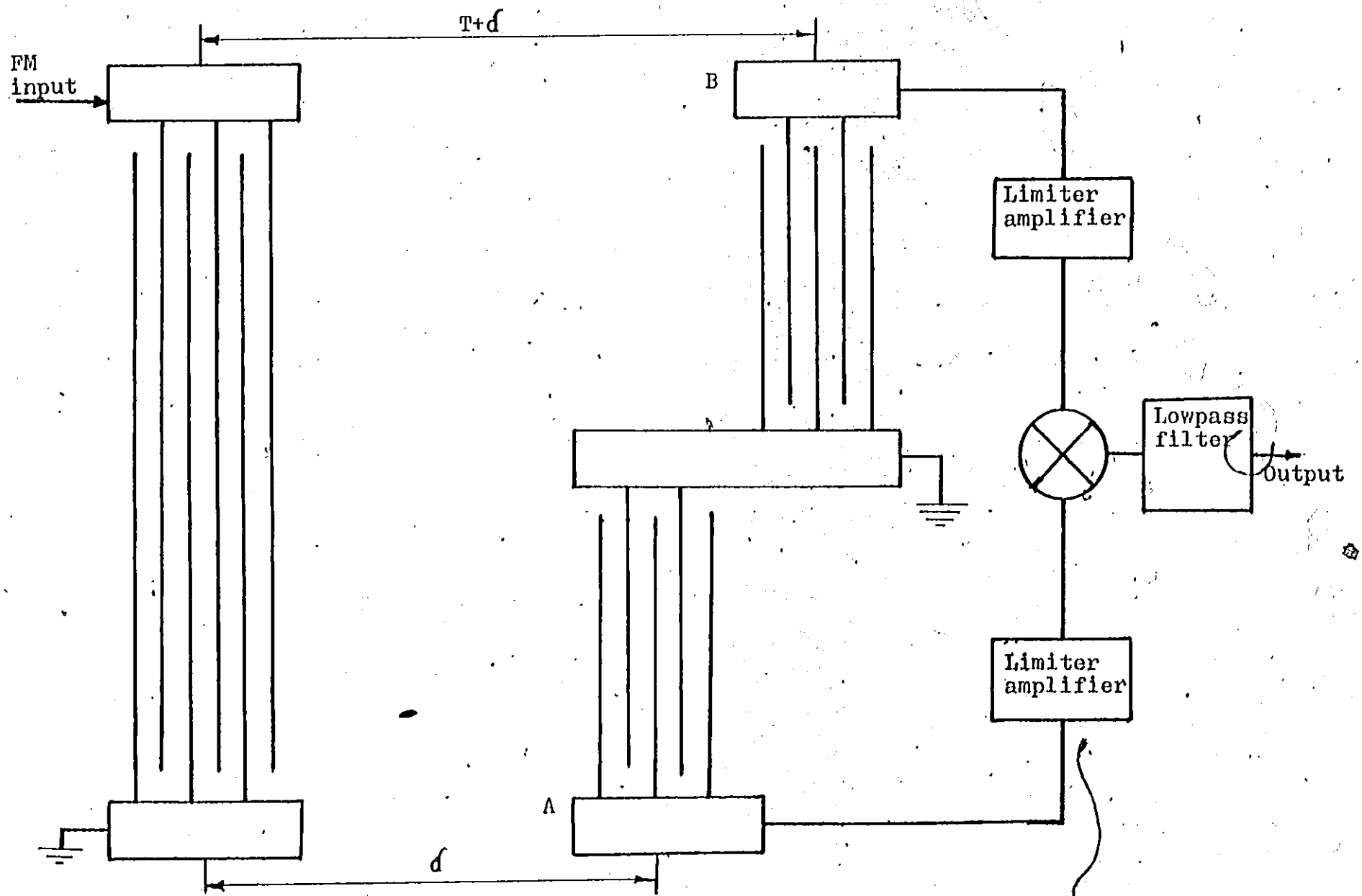


Figure 2.6 SAW differential delay line discriminator.

signal as

$$S_i = \cos\,[(\omega_o + \omega_m)t] \tag{2.3}$$

where ω_o = carrier frequency

and ω_m = modulation frequency.

The amplitude of the signal is of no consequence to us, since any variation in amplitude would be removed by the limiter. Therefore, for convenience we have omitted the amplitude term in the equation.

The output from the transducer A is given by

$$S_A = \cos\,[(\omega_o + \omega_m)\,(t + \delta)] \tag{2.4}$$

where δ is the delay due to the separation between the input transducer and transducer A. Similarly the output signal from transducer B is

$$S_B = \cos\,[(\omega_o + \omega_m)\,(t + T + \delta)] \tag{2.5}$$

where $T + \delta$ is the delay between the input transducer and output transducer B.

These two signals are fed to two limiters to limit the amplitudes, and the outputs from the limiters are fed to a mixer. The output of the mixer is the product of S_A and S_B and can be expressed as

$$2 \cos x \cos y = \cos\,(x + y) + \cos\,(x - y) \tag{2.6}$$

The difference term can be expressed as

$$\cos\,(x - y) = \cos\,[(\omega_o + \omega_m)T] \tag{2.7}$$

We can choose the lowpass filter to respond to this term and reject the other terms. On careful investigation of this

term we find that both ω_o and T are constants, where ω_o is the carrier frequency and T is a fixed term due to differential delay. Therefore, the only variable term in the equation is $\omega_m T$. Thus, after filtering out the carrier frequency we find that the output varies as $\omega_m T$ varies. This is the required function of the discriminator.

However, in order for the mixer to operate in the linear region, it is desirable to arrange the differential delay such that

$$\omega_o T = n\ \pi/2 \tag{2.8}$$

where n is any odd integer.

Furthermore, for good linearity, it is desirable to choose T such that

$$\omega_m T < \pi/6 \tag{2.9}$$

It is quite apparent from equations (2.8) and (2.9) that the restrictions imposed in the choice of T place a restriction on the maximum value of modulating signals that can be processed with good linearity. In other words, we have a bandwidth limitation in this device. That is to say that even though the delay line discriminators are less complicated and easy to produce, we have no real benefits as far as bandwidth of operation and linearity are concerned.

It is difficult to predict the FM threshold performance of the device. We note that equation (2.8) indicates that the fixing T acts as biasing of the discrimi-

nator. As such, the operation of the circuit is somewhat similar to the second-order phase-locked loop operation even though this discriminator does not have any feedback circuit. What the second-order phase-locked loop achieves with its feedback network, we have achieved by biasing. One might be tempted to assume that the FM threshold of the device is similar to that of the second order phase-locked loop. Such a guess or assumption could easily be wrong since the threshold extension in the second-order phase-locked loop comes about from its ability to suppress response to transients, and we do not know how this particular device responds to transients.

Another aspect of this particular device which we have not yet addressed relates to the spurious efects of the SAW device and their influence upon the overall response of the discriminator. However, if one is to design and produce this device, there are techniques that can be used to overcome these effects. In concluding the discussion on delay line discriminators, SAW differential delay line discriminators may offer much to improve device performances, especially when narrow band discriminators are required.

2.3 FEEDBACK CONTROLLED FM DISCRIMINATORS

It is well known that FM demodulation offers better signal-to-noise ratio performance than does AM demodulation.

Also we are aware of the bandwidth requirements for FM are greater than for AM operation. Thus, when we choose FM over AM to transmit our signal, we are paying a price in bandwidth to achieve a better signal-to-noise ratio. Therefore, it is extremely disappointing to find that there is an FM threshold which forbids one to exploit the advantages of FM. For this reason, considerable studies have been conducted by many investigators on methods of extending the FM threshold. Feedback-controlled FM discriminators are one outcome of such studies.

Here, we shall review two types of feedback-controlled discriminators. They are:

a) Phase-locked loop (PLL) discriminators

b) FM Feedback (FMFB) discriminators

The feedback component that is common to both of these discriminators is a voltage-controlled oscillator. In what follows below, we will review both of these discriminators, and also try to identify the threshold extension mechanism.

2.3.1 Phase-Locked Loop (PLL)

The basic first-order phase-locked loop, as illustrated in Figure 2.7, has two major components:

a) Phase Comparator

b) Voltage-Controlled Oscillator

To obtain a better understanding of the PLL, let us look at

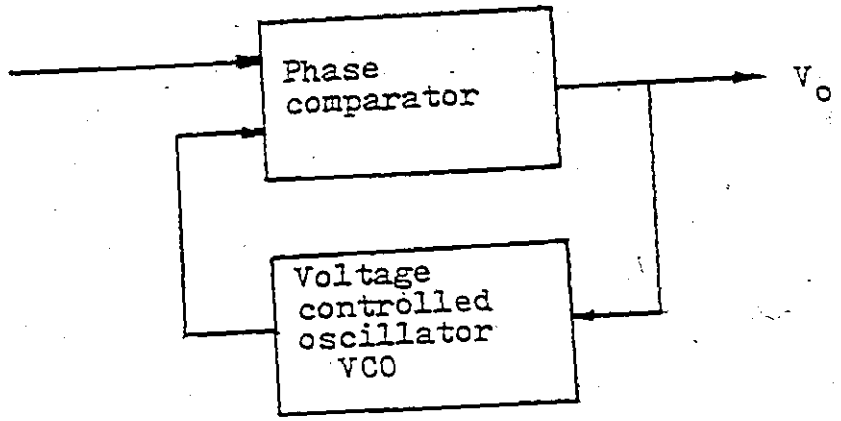


Figure 2.7 First order phase-locked loop

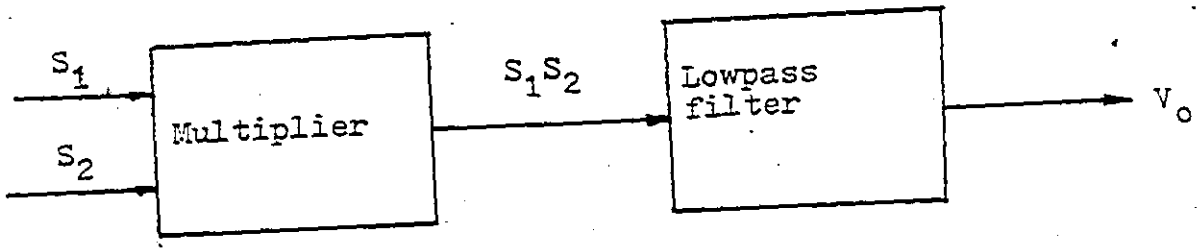


Figure 2.8 Phase comparator.

the principles of operation of these components.

2.3.1.1 Phase Comparator

The phase comparator is a 3-port device. As the name suggests, the device compares the phase between two input signals and produces an output which depends on the phase difference. As indicated in Figure 2.8, one way of constructing a phase comparator is by combining a multiplier and lowpass filter. Let us assume S_1 and S_2 are two sinusoidal signals given by

$$S_1 = A \sin [\omega_c t + \theta_1(t)] \tag{2.10}$$

$$S_2 = B \sin [\omega_c t + \theta_2(t)] \tag{2.11}$$

where $\theta_1(t)$ and $\theta_2(t)$ are time-varying phase angles. It can be shown that the output from the multiplier consists of a low frequency term given by

$$\begin{aligned} S_{mo} &= AB/2 \sin [\theta_1(t) - \theta_2(t)] \\ &= AB/2 \sin \psi(t) \end{aligned} \tag{2.12}$$

and other high frequency terms. For convenience, we shall assume that the multiplier gain is to be unity. We eliminate all the high frequency terms by passing the multiplier output through a lowpass filter. Thus the output of the lowpass filter becomes

$$S_o = ABG_p \sin \psi(t) \tag{2.13}$$

where the gain of the lowpass filter is equal to $2G_p$. The output characteristics of the phase comparator as a function of ψ are plotted in Figure 2.9(a).

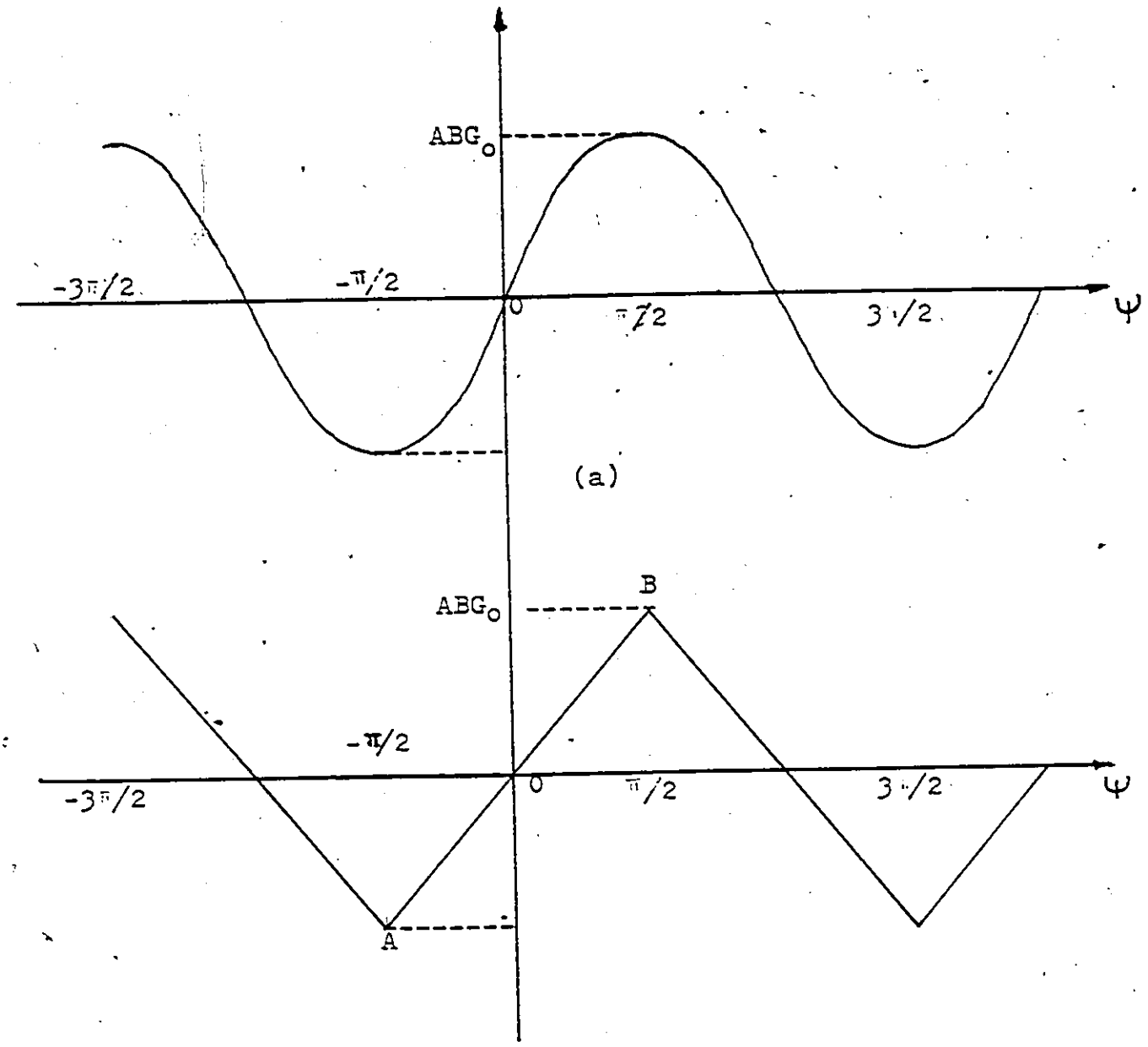


Figure 2.9 Phase characteristics of the comparator.
(a) With the sinusoidal inputs.
(b) With the square wave inputs.

In FM reciever circuits, usually one uses hard limiters prior to a discriminator. The output from the hard limiter is almost a square wave. The output response from the phase comparator as the response to two such square wave inputs is shown in Figure 2.9(b). As illustrated, we find the response displays a piece-wise linear variation with ψ.

2.3.1.2 Voltage-Controlled Oscillator (VCO)

As the name suggests, the VCO is an oscillator whose output frequency is controlled by an input voltage. Thus we recognize that the device can be used as a frequency modulator.

Suppose, the VCO generates a sinusoidal signal of amplitude B at an angular frequency of ω_c, when the input voltage is zero volts. We shall define the sensitivity of the VCO as

$$G_o = d\omega_i/dv \tag{2.14}$$

where $d\omega_i$ is the change in instantaneous frequency due to a change in voltage dv. Now if the input to the VCO is a time-varying voltage then the output of the VCO can be expressed as

$$V_o = B \cos [\omega_c t + G_o \int_o^t v(t)\, dt] \tag{2.15}$$

in this case the instantaneous frequency is

$$\omega_i = \omega_c + G_o\, v(t) \tag{2.16}$$

we can rewrite the output of the VCO as

$$S = B \cos [\omega_c t + \theta_2(t)] \tag{2.17}$$

where

$$\theta_2(t) = G_o \int_o^t v(t)\, dt \tag{2.18}$$

With this information let us investigate the operation of a first-order phase-locked loop.

2.3.1.3 PLL Demodulator

To explain the principle of operation of the PLL demodulator we refer to Figure 2.7. The input signal to the modulator is

$$S_1 = A \cos [\omega_c t + \theta_1(t)] \tag{2.19}$$

where

$$\theta_1(t) = K_f \int_o^t m(t)\, dt \tag{2.20}$$

in which m(t) is the baseband signal and K_f is a constant. Now we assume that initially $\theta_1(t) = 0$, and we have adjusted the VCO so that, when the input to VCO is zero volts, the output frequency is exactly ω_c. Also, we have adjusted the VCO output signal to have 90° phase shift relative to the carrier. This phase shift is required to make the comparator output to be zero when $V_o(t) = 0$. Now at this point the PLL demodulator is in an equilibrium condition.

Now, at a time t = 0 the input signal frequency is changed abruptly. This change will be sensed by the comparator as a change in $\theta_1(t)$. The output of the

comparator will generate a voltage; corresponding to the difference in phase of the input signals. This output voltage from the comparator, which is also the input to the VCO, will increase the output frequency of VCO. Thus, eventually when the two input frequencies are equal, the whole system will see a new equilibrium condition. Note that this new equilibrium condition is not the same as the initial equilibrium condition, since the phase difference between the signals is no longer 90°. As a result, the output voltage of the modulator is no longer zero volts. The change in frequency $\omega(t)$ is seen by the comparator as a change in $\theta_1(t)$. This change has produced change in $\theta_2(t)$ at the output of the VCO

$$\frac{d\ \theta_1(t)}{dt} = \frac{d\ \theta_2(t)}{dt} \tag{2.21}$$

Substituting equation (2.18) in the above, we get

$$\frac{d\ \theta_1(t)}{dt} = \frac{d}{dt}\ [G_o \int_o^t v(t)\ dt]$$

Therefore, since the L.H.S. of (2.21) is $\omega(t)$,

$$V_o(t) = \frac{\omega(t)}{G_o} \tag{2.22}$$

Thus the output voltage is proportional to the input frequency which is the required function for an FM demodulator.

We find as long as the rate of change of frequency is slow compared to the time taken by the phase-locked loop to

re-establish its new equilibrium position, the PLL demodulator will track the change in frequency of the input FM signal. In other words, the PLL demodulator is a tracking filter which tracks the baseband signal. This tracking capability is an important feature of the first-order phase-locked loop demondulator, and it must be emphasized that the operating point continuously changes in the device.

Haykin [7] has detailed the theory of the phase-locked loop. We shall not undertake to deal with this theory here. However, we shall look at some of the important properties of the PLL demodulator to make an assessment of its performance.

The time constant of the PLL demodulator, which is proportional to the time taken by the PLL to re-establish its equilibrium condition, is the most important parameter in the circuit. This parameter is given by

$$\tau = \frac{\pi}{2\,A\,B\,G_p\,G_o} \tag{2.23}$$

where τ = Time constant of the phase-locked loop

G_o = Gain of the lowpass filter in the phase comparator

G_p = Sensitivity of the VCO, given by $d\omega_i/dv$

A and B are amplitudes of FM signal and the VCO output signal respectively.

The parameter τ determines the entire behaviour of the PLL demodulator. With reference to Figure 2.9(b), the range of the PLL is defined as the change in angular frequency ω of the input carrier signal which will just carry the steady state operating point from 0 to B or 0 to A. This range is given by

$$R = \omega_{max} = \pi/2\tau \qquad (2.24)$$

where ω_{max} is the maximum frequency deviation of the FM signal. Thus, if the bandwidth of the PLL is 2 ω_{max} and if the bandwidth of the PLL is to be increased, then τ should be decreased. Thus we see that τ determines both range and frequency. Hence, in the first-order PLL demodulator we are not free to vary these parameters independently.

Time constant of the PLL has other consequences. First of all, it dictates the maximum baseband frequency that can be modulated. Secondly, it determines the capability of the PLL to suppress noise. If the time delay τ is large, then the PLL will hardly respond to fast changing frequencies. However, noise in general has slowly varying frequencies, as well as fast changing frequencies. Thus the device tends to respond to these slowly varying frequencies components. In other words, the device will respond to certain spectral components of the noise and reject the others.

This noise rejection property is the reason for

threshold extension. Obviously, if we want to extend the threshold further, we must find a way to increase the time constant of the PLL demodulator. We cannot continue to increase τ indefinitely in the first-order demodulator for the reasons mentioned above. Thus there is a limit to the noise suppression in the PLL demodulator even though its performance is better than the conventional demodulator. However, some performance improvement can be achieved by using second-order PLL discriminators. We shall briefly review the operation of this device next.

2.3.1.4 Second-Order Phase-Locked Loop

The circuit arrangement for the second-order phase-locked loop is depicted in Figure 2.10. Comparing this to the first-order phase-locked loop we find the only difference is that this circuit has an extra filter. Thus, the second-order phase-locked loop has two filters including the one inherent in the phase comparator. It is interesting to note that the function of the filter in the phase comparator is to suppress the carrier and its harmonics. Hence, it has absolutely no influence on the first-order PLL performance. On the other hand, in the second-order phase-locked loop, the inclusion of the second filter has a definite influence on the performance of the device.

Haykin [7], and also Taub and Schilling [13], have

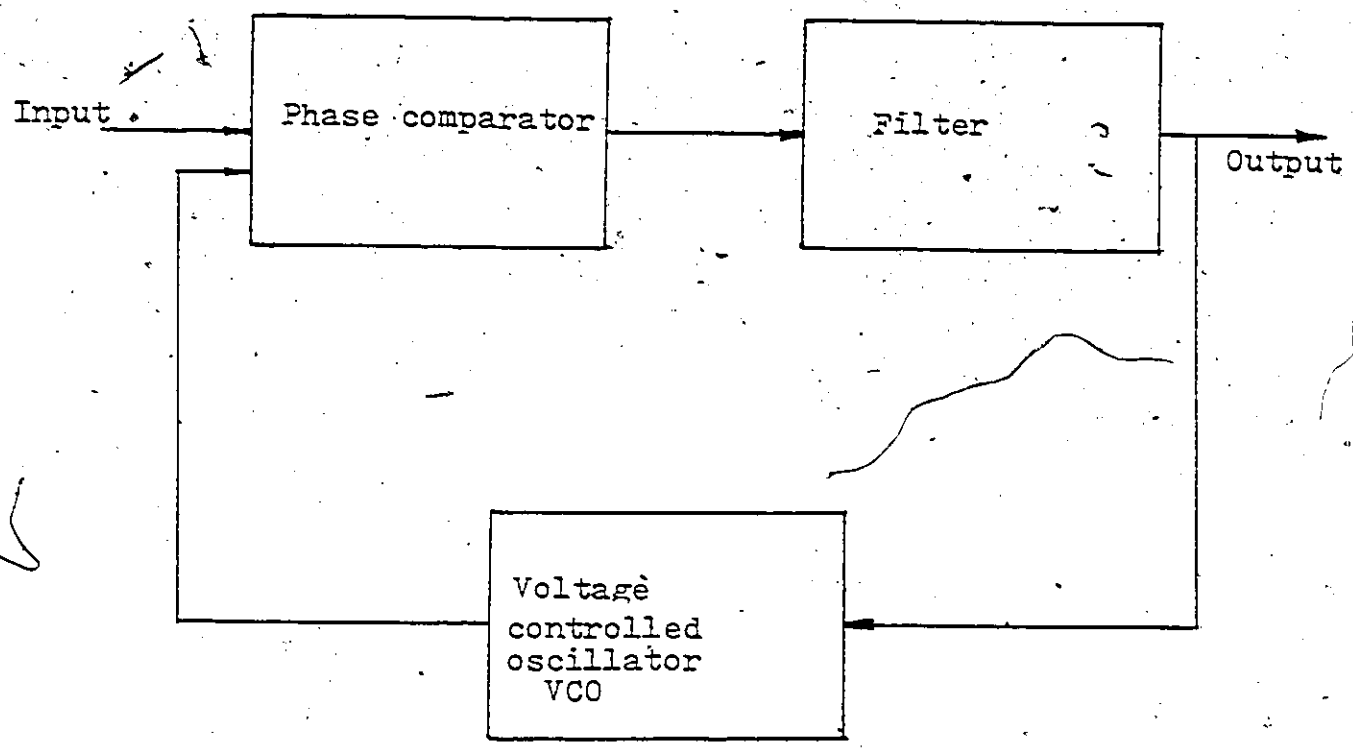


Figure 2.10 Second order phase-locked loop.

dealt with the theory of this device. Here, we shall only deal qualitatively with the principles of operation. The major difference between the first-order PLL and second-order PLL is that in the first order PPL the operating point makes a transition from the initial equilibrium position and stays at the new position until it establishes an equilibrium condition at that point. However, in the second-order phase-locked loop filter, although the operating point makes a transition, it does not stay permanently in the new position. Instead, it comes right back to the initial condition. Thus, the operating point of the phase comparator always hovers close to its initial condition. This difference between these two filters can be shown theoretically as the difference between the two transfer functions [13]. In the case of the first-order PLL, the transfer function resembles the transfer function of a lowpass filter. In the case of the second-order PLL, the transfer function resembles that of a bandpass filter. This difference is believed to enable the second-order phase-locked loop demodulator to be less sensitive to noise spikes than the first-order phase-locked loop.

In general, both the first-order and the second-order phase-locked loop demodulators achieve noise suppression capabilities through the feedback mechanism. Another type of discriminator which uses feedback technique is the FMFB

discriminator. Let us now look at this discriminator.

2.3.2 FM Feedback (FMFB) Demodulator

The circuit block diagram of a FMFB demodulator is illustrated in Figure 2.11. We notice that similar to the PLL demodulators, the FMFB demodulator also uses a VCO as the feedback element. We shall show that this arrangement also recovers the baseband signal, and give a qualitative explanation as to how it extends the threshold.

To explain the demodulation principle, let us consider the FM input signal which is composed of the signal plus noise. It is represented by

$$V_i(t) = R(t) \sin [\omega_c t + \phi_s(t) + \phi_n(t)] \qquad (2.25)$$

where $R(t)$ is the envelope of the carrier signal plus noise

$\phi_s(t)$ is the angular modulation due to the signal

$\phi_n(t)$ is the angular modulation due to noise, and

ω_c is the carrier signal

The other input signal to the multiplier is the output of VCO which is given by

$$V_{osc} = B \cos [(\omega_c - \omega_o)t + G_o \int_o^t v(t)\, dt] \qquad (2.26)$$

where ω_o is the centre frequency of the bandpass filter

G_o is the sensitivity of the VCO, and

B is the amplitude of the output signal from VCO.

We have assumed that the bandpass filter only responds to

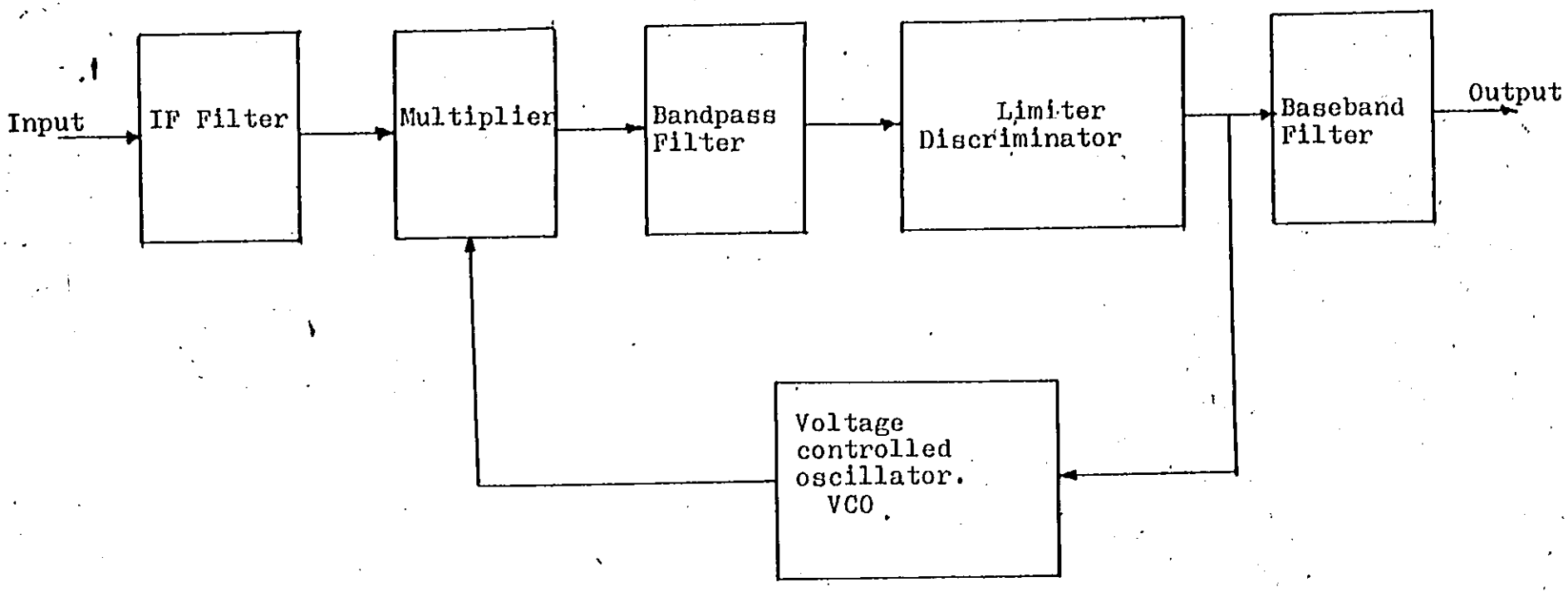


Figure 2.11 FMFB discriminator

the difference - frequency components of the multiplier. Also we shall not consider the amplitude term since it is of no consequence to the operation of the discriminator. Thus, the output of the bandpass filter can be written as

$$V_m(t) = \cos\left[\omega_o t + \phi_s(t) + \phi_n(t) - G_o \int_o^t v_o(t)\, dt\right] \tag{2.26}$$

This is the input to the discriminator whose function is to differentiate the phase angle with respect to time. Since the discriminator response to the frequency ω_o is zero, the discriminator output is,

$$V_o = K\left[\frac{d\,\phi_s(t)}{dt} + \frac{d\,\phi_n(t)}{dt} - G_o\, v_o(t)\right] \tag{2.27}$$

where K = discriminator constant

Solving for $V_o(t)$ we find

$$V_o(t) = \frac{K}{1 + K G_o} \frac{d}{dt}\left[\phi_s(t) + \phi_n(t)\right] \tag{2.28}$$

The output from the conventional discriminator is given by

$$V_{oc}(t) = K \frac{d}{dt}\left[\phi_s(t) + \phi_n(t)\right] \tag{2.29}$$

With these results we find the circuit arrangement shown in Figure 2.11 does indeed demodulate the FM signal. Comparing equation (2.28) and (2.29) we find the output of the FMFB discriminator is scaled down by a factor $(1 + K\, G_o)$. Since both signals and noise have been reduced by the same amount, there is no change in the signal-to-noise ratio between the

conventional FM demodulator and the FMFB demodulator.

Now let us pay some attention to the bandpass filter. We want to know whether we can reduce the bandwidth of the device without introducing any undue distortions. Using equation (2.28) we may write equation (2.26) as

$$V_m(t) = \cos \{\omega_o t + \frac{1}{1 + K G_o} [\phi_s(t) + \phi_n(t)]\} \quad (2.30)$$

If we compare this equation with equation (2.25), we observe that the feedback has suppressed the frequency deviation produced the signals ϕ_s and ϕ_n by a factor $1/1+KG_o$. Thus, the bandwidth of the bandpass filter can be reduced. To obtain the some feeling towards the bandwidth of this filter let us consider the following example. Suppose the modulation is sinusoidal, then

$$\phi_s(t) = \beta \sin \omega_m t \quad (2.32)$$

where β is the modulation index. Therefore, the phase of the signal present in the multiplied signal is

$$\frac{\phi_s}{1 + K G_o} = \frac{\beta}{1 + K G_o} \sin \omega_m t \quad (2.32)$$

Carson's rule gives us the minimum bandwidth required to process the signal without any undue distortion for a given modulation index. Upon applying this rule to the new modulation index in equation (2.32), we may calculate the bandwidth of the post multiplier bandpass filter. This is given by

$$B_p = 2\left[\frac{\beta}{1 + K G_o} + 1\right] f_m \qquad (2.33)$$

If we were to use a conventional discriminator, the bandwidth requirement of the IF filter would be

$$B_c = 2\,[\beta + 1]\, f_m \qquad (2.34)$$

Thus we express B_p as

$$B_p = \frac{[\beta/(1+KG_o)] + 1}{\beta + 1} B_c \qquad (2.35)$$

for example, $\beta = 9$ and $KG_o = 8$, then $B_p = 1/5\ B_c$. This reduction in bandwidth of the post multiplier filter has definite noise rejection properties.

Now let us look at these noise rejection properties carefully. If we look at the spectral density of the narrow-band noise out of the IF filter, we find that the bandwidth of narrow-band noise occupies the same bandwidth as the bandwidth of the IF filter. Now the effect of having a post-multiplier filter whose bandwidth is much narrower than the pre-multiplier filter, tends to suppress the noise components found in the band edges of the pre-multiplier filters. Thus under these conditions this discriminator has much less noise than the conventional FM discriminator. This is the reason for the extended threshold in the FMFB demodulator.

Using statistical analysis of noise, it is possible to calculate the threshold extension of the FMFB discriminator. Instead of getting involved in deriving these

formulae, we shall just use the results [13] to show how the threshold extension is achieved. The total number of noise spikes per second appearing at the output of a conventional discriminator is given by

$$N_{CON} = \frac{B_c}{2\sqrt{3}} \operatorname{erfc} \left(\frac{f_m}{B_c} \frac{S_i}{N_m}\right)^{1/2} + \frac{2\Delta f}{\pi} \exp \left[-\left(\frac{f_m}{B_c}\right) \left(\frac{S_i}{N_m}\right)\right] \qquad (2.36)$$

where B_c = bandwidth of IF filter

f_m = modulation frequency

S_i = input signal power

N_m = baseband noise power

In the case of the FMFB demodulator, the total number of spikes appearing per second is given by

$$N_{FMFB} = \frac{B_p}{2\sqrt{3}} \operatorname{erfc} \left(\frac{f_m}{B_p} \frac{S_i}{N_m}\right)^{1/2} + \frac{2\Delta f}{\pi(1+KG_o)} \exp \left[-\left(\frac{f_m}{B_p}\right) \left(\frac{S_i}{N_m}\right)\right] \qquad (2.37)$$

Now we have shown that B_p can be much less than B_c, and also we have noticed $(1 + K\ G_o) >> 1$. Thus we find $N_{FMFB} << N_{CON}$. This means that the number of noise spikes per second available at the output of the FMFB demodulator is much less than the conventional demodulator and thereby the threshold extension is achieved.

2.4 LIMITATIONS COMMON TO THESE DISCRIMINATORS

Having reviewed the operational principles of these discriminators, we can now outline some of the limitations

common to them. In Table 2.1 we have summarized these limitations.

The first common limitation to all the discriminators we have discussed so far, is that they suffer from bandwidth limitations. By bandwidth we refer to bandwidth up to 100%. None of the principles of operation, discussed so far, allows one to produce devices with such large bandwidths in the high-frequency range. By high-frequency range we mean the frequency range of 10 MHz - 1 GHz.

The second limitation is the linearity of these devices over their bandwidths. We have seen that it has been difficult for these devices to have reasonably good linearity over the device bandwidths. Consequently, their good linearity is maintained over a fraction of the bandwidth of the device. Our proposal is to produce linearities better than 5% over the entire bandwidth of the device.

The third limitation is the frequency of the operation of these devices. With the exception of SAW device-based discriminators, other types use inductors (L) and capacitors (C) in their networks to achieve the discrimination function. It is difficult to produce accurate LC networks in the frequency range of 10 MHz to 1 GHz. Thus, one always uses additional tuning ciruits to combat these inaccuracies. Furthermore, in this kind of

TABLE 2.1

PROBLEMS ASSOCIATED WITH EXISTING DISCRIMINATORS

1. Narrow Bandwidths
2. Poor Linearity over the Device Bandwidths
3. a. *Limitation of Frequency of Operation
   b. Limitation of Bandwidths at Higher Frequency of Operation
4. Limitation in FM Threshold
5. *Reproducibility
6. *Reliability
7. *Economy

* limitations from which SAW discriminators do not suffer.

situation, it is almost impossible to produce wideband frequency discriminators at these frequencies, using the above mentioned principles. Even the SAW device discriminators explained in this chapter are confronted with this problem, even though they do not face the problem of frequency of operation and the reliability problem that LC network type discriminators face.

The fourth limitation relates to the signal-to-noise performance. As we have pointed out, it is difficult to make a signal-to-noise performance assessment on all these devices since there is no data available. This is the case with SAW discriminators. However, we have some results to make a comparative assessment on the conventional discriminators, phase-locked loop discriminators and the FMFB discriminators. As far as the output SNR above the threshold is concerned, there is no real advantage to be gained by choosing one from another. However, using PLL demodulators or FMFB demodulators can give a threshold reduction of 5-7 dB over the conventional discriminators. The FM threshold of conventional discriminators occurs when the input carrier-to-noise ratio is about 11 dB. Therefore, both PLL and FMFB shows a significant improvement over the conventional discriminators. Nevertheless, this is a limitation especially if we can extend it by some other method.

The fifth, sixth and seventh limitations - namely reproducibility, reliability and economy - are not shared by SAW discriminators. However, the others do face these problems. When we dealt with the second limitation, we explained the reliability problem of LC networks. This problem has triggered problems of reproducibility, reliability and economics. The uncontrollable nature of the inaccuracies of the LC components creates the problem of performance reproducibility. Thus, one has to use wide tolerances in producing these devices. Furthermore, it is also difficult to guarantee their reliability of performance. Temperature, humidity, physical shocks and aging can change the tuning of these discriminators. Since at high frequencies these devices are generally narrowband, these changes in tuning conditions can have drastic performance effects. Finally, since these devices are produced on a one to one basis, and also due to expensive production and testing procedures, the cost of production becomes excessive. Thus, they are very expensive to buy. Furthermore, there is no room for cost reduction in the case of large scale production.

CHAPTER 3

PRINCIPLE OF SAMPLING-TYPE DISCRIMINATORS

In Chapter 2, we reviewed various discriminator principles, and outlined various limitations to which these discriminators are subjected. In a very broad sense we may divide these limitations into two categories. They are:

1. Limitations which are inherent to the principle of operation, i.e. linearity, bandwidth and noise performance
2. Hardware limitations, which are mainly device problems, i.e. frequency of operation, reproducibility, reliability and economy.

Under these conditions, whatever the solution that we adopt, it must be a dual solution. In other words, any new principle of operation must be realistic so that we should be able to realize this operation with the existing devices. In this Chapter, we shall develop another principle operation and define certain hardware requirements. In doing this, we shall outline how this principle overcomes some of the limitations encountered in the above mentioned first category.

3.1 BASIC CONCEPTS

The discriminators we discussed in Chapter 2 are continuous wave discriminators. By this we mean that the discriminators respond to a continuous wave by performing a differentiation on the phase angle with respect to time. Thus, they produce an output voltage proportional to the input frequency. As such, the discriminator will respond to noise as well as to signal. The result is that we find the output of the discriminator has two terms. The output is given by

$$V_o(t) = K_f\, m(t) + n_d(t) \tag{3.1}$$

where the first term $K_f\, m(t)$ is a demodulated signal term in which K_f is a constant, and $m(t)$ is the baseband signal. The second term $n_d(t)$ is that term produced by the discriminator due to narrow-band input noise. It is this term which is responsible for the deterioration of the output signal to noise ratio, and for the existence of the FM threshold. Naturally, if we minimize this term, then we will minimize its effect on the output SNR and consequently, extend the threshold. An obvious way to reduce $n_d(t)$ is to reduce the noise present at the input of the discriminator. The question is, "What is the best way to do this?". In Chapter 2 we have seen one method - namely the feedback method. Both the phase-locked loop and the FMFB demodulators use this method to extend the threshold even

though the details by which they achieve this are different. The answer to the question as to which is the best method is that we do not know. We cannot compare methods until we have another method of comparison.

The method that we are about to explain employs an optimization procedure. Basically, we argue that if we can optimize the input signal-to-noise ratio prior to demodulation then the demodulator sees an optimum signal and very little noise. Therefore, we should be able to extend the FM threshold to an optimum value. Before we embark on this argument in great depth we shall explain some communication concepts which will make this argument easier to understand.

The process of transmission is seen as a mapping process. Therefore, the process of receiving is viewed as an inverse mapping process. In Figure 3.1 we have illustrated the communication system. Here we map the message signal m(t) to a transmitted signal s(t). In the course of the transmission of signal through the channel noise is added to the transmitted signal. The noise is regarded as an additive white Gaussian noise of zero mean. The received signal r(t) consists both of signal and noise and is given by

$$r(t) = s(t) + n(t) \tag{3.2}$$

where n(t) is additive noise. In the receiver, the signal

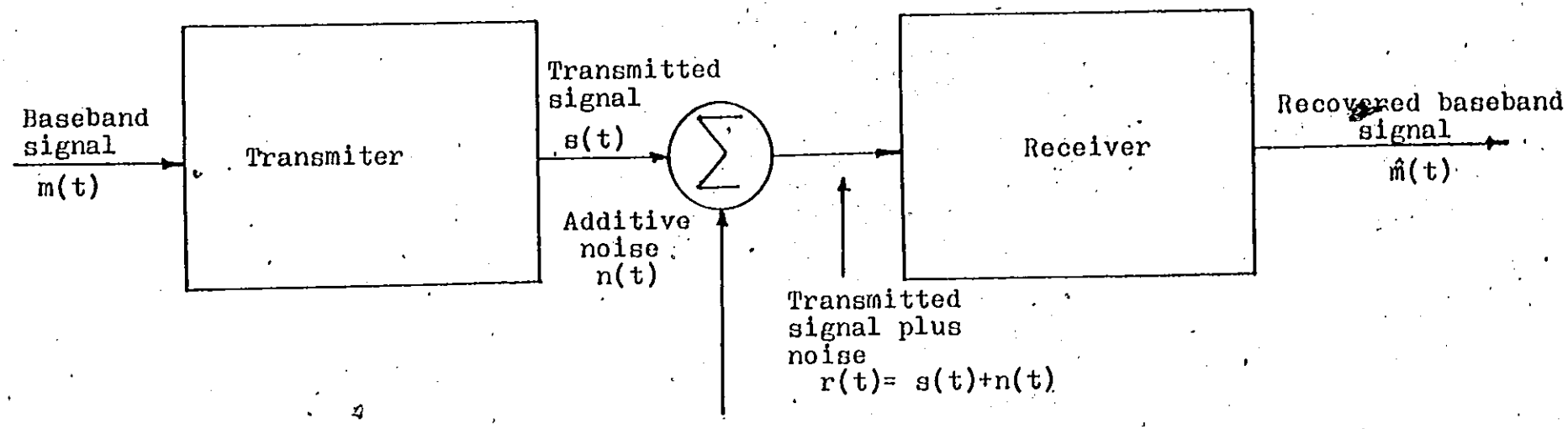


Figure 3.1 A communication system.

r(t) is processed and maps s(t) back to the message signal $\hat{m}(t)$.

Coming back to the discriminator principle we look at the discriminator as an inverse mapper. In contrast to the continuous mapping procedure the CW discriminators use, we shall employ a point-by-point mapping procedure in our sampling discriminator. In other words, we will sample the input signal, and process the signal sample by sample. We have illustrated the basic concepts of the discriminator principle in Figure 3.2. This piece-wise processing offers us many advantages. First, it allows us to optimize the signal-to-noise ratio of the sample effectively, assuming that within a sample there is only one frequency. We shall see how we achieve this condition later. Second, we can assemble the output signal in a suitable way to achieve good linearity. Third, we do not have to worry about bandwidth limitations such as we found in other discriminators. Since we are processing sample by sample, the FM signal bandwidth can be large as long as we know how to design the processor. On the assumption that we can take care of the know-how problem, we find that this arrangement will overcome the limitations of the first category, namely linearity, bandwidth and noise performance. Now, we have a general functional diagram (Figure 3.2) which can be used as a guide to build the operational principle.

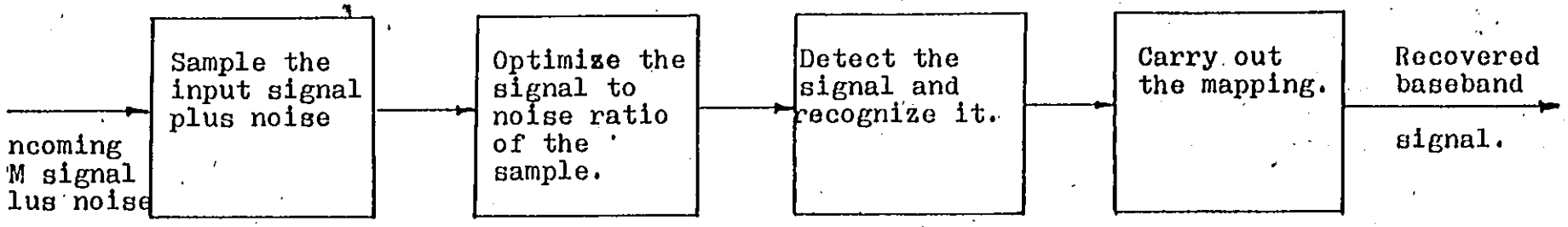


Figure 3.2 Functional diagram for illustrating the basic principle of the demodulator.

3.2 DEVELOPMENT OF THE PRINCIPLE OF OPERATION

We shall initially consider the input signal in the absence of noise. Our first step is to sample the incoming FM signal, by multiplying it by a pulse train with pulse duration T and period $1/f_s$. The quantity T is several times larger than the period of the carrier. Also, the sampling frequency f_s is much greater than the highest frequency of the baseband signal. The frequency of the FM signal within the sample is assumed to be constant. Thus we have transformed our FM signal into a discrete set of signals - $S_1, S_2, S_3, \ldots, S_i$ where each S_i has a different frequency. Thus, we find them to be orthogonal signals. That is to say

$$\int_0^T S_i \, S_j \, dt = 0 \qquad \text{when } i \neq j$$

and (3.3)

$$\int_0^T S_i \, S_j \, dt = \int_0^T S_j^2 \, dt = E_j \quad \text{when } i = j$$

where E_j is the energy of the jth signal.

Our next problem is that having received a particular signal S_i, how are we to recognize the signal? The solution to this problem is indicated in equation (3.3). First we must generate a set of replicas of the signals. Then we must multiply each incoming signal with its replica and integrate. At the output of the integrator we find E_j which is the energy of that particular signal. Assuming for the moment, that the energy of each signal is different, we have

been successful in translating the input signal to a discrete set of energy levels. In other words we have the means of recognizing the signals by recognizing the energy levels. Assuming, of course, we have prior knowledge of signals and corresponding energy levels. We shall illustrate this signal processing scheme in Figure 3.3. Accordingly, we find that if we have M signals to be processed, we need M multipliers and M signal replicas stored in the system. Also we need M integrators. Since we are processing only one signal at a time we may reduce this to a single multiplier and an integrator as shown in Figure 3.4. However, reducing the multipliers to one generates another problem. That is, how are we to choose the particular replica signal to compare with the incoming signal? In other words, we have a synchronization problem.

Let us leave these problems aside and look at this structure. We recognize this as a correlation receiver. It has been shown [14] that an equivalent of a correlation receiver is a matched filter receiver. In Figure 3.5 we have illustrated the matched-filter receiver. Here we feed the input signal plus the noise into a bank of matched filters. The matched filter is an optimum filter. As such, it will optimize the output signal-to-noise ratio. The output of the matched filter depends only upon the ratio of the signal energy to the spectral density of white noise

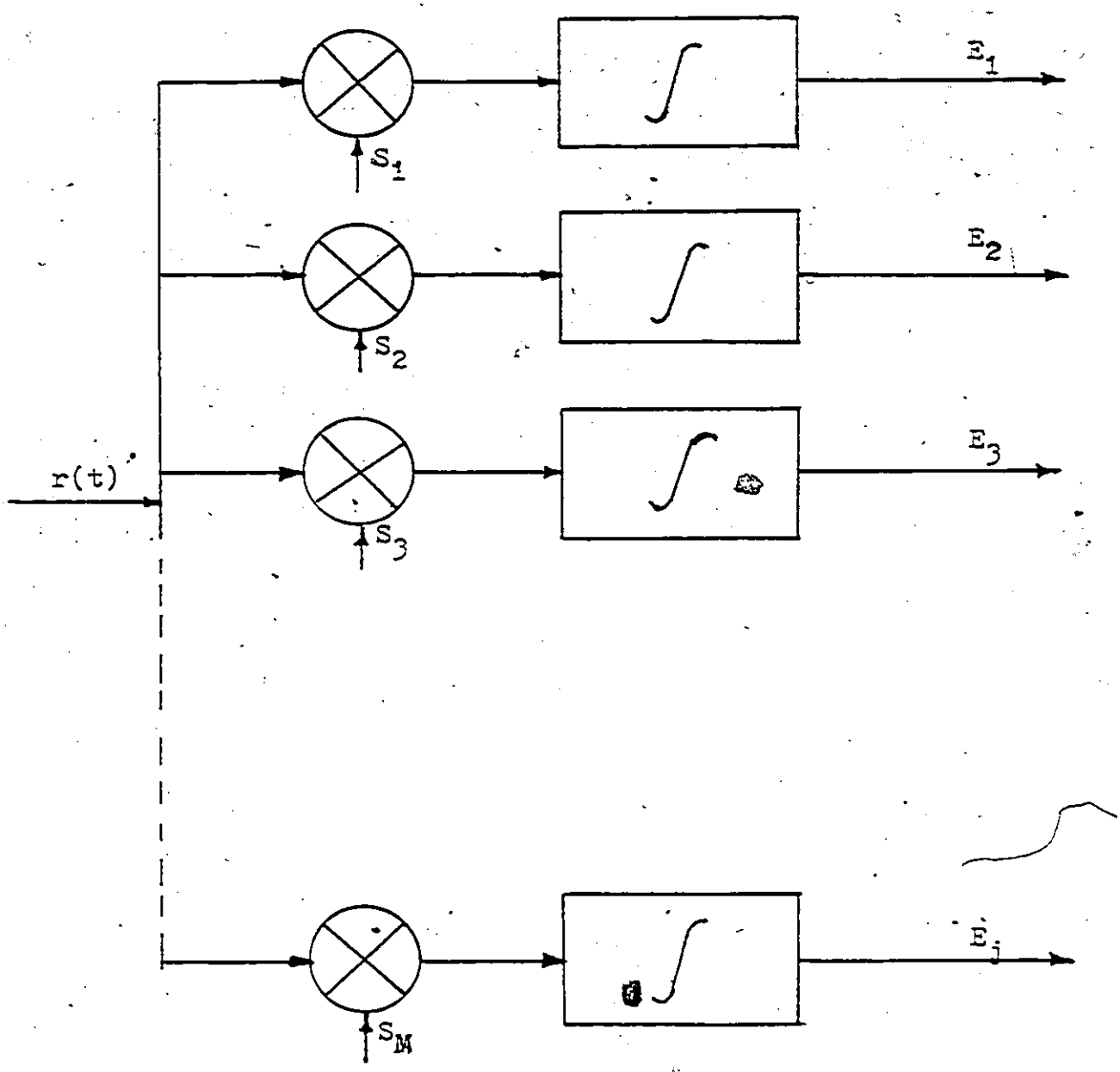


Figure 3.3 Signal processing scheme using correlation.

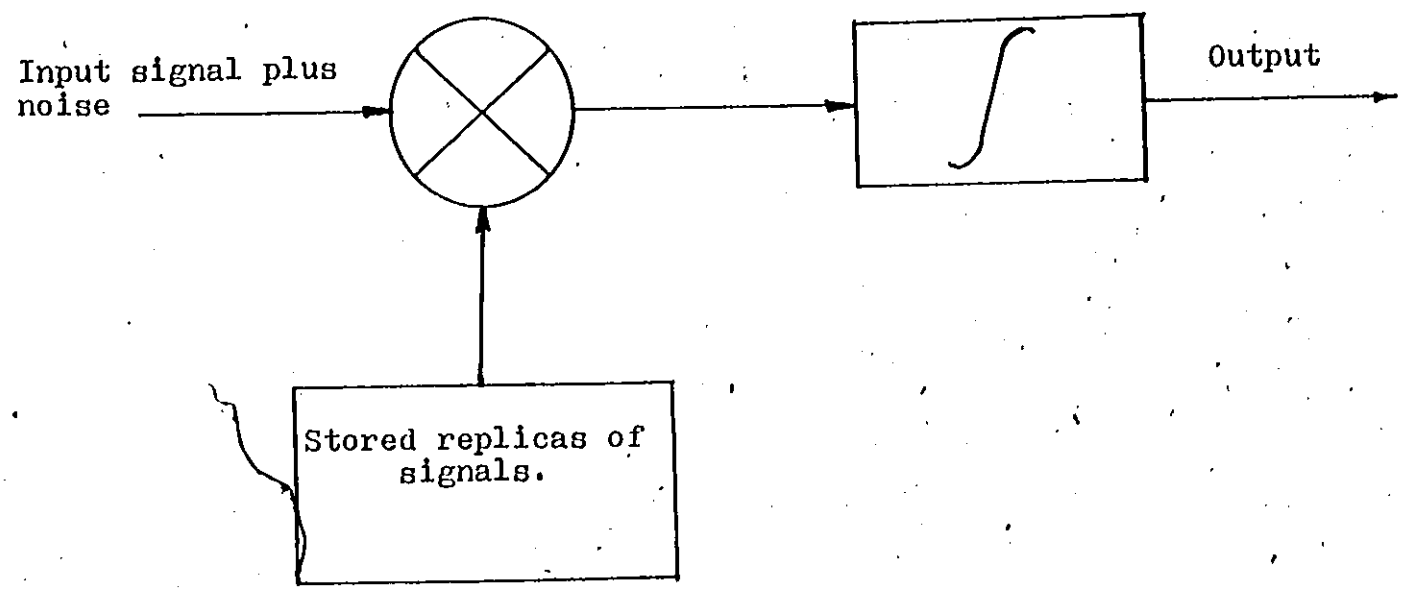


Figure 3.4 Reduced correlation receiver.

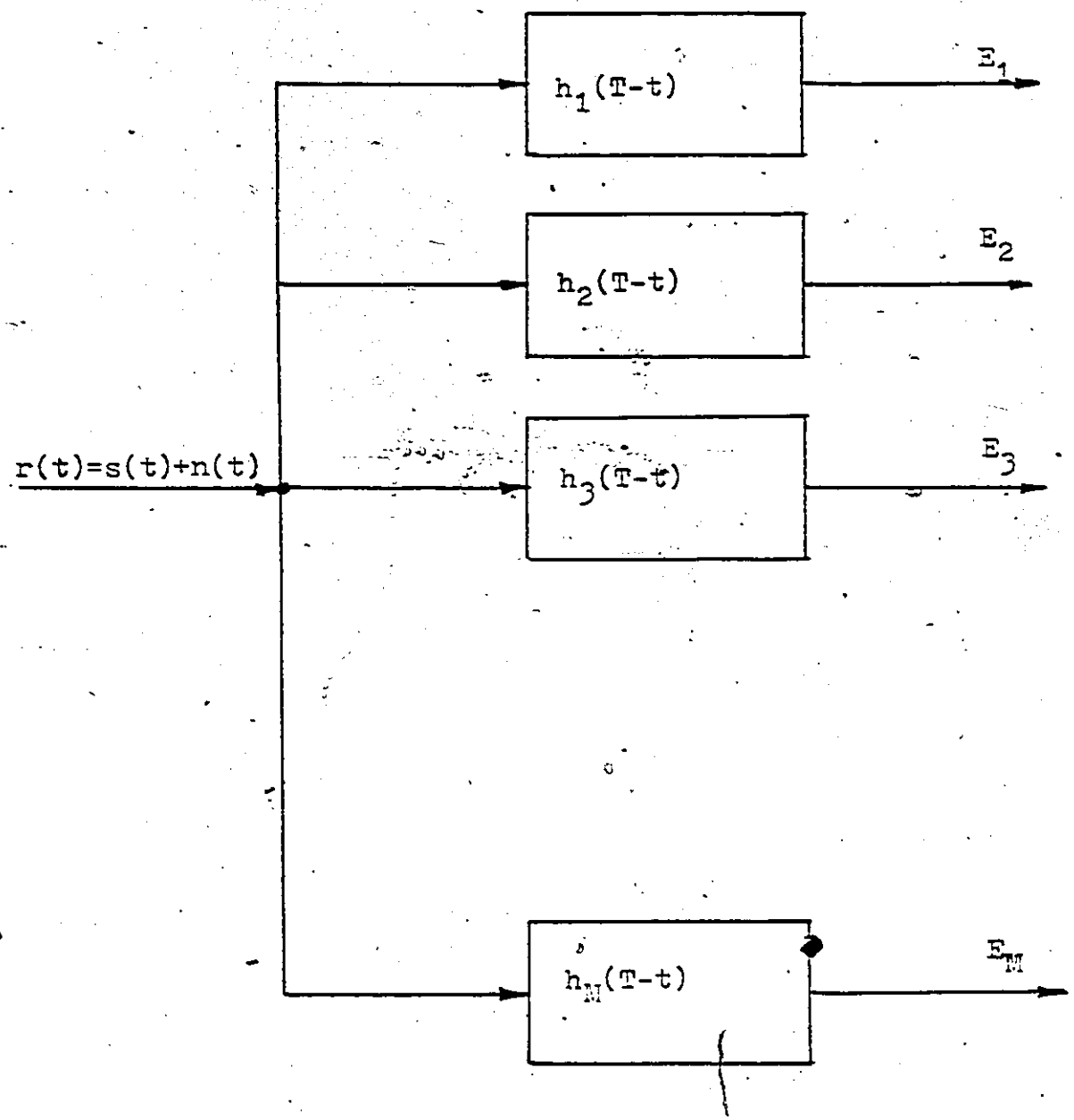


Figure 3.5 Matched filter receiver.

present at the input of the filter. Thus it is given by

$$(SNR)_{o_{max}} = \frac{2E}{N_o} \tag{3.4}$$

where a spectral density of zero mean Gaussian white noise is $N_o/2$. In this receiver, we have M matched filters. The reason why we have M matched filters is that as the name implies, each filter is matched to a particular input signal. That is to say, that the impulse response of a matched filter is a time-reversed and delayed version of the input signal $S_i(t)$. Therefore, the impulse response of a matched filter is given by

$$h(t) = S_i(T-t) \tag{3.5}$$

It is shown, when all the input signals $S_i(t)$ enter the filter by the time t=T, maximum signal amplitude is delivered to the output.

Both the correlation receiver and the matched filter receiver give an output corresponding to the energy in the signal. We said that we can solve the problem of the signal identification by using a correlation receiver or an equivalent matched filter receiver, providing that the signal energy is different from pulse to pulse. This means we have assumed some sort of amplitude scaling in the input signal. However, our input signals do not have any such amplitude variation. It is a constant amplitude signal. As such, the output of the matched filter receiver or the

correlation receiver will be constant amplitude signals. This means that the amplitude scaling we wanted to apply as the means of recognizing the signal is not possible. The obvious question is "Can we use some sort of phase measurement to recognize the signal?" The answer to that is also no, since both the matched filter and the correlator lose phase information. However, there is another way to solve the signal recognition problem. That is by time scaling the signal.

3.2.1 Time Scaling the Signal

It is possible to arrange our signals to arrive at the matched filter at different times. In effect, what we do is to delay each signal by a unique delay factor assigned to the particular signal. This can be easily accomplished by incorporating a set of delay lines into the matched filter receiver or to the correlation receiver. The modified matched filter and correlation receivers are shown in Figures 3.6 and 3.7. The delays of the signal are given by τ_i and the delayed outputs from the matched filters are given by $E_j(\tau_i)$. Now we have a definitely recognizable signal set. We recognize the particular signal by detecting amplitude, and measuring time delay.

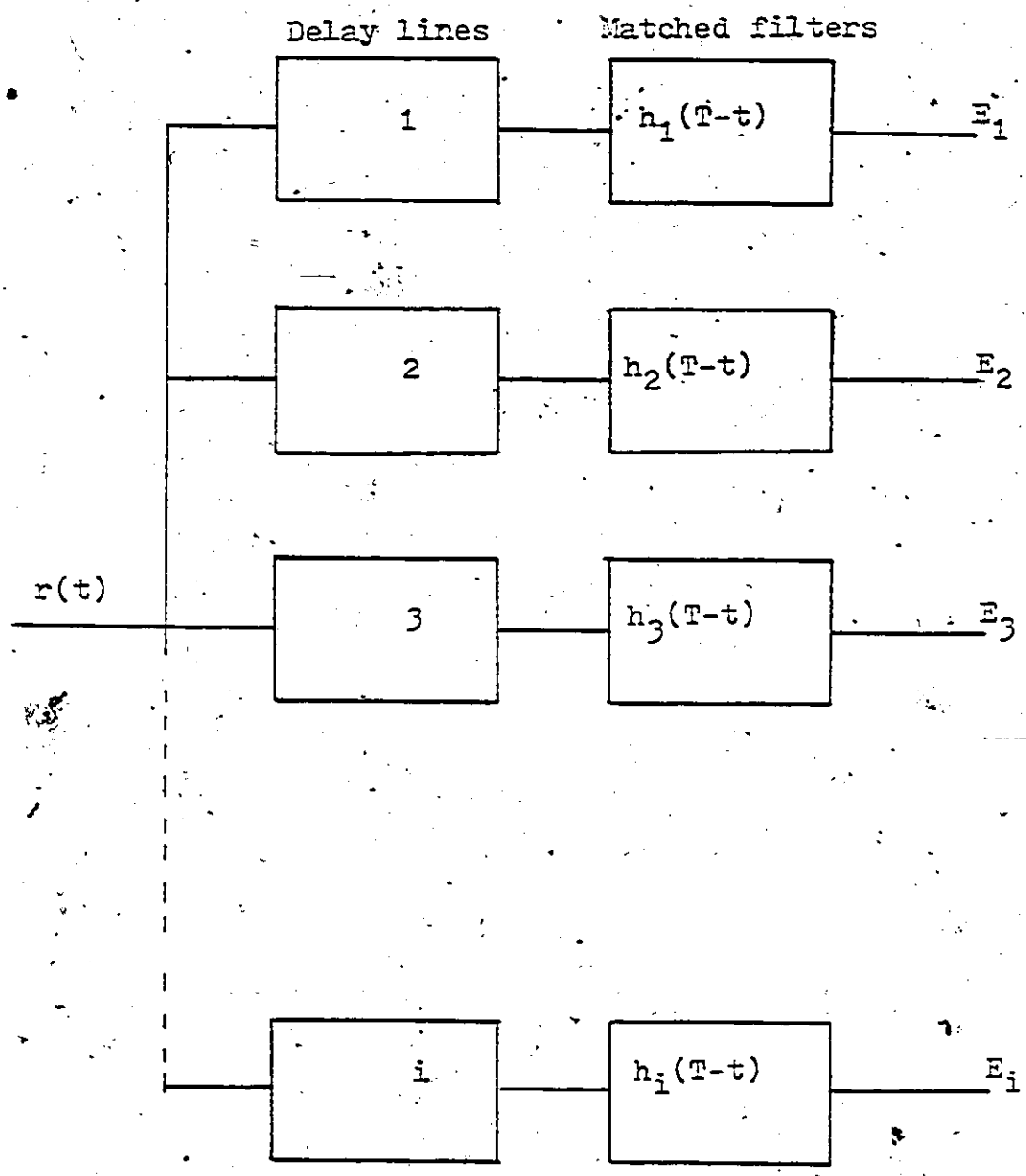


Figure 3.6 Modified matched filter receiver

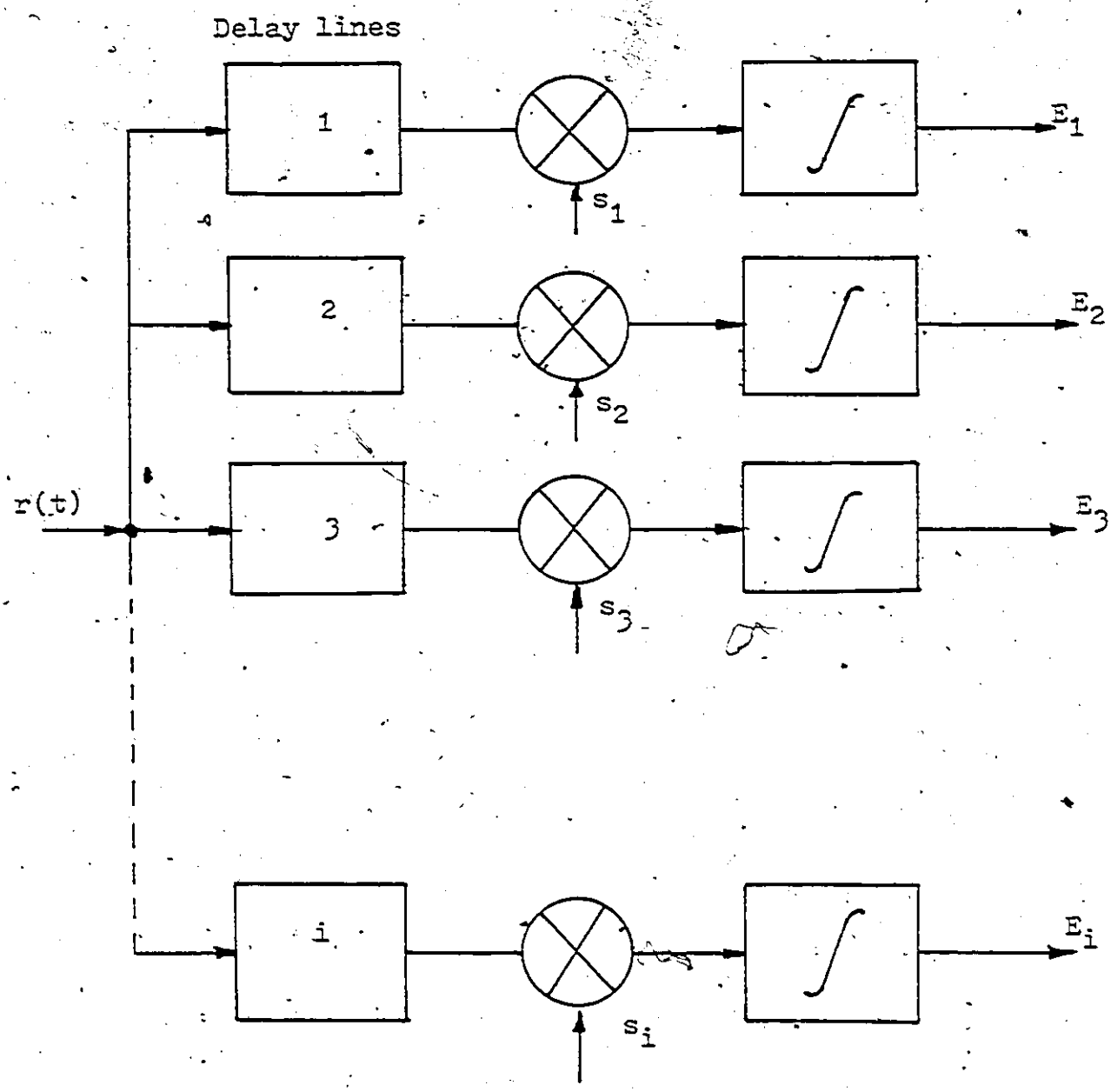


Figure 3.7 Modified correlation receiver.

3.2.2 Baseband Signal Mapping at the Output

Having optimized the signal-to-noise ratio of the received signal sample and identified the signal, our next problem is, how do we assign a unique value to the output so that we can reconstruct the baseband signal? One quality which is required for minimum distortion signal reconstruction, is the linearity. This means that the input frequency versus the output voltage transfer characteristics should be linear. Since we are processing sample by sample, it is possible to measure the time difference between the time that the signal enters the matched filter receiver or the correlation receiver and the time that it takes to get out of the receiver. Having done this measurement, we can allocate a particular voltage value to the output to preserve the linearity.

One method of doing this is initially to generate a constant amplitude rectangular pulse whose pulse width is proportional to the time measurement. Assuming that the time measurement started at a time $t = 0$, our rectangular pulse is given by

$$f(t) = \begin{cases} A & 0 < t < \tau_i \\ 0 & \text{elsewhere} \end{cases} \tag{3.6}$$

where A is the amplitude of the pulse. Let us now integrate this pulse. The output voltage is

$$V_o = \int_o^{\tau_i} f(t)\, dt = K\, A\, \tau_i \qquad (3.7)$$

where K is the constant of the integrator. We found that the output voltage is a linear function of the time delay.

Another way to solve this problem is to generate a ramp whose amplitude is linear with the time starting at t=0. We then use the generated rectangular pulse to sample the linearly varying amplitude wave form.

We have now completed the development of the principle. We have sampled the incoming FM waveform and succeeded in optimizing the signal noise ratio of this sample with an optimum filter. We have recognized the particular signal and assigned the baseband signal component which corresponds to that signal. We find that this principle of operation should produce an extension of threshold since the principle is very similar to the optimum receiver principle [14]. In fact, it has a close resemblance to the M-ary signal receiving system.

3.3 DISCRIMINATOR AS AN M-ARY RECEIVER

A considerable amount of effort has been spent by numerous authors [15,16,17], to study the theoretical aspects of M-ary frequency-shift keying (FSK), signal transmission and detection problem. We will outline the concepts to show what we have developed is in fact very

similar to the M-ary receiver principle. If we can draw up a one-to-one correspondence between the M-ary receiver principle and our discriminator principle, then we shall be able to use all the theoretical analyses developed for the M-ary principle to our discriminator principle.

In the basic principle of operation of an M-ary transmission system, at the transmitter a stream of equally likely message signals M_1, M_2, M_3, ..., M_m is modulated to generate M distinct wave forms of S_1, S_2, S_3, ..., S_m. These signals are then transmitted through the channel. The noise in the channel disturbs the transmitted signal by adding Gaussian white noise. Thus at the receiver input we have

$$\bar{r}_i = \bar{s}_i + \bar{n}_i \tag{3.8}$$

where $\bar{n}_i$ is the random noise vector. It is the responsibility of the receiver to process $\bar{r}_i$ and recover the base-band message. Now the signal $\bar{s}_i$, similar to our sampled signal, has a finite duration. Also, like our signal it is an orthogonal signal defined by the same set of equations as equation (3.3) which is

$$\int_0^T S_i \, S_j \, dt = 0 \qquad \text{when } i \neq j$$

$$\int_0^T S_i \, S_j \, dt = \int_0^T S_j^2 \, dt = E_j \qquad \text{when } j = i$$

Having defined this common identity between the two signal sets, one can proceed to use the same principle to solve the receiver problem. Usually, an M-ary FSK receiver is a matched filter receiver.

In fact, this identification of the equivalence between principles shall give us some other important information. M-level FSK is a M-dimensional signaling scheme in signal space. It is interesting to compare M-level FSK with other M-level signaling schemes. M-level amplitude modulation, such as M-level PCM, is a one-dimensional system in signal space. In other words, the signal points are directly spaced along one line. As M increases, the spacing between points decreases, resulting in an increase in probability of error. M-ary phase shift keying (PSK), on the other hand, is a two-dimensional system with M points distributed over a 2π radian range. Obviously, with an increase in M, closer packing of signals takes place. Therefore, the probability of error increases with increasing M in an M-level system. In contrast to both of these, M-level FSK does not increase the probability of error when M is increased. Increasing M in FSK results in an increase in dimensionality. It is still possible to maintain the spacing between points even when one increases M. Thus processing M signals in M-level FSK does not increase the probability of error.

3.3.1 Probability of Error in an M-ary Receiver

Considering the receiver structures, we wanted to know which is the optimum receiver structure that will process the received signal most effectively. The received signal vector was given by equation (3.8) which is

$$\bar{r} = \bar{S}_i + \bar{n}_i$$

where $\bar{S}_i$ is the signal vector and $\bar{n}_i$ is the random noise vector. It is possible, that at times, the noise vector is sufficiently large enough to force the receiver to make an error. What we want to find out is, if given a signal set, what is the probability that the receiver will make an error? Thus, the probability of error gives us a measure of the accuracy of the operation of the reciever.

The probability of error calculation for both coherent and incoherent M-ary FSK signal receivers has been studied by Arthurs and Dyms [16]. Their findings show that there is hardly any difference between the probability of error in the incoherent M-ary FSK and the probability of error in the coherent M-ary FSK. For the coherent FSK they have derived the upper bound and lower bound. For given M message wave forms, perturbed by zero mean additive white Gaussian noise with two sided spectral density of $N_o/2$, the average probability of error is bounded by

$$\frac{1}{2} \operatorname{erfc} \frac{E}{2N_o} \leq P_e \leq \frac{M-1}{2} \operatorname{erf} \frac{E}{2N_o} \tag{3.9}$$

As can be seen, the probability of error depends solely on the ratio of signal energy to spectral density of noise. Nuttall [17] compared the M-ary transmission scheme with an equivalent binary transmission scheme. His comparison indicated a reduction in overall message probability of error as M increases. He claims that as M increases beyond the bound, the overall probability of error can be shown to approach zero. This is based on the assumption that the channel capacity is not exceeded.

3.4 HARDWARE CONSIDERATIONS AND REQUIREMENTS

The critical signal processing area of the discriminator is the SNR optimization section. Also it is the difficult section to implement. We are concerned with discriminator operation frequencies above 10 MHz and up to 1 GHz, with bandwidths up to about 100%. The linearity of the transfer characteristic is to be less than 5% over the total bandwidth of the device. As far as the principle of operation is concerned, large bandwidths mean large signal sets or large M. This in turn creates the difficulty in implementation.

In the case of the M-ary matched filter receiver we need M matched filters. If we use conventional LC networks, our receiver will be complex and cumbersome. Furthermore, it is difficult to produce such banks of matched filters in

the frequency range of interest. Quite apart from these, we will have problems of accuracy, reproducibility, reliability and economy.

In the case of correlator receivers, as we pointed out earlier, we would need M multipliers. It is possible to reduce this to a single multiplier. However, such reduction would introduce a synchronization problem. Attempting to solve this synchronization problem, using methods such as recirculating loop methods would introduce long processing time. We wish to avoid long processing time since it means slower sampling rates. Slower sampling rates result in limiting the baseband frequencies that can be sampled. We have the additional problem of generating M signal replicas. These replicas should be accurate and need to be stored in the system by some means. The access time required to retrieve these signals from the store is yet another factor which will affect the process time.

Both types of receivers need M delay lines. These delay lines should be capable of operating in the frequency range specified. Electromagnetic delay lines in this range would be too long and bulky. In addition, stray fields may distort the received signals further. We want to avoid complicating matters any further with these distortions since the problem is complicated enough as it is. If the insertion loss and propagation losses vary from delay line

to delay line, we would need M amplifiers with different gains to overcome these losses. Having such amplifiers would again complicate matters. It is quite clear that electromagnetic wave delay lines are not the most suitable devices for this application.

Ideally, it would be desirable and convenient indeed, if we could build all these functions into one device. In other words, in the case of matched filters, if we can incorporate all the impulse responses and delay lines into one device, we should be satisfied. In doing so, we should like to keep the losses constant, or at a minimum level. In the case of correlator receivers, it would be desirable to have a self-synchronizing minimum process time correlator. Furthermore, it would help us a great deal if we could have the bank delay lines of minimum losses or constant losses incorporated into the device, and of course, we would want these devices to operate in the frequency band 10 MHz - 1 GHz. Also we would want them to be reproducible, reliable, and economical. Having demanded the above mentioned features, we would further demand that the device should be compact and small.

It is quite clear that these demands are not the easiest to meet. In fact, if we were to use conventional network devices, one might go as far as to say that it is impossible to meet the vast majority of these demands, let

alone all of them. Fortunately, one could use surface acoustic wave technology to produce devices that will meet all these demands. This topic shall be considered in the next Chapter.

Having defined the requirements of the central signal processing section of the discriminator, we shall specify one single requirement for other peripheral electronic circuits such as sampling circuit, signal detector circuit, and baseband signal assignment circuits. These circuits should be built with non-critical circuit elements and should be able to integrate all these circuits into one chip eventually. Thus, we may be able to exploit the cost reduction qualities of integrated circuit technology.

CHAPTER 4

SURFACE ACOUSTIC WAVE DEVICES

In Chapter 3 we formulated the principle of operation of the discriminator. Having done this, we considered the hardware requirements for implementation of the discriminator. We realized that these demands cannot be met with the conventional network type devices and said that surface acoustic wave devices can meet these demands. In this Chapter we shall review the surface acoustic wave device principles to show how these demands can be met.

4.1 FUNDAMENTALS OF SURFACE ACOUSTIC WAVES (SAW)

Surface acoustic waves are mechanical or elastic waves that travel on the surface of any solid. The propagating wave is confined to the region of the solid immediately below the free surface. The frequency range of interest for surface acoustic waves is from 10 MHz to 2 GHz. If the material that the SAW propagates in is piezo-electric, then an electric field associated with the propagating stress wave will be in phase with the SAW. Since the wave is on the surface, it can be easily tapped [18], manipulated [19], and guided [20,21,22]. In general,

the large majority of materials used for SAW devices are single crystal materials. Propagation losses in these materials are negligible up to a few hundred megahertz. In general, the losses in single crystals are much less than the losses of polycrystalline materials of the same kind, and the loss increases as the square of the frequency.

In an unbounded isotropic material, there are three elastic plane waves that can propagate in any direction. One is polarized along the propagation vector, called the longitudinal bulk wave, and the other two are polarized perpendicular to the propagating vector, and are called shear waves. Shear wave velocity V_t is less than the longitudinal bulk wave velocity V_l. In a bounded solid these modes can be reflected or excited at the bounding surfaces (mode coupling). However, the surface wave velocity on the free surface is much less than the shear wave velocity. Due to this, the reduced velocity of the SAW, the wave travelling on the free surface cannot phase match, and thereby couple into bulk waves. However, if the surface is perturbed, then such coupling could occur. A velocity diagram for the complete range of possible ratios is illustrated in Figure 4.1. Surface acoustic wave velocities for the materials of interest are in the range from 1000 to 6000 meters/second. This is about 5 orders of magnitude slower than electromagnetic wave velocities in

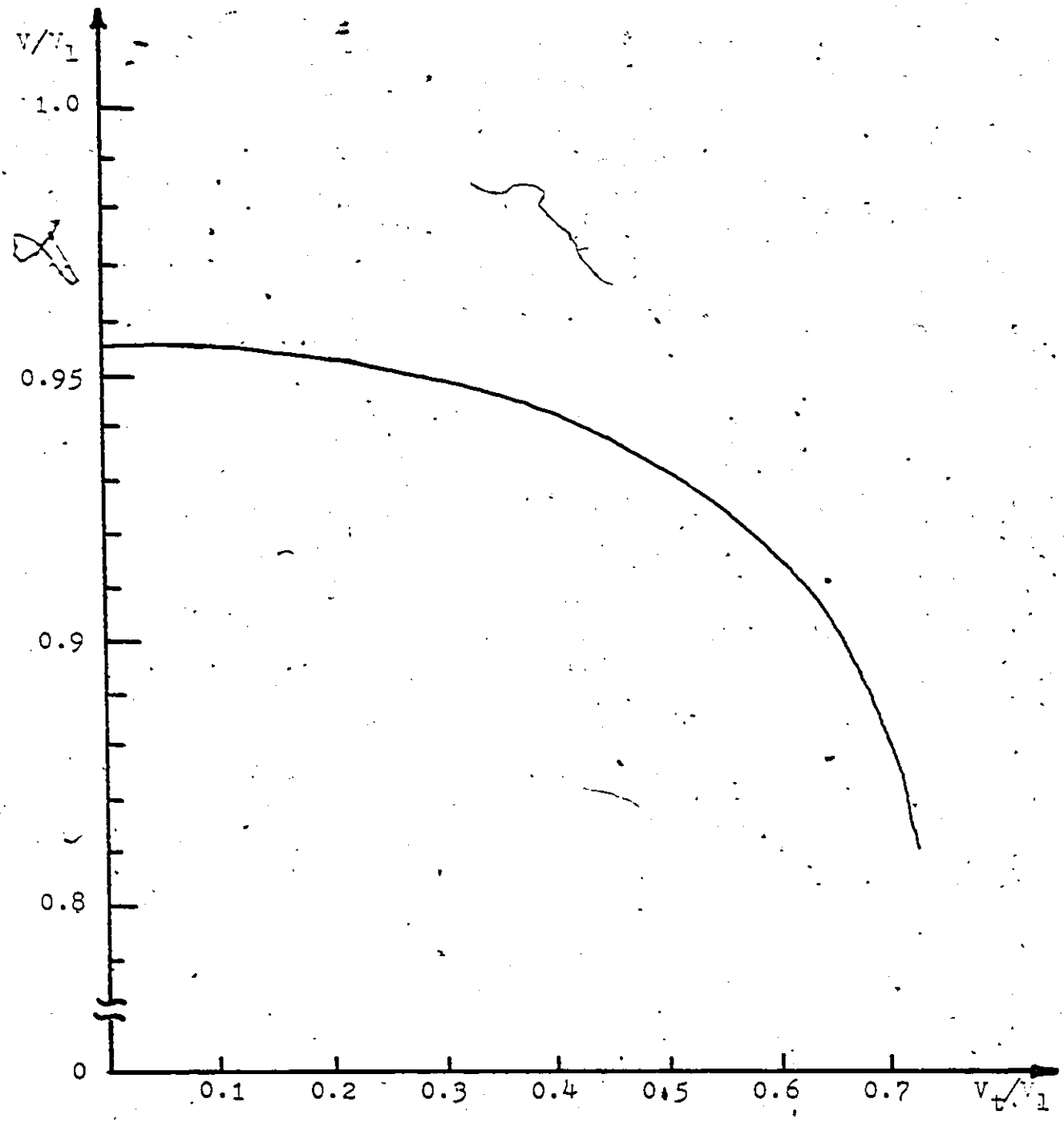


Figure 4.1 Normalized SAW Velocity
vs
Normalized Shear Wave Velocity (41)

air. Furthermore, the SAW velocity is independent of frequency.

Single crystal materials that are used for SAW devices are inherently anisotropic to some degree. In other words, the surface acoustic wave velocity changes with the crystal cut and with the direction of the propagation vector on that surface. The direction of the propagation vector can be fixed by the orientation of the source transducer, and it is usually chosen along the extreme of the phase velocity vs. angle curve. This direction is known as the pure-mode axis. Along this axis, energy transport is in the same direction as the propagation vector. The energy transportation by surface waves is expressed by the mechanical Poynting vector [23] which is the acoustic equivalent to the electromagnetic wave Poynting vector. The variation of acoustic Poynting vector with depth is shown in Figure 4.2. We notice that all of the energy in the surface wave is carried close to the surface. The depth of penetration is only of the order of a wavelength.

These are the fundamental properties of surface acoustic waves. Next, we shall consider the way in which surface acoustic waves are generated and received.

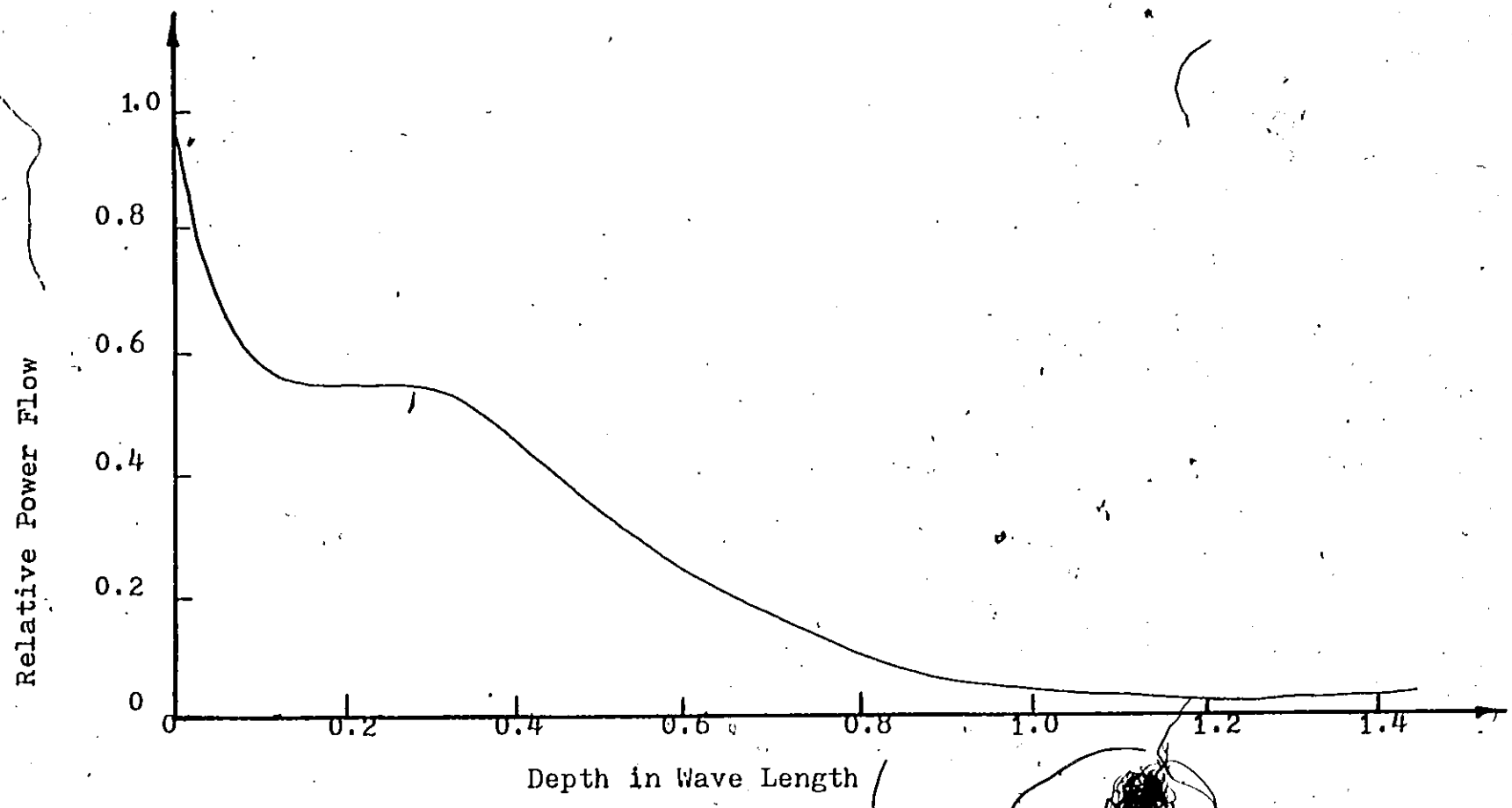


Figure 4.2 Power flow per unit area normal to the sagittal plane as a function of depth below surface (41)

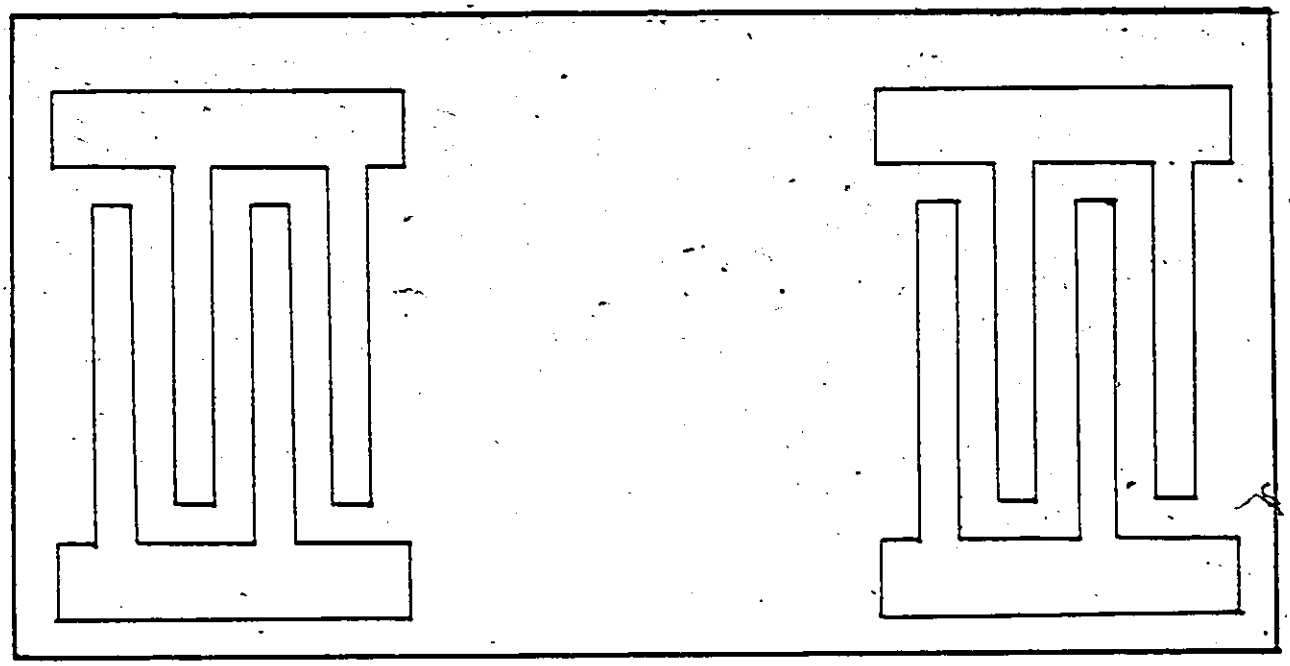
4.2 INTERDIGITAL TRANSDUCER [5]

The most efficient and frequently used means of generating and detecting surface acoustic waves for electronic application is the interdigital transducer (IDT). As illustrated in Figure 4.3, the simplest IDT consists of electrodes placed on the surface of a piezo-electric substrate. The alternate electrodes are spaced periodically and are interconnected by two bus-bars. Let us now see how this device generates and detects surface acoustic waves.

4.2.1 Generation and Detection

Suppose we apply a DC voltage between the bus-bars. As shown in Figure 4.4, an electric field will be induced in the material due to this appled voltage. Since the material is piezo-electric, due to piezo-electric coupling a periodic stress of periodicity λ will be set up in the material. It should be pointed out that in a piezo-electric material, stress induced due to an electric field or vice-versa is linearly proportional to that field strength.

If we now replace this DC voltage with a time-varying voltage with the same mechanism, a time varying periodic perturbation will be induced in the substrate. This perturbation will generate surface acoustic waves. When the frequency of applied voltage is at some frequency given by



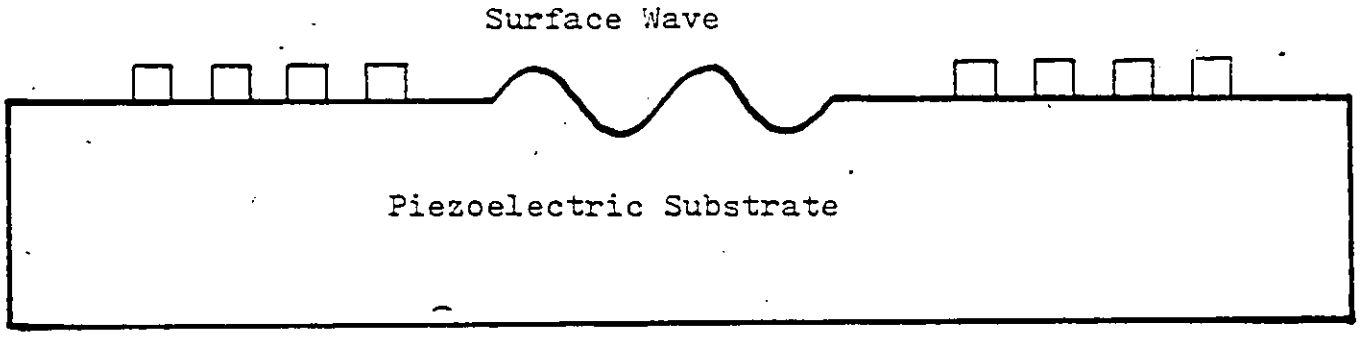


Figure 4.3 Schematic Representation of a SAW Device

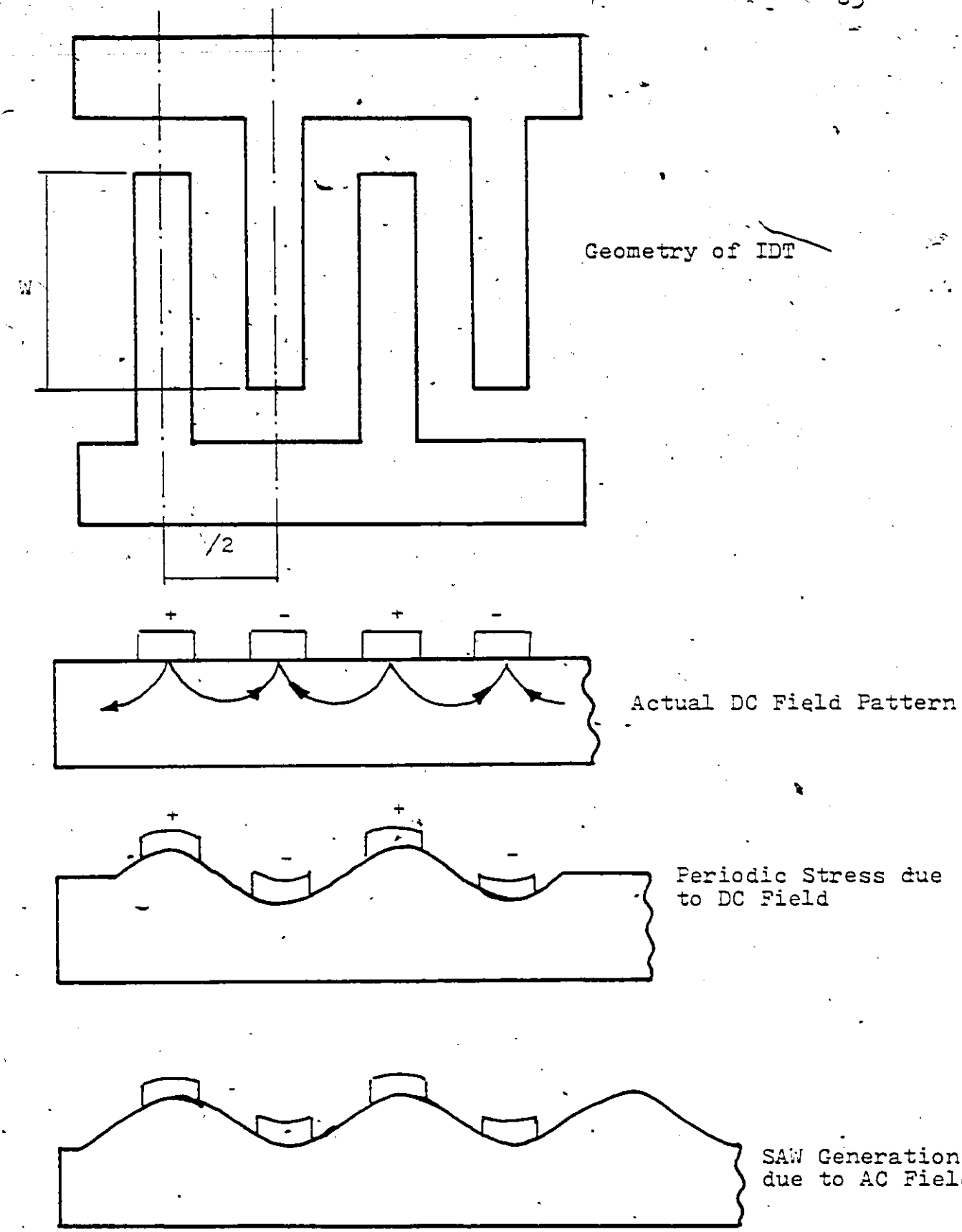


Figure 4.4 Physical Mechanism of SAW Generation by IDT

$$f_o = \frac{V}{\lambda} \tag{4.1}$$

where f_o is the frequency and V is the surface wave velocity, the waves from each source will add in phase to generate a surface wave of width W. The wave that propagates from left to right will add in phase to increase in amplitude as it travels across the transducer as it leaves the transducer from the right, it continues to travel as a free surface wave. A similar wave will be generated in the same way, and will travel to the left. Thus IDT is a bi-directional device.

The receiver transducer will operate in the inverse mode. The incoming elastic stress wave will induce a periodic electric field across the receiver transducer fingers. These fields will then be integrated by the pads to generate the original time varying voltage of frequency f_o.

4.2.2 Models for Interdigital Transducers

The electric field produced by the fingers have a very complex field structure. Tangential and vertical components of this field have different spatial distributions. Both these field components are large at the finger edges. Through the various piezo-electric coupling constants involved, these field components couple to a Rayleigh wave which itself has a complicated mechanical

displacement distribution. Therefore, it is not very obvious what sort of model one must use for effective source distribution. To complicate matters even further, there are various second order effects present. We shall discuss some of these later. Because of this complicated situation, there is no complete transducer model in existence which takes every aspect into account. Much effort has been spent on this problem ever since the birth of the interdigital transducer device. However, these efforts have produced various approximations and models [24, 25,26] which fulfill partial requirements. Choosing the approximations and the model is completely dictated by the dominant specification of the device being designed and the kind of second order effects that need to be included in the design. Frequently, the design routine involves more than one model; one for analysis, and the other one for synthesis. Another factor which affects the choice of the model is the affordable computer capacity for related calculations. The required computer capacity increases rapidly with the sophistication of the model.

One of the models widely used is to so-called Mason model or the equivalent circuit model [27]. With this model the IDT can be visualized as a 3-port network. The circuit consists of two acoustic ports and one electrical port. We shall review this model now.

4.2.2.1 The Equivalent Circuit Model

The field patterns of an IDT are illustrated in Figure 4.5. These fields can be separated into two components; one vertical component and the other a horizontal component. We may approximate the particular device using either of these components. In other words, we have two field models to choose from. These models, known as crossed-field and in-line models, are illustrated in Figures 4.5(a) and (c). The criteria for choosing a particular model depends on the particular field component dominant in coupling to the Rayleigh waves. The dominant field component depends on the metallization ratio and the electro-mechanical coupling constant. The metallization ratio is given by

$$\eta = \frac{a}{\ell} \tag{4.2}$$

where a = width of the finger and $\ell = \lambda/2$. The electo-mechanical coupling constant is given by

$$K^2 = 2\,\frac{\Delta V}{V} \tag{4.3}$$

where ΔV is the difference in velocity between the free surface wave velocity and the velocity of the same material when the piezo-electric fields are shorted with a thin metal layer. The most widely used piezo-electric material is

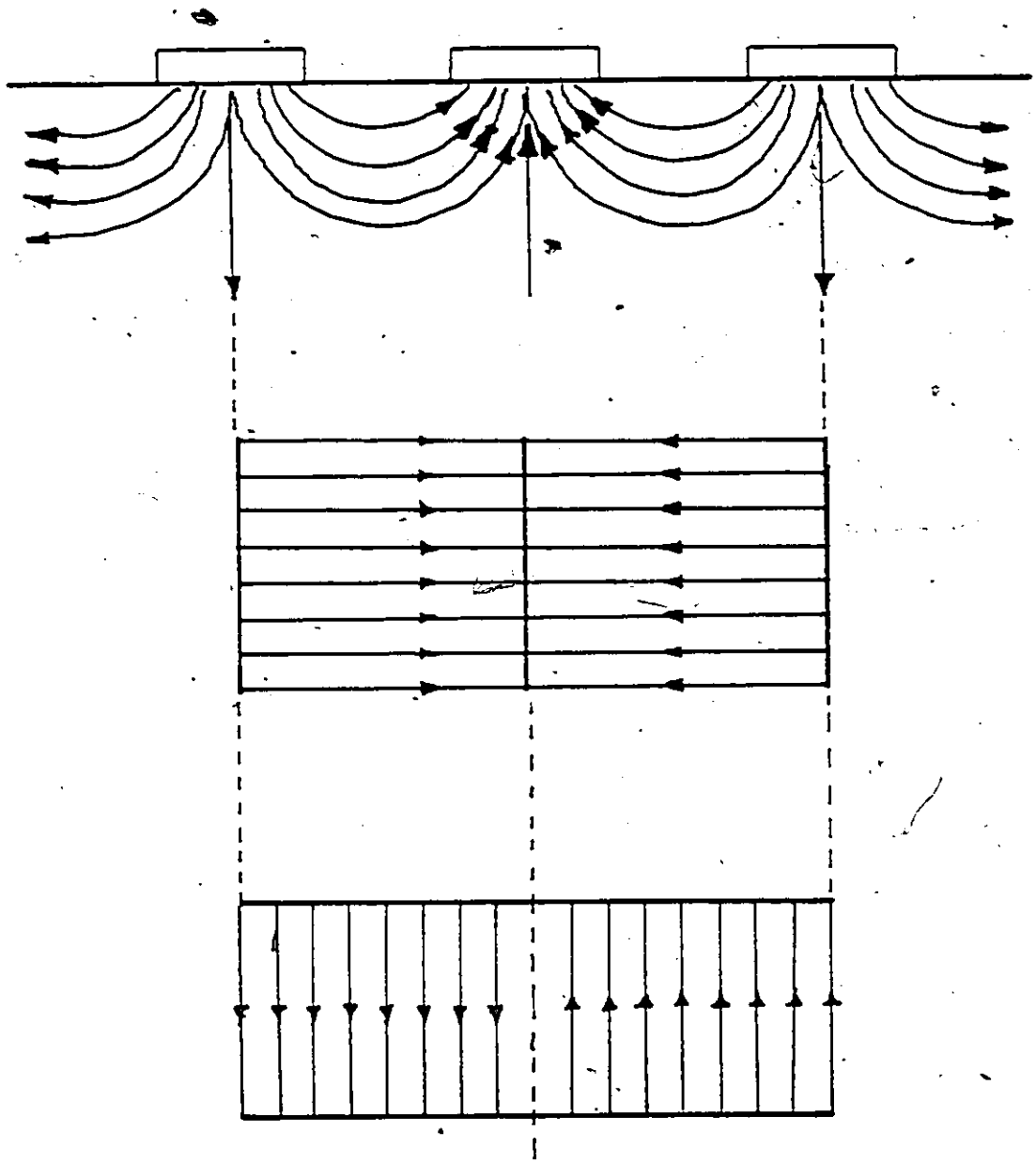


Figure 4.5 Field Patterns of an IDT

(a) Actual field pattern
(b) In-line field model
(c) Crossed field model

lithium niobate ($LiNbO_3$) for reasons of its high coupling constant. The model used for this material is the crossed-field model. We shall look at this model in some detail. The crossed-field model allows us to make use of an already developed Mason equivalent circuit for bulk wave transducers [28]. In this equivalent circuit model for the SAW transducer, and each finger is represented with a bulk wave transducer, as such, each finger has an equivalent circuit. Such an equivalent circuit is shown in Figure 4.6. This equivalent circuit can be converted to a completely electrical one by introducing the electrical port equivalent to the acoustic port with $e_i' = F_i/\phi$ and $i_i' = \phi\ U_i$ and an equivalent characteristic impedance $R_o = Z_o/\phi^2$. The equivalent electrical circuit is shown in Figure 4.7 where R_o can be expressed as

$$R_o = \frac{\pi}{\omega_o\ C_L\ K^2} \tag{4.4}$$

where $\omega_o = 2\pi f_o$ resonance angular frequency and C_L = electrode capacitance.

The symbol C_L is the capacitance per unit length of a section of length L, and can be calculated by using the empirical formula given by [29]

$$C_L = (\epsilon_s + 1)\ [6.5\ (a/L)^2 + 1.08\ (a/L) + 2.37 \tag{4.5}$$

where a = finger width and $L = \lambda/2$ and E_s is the effective dielectric constant of the chosen plane. For the pure mode

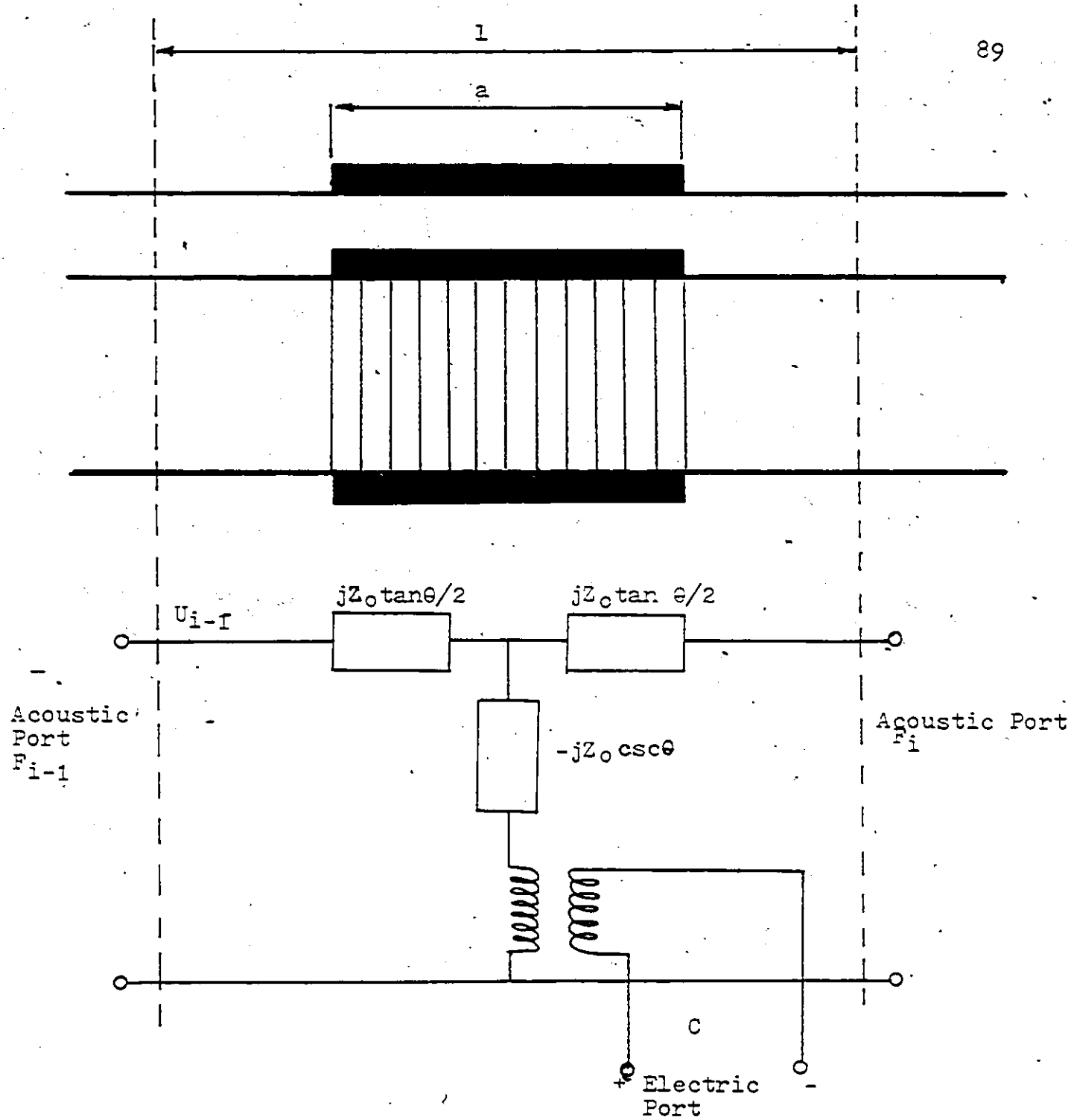


Figure 4.6 Mason's Equivalent Circuit for Bulk Wave Transducer

Z_o = Acoustic impedance
θ = Transit Angle

C = Static Capacitance

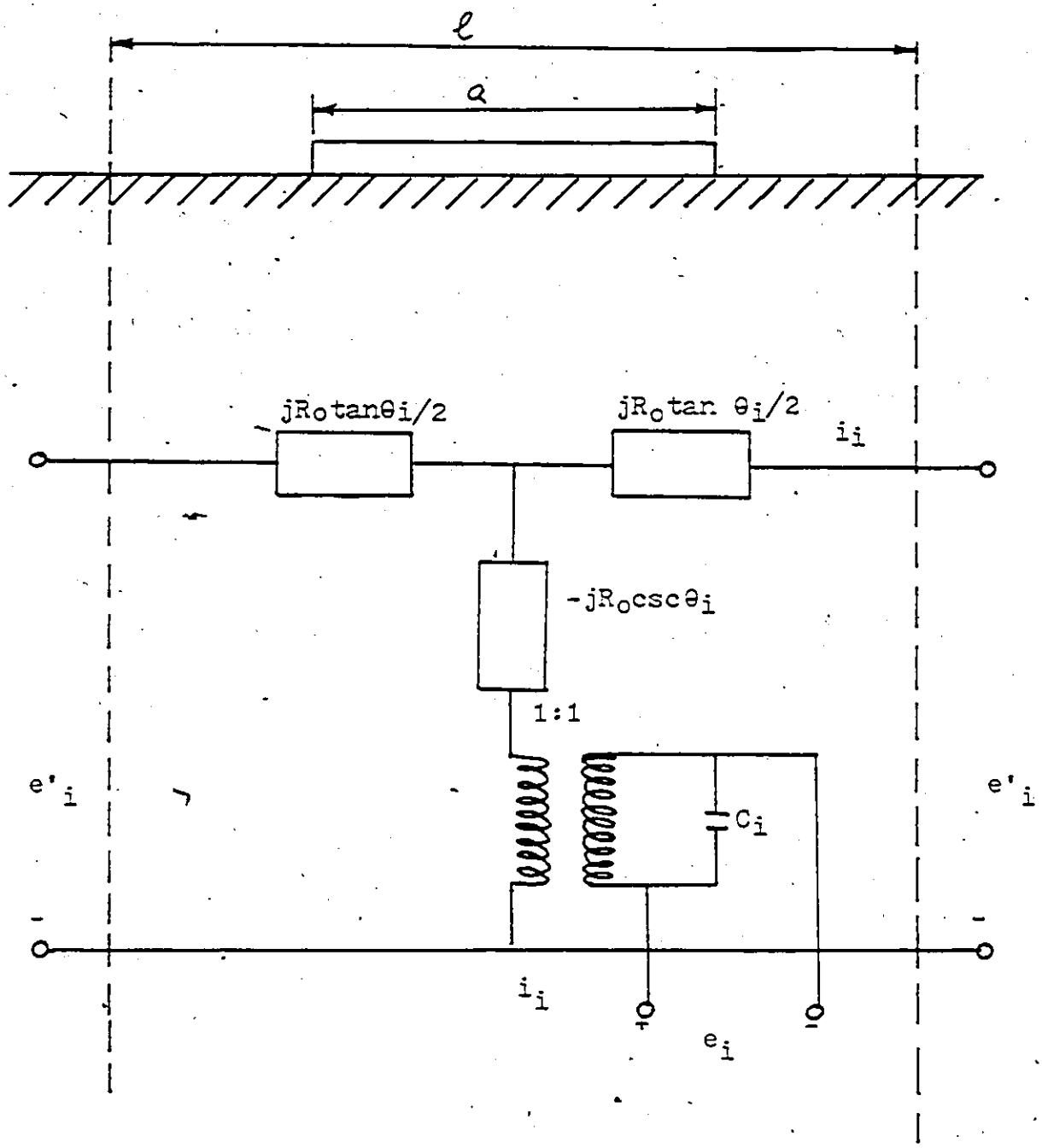


Figure 4.7 Complete Electrical Equivalent Circuit for One Section of the IDT.

axis on this plane, E_s is given by

$$\varepsilon_s = [\varepsilon_{11}\,\varepsilon_{33} - \varepsilon_{13}^2]^{1/2} \tag{4.6}$$

where ε_{ji} are dielectric constant tensor elements.

The admittance matrix for this equivalent circuit is given by

$$\begin{bmatrix} i'_{i-1} \\ i'_i \\ i_i \end{bmatrix} = [Y] \begin{bmatrix} e'_{i-1} \\ e'_i \\ e_i \end{bmatrix}$$

where

$$[Y] = \begin{bmatrix} -\cot\theta & \csc\theta & -\tan\theta/2 \\ \csc\theta & -\cot\theta & -\tan\theta/2 \\ -\tan\theta/2 & \tan\theta/2 & 2\tan\theta/2 + \omega C_L/G_o \end{bmatrix} \tag{4.7}$$

and

$$G_o = \frac{1}{R_o}$$

For a uniform transducer of width (aperture) W, there are N identical sections for each bank of fingers of the same polarity. Therefore, we need 2N section cascades. The resultant admittance matrix for the IDT is given by

$$\begin{matrix} i'_1 \\ i'_2 \\ i \end{matrix} = j\,G_o \begin{bmatrix} -\cot 2N\theta & \mathrm{CSC}\ 2N\theta & -\tan\theta/2 \\ \csc 2N\theta & -\cot 2N\theta & -\tan\theta/2 \\ \tan\theta/2 & -\tan\theta/2 & 4N\tan\theta/2 + \omega C_T/G_o \end{bmatrix} \tag{4.8}$$

where $C_T = 2NC_L$

and $\theta = 2\pi L/\lambda = \pi\omega/\omega_o$ = transit angle.

The transducer can now be represented by a three-port device as illustrated in Figure 4.8(a). Assuming that the transducer radiates into an infinite medium or equivalently into two acoustic ports which are terminated by the characteristic impedance R_o, the input admittance Y of the transducer as seen from the electrical port is given by

$$Y = G_a(\omega) + j\,[B_a(\omega) + \omega\, C_T] \qquad (4.9)$$

where $G_a(\omega) = 2\, G_o\, [\tan \theta/2 \sin N\theta/2]^2$ (4.10)

and $B_a(\omega) = G_o \tan \theta/2\, [2N + \tan \theta/2 \sin N\theta]$ (4.11)

For frequencies near the synchronous frequency $\omega_o = 2\pi V/\lambda$ the radiation conductance is

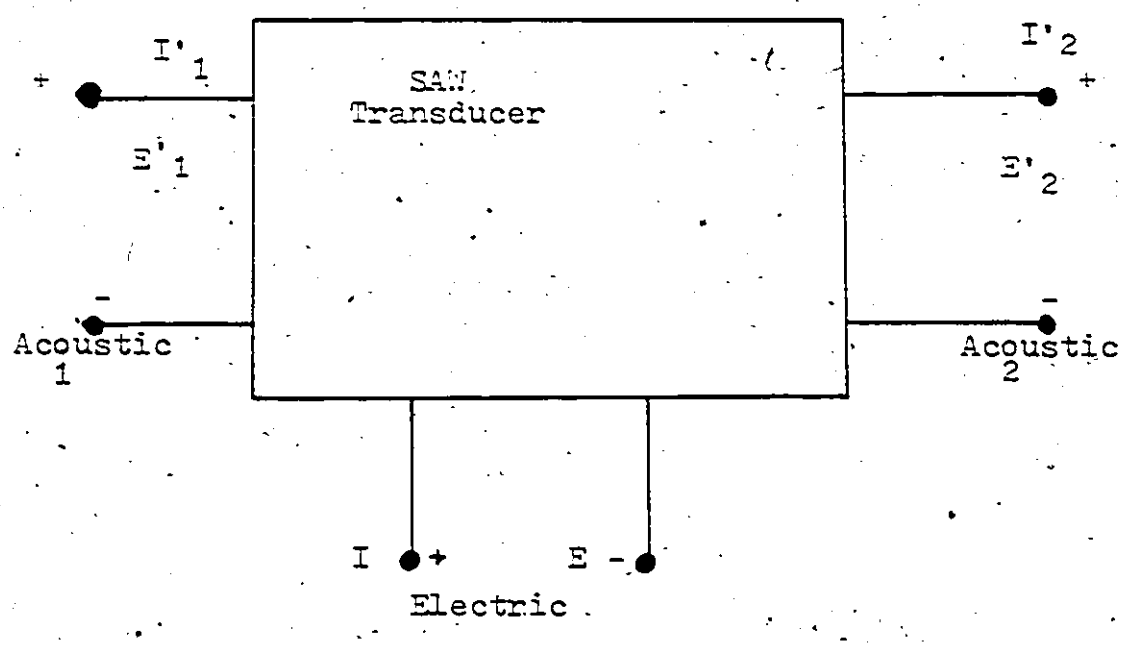
$$G_T(\omega) = 8\, N^2\, G_o\, (\sin X/X)^2 \qquad (4.12)$$

$$B_T(\omega) = 8\, N^2\, G_o\, (\sin 2X - 2X/2X) \qquad (4.13)$$

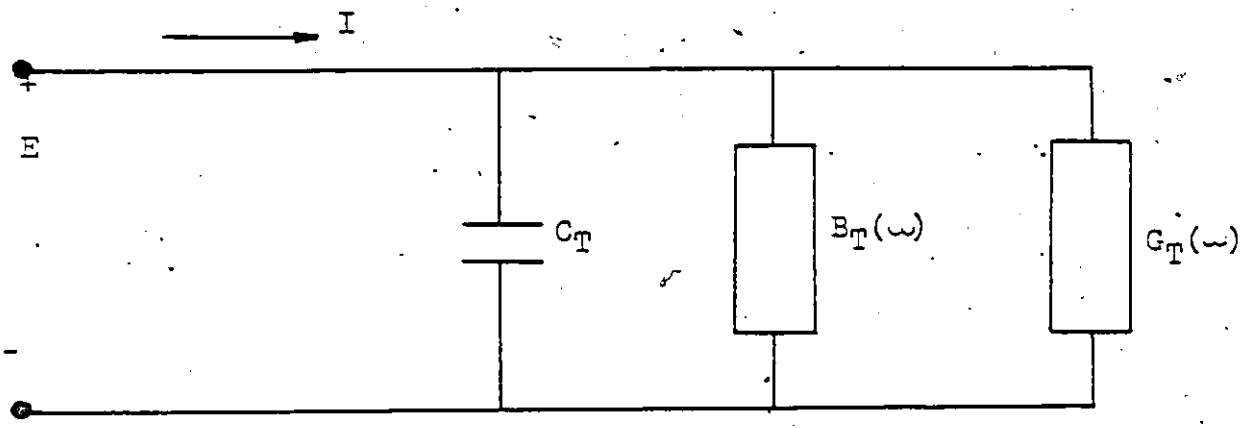
where $X = N\pi\, (\omega - \omega_o/\omega_o)$ (4.14)

The equivalent circuit of the transducer as seen from the electrical port is illustrated in Figure 4.8(b). Conductance and susceptance as a function of normalized frequency are shown in Figure 4.9.

The power carried by the surface wave is equivalent to the power dissipated in G_T. The radiation susceptance is zero at resonance, and to obtain the total input susceptance ωC_T must be added. The bandwidth of the transducer is inversely proportional to the length of the transducer. In other words the bandwidth of the transducer is proportional to 1/N.



(a) 3 Port model for SAW transducer



(b) Equivalent circuit as seen from electrical terminals

Figure 4.8 3 Port Model and Equivalent Eircuit for SAW transducer

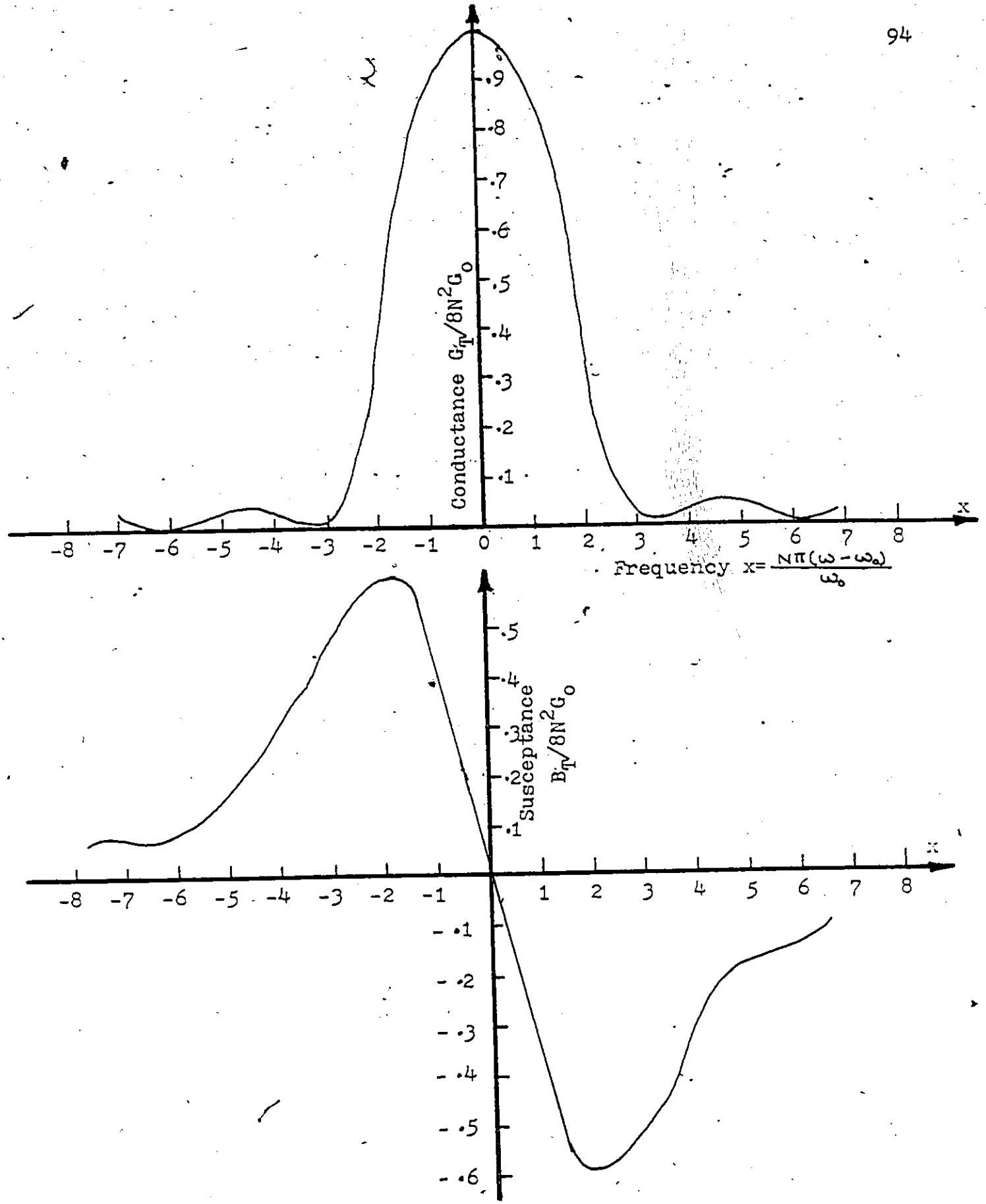


Figure 4.9 Input admittance of a uniform IDT as a function of normalized frequency

4.3. SAW FILTER

We recognized that the uniform transducer is a bandpass filter whose electrical parameters can be calculated by using an equivalent circuit model. We now want to show that the SAW filter is a member of the class of transversal filters. To bring about this analogy let us first look at the transversal filter concept [30].

4.4 TRANSVERSAL FILTERS

Suppose that the incoming propagating wave passes through different delays D_n and is added as shown in Figure 4.10. It is possible that under certain conditions these delayed signals add constructively, and yet another set of conditions these signal add destructively. The constructive addition will produce the bandpass signals and destructive additions will produce stop band signals. The filter shown in Figure 4.10 consists of N taps separated by D_n and contributions from each to the summing point are weighted by a coefficient A_n. Thus the input signal is repetitively delayed and multiplied by the weighting coefficients and added. Thus, if the input signal in the complex notation is given by

$$V_i(t) = V_{in}(f)\ e^{j2\pi ft}\ , \qquad (4.15)$$

where $V_{in}(f)$ is the input excitation vector, the output signal voltage can then be given by

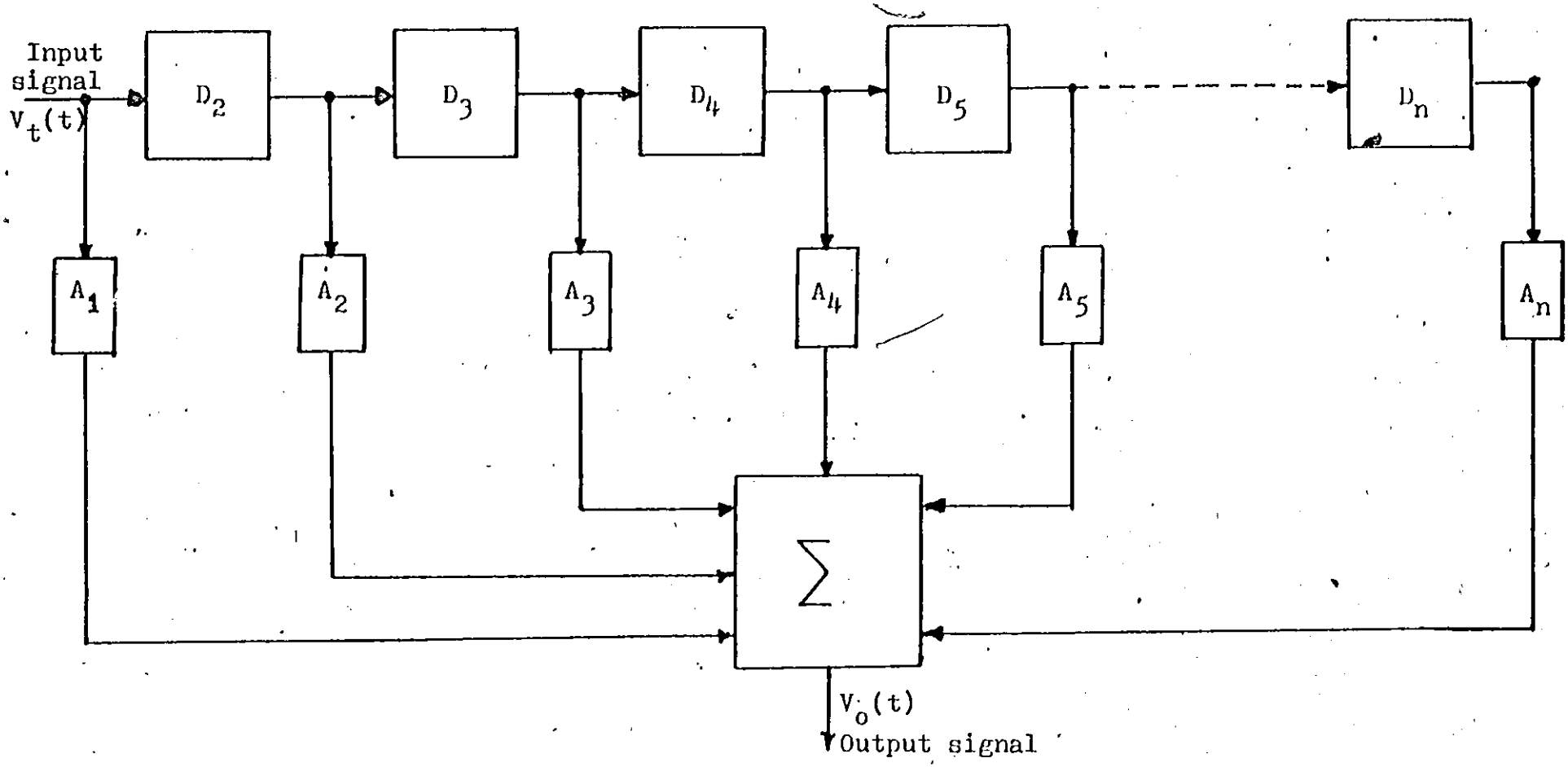


Figure 4.10 Schematic Diagram of A Transversal Filter.

$$V_o(f) = V_{in}(f) \sum_{n=1}^{N} A_n e^{-j2\pi f D_n} \quad (4.16)$$

where exponent $(2\pi f D_n)$ is regarded as the phase angle of the input signal at the nth tap. The transfer function is defined as

$$H(t) = \frac{V_o(f)}{V_{in}(f)} = \sum_{n=1}^{N} A_n e^{-j2\pi f D_n} \quad (4.17)$$

Now, we see that the transfer function is a function of two variables A_n and D_n. Evidently changing these variables will change the transfer function.

4.5 THE SAW FILTER AS A TRANSVERSAL FILTER

Recalling the receiver transducer mechanism, we find that the incoming surface acoustic wave generates fields across the fingers and the pad integrates to produce the output voltage. If we now consider the field distribution in the fingers we find the fields are high at the finger edges [29]. This approximation is true especially if the materials are low coupling constant materials (i.e. ST Quartz). However, even for high coupling constant materials this is not such an unreasonable assumption. With this assumption we may proceed to represent the fields by delta functions [31] as shown in Figure 4.11. The polarity of the delta functions alternate because the polarity of each finger of the transducer alternates.

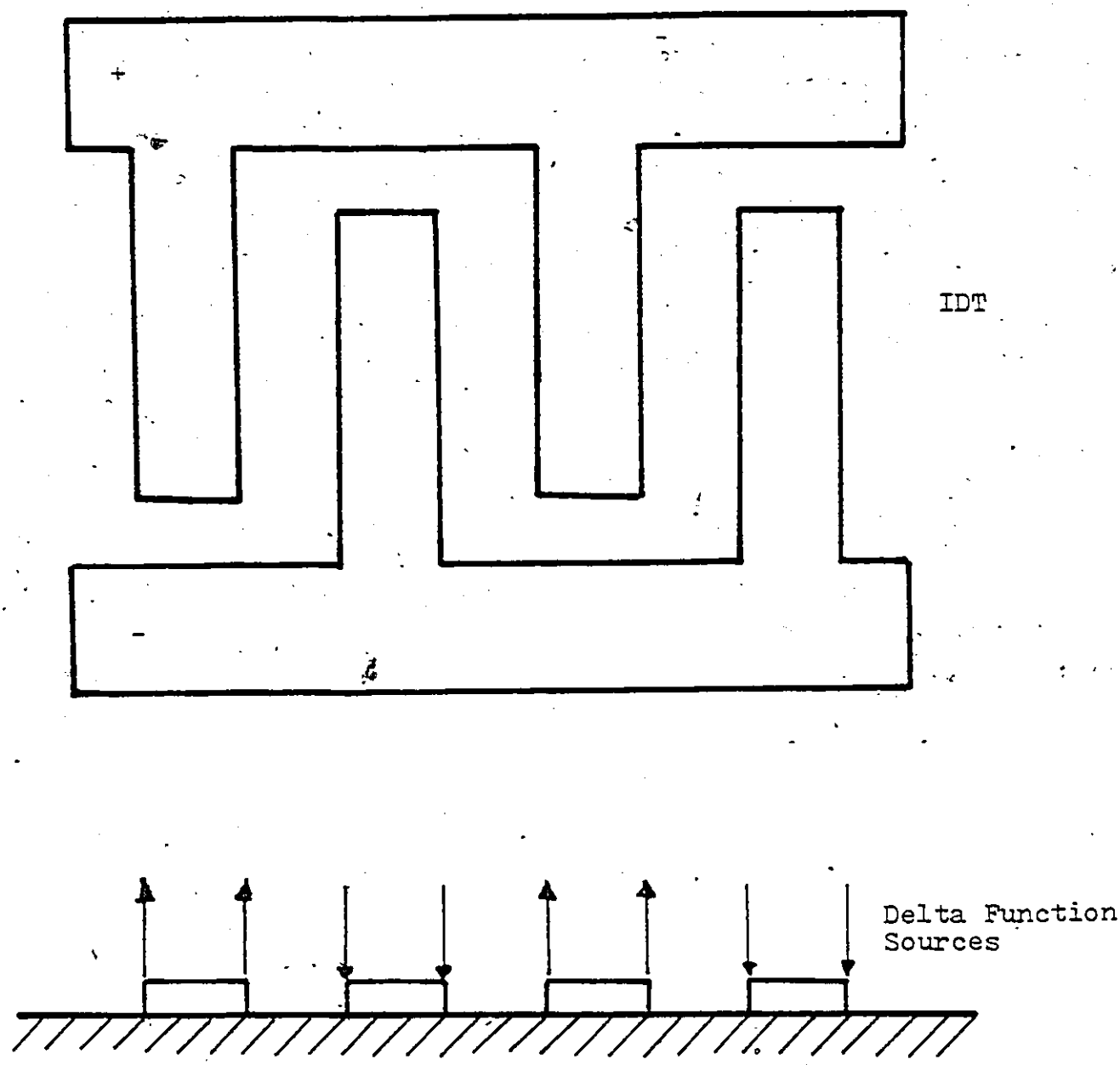


Figure 4.11 Delta Function Representation Model

Now let us turn our attention to the two transducer structures as shown in Figure 4.12. The transducer on the right is the SAW generator and the other is the receiver. The generator generates two waves - one to the right and the other to the left. Considering the wave propagating to the right, we find that the acoustic wave generated by the entire transducer is the sum of the contributions from each delta function source. Since all electrodes are connected to the same bus-bars and driven by the same voltage source the wave has the same phase all along the array. However, electrodes have different positions along the transducer. Hence, each wave has a different phase factor when it reaches the end of the transducer. Therefore, the acoustic wave at the position x is given by

$$A(f) = \sum_{n=1}^{N} I_n \exp j \frac{2\pi f}{v} (x - x_n) \qquad (4.18)$$

where I_n has a value of plus or minus to indicate the sign of the source. x_n is the position of the nth source and N is the total number of sources in the transducer.

Now consider the receiver transducer. As the wave travels under this transducer, each edge detects the wave under it. The detected voltage at the first edge (at Y_1) is given by

$$S_1 = \sum_{n=1}^{N} I_n \exp j \frac{2\pi f}{v} (Y_1 - x_n) \qquad (4.19)$$

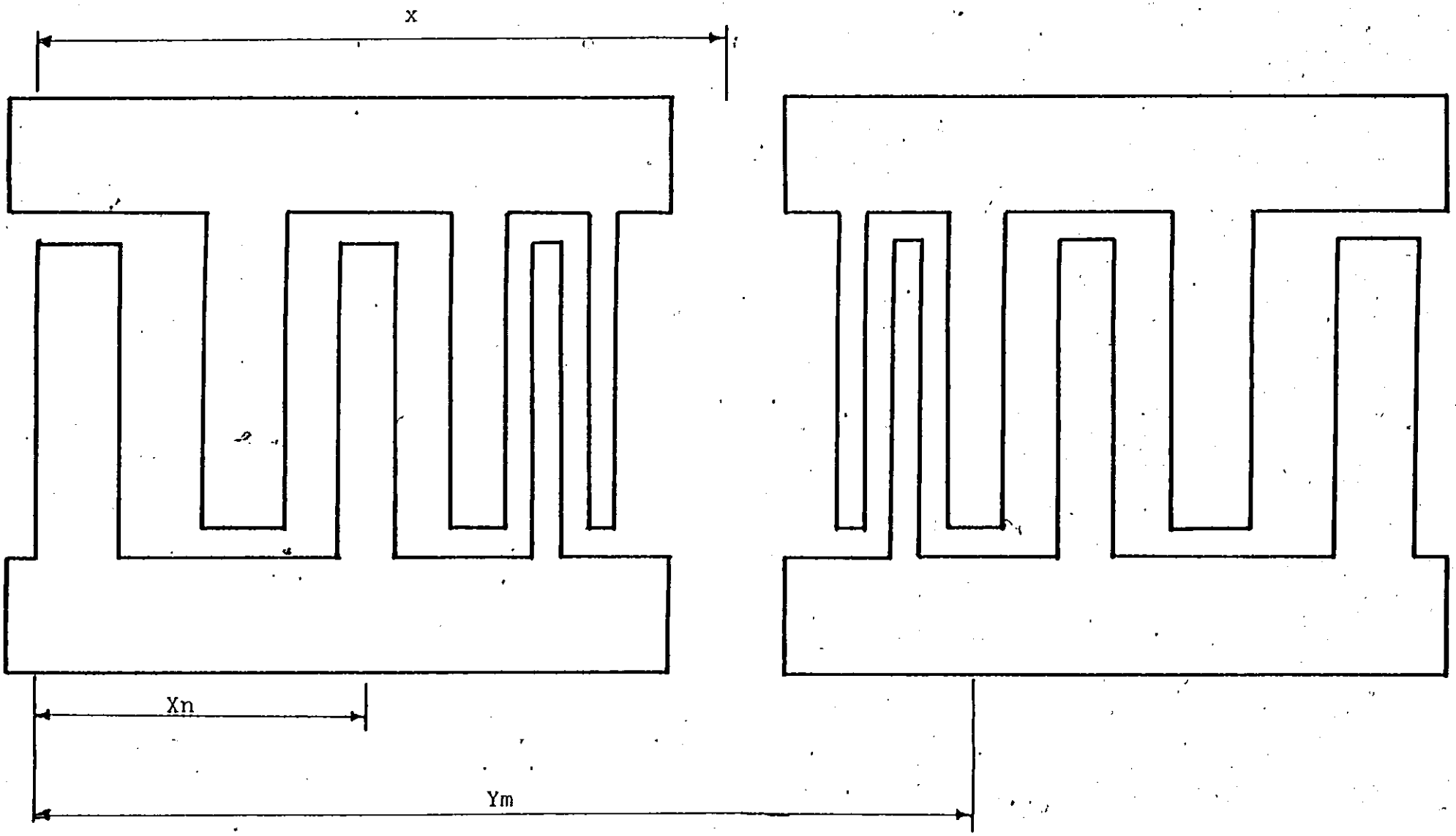


Figure 4.12 Two IDT Structure for SAW Filter

Therefore, the total output voltage is the sum of all the voltages detected at each electrode, since they are connected to a common bus-bar. Similar to the generating transducer, each delta function on the output transducer is associated with a sign (±1) indicating which bus-bar the electrode is attached to. The total transfer function of the filter is a sum of terms like equation (4.19) for each edge. When the input transducer has N edges, and the output transducer has M edges then the total transfer function is

$$H(f) = \sum_{m=1}^{M} \sum_{n=1}^{N} I_n I_m \exp\left[j \frac{2\pi f}{v} (y_m - x_n)\right] \qquad (4.20)$$

$$= \sum_{n=1}^{N} I_n \exp\left(-j \frac{2\pi f}{v} x_n\right) \sum_{m=1}^{M} I_m \exp\left(j \frac{2\pi f}{v} y_m\right)$$

$$= H_i(f)\, H_o^*(f) \qquad (4.21)$$

This shows us the total transfer function is the product of the transfer function of the input transducer and the complex conjugate of the transfer function of the output transducer.

It is important to see that the delta function representation of the transducer is the impulse response of the transducer. If we energize the transducer with an impulse due to piezo-electric coupling, the surface distorts to produce an image of the transducer source pattern. This distortion propagates away from the transducer, and when it

travels under the output transducer, each delta function distortion induces an output voltage. If the impulse response of the input transducer is $h_1(t)$ and the output transducer is $h_o(t)$ then the total impulse response $h(t)$ is given by the convolution of the two responses

$$h(t) = \int_{-\infty}^{\infty} h_i(t-\tau)\ h_o(\tau)\ d\tau \qquad (4.22)$$

The frequency response can be found by taking the Fourier transform of $h(t)$. This means taking the Fourier transform of the convolution integral. That is

$$H(t) = H_i(t)\ H_o^*(t) \qquad (4.23)$$

One distinction between SAW filters and general transversal filters is that in surface wave filters one cannot have all the delta sources of the same sign. This means that the filters must be bandpass rather than lowpass centered at zero frequency. The other distinction between the general transversal filter and SAW filter is that only one array of taps are used in the general transversal filter while SAW filters use two arrays. This gives the SAW filter designer an added advantage. He may choose both $H(t)$ and $H^*(t)$ independently to realize the overall filter response. SAW filters are in general two transversal filters in tandem. Thus it is possible to produce correlators. The uniform transducer pair is in fact a correlator. Producing a matched filter is also possible. For example, one

transducer generates a particular impulse response while the other has its time-reversed delayed response. We shall come back to this subject a little later.

We shall now look at the parameters in SAW filters that can change or modify the frequency response of the filters. In the general transversal filter, parameters available for changing the transfer function of the device were amplitude weighting coefficients A_n and delays D_n. With SAW filters we can do the same. For example the weighting coefficients A_n will correspond to the overlap distances or aperture lengths of the fingers. We may change this in a particular fashion to obtain the required transfer function. Delays correspond to finger spacing. This too can be changed to achieve the required transfer function. There are other methods (such as finger withdrawal [34], series weighting methods [37], etc.) that can be used to achieve the required transfer function. The most popular method which has been used to date for bandpass filters is the amplitude weighting techniques. Using any one of them is a costly affair. Every time one wants to change the weighting function a new mask must be made. As we shall see later, this would require a series of sophisticated steps. We have developed an inexpensive thin film technique [32] that can be used to modify the aperture of a transducer. We shall discuss these when we deal with fabrication

technology.

There are various techniques [31,32,33] available for designing SAW filters. Originally, all these techniques evolved from the search for the optimum impulse response to realize a given transfer funtion. For a given frequency response it is a simple matter to find the impulse response. This impulse response is then made to correspond to the source distribution of the transducer. Generally, it is difficult to realize this source distribution because it extends to infinity. Thus it is necessary to truncate the impulse response. Generally, all the techniques developed are to find the best method to reduce the truncation errors and find the realizable finite impulse response of the filter. We shall not review these techniques here. However, we shall consider some higher order effects, which influence the performance of these devices. These effects are briefly discussed.

4.6 HIGHER-ORDER EFFECTS

The first higher-order effect is the distortion due to the fact that the generator and load are connected to the transducers whose impedance vary with frequency. Mismatches due to this variation can cause distortions in the phase response and also in the amplitude response of the SAW filters.

The second higher order effect is the triple-transit echo. This is caused by the regeneration of acoustic waves within the transducer. An acoustic wave travelling under a transducer will induce a voltage across the bus-bars. This voltage in turn causes another acoustic wave within the transducer. This wave will travel back to the input transducer. In the same manner another wave can be generated by the input transducer and sent to the output transducer. Since the first echo has travelled 3 trips across the substrate, this is known as the triple-transit echo. The reduction of the triple-transit echo by designing a matching network is discussed by Snow [38]. The penalty one pays for this is an increase in insertion loss which of course is a disadvantage. It is possible to reduce the triple transit while maintaining a low insertion loss by using other techniques. However, such techniques increase the complexity of the filter design.

The third higher order effect is the finger reflection. Directly under metal electrodes the SAW velocity is less than the free surface velocity due to two reasons. They are:

1. Due to shorting of the piezo-electric fields, sometimes called piezo-electric stiffening
2. Due to mass loading caused by the deposited metal

The effect of mass loading can be reduced to a negligible

amount by reducing the thickness of the material. However, one needs to be careful in doing this because the reduction in the thickness can increase the ohmic losses in the transducer. One can overcome this problem by choosing a high-conductivity light metal such as aluminium. On the assumption that aluminium is used, and that the effects mass loading are negligible, then the change in acoustic impedance due to piezo-electric stiffening is given by

$$\frac{\Delta z}{z_o} = \frac{\Delta v}{v} = \frac{1}{2} K^2 \qquad (4.24)$$

where z_o = acoustic impedance in free surface

Δz = change in impedance

Δv = change in velocity

The change in acoustic impedance is seen as a discontinuity by the surface wave, thus giving rise to reflections. These reflections can cause distortions in the transfer function or the driving point impedance. The other effect caused by the change in velocity is a slight shift in frequency. Thus it is necessary to compensate for this frequency shift in the design of the transducer. An effective method of reducing the effect of the finger reflection is to use a split-electrode transducer [39]. Here, the geometry is such that the finger reflection is cancelled out. Both the normal transducer geometry and the split electrode geometry are shown Figure 4.13.

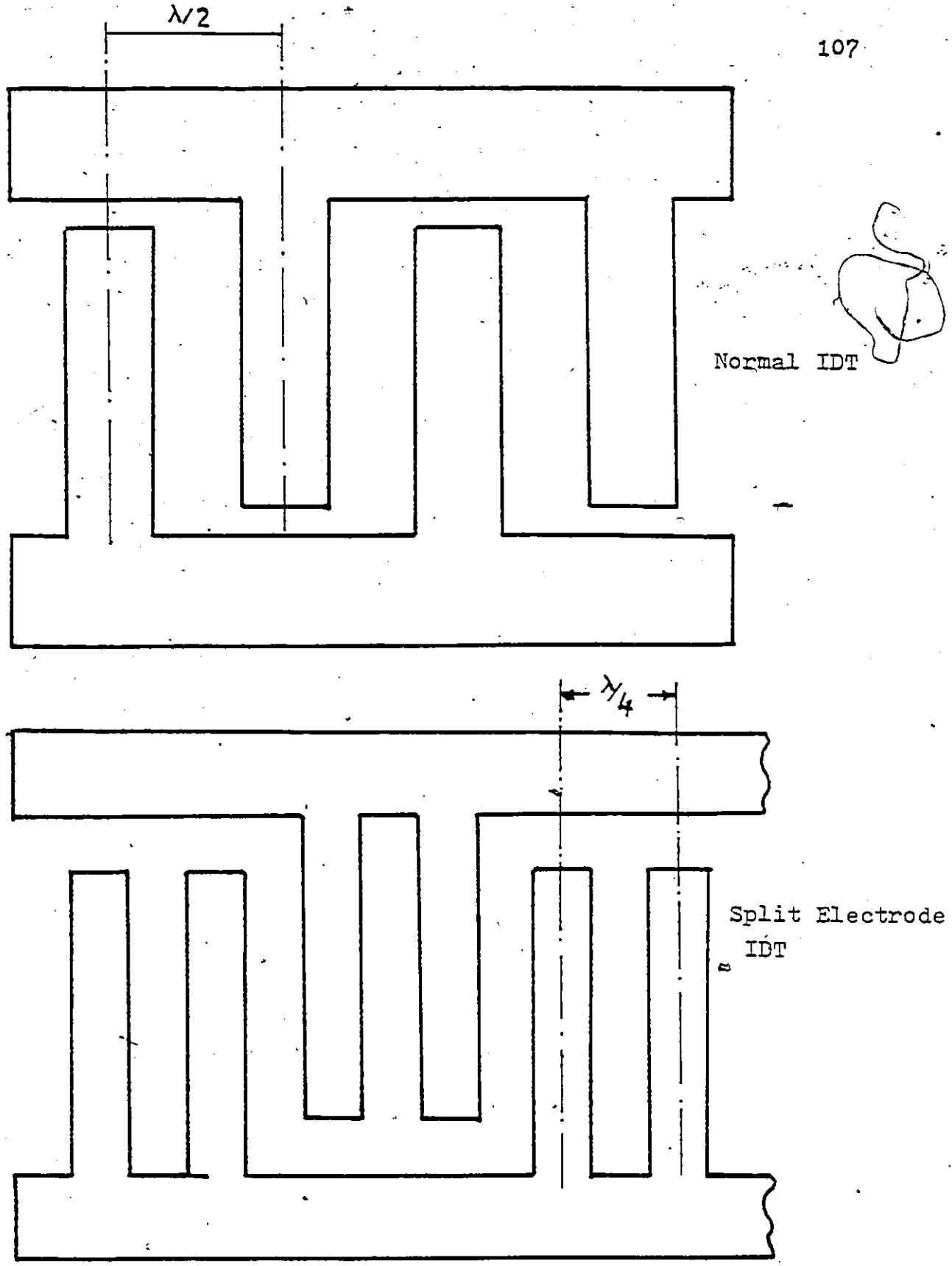


Figure 4.13 Normal and Split Electrode IDT

The fourth higher-order effect is the bulk wave generation [40]. When we discussed the fundamentals of SAW, we pointed out that the mode coupling does not occur due to the difference in the velocities between bulk modes and SAW, providing there is no perturbation on the surface. However, the presence of the IDT implies surface perturbation. Thus, it is natural to expect the bulk modes to be generated. The bulk modes that are generated radiate into the crystal, bounce off the bottom surface and are detected by the output transducer. It is possible for these bulk waves to take several bounces between the two surfaces and to generate several plate modes as the frequency is swept. The frequency of the bulk wave generated and the velocity of the bulk wave is given by

$$f_B = \frac{V_b}{d \cos \theta} \tag{4.25}$$

where f_B = frequency of the bulk wave

d = one period of the electrode spacing

θ = at the angle that the bulk wave is generated

V_b = bulk wave velocity

Because the wave velocity of the bulk waves are usually higher than the surface wave velocity, the bulk wave frequencies occur at higher frequencies than the resonance frequency of the surface wave. The effect is to prevent the filter from having a good stop-band rejection. There are

various methods [41] that can be used to reduce the effects of the bulk waves. For example, one may roughen the bottom of the crystal and coat it with a conductive material which will absorb the bulk waves. A material commonly used is silver epoxy. Another method is to use a track-changing multistrip coupler [42(a)]. We shall review the multistrip coupler later in this Chapter.

The sixth spurious effect is electromagnetic feed through between the input and the output transducers that can cause distortions in the response. However, this can be reduced considerably by proper packaging [38].

The seventh spurious effect is caused by the reflections from the edges of the crystals. They can be reduced by either incorporating the absorber pads (i.e. black wax) or rounding off the edges.

Finally, the diffraction effects [43] must be discussed. The diffraction effects can generally be neglected since diffraction is a function of the aperture width of the transducer and is usually many times larger than the wave length. However, whenever small apertures are involved such as in a apodized transducer, diffraction has to be taken into account. One very convenient way of overcoming the diffraction problem in high K^2 material is to incorporate a multistrip coupler.

4.7 MULTISTRIP COUPLER

We have considered many second-order effects, and have pointed out some methods used to reduce these effects. One such method was incorporating a multistrip coupler (MSC). A MSC can offer extra convenience to a SAW filter designer when he uses high coupling coefficient materials, such as LiN_bO_3. Here we shall review the principle of operation of this device and its advantages [42(b)].

4.7.1 Principle of Operation of MSC

As illustrated in Figure 4.14, the MSC is an array of metal strips deposited on the surface of the piezo-electric substrate. Hence, we have divided the SAW propagation path into two tracks. Furthermore, we have four ports designated by A_1, B_1, A_2 and B_2. Evidently, the geometry is symmetrical about the centre line between the two tracks. Thus the modes of propagation can be separated into two modes.

Let us consider the case of the uniform wave generated along the track 1. The wave can be resolved into two modes. These modes are known as the symmetric mode and anti-symmetric mode and are illustrated in Figure 4.15. On the free surface these modes have the same velocity which is equal to the Rayleigh wave velocity of the substrate. However, within the MSC the velocities of these modes are

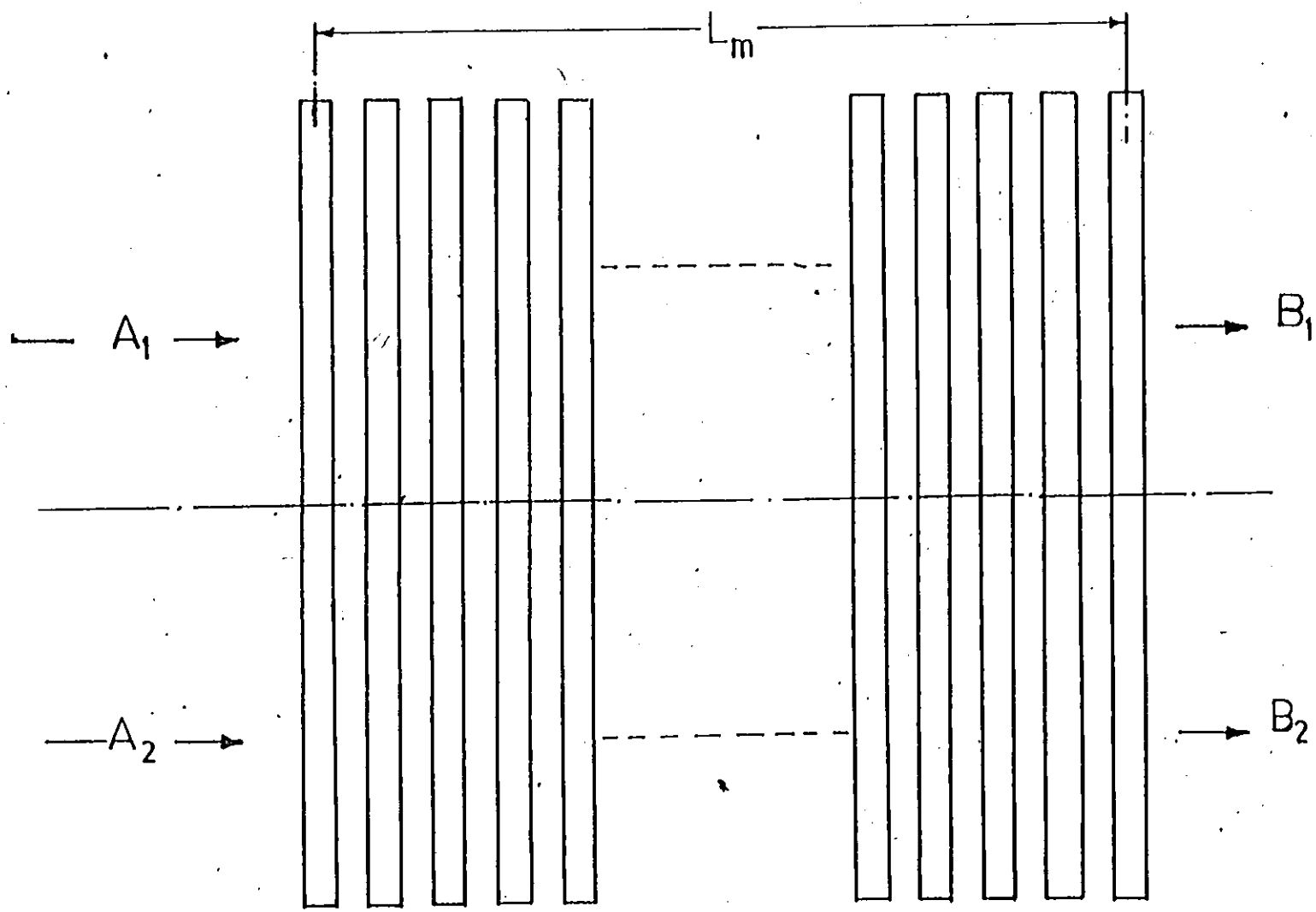


Figure 4.14 A Multistrip Coupler

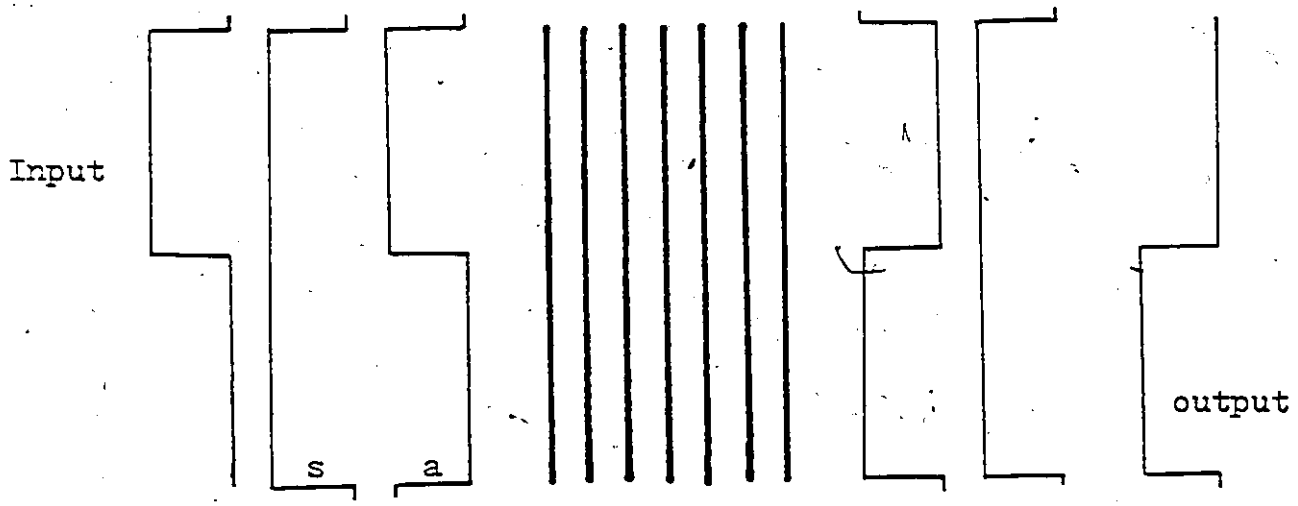


Figure 4.15 Modes of the MSC along with the Input and Output Waveforms.
a:- Antisymmetric Mode.
s:- Symmetric Mode.

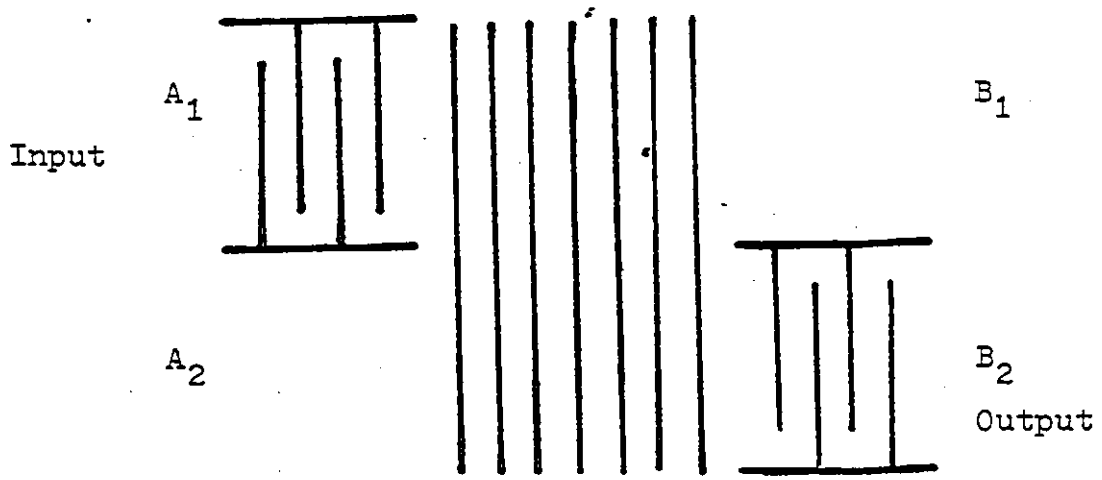


Figure 4.16 Track Changing MSC

different. The total SAW at any point on the surface is the sum of these two modes. Obviously, the length of the coupler or equivalently the number of strips in the coupler will determine the total SAW received at the other side of the tracks.

Considering these modes, when a uniform SAW is excited in the track 1, we find an equipotential condition along the metal strips, assuming there are no ohmic losses along the strips. For the symmetrical mode the particle displacements are in phase on both tracks and the electrical potentials of a strip oscillate in synchronism with these displacements. Thus there is no current associated with these modes. The velocity of these modes is slightly less than the free surface velocity due to piezo-electric shorting effects caused by the presence of metal strips.

On the other hand, the particle displacement for the anti-symmetric mode is 180° out of phase on the two tracks resulting in a zero potential at all times. As the wave progresses, currents in the strip flow back and fourth. If the width of the strips were comparable to their period, the velocity of this mode would be the same as the velocity of the anti-symmetric mode. However, since it is not the case, this velocity is slightly greater than the stiffened velocity.

Since the velocities of the two modes are different,

there will be a phase difference between the modes as they travel along the MSC. This phase difference between the symmetric and anti-symmetric mode is proportional to the length of the coupler and is given by

$$\theta_L = (K_a - K_s)\, L_m = \frac{R_m}{2} K^2 K_f L_m \qquad (4.26)$$

where K_f is the wave number on the free surface

L_m is the length of the multistrip coupler, and

R_m is the geometric factor for the finite strip width.

Let us suppose that we have two input complex waves in track 1 and track 2. The outputs at B_1 and B_2 can be found from

$$\begin{bmatrix} B_1 \\ B_2 \end{bmatrix} = \begin{bmatrix} \cos\left(\frac{\theta_L}{2}\right) & -j \sin \frac{\theta_L}{2} \\ -j \sin\left(\frac{\theta_L}{2}\right) & \cos \frac{\theta_L}{2} \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \qquad (4.27)$$

4.7.2 Track Changer

Let us suppose that we want to build a track changer. That is to say that we want to produce a MSC so that we can send our surface wave along track 1 and pick up the signal back from track 2. In doing so, we want to transfer all the energy from track 1 to track 2. Figure 4.16 illustrates the geometry of this track changer. To do this we have to choose the length of the coupler such that the phase shift

$\theta_L = \pi$. This can be done using equation (4.26). Having done so we can easily see from equation (4.27) that the output of the first track B_1 is zero and $|B_2| = |A_1|$.

This track changer is particularly useful in avoiding generated bulk waves. Bulk waves generated by the input transducer will propagate along the 1st track, and thereby the receiver transducer avoids picking up these bulk waves.

The other advantages of the MSC is that it can reduce the diffraction effects which exist on otherwise apodized filters. The aperture width of certain areas of the apodized transducer is small and can cause diffraction. The problem can be even more significant if one uses two apodized transducers. Incorporating the multistrip coupler allows one to overcome this problem and also reduces the triple transit echo, since the reflection from the MSC outside the stop band is small. The MSC is a wide band coupler and has other useful applications such as an undirectional transducer, multipath delay lines, power dividers, bandwidth compressors, etc. We shall not review these here.

CHAPTER 5

HARDWARE IMPLEMENTATION OF THE DISCRIMINATOR

In Chapter 4 we reviewed the fundamentals of surface acoustic waves, coupling mechanisms, and we pointed out how surface acoustic wave principles can be applied to produce transversal filters. Finally, in that Chapter we discussed various higher order effects and some methods used to overcome these effects. In this Chapter we shall discuss how surface acoustic wave devices can be adapted to approximate a bank of correlators, and we will explain how the hardware implementation is carried out..

5.1 TRANSDUCER CORRELATION MECHANISM

Let us consider two uniform transducers as shown in Figure 5.1. The periodicities of these two transducers are identical, as are their lengths. Now we excite the input transducer with an RF pulse whose pulse length is equal to the propagation time within the transducer.

$$T = \frac{L}{V} \tag{5.1}$$

where L = physical length of the transducer

V = SAW velocity

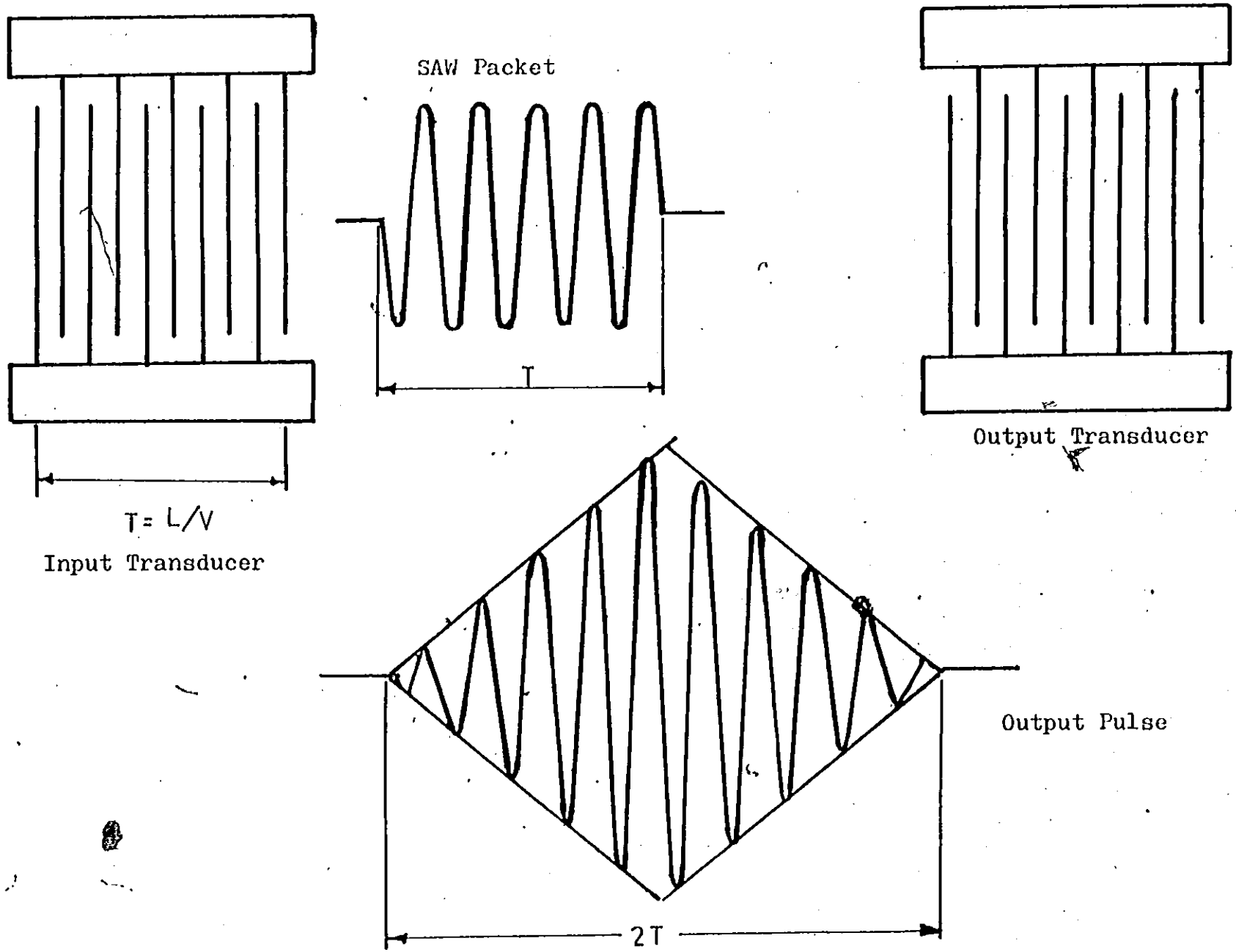


Figure 5.1 IDT Pair as a Correlator

The frequency of the RF signal is also equal to the resonance frequency of the transducer. This RF pulse will generate a surface acoustic wave packet as shown in Figure 5.1. (Here we will consider only the wave travelling to the right and ignore the other wave which is travelling to the left. Also we will assume that there are no losses and ignore all the higher order effects.) When this wave travels under the receiver transducer, induced electric fields between fingers will be integrated to produce a voltage. If we look at the output voltage of the receiver transducer we will find that the maximum output occurs when the surface wave packet sits right under the output transducer. In other words, when the spatial distribution of the surface acoustic wave which is equal to the spatial distribution of the RF pulse is also equal to the spatial distribution of the output transducer fingers, then the electrical structure will produce a maximum voltage. This output voltage is shown in Figure 5.1. We recognize this as correlation. Expressed in another way we see that the two transducers act as a correlator in which the reference signal is built into the output transducer.

5.2 M-ARY CORRELATORS WITH TIME SCALING

Our requirement is to build M-ary correlators to detect M signals. This can easily be done now. If our

signal frequencies are $f_1, f_2, \ldots, f_m$, then we can build M transducer pairs to respond to these frequencies. Each pair is to resonate at a particular frequency f_i. Since the signal energy in every signal is the same, our next requirement is time scaling. This can be easily achieved. All that we have to do is scale the separation distance between the transducers. Thus we have different delays allocated for different transducers. Such an M-ary correlator is shown in Figure 5.2(a). The M signals, each having a time length of T_i are shown in Figure 5.2(b), and Figure 5.2(c) shows the frequency time delay characteristics for this device. Although it is possible to produce such a device for our application, for a large M this arrangement becomes a cumbersome device.

What we can do is to find a SAW device whose frequency versus time delay characteristic is linear, and see whether we can adapt that device to behave as an M-ary correlator. An example of such a SAW device is the FM linear chirp filter. We shall explain the operation of this device in the next paragraph.

5.3 SAW LINEAR CHIRP FILTER PRINCIPLE

In considering SAW transversal filters we said that the impulse response of the filter is physically realized on the transducer. Therefore, varying the tap positions will

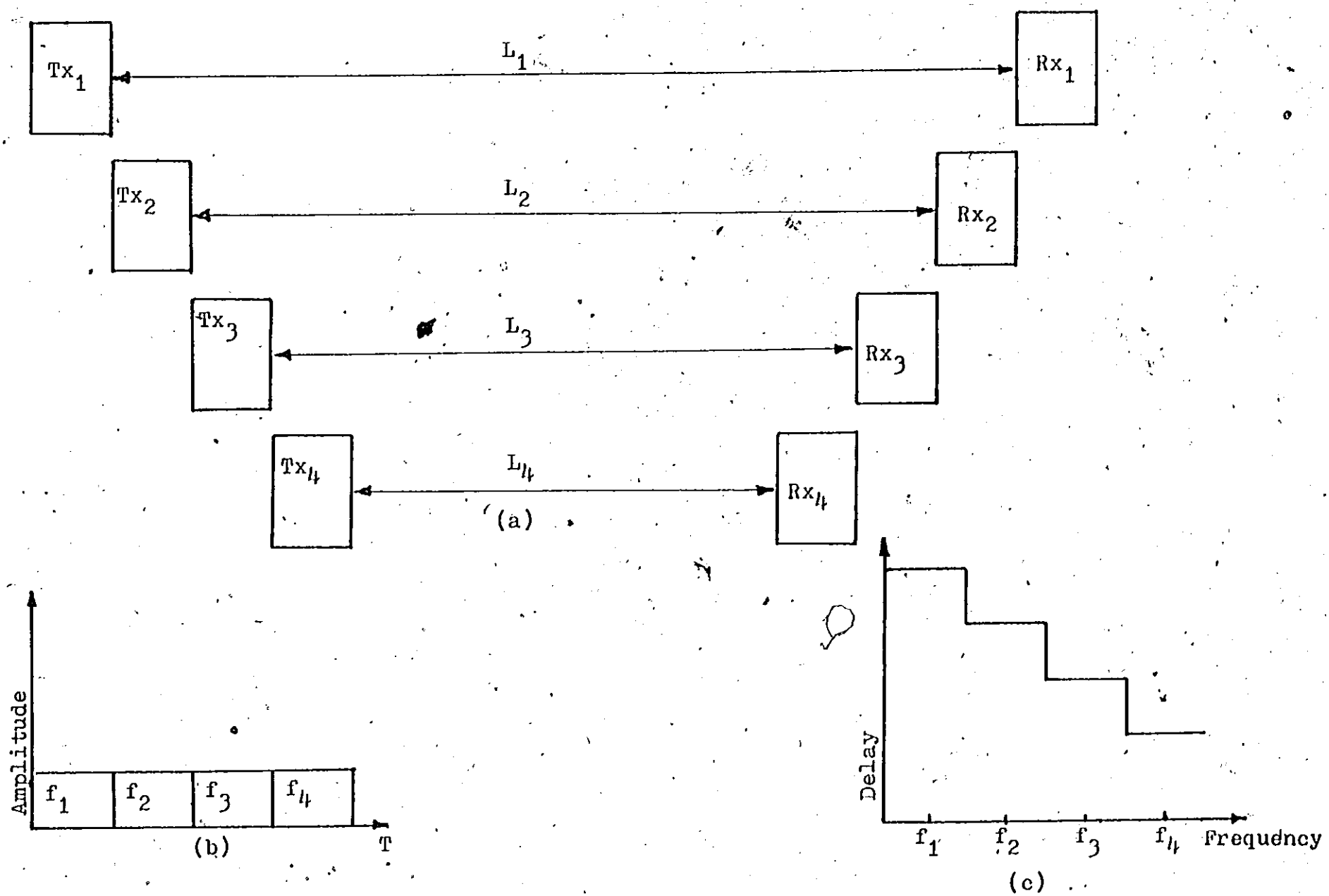


Figure 5.2 (a) M-Ary Correlator with Uniform IDT
(b) Input Signals
(c) Frequency vs Time Delay Characteristics

vary the impulse response. In other words, we have a one-to-one correspondence between the tap positions and impulse response. Suppose we have a transducer whose tap positions vary linearly along the length of the transducer as shown in Figure 5.3, the aperture of this transducer is kept constant. We shall neglect all the second-order effects for the moment. Now let us excite this transducer with an impulse. The output SAW to the right is shown in Figure 5.3(b). We recognize this waveform as the well known linear FM chirp signal.

Now let us fabricate a wideband transducer on the right hand side of the chirp transducer as illustrated in Figure 5.4(a), and measure the time of arrival of the various frequencies. We find that the frequency-time characteristic is a nonconstant function of the instantaneous frequency of the input signal. In Figure 5.4(b) we have shown these characteristics. Let us suppose that we now have two transducers. As shown in Figure 5.5(a), one transducer is the mirror image of the other. This arrangement is popularly referred to as a down chirp configuration.

The frequency vs. time delay characteristics of the two transducers are shown in Figure 5.5(b). Suppose we excite the input transducer with a linear FM waveform as shown in Figure 5.5(c). Low frequency components of the

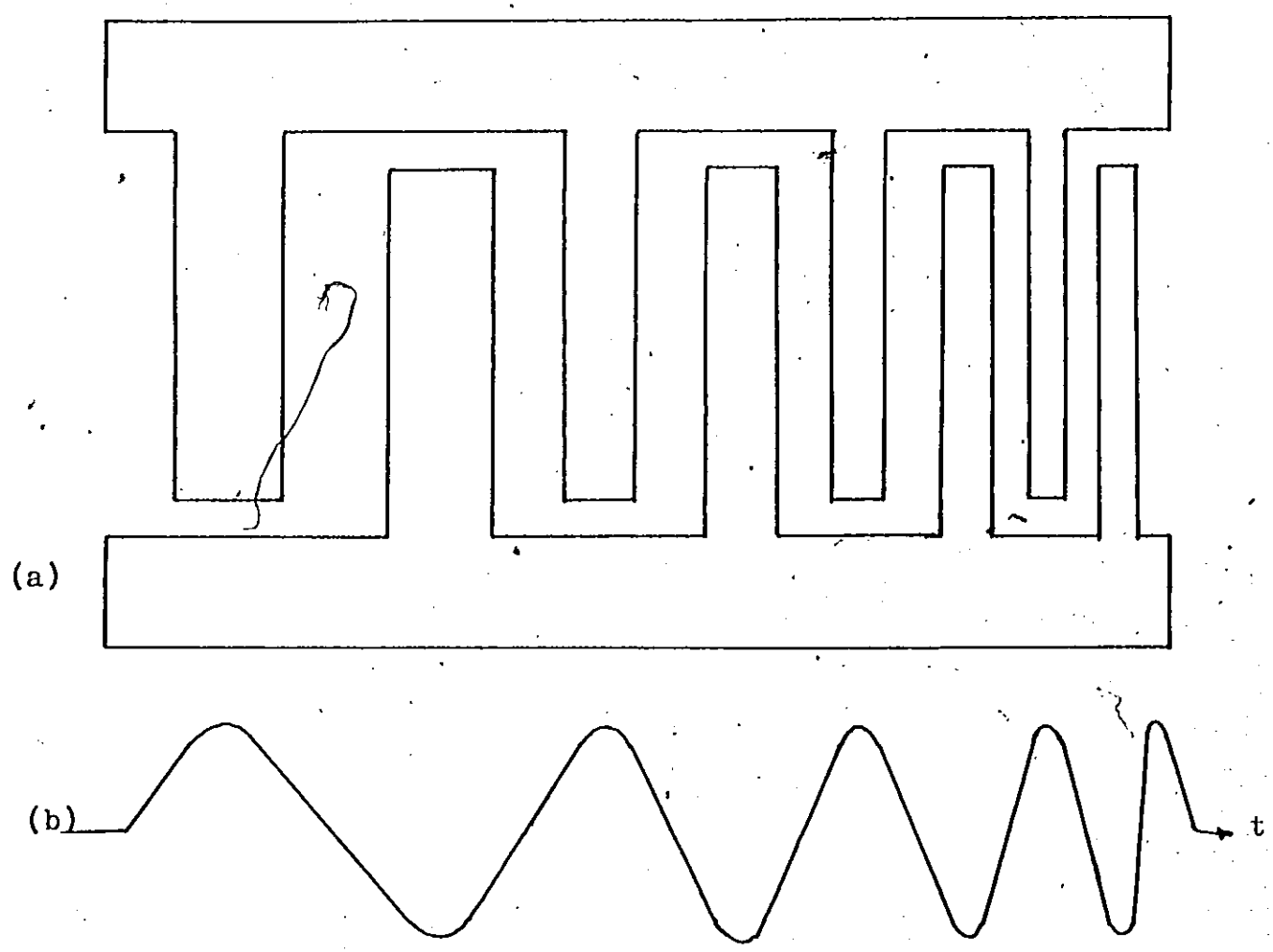


Figure 5.3 (a) Linear Chirp Transducer

(b) Linear FM Chirp Signal

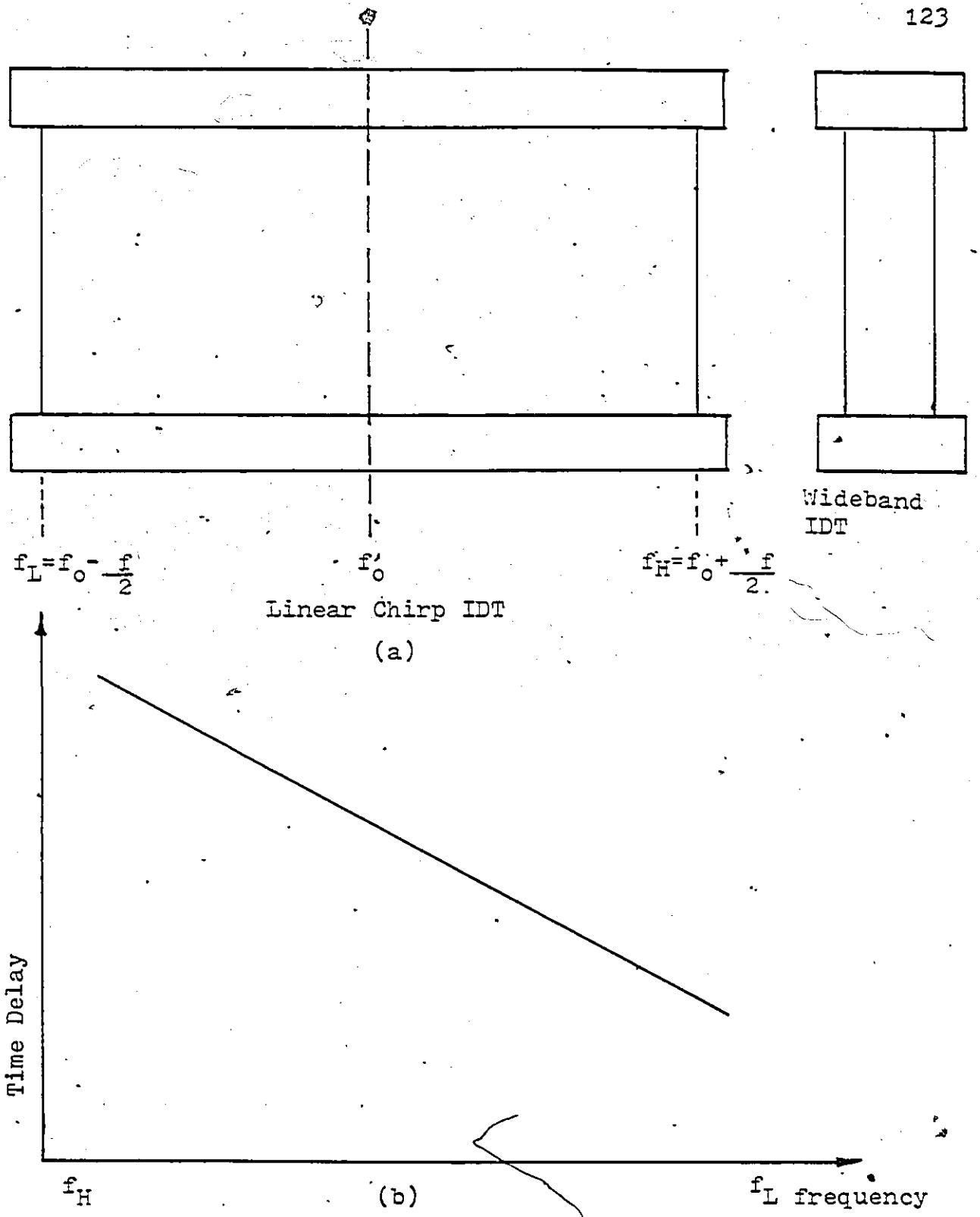


Figure 5.4 (a) Linear Chirp and Uniform Wideband IDT
(b) Frequency vs Time Delay Characteristics

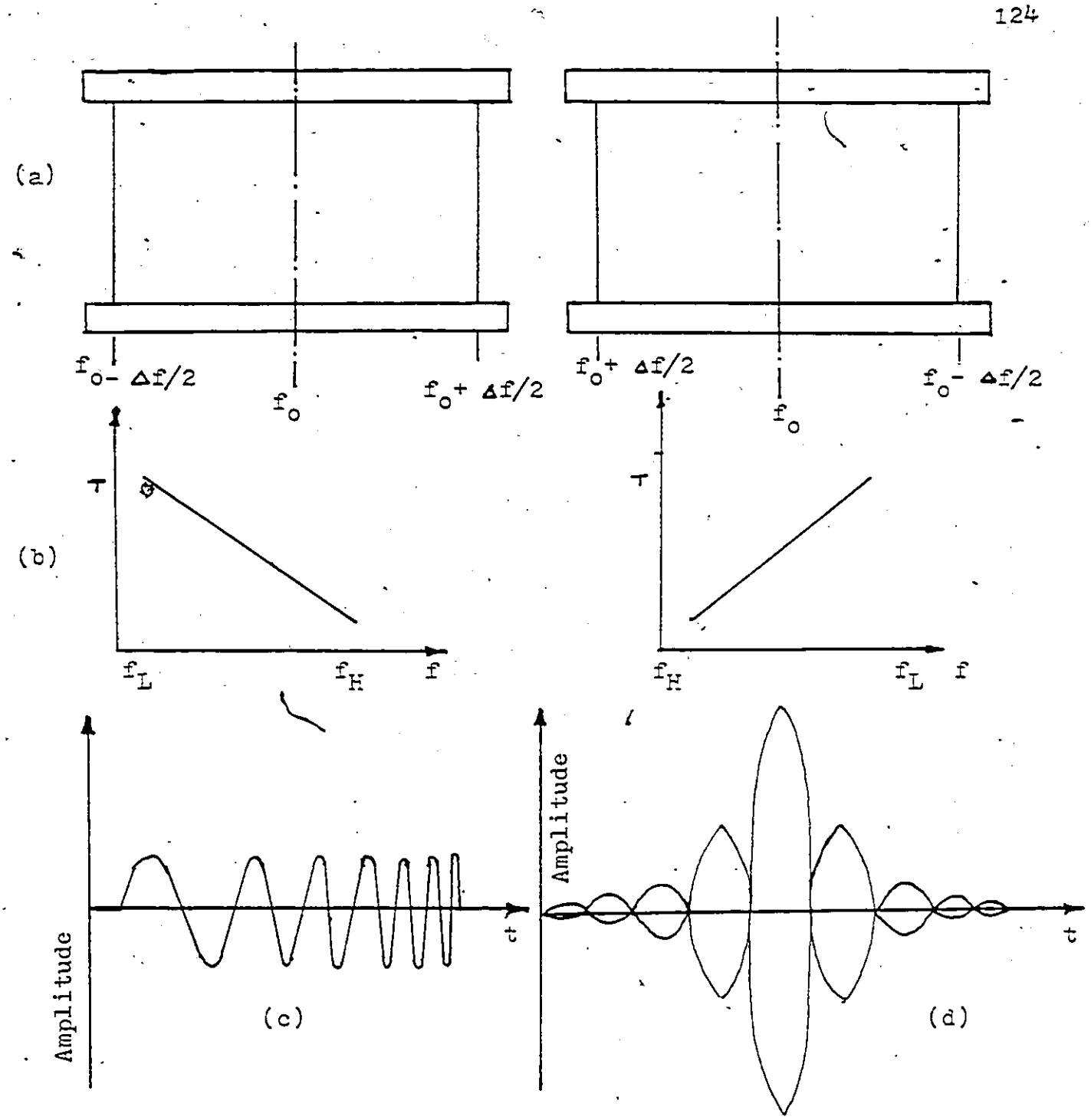


Figure 5.5 (a) Linear Chirp and its Mirror Image Configuration
(b) Frequency vs Time Delay Characteristics
(c) Input Chirp Signal
(d) Output Correlated Signal

signal will set in resonance the low frequency end of the transducer, and as the signal progresses in time, other areas of the transducer will resonate corresponding to the incoming signal frequency. Therefore, this signal will generate a SAW signal identical in time distribution to the input signal. Lower frequency signals will travel in front, and higher frequency signals will follow these lower frequencies. Now as this SAW travels under the receiver transducer, it will continue to generate fields and integrate to produce voltages at the pads. Since the receiver transducer is a mirror (or time reversed) image of the input transducer and the SAW signal, there occurs a time when the whole wave form sits right under the transducer. When the spatial distribution of the signal matches the spatial distribution of the receiver transducer, at that instant, the maximum output is found at the contact pads. In Figure 5.5(d), we have plotted the response of the output transducer. We can now note a few facts about this kind of transducer.

a. The transducer arrangement provides a facility to correlate two signals. By that we mean that the incoming signal from one transducer is compared with a reference signal built into the second transducer.

b. By definition this is a matched filter. Suppose the input transducer generates the signal S(f). Then the

output transducer transfer function is the complex conjugate of the input signal, i.e., $S^*(f)$. In the time domain, the output filter time response is the time-reversed impulse response of the input signal.

Our next task is to recognize this as an M-ary correlator or M-ary matched filter.

5.4 CHIRP FILTER AS AN M-ARY CORRELATOR

In Chapter 2 we developed the principle and called for an M-ary correlator. In this Chapter we have pointed out the possibility of producing an M-ary correlator with uniform transducers, and have decided not to use it since it is cumbersome. Now we want to point out that down chirp arrangement, in fact, can be used as an M-ary correlator.

Suppose that we have sampled an FM signal in which only one frequency is contained, and this frequency is within the bandwidth of the chirp filter transducer. Corresponding to this frequency there will be a certain number of fingers that have the correct spacing for resonance, and will all contribute in phase. However, at some distance on either side, the phase of the signal contributions of these fingers will have shifted by 90°. Using this 90° phase estimate, the effective number of fingers (or length of the transducer) can be calculated by the equation given by [44]

$$N_e = 2f \left(\frac{T}{\Delta f}\right)^{1/2} \tag{5.2}$$

where f = the frequency of interest

T = the total dispersion of length of the transducer

Δf = bandwidth of the transducer

Assuming the pulse length of the sample is less than the equivalent length of the transducer which contains these effective fingers, the input transducer will emit a SAW of the same frequency as the signal. Now, the receiver will have a similar portion of transducer which will respond to the incoming signal. When the signal sits right under these fingers, it will correlate and give out a pulse. Thus for different signal frequencies, different sections of the input and output transducers will be in resonance. In Figure 5.6 we have explained this phenomenon. Furthermore, they are separated by different distances. If we rearrange these different sections as uniform transducers, we note that we have an identical situation as in M-ary correlator.

However, there is a subtle difference between the two. This difference is small and in fact is favourable to our system. However small the number of effective fingers are, their spacing varies linearly along the section of the transducer. In that sense one may regard these sections of the transducers as chirp matched filters. As such, the output will be a sinc function with some side lobes, yet the position of the peak of the main lobe will be governed by

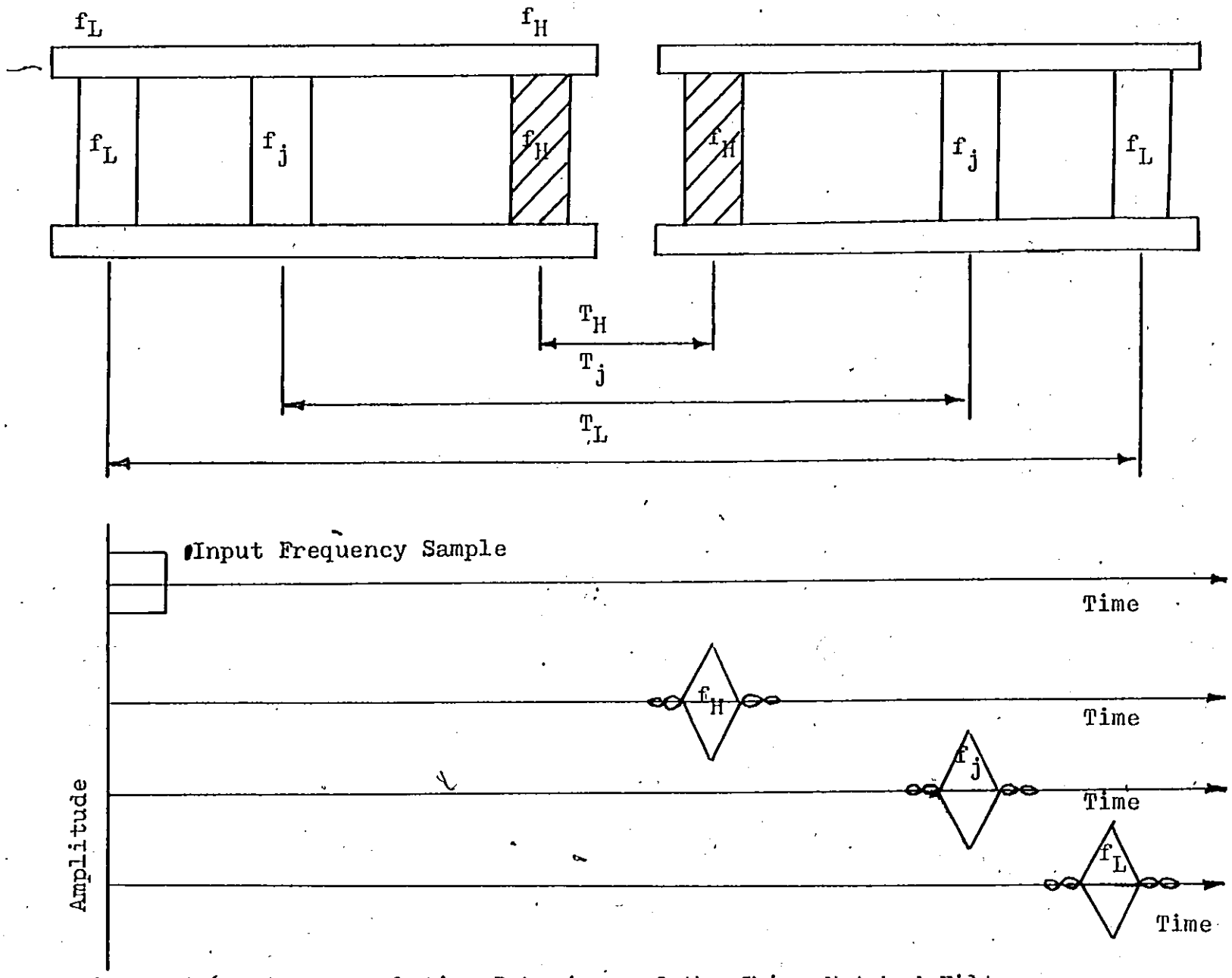


Figure 5.6 M-Ary Correlating Behaviour of the Chirp Matched Filter.

the main frequency of the sample. This situation is favourable because even if the sample signal frequency has a small deviation from the dominant frequency, the situation would help in providing a little more energy to the correlated pulse. Furthermore, we can relax the assumption that we took at the beginning concerning the frequency within the sample. A slight deviation can now be tolerated. In fact, the degree of deviation is determined by the time-bandwidth product of the device. Now that we have completed the development and the explanation of the physical principle of operation of the device, let us consider how to design a chirp transducer.

5.5 CHIRP FILTER DESIGN

In designing the surface acoustic wave chirp filter one may take different approaches. We can take all the second order effects into account at the beginning of the design procedure itself. This procedure would be extremely complicated and very expensive because of the required computer times. Even then, one could never be sure of the accuracy in materializing the prediction. Although some researchers claim they have written programs where this can be done within certain limits of accuracy, such programs have never been published for the general public. However, one reason why it is very difficult to achieve accurate

design programs which predict the performance is that fundamentally, all the second-order effects are not completely understood and modelled. Fortunately, there are various techniques developed to minimize their effects on the performance. Another way the design can be carried out is to carry out the initial design procedure neglecting the second-order effects and then use various techniques to minimize these effects on the performance. This procedure simplifies design routine. We will outline this procedure here.

The first problem is to decide on the initial parameters, such as the centre frequency of operation. f_o, the bandwidth of the device Δf, and the dispersive time T. Having done this, our next step is to find the proper positioning of the electrodes.

The impulse response of the transducer can be written in the complex notation in the form

$$h(t) = a(t) \cdot e^{j\theta(t)} \tag{5.3}$$

where $a(t)$ is the time-dependent envelope function

$\theta(t)$ is the time-dependent phase function.

In the FM chirp filter case, we take $a(t)$ as having a constant amplitude. There we may define the envelope function as

$$a(t) = \begin{cases} 1 & -T/2 < t < T/2 \\ 0 & \text{elsewhere} \end{cases} \tag{5.4}$$

In a linear FM modulated signal, the instantaneous frequency is given by

$$f_2 = \frac{1}{2\pi}\frac{d\theta}{dt} = f_o + \frac{\Delta f}{T}\, t \tag{5.5}$$

Therefore, the phase function becomes

$$\theta(t) = 2\pi \int_o^T (f_o + \frac{\Delta f}{T}\, t)\, dt$$

$$= 2\pi\, [f_o t + \frac{1}{2}\frac{\Delta f}{T}\, t^2] \tag{5.6}$$

We place the fingers at the zero-crossing points of the impulse response [45]. Therefore, the equation governing the position of electrodes is given by

$$\theta(t) = n\pi \tag{5.7}$$

where $n = 0, \pm 1, \pm 2, \ldots$

Applying this condition to equation (5.6), the time position of the fingers is then given by

$$t_n = f_o \frac{T}{\Delta f}\, [-1 + (1 + n\, \frac{\Delta f}{T f_o^2})^{1/2}] \tag{5.8}$$

The limits of n can be found from the boundaries of the transducer given by

$$n_H = T\, (f_o + \frac{\Delta f}{4}) \tag{5.9}$$

$$n_L = T\,(f_o - \frac{\Delta f}{4}) \tag{5.10}$$

where n_L is the limit of the low-frequency side of transducer, and n_H is the limit of the high-frequency side of the transducer. The total number of fingers is given by

$$N = n_H + n_L = 2\,f_o\,T \tag{5.11}$$

Having found the time dependency, the spatial positioning of the fingers can then be determined as

$$Z_n = V\,t_n \tag{5.12}$$

where V is the velocity of the SAW and Z_n is the spatial position of the electrodes.

Now the velocity in the free surface is different to that under the metal strip. Therefore, some correction must be made. A rule of thumb used for calculating the effective velocity in the case of equal mark to space ratio is

$$V_e = V\,(1 - \frac{dV}{2V}) \tag{5.13}$$

where V_e is the effective velocity and dv is the reduction in velocity due to piezo-electric stiffening. The effective velocity approximation is taken on the fact that half the transducer area is covered by metal. Therefore, the net velocity is the average velocity between the two velocities. Taking the new effective velocity into account, equation (5.12) modifies to be

$$Z_n = V_e\,t_n \tag{5.14}$$

Having found the electrode positions, our next problem is to

find the electrode length to obtain a frequency response that we require. Our ideal frequency response is shown in Figure 5.7(a). One method of doing this is to go through an optimization routine and find the apodization function required. Another method is to use a window function to multiply the impulse response and find the various aperture lengths. Yet, another method is to find an approximate apodization function and correct for errors experimentally. No matter which method we adopt, we finally end up with experimental corrections unless we incorporate the second-order effects into our computer program.

We shall look for a simpler method to do this. In Chapter 2 we used an equivalent circuit model to analyze the transducer, namely, the crossed-field model. We shall make use of that model. Let us suppose that we have a uniform acoustic wave of width W arriving at the chirp transducer. We want to know the strengths of the taps that we need to have along the chirp transducer to produce the transfer function we require. As we mentioned in Chapter 2, each finger is represented by a Mason equivalent circuit, and the complete transducer is represented with N equivalent circuits. This circuit can be reduced to a single equivalent circuit as shown in Figure 5.7(b). Slobodnik has developed a computer programme to do this type of analysis

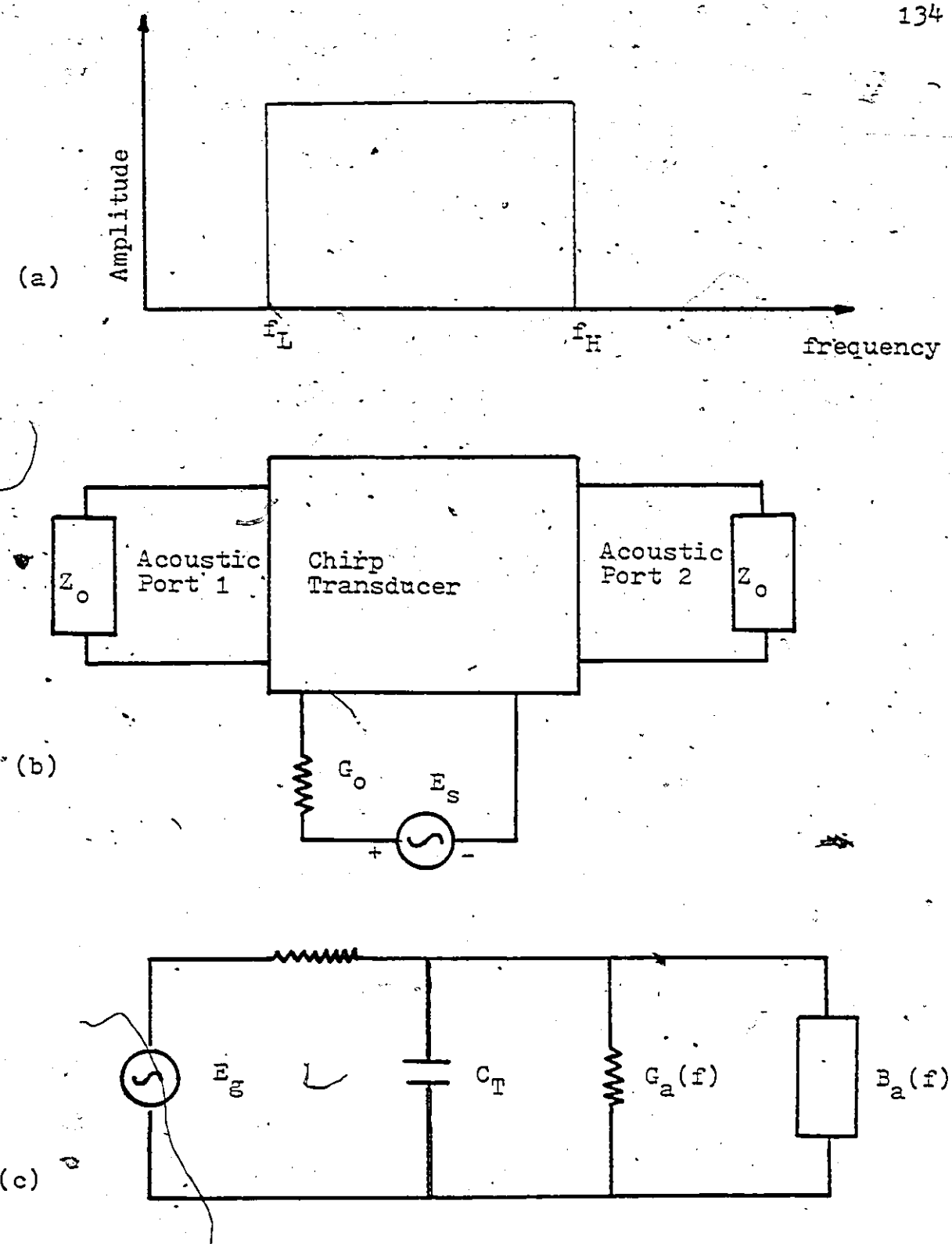


Figure 5.7 (a) Ideal Chirp Filter Response
(b) Equivalent 3 Port Network of the Chirp Filter
(c) Electrical Equivalent Circuit

[46]. Smith, Gerard, and Jones [45] carried out this analysis and proposed an apodization function given by

$$\frac{w_n}{W} = \left[\frac{f_n}{f}\right]^{-3/2} \frac{a(t_n)}{A} \tag{5.15}$$

where w_n is the overlap length of the nth electrode

W is the acoustic wave width arriving at the chirp filter

f_n is instantaneous frequency of the nth electrode

A is the maximum value of the impulse response, and

f is the normalization constant.

f is determined such that $W = w_n$ when the maximum value of w_n occurs. The transfer function for a chirp transducer which has this apodization function is given by

$$T_{13}(f) = \frac{j2\pi\ (2\ G_o\ K^2\ C_o\ W)^{1/2}\ F\ \{h(t)\}}{G_o + Y_{in}(f)} \tag{5.16}$$

where G_o = source conductance as shown in Figure 5.7(b)

$Y_{in}(f)$ = input admittance of the transducer

$F\{h(t)\}$ = Fourier transform of the desired impulse response

K^2 = coupling constant of the material

C_o = capacitance per unit length of one electrode.

Similar to the case of uniform transducer discussed in Chapter 4, the three-port network for the chirp filter shown in Figure 5.7(b) can also be translated completely into the

electrical circuit form as shown in Figure 5.7(c). With this circuit we can calculate the electrical admittance as given by

$$Y_{in}(f) = G_a(f) + j\,[w\,C_T + B_a(f)] \qquad (5.17)$$

where C_T is the capacitance of the entire transducer

$G_a(f)$ is the radiation conductance, and

$B_a(f)$ is the radiation susceptance.

Usually $B_a(f)$ is negligible compared to jwC_T. Thus we can ignore this term. Therefore, equation (5.17) becomes

$$Y_{in}(f) = G_a(f) + jwC_T \qquad (5.18)$$

Now, the conductance is related to the position and the geometry of the electrode. Thus, the conductance is given by

$$G_a(f) = K^2 \left[\sum_{n=1}^{N} (-1)^n (f_n C_n)^{1/2} \sin\left(\frac{\pi}{2}\frac{f}{f_n}\right) e^{-j2\pi f t_n} \right]^2 \qquad (5.19)$$

where f_n is the instantaneous frequency

C_n is the capacitance, and

t_n is temporal position of the nth electrode.

The next parameter that needs to be found is C_T which is equal to the sum of all the electrode capacitances and is given by

$$C_T = \sum_{n=1}^{N} C_n = C_o\, W \sum_{n=1}^{N} \frac{w_n}{W} \qquad (5.20)$$

Now we have all the parameters to find $y_{in}(f)$. With this information we can go ahead and design the chirp filter transducer. With equation (5.8) and (5.14) we can find the spatial position of the fingers. Equation (5.15) gives us the length of the fingers and equation (5.16) enables us to check the transfer function of the transducer.

The device that we are concerned with employs two apodized transducers. If we were to produce an in-line dispersive chirp filter, we would require two chirp transducers - one arranged as the mirror image of the other. If the transfer function of one filter is $H_1(f)$ then the transfer function of the other filter is $H_1^*(f)$ which is the complex conjugate of $H_1(f)$. Therefore, the overall transfer function is given by

$$\begin{aligned} H(f) &= H_1(f)\, H_1^*(f) \\ &= |H_1(f)|^2 \end{aligned} \tag{5.21}$$

The consequence of this is that the amplitude of the ripple in the passband is squared. The passband ripple is caused by second-order effects such as triple-transit echo, inter-electrode coupling in the dispersive transducers, bulk wave generation, diffraction effects, etc. In the case of the in-line doubly dispersive chirp filters, perhaps the most detrimental second-order effects would be bulk wave generation, diffraction and triple-transit echo. We

mentioned earlier that for a given frequency, only certain parts of the transducers will resonate. In an in-line doubly-dispersive geometrical structure, the SAW travels right along the length of the two transducers, and while it travels along the non-resonant areas of the transducer the SAW interacts constructively to produce a large reflection or large bulk wave generations. The diffraction effect is a function of aperture and the distance the SAW travels. In the in-line doubly-dispersive case, this could be significant especially for small aperture SAW components. Both these effects can be minimized by using a multistrip coupler [43]. Perhaps the bulk wave suppression would be the most significant contribution that the multistrip coupler could make in this particular case. This is because it would suppress bulk wave generation due to one transducer completely [43].

5.6 MULTISTRIP COUPLER DESIGN

The behaviour of the multistrip coupler has been dealt with in Chapter 4. Here we shall explain a very simple rule of thumb method to design a multistrip coupler track changer for LiN_bO_3. We need to find the correct length of the coupler if 100% of the energy is to be coupled from one track to the other. At a frequency f_s, the coupler becomes a resonant structure and behaves like a reflector.

This frequency is known as the stopband frequency and is given by

$$f_s = \frac{V}{2d} \tag{5.22}$$

where V is Rayleigh wave velocity, and

d is the repeat distance of the periodic multistrip coupler.

The device is broadband and below the stopband, energy transformation from track to track can be achieved. For an equal gap-to-width of the metal strip, the total number of strips required for 100% energy transform is given by [47]

$$N_T = \frac{195\ f_o/f_s}{1 = \cos\ (153\ f_o/f_s)} \tag{5.23}$$

As a rule of thumb the working frequency f_o is set to be 0.75 f_s. For LiN_bO_3, N_T is approximately 104 strips. The length of the coupler is twice the maximum aperture width of the chirp transducer.

5.7 COMPLETE DISCRIMINATOR HARDWARE IMPLEMENTATION [48]

To explain the hardware implementation of the discriminator principle we use the system block diagram as shown in Figure 5.8. The functional diagram of the discriminator illustrated in Figure 3.2 has been super-

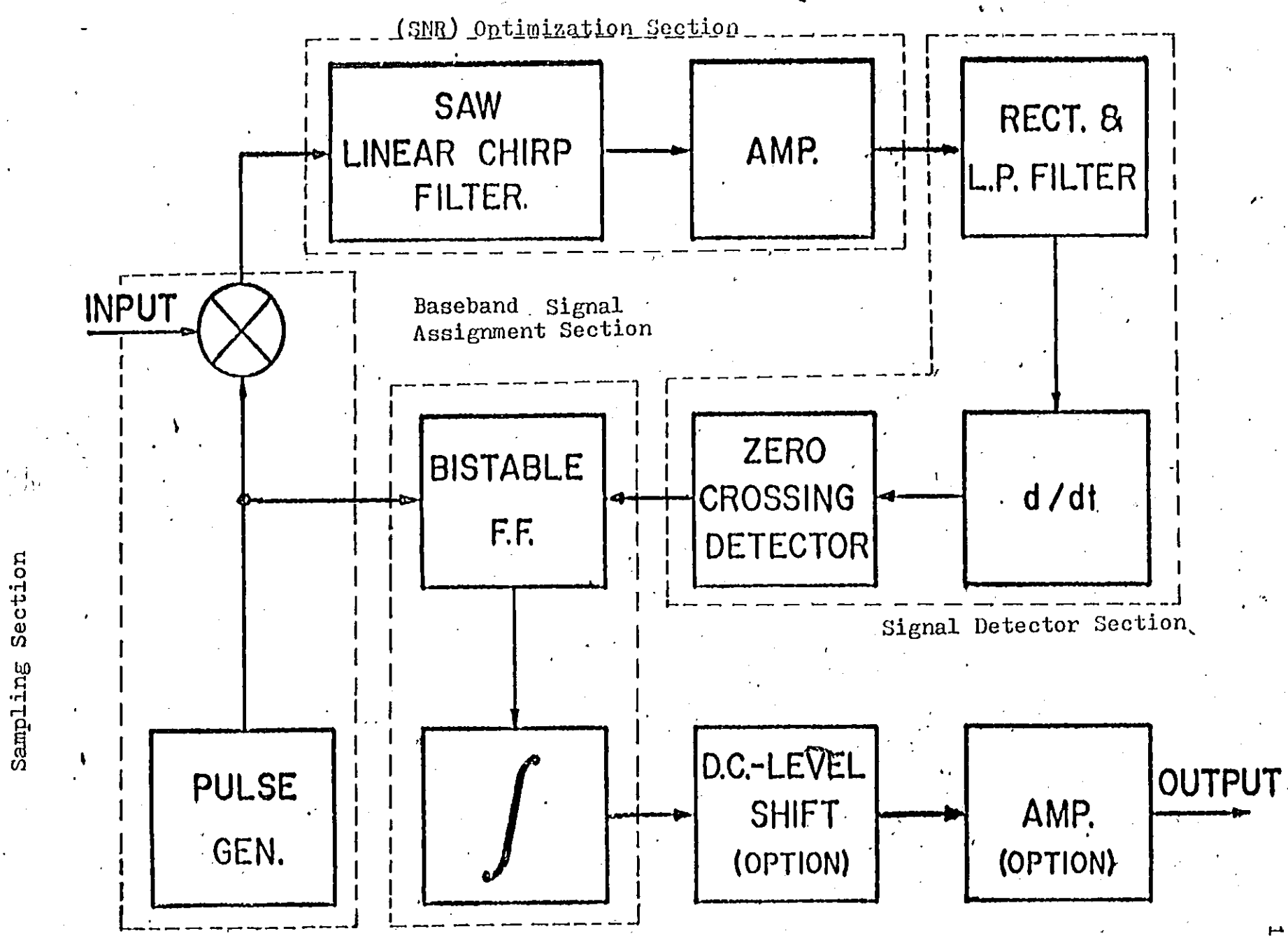


Figure 5.8 System Block Diagram of the Discriminator

imposed on Figure 5.8 and shown in dotted lines. In the sampling section, the incoming FM signal plus noise is sampled using a mixer and a pulse generator. The sampling signal frequency generated by the pulse generator is set to f_s and the pulse width of the sampling pulse is T_s. The leading edge of the sampling pulse is used as a reference to set the bistable flipflop to "high" state.

The sampled received signal is applied to the SAW linear chirp filter. In this section we optimize the signal to noise ratio. The output of the SAW chirp filter is a delayed correlated pulse which depends on the input signal frequency. We amplify this pulse to overcome the signal loss in the SAW device due to its insertion loss. Our next task is to detect the signal.

5.7.1 Zero-Crossing Detector Technique

The frequency response of the chirp filter will have some ripples in the passband. These ripples will cause the received correlated pulses to vary in amplitude. In other words, pulses corresponding to different frequencies will have different amplitudes. The information regarding the received signal is in the position of the peak of the pulse. Therefore, the accuracy of detecting the peak will determine the accuracy in recovering the baseband signal. Another way

of saying this is that any error in peak position detection caused by amplitude variation will show up as non-linearity of the transfer function of the discriminator. Therefore, it helps us a great deal to keep these errors to a minimum.

What we need is a peak detector which is insensitive to any amplitude variation of the correlated pulses. Voltage threshold type peak detectors cannot help us here. In fact, they will cause more errors. The technique we adopted for peak detection is shown in the signal detection section of Figure 5.8. The amplified output of the SAW chirp filter is rectified and lowpass filtered to obtain the positive part of the envelope. This wave form is a pseudo-triangular waveform. If we now differentiate this wave form, we get a waveform as shown in Figure 5.9. Notice the zero crossing occurs right at the peak of the pulse, regardless of the amplitude of the pulse. Now we feed this signal into a zero-crossing detector circuit. The edge corresponding to the zero-crossing point of the output pulse of the zero-crossing detector is being used to reset the bistable flipflop to a low position. Thus, we have a peak position detector circuit which is insensitive to variation in signal amplitude.

The bistable flipflop circuit belongs to the baseband signal mapping section or signal demodulation section. The output of the flipflop is a constant amplitude rectangular

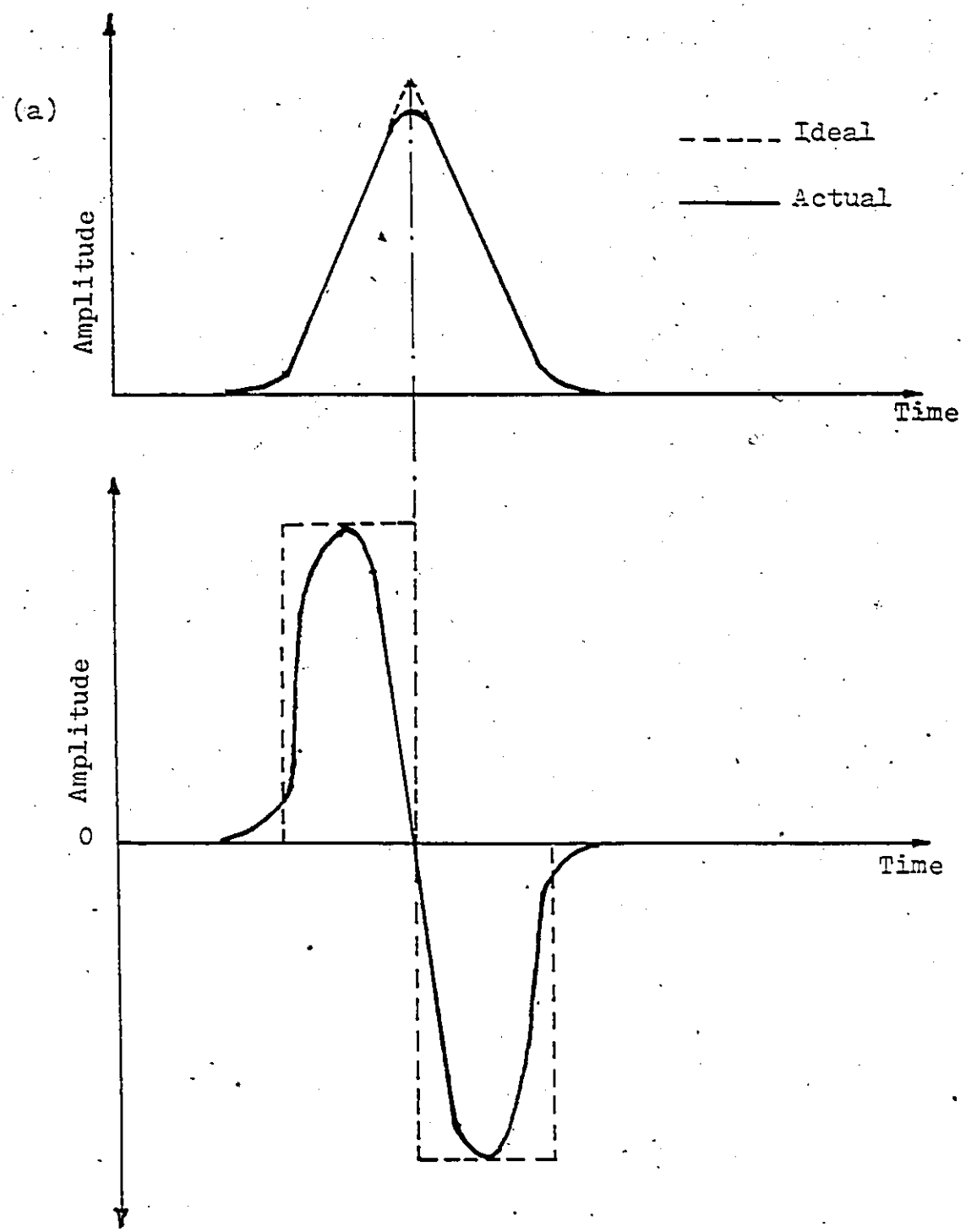


Figure 5.9 (a) Output from Rectifier and Low-Pass Filter Circuit
(b) Differentiated Signal Indicating Zero Crossing Time Corresponding to the Time of the Peak of the Output Signal from the Rectifier and Lowpass Filter Circuit

pulse whose pulse width is equal to the time delay of the SAW chirp filter. As we mentioned before, this time delay is linearly proportional to the signal frequency of the sample. Therefore, the pulse width of the bistable output pulse is linearly proportional to the frequency. Now if we integrate this pulse we find the output voltage of the integrator is proportional to pulse width. Therefore, we have the output voltage of the complete discriminator proportional to the input frequency, which is the required function of the discriminator.

We can explain the operation of this discriminator in another way. The input FM modulated signal is converted to a pulse-duration-modulated (PDM) signal. Then the pulse duration modulated signal is demodulated to recover the baseband signal.

CHAPTER 6

FABRICATION OF SAW DEVICES

In Chapter 5 we discussed the hardware implementation of the sampling FM discriminator. There, we pointed out the physical implementation of functions which were defined in the fucntional diagram. The most important element in this discriminator is the SAW linear FM chirp filter. We have discussed its behaviour and outlined the design procedure. Since SAW devices are relatively new in communication, and the advantages such as reproducibility, reliability and production economy come about because of their fabrication process, it is appropriate to discuss fabrication in this Chapter. Furthermore, we will describe a simple thin film technique [49] to modify the aperture of an interdigital transducer.

6.1 SAW PLANAR TECHNOLOGY

If there is a single discovery which is responsible for the rapid development of SAW technology, it would be that of the IDT. This coupling structure is planar and thus planar fabrication techniques developed for integrated circuit construction can be applied.

In general, a SAW device consists of some sort of interdigital electrode pattern on a single crystal of piezo-electric material. The fabrication process steps required to produce these devices are fewer than what normal integrated circuits require. SAW planar technology can be divided into the following areas:

1. Substrate preparation
2. Film deposition
3. Lithography

6.2 SUBSTRATE PREPARATION

As mentioned earlier, in general, SAW devices employ single crystal piezo-electric substrates. Thus, substrate preparation includes techniques such as crystal growth, orientation, cutting, polishing and cleaning. Generally, it is possible to buy a prepared substrate for most purposes. However, often one finds the quality of polishing and cleaning is not up to the desired standards. Surface quality of the substrate is extremely important to SAW devices. Defects on the surfaces such as scratches, roughness and other work damage, increase the losses in the device [50].

Polishing is at the present time more of an art than a science. This lack of refinement in the process may result in inconsistent surface conditions from batch to

batch. It is difficult to make an estimate of the reproducibility of these techniques because most manufacturers keep their techniques closely guarded. However, some efforts have been spent on characterizing polishing techniques. Deitz and Bennett [51] introduced the so-called "bowl feed polishing technique". The technique involves polishing the material in a mixture of water and polishing compound. Polishing is continued while the work piece is immersed in the slurry and an agitator continues to stir the slurry while polishing. They claim that a surface roughness of less the 5 A can be obtained by this technique.

There are other polishing techniques, including the chemical-mechanical polishing technique, ion bombardment polishing technique and polishing by chemical etching. We shall not review these techniques here. However, we point out that smoothness of the surface rather than flatness of the surface is more desirable in these preparations.

Another area of substrate preparation that is extremely important and often creates a great deal of problems is the substrate cleaning. An improperly cleaned substrate can cause problems such as poor adhesion, difficulties in etching, open-circuit fingers, and the creation of scattering centers or lossy media to the SAW, etc. This is another process which varies from laboratory to laboratory. Furthermore, there is no universal cleaning

procedure common to all materials. One needs to adapt the cleaning procedure to suit the situation and material. The step by step procedure we follow for fabrications relating to this thesis is given in Appendix A.

6.3 FILM DEPOSITION

Thin metal films are required in the fabrication of the IDT. It is necessary, therefore, to coat the substrates with metal at some stage in the process. Usually, metal depositions are carried out in a suitable vacuum system. The most popular metals used are aluminium and gold. The latter cannot be deposited by itself due to its poor adhesion properties on most SAW materials. Thus, a seed layer of chromium, or nickel chromium alloy or titanium is used as an interface layer to aid adhesion.

The method used for the deposition of metal includes thermal evaporation, electron beam evaporation, RF sputtering and DC sputtering. The choice of the method used may depend on the process, adhesion requirements and availability of the equipment. For example, the lift-off process [52] which we will discuss later in this Chapter, requires a method which has better adhesion properties than that offered by thermal evaporation. In terms of adhesion properties the best method is sputtering; the next best is the electron beam evaporation. Thermal evaporation is the

poorest technique for adhesion. Adhesion properties do not depend solely on the metal evaporated and the method employed. They also depend on other parameters such as vacuum conditions and the cleanliness of the subtrate. Often it helps to preheat the substrate and/or ion clean it prior to evaporation. However, this cannot be done if one is using the lift-off process for pattern definition.

For SAW devices the thickness of the metal required is usually only about 2000 Å, and with such thick layers one has to be careful about their quality. Uniformity of thickness and pinhole density are two factors that could affect the subsequent processing and performance of the device. Non-uniformity of metal layers could produce non-uniform patterns due to differential etching. In fact, this could have severe consequences in large area devices such as large time-bandwidth product filters. Sometimes with these devices it pays to use rotating substrate holders. An increase in pin-hole density can cause non-uniform etching patterns and possibly create open circuits in the IDT. Poor cleaning of the substrate, and the roughness of the substrate, are some of the factors that could affect the pin-hole density. Thus we see how the preparation of the substrate can affect the metal deposition. If the process to be used is a conventional contact printing process for pattern generation, then it helps to preheat and/or ion

clean the substrate prior to the evaporation of the metal.

There are SAW devices that would require thicker metal layers. For example, surface acoustic wave guides [53] would require these metal layers. Such devices usually use gold layers. For thicker layers it is difficult to use thermal evaporation due to the poor adhesion characteristics, and of course, the difficulty in melting the quantity of metal needed to obtain the thickness of the layer. With the standard equipment found in many laboratories, the upper limit of the thickness that can be achieved is about 1 m.

Metal films are not the only films SAW devices may require. There are other dielectric materials that have been used as guides for these devices. One particular area of SAW devices that need non-metallic films relates to the so-called integrated SAW devices [54]. Here the general concept is to integrate SAW devices with the necessary integrated electronics onto one substrate. Basically, the thin piezo-electric material is deposited on a nonpiezo-electric substrate for the SAW devices and, in another section of the same substrate, thin film semiconductors are used for the integrated electronic circuits. This is an emerging technology. For most of this work, thin film sputtering techniques are being developed [55]. One very promising area which may prove to be successful would be to couple these techniques with silicon-on-sapphire technology.

Piezo-electric materials used with some success in sputtering are zinc oxide [56] and aluminium nitride [57]. The latter has been deposited on sapphire by thermal evaporation as well. There are indications that this field of SAW device technology may be most promising.

6.4 LITHOGRAPHY

The basic principles of lithography involve generating a pattern using a mask. The function of the mask is to define the boundaries of the pattern in the process of pattern generation. In Figure 6.1 we have illustrated the most widely used lithographic process. Here we coat the substrate with a metal layer as shown in Figure 6.1(a). Next we coat the metal layer with a radiation sensitive polymer - see Figure 6.1(b). Then a mask which contains radiation absorbers is used to expose, with radiation, the desired areas of the polymer. The absorbers are arranged in such a way to define the boundaries of the required device pattern. This step is illustrated in Figure 6.1(c). The radiation can either enhance the polymer by cross linking (in the case of negative resist) or make it unstable by polymer chain scission (in the case of positive resist). The unwanted polymer is removed in the next step, as shown in Figure 6.1(d).

Finally the unwanted metal is taken away from the

Figure 6.1 Lithographic Process

substrate by some form of etching. The polymer is then removed after completion of the etching. The final pattern in the metal is shown in Figure 6.1(e).

6.4.1 Lift-Off Technique

The other method of lithography is the so-called lift-off technique. The principle of this method is shown step by step in Figure 6.2. The first step is to coat the substrate with a polymer. Then the polymer is exposed through a mask in order to define the desired pattern. We next remove the unwanted polymer, thus leaving uncovered areas of substrate in the appropriate areas. Deposition of metal onto the entire substrate is carried out as the next step. Finally, the unwanted metal and polymer are removed, leaving behind the metal which has adhered directly to the substrate. The final pattern is thus obtained.

The basic principles of both of these methods were known to man as far back as the pre-Christian era. About 300 years BC, Arizona Indians used etching methods to produce artifacts. Over 2,500 years ago the Ceylonese used generating patterns through masks and lifting off masks in their art of Batik. However, credit must be given to the ingenuity of modern man in developing these techniques to a powerful tool in the electronics industry. The relative advantages and disadvantages of these two techniques would

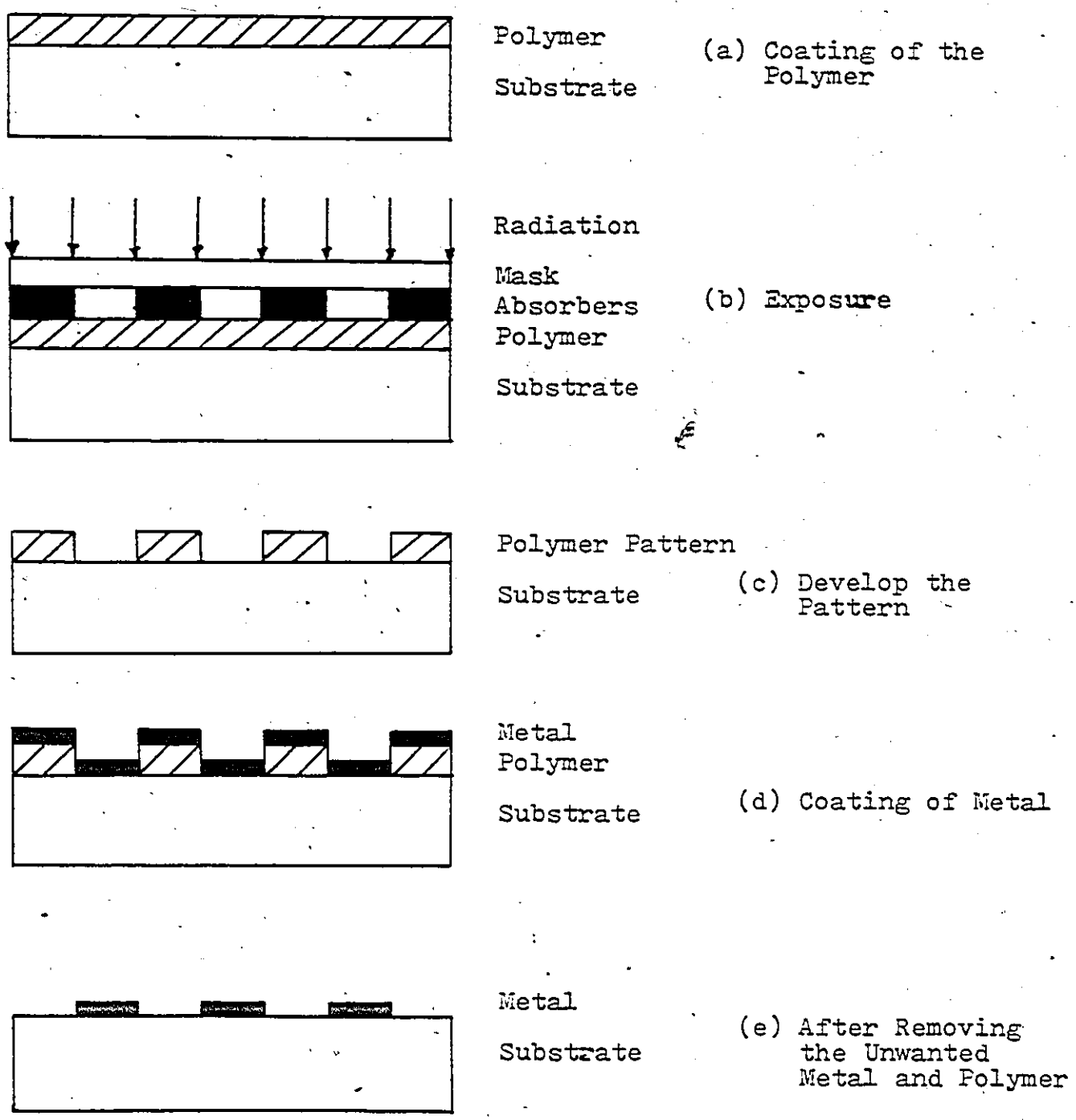


Figure 6.2 Lift-off Technique

become evident in the rest of the discussion in this Chapter.

Lithography can be divided into 3 catagories on the basis of radiation used for pattern definition. They are:

1. Photo-lithography
2. Electron lithography
3. X-ray lithography

6.4.2 Photo-Lithography

Light is used as the radiation source in photo-lithography. The polymer used here is a light sensitive polymer, popularly known as photoresist. There are many varieties of photoresists available. The sensitivity of these resists is a maximum when the light source used is an ultra violet one. There are two types of resists, called positive photoresists and negative photoresists. For SAW device fabrication, positive resists such as Shipley Type AZ1350 are widely used for their high resolution capabilities. This photoresist, when exposed to ultra violet light, selectively creates a reaction that leads to selective solubility in a developer. Selective exposure can be achieved by two techniques, namely

1. Contact printing
2. Projection printing

6.4.2.1 Contact Printing

In contact printing for exposure, a mask is used for selective exposure. The mask may be made of a thin film of chromium or photo emulsion on a glass plate. The pattern on the glass is identical to the pattern required to be produced. Such a mask is called a positive mask. It has to be used with positive photoresist if conventional lithography is being used. Here, we will consider a conventional contact technique first. The mask is placed in contact with a substrate coated with metal and photoresist. The exposed photoresist is removed by developing, and the metal is etched away chemically or by other means.

The most serious problem in this technique is diffraction. As illustrated in Figure 6.3 the incident ultra violet light diffracts at the edges of the pattern and exposed areas under the mask. The result is that, after developing, a fraction of the area under the mask is also being exposed. After etching, one finds a net reduction in line widths. This reduction is called under-cutting. The diffraction is mainly caused by the distance between the mask and the substrate.

The problem can be minimized by using conformable masks and the lift-off techniques [52]. Conformable masks are made with very thin (≈ 0.2 mm) glass plates. The cross section of the substrate and mask holding jig used for

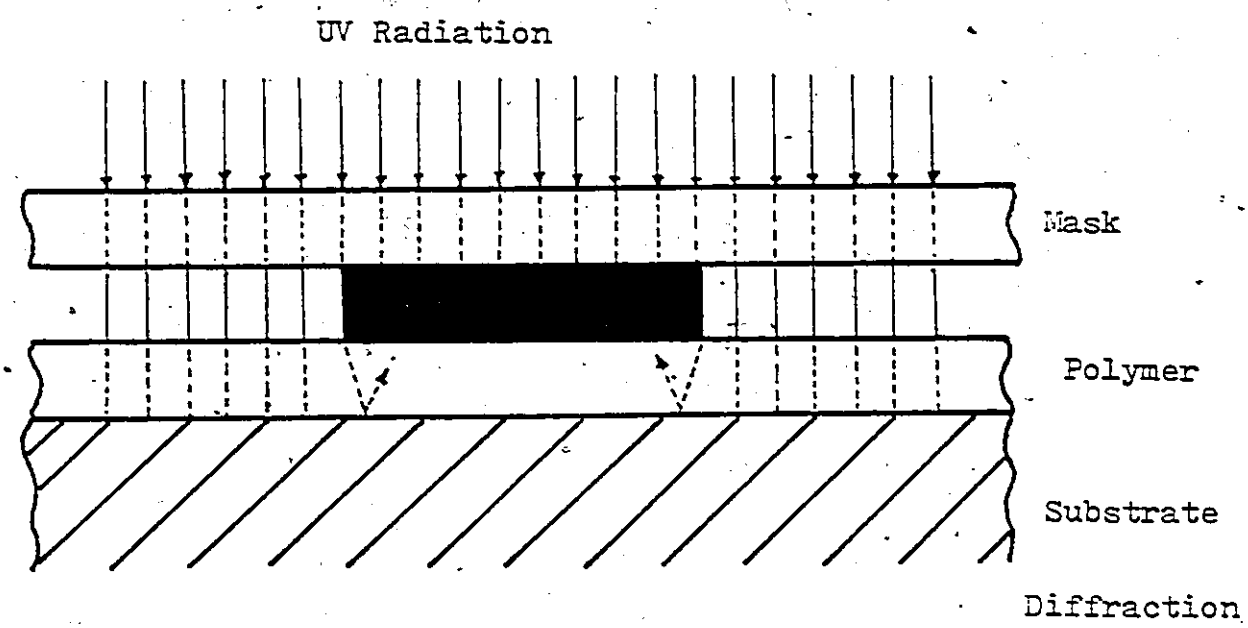


Figure 6.3 Diffraction Effect

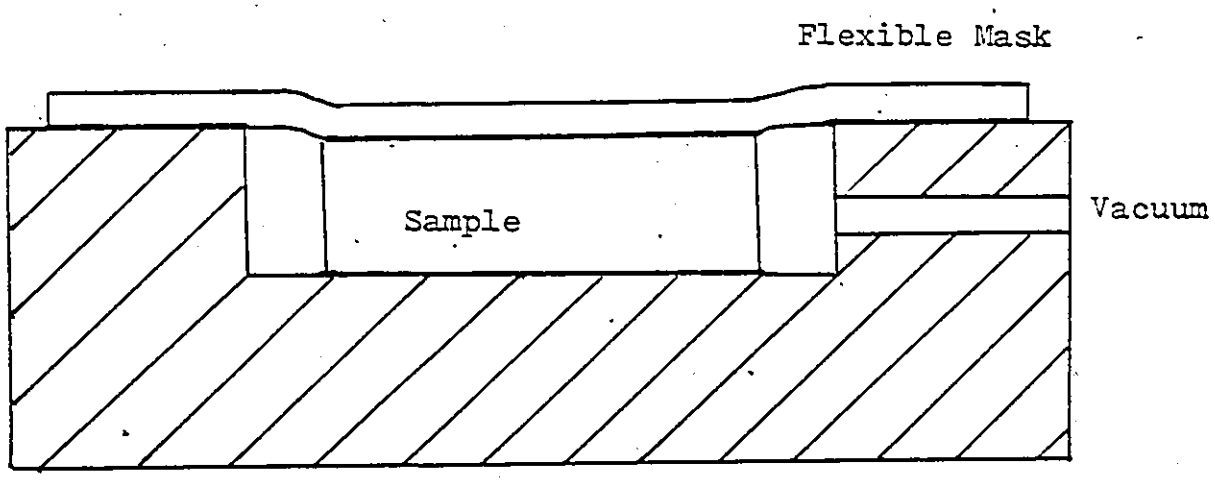


Figure 6.4 Cross Section of the Vacuum Jig used for Conformable Mask Technique

conformable mask techniques are shown in Figure 6.4. The major advantage of this technique is that one can minimize the diffraction effects. A flexible mask conforms to the contours of the substrate, minimizing the distance between the substrate and the mask and reducing diffraction effects. Even if there is some diffraction resulting in a slightly under cut photoresist profile as shown in Figure 6.5(a), the profile helps the lift-off process. The success of the lift-off process depends on the photoresist profile. Vertical wall profiles prevent the metal from bridging providing that the thickness of the metal is less than the wall height, thereby enabling one to remove the unwanted metal and photoresist easily. Even if there is a small metal bridge, as long as its thickness is small, the photoresist may be removed. This is because, at the removal stage, the solvent expands the photoresist, and cracks open the bridge. However, if there is a slight under-cut as shown in Figure 6.5(a) one can be sure that the thickness of the metal in the bridge is very small. If not, there is an absence of a bridge. In Figure 6.5(b) we have illustrated this point.

The masks for contact printing are made through a series of reduction procedures. Initially the art work is carried out on a much larger scale on Rubylith. This art work is then reduced (usually 20X) to obtain the final mask

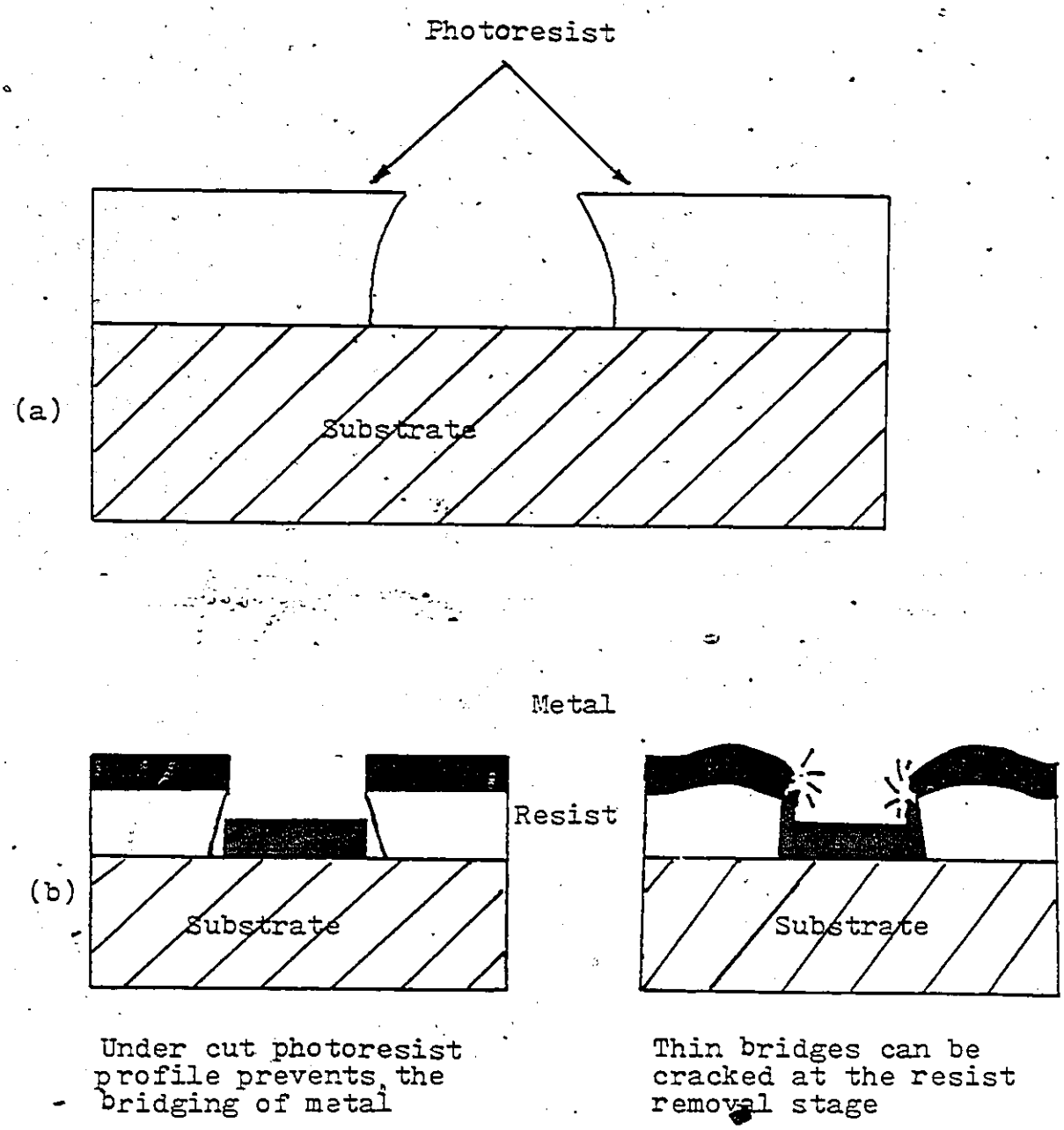


Figure 6.5 (a) Under cut Photoresist Profile
(b) The Effect of Photo Resist Profile on the Process of Metallization and Lift-off

size. Conformable masks are made use of in the lift-off technique from the original photo mask. A similar jig to the one shown in Figure 6.4 is used in making the conformable mask. As we may notice there is a phase reversal when the lift-off technique is used. Thus one has to use a negative mask in the lift-off technique to produce the pattern.

The normal contact lithography which uses etching can produce ultimate linewidth resolutions of about 1μm. Actually this figure is very generous. A more practical figure would be 3 μm. In contrast to this lift-off technique one can produce linewidths below 1μm with relative ease.

6.4.2.2 Projection Printing

In projection printing, the image of a mask is projected onto a photoresist-coated substrate. Often the projection printing optical systems have demagnification stages. These systems are very expensive due to the precision lenses that they use. It is possible to generate sub-micron lines using this kind of system. However, with such orders of linewidths, the field of view is restricted. The method does not eliminate diffraction. One finds diffraction effects at the focal plane. The other problems associated with this technique include pattern distortions and difficulty in focussing. In general, this method of

photolithography is not widely used. With proper improvement however, it may prove to be a useful method for producing masks.

6.4.3 Electron-Beam Lithography

Here the radiation source used for pattern definition is electrons. The substrate is covered with electron resist, and the pattern is defined by exposing a selective area to electrons. These areas will undergo cross linking of the polymer (in the case of negative resist) or chain scission of the polymer (in the case of positive resist). Upon removal of the unwanted polymer, the pattern can be generated by the normal lithographic methods discussed. The most popular electron beam technique to date, has been the so-called scanning electron beam lithographic system. The basics of this system are illustrated in Figure 6.6. Here a focussed beam of electrons is being used to write the pattern on the substrate. The pattern is generated by a computer and the beam and the substrate positions are controlled by the computer via beam control electronics and substrate stepping control electronics, respectively.

Since an electron beam is used to write the pattern, the ultimate resolution of the system will be governed by the diameter of the electron beam. Although it is possible to focus the beam down below 100 Å diameter, it is extremely

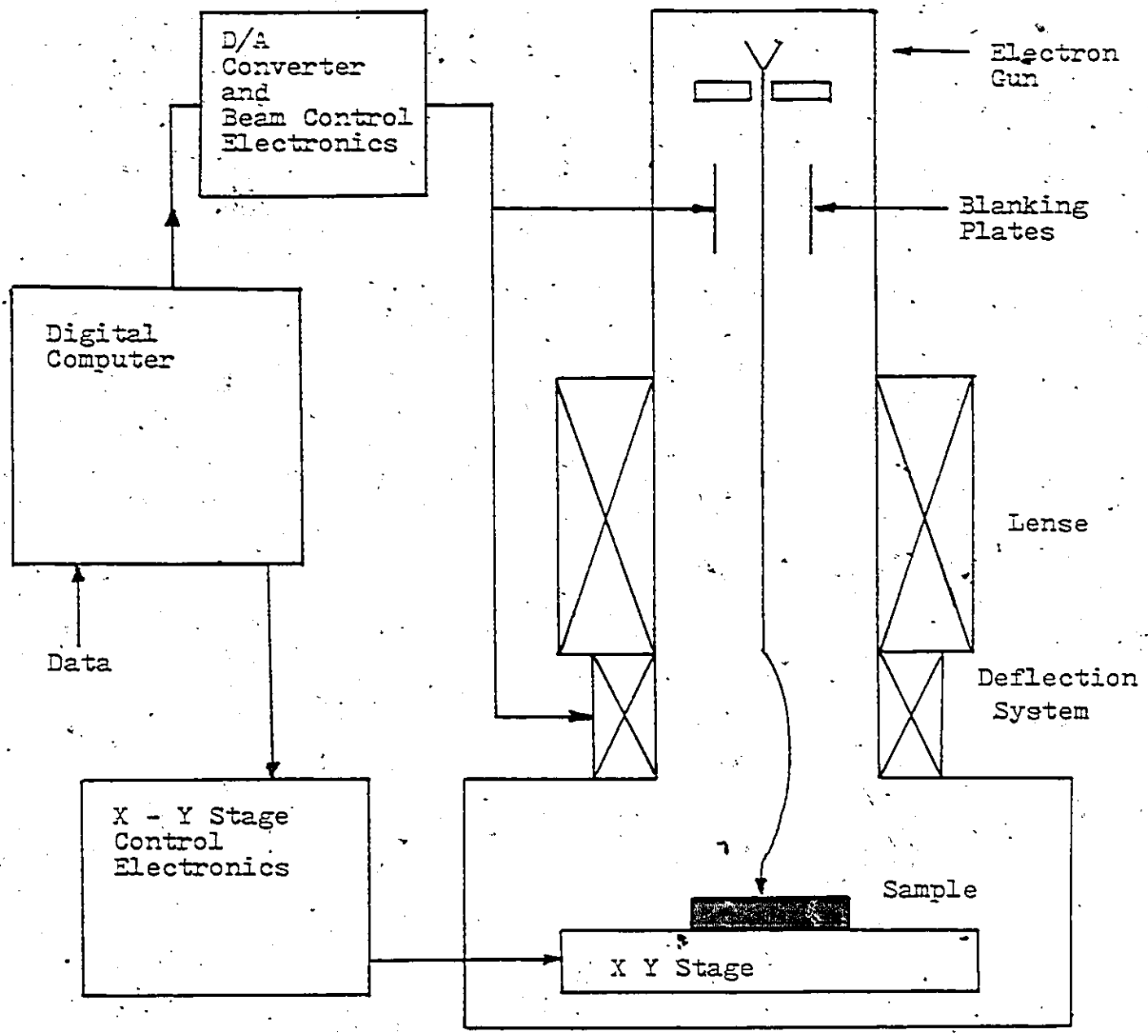


Figure 6.6 Basic Electron Beam Lithographic System

difficult to produce line widths of this dimension. This is because back scattering of electrons prevents the achievement of this resolution. The maximum resolution that can be achieved in practice is 1000 Å.

It is necessary to have extremely accurately controlled substrate stepping control systems in electron beam microscopes due to the limited field of view. For beam diameters less than 1000 Å the practical maximum field of view is only about 1 mm^2. A considerable amount of work has been carried out developing an accurate control system. As a result there are electron beam lithographic systems in the market that use laser interferometers to monitor and control the motion of the substrate by the computer.

Pattern generation is carried out using computer control. In principle there is no restriction on using analogue or digital computers. However, since, the prime concern here is high resolution patterns, the accuracy of the pattern generator becomes a significant factor, and for this reason digital computers are being preferred over analogue ones.

The electron resists used are polymers. They are dielectric materials, as are the substrates. Constant bombardment of electrons can charge these dielectrics. These charging effects can cause distortions. However, this can be minimized by providing a short circuit path to the

beam current. The standard technique used is to evaporate a thin layer of metal film over the polymer and make a connection between the metal layer and the ground.

The most penalizing disadvantage in electron beam lithography is the exposure time. The pattern is written on the substrate point by point; therefore, one may look upon the system as a serial system. If the device required is a complex structure, the time required for exposure can be hours. Such a system is not practical in the production environment. It is difficult to see how this problem can be overcome with a serial processing technique.

As an effort toward solving the exposure time problem, three other techniques have been suggested. They are:

1. Transmission electron lithography
2. Photo-emitted electron projection lithography
3. X-ray lithography

All these techniques are pattern replication techniques, and they are parallel processing techniques. The basic philosophy adopted in these techniques is to produce the original mask using a scanning electron beam system, and then use this mask in these pattern replicators to produce the subsequent devices. Here we will very briefly outline the basic principles of these ideas.

6.4.4 Transmission Electron Lithography [88]

The basic principle of operation is illustrated in Figure 6.7. The electrons are transmitted through an electron opaque mask, and the image is focused onto a substrate. The masking is achieved with electron absorber pads arranged in a pattern. Initial papers on this system do not go into any details other than to explain the very general principle [58]. The results they published were promising. Exposure times have been cut down to less than a minute. The idea is still in the research stages, and very little is being discussed on the subject in the open literature.

6.4.5 Photo-Emitted Electron Projection Lithography [59]

The basic components of this system are illustrated in Figure 6.8. The mask is made with ultra violet transparent material and the masking of the pattern is carried out with ultra violet absorbent materials. Next a thin layer of photo conductive material is coated on the surface. The mask is loaded into the vacuum system which contains the substrate at the bottom. Solenoids are used to obtain parallel paths for electrons. An ultra violet light is used to excite electrons from unmasked areas, and these electrons travel onto the substrate and form an image of the pattern. In principle, this is very simple and

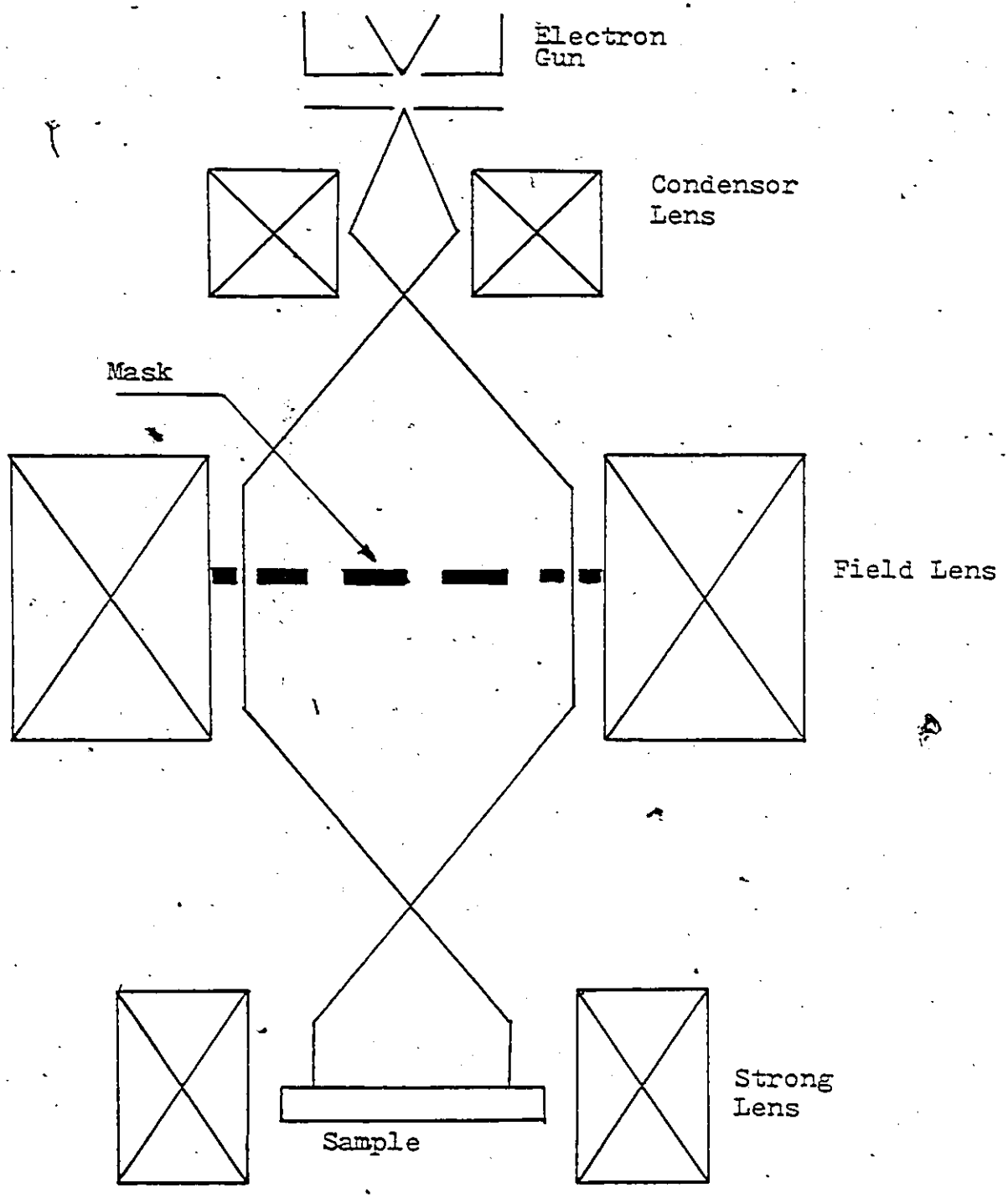


Figure 6.7 Basic Transmission Electron Lithographic System

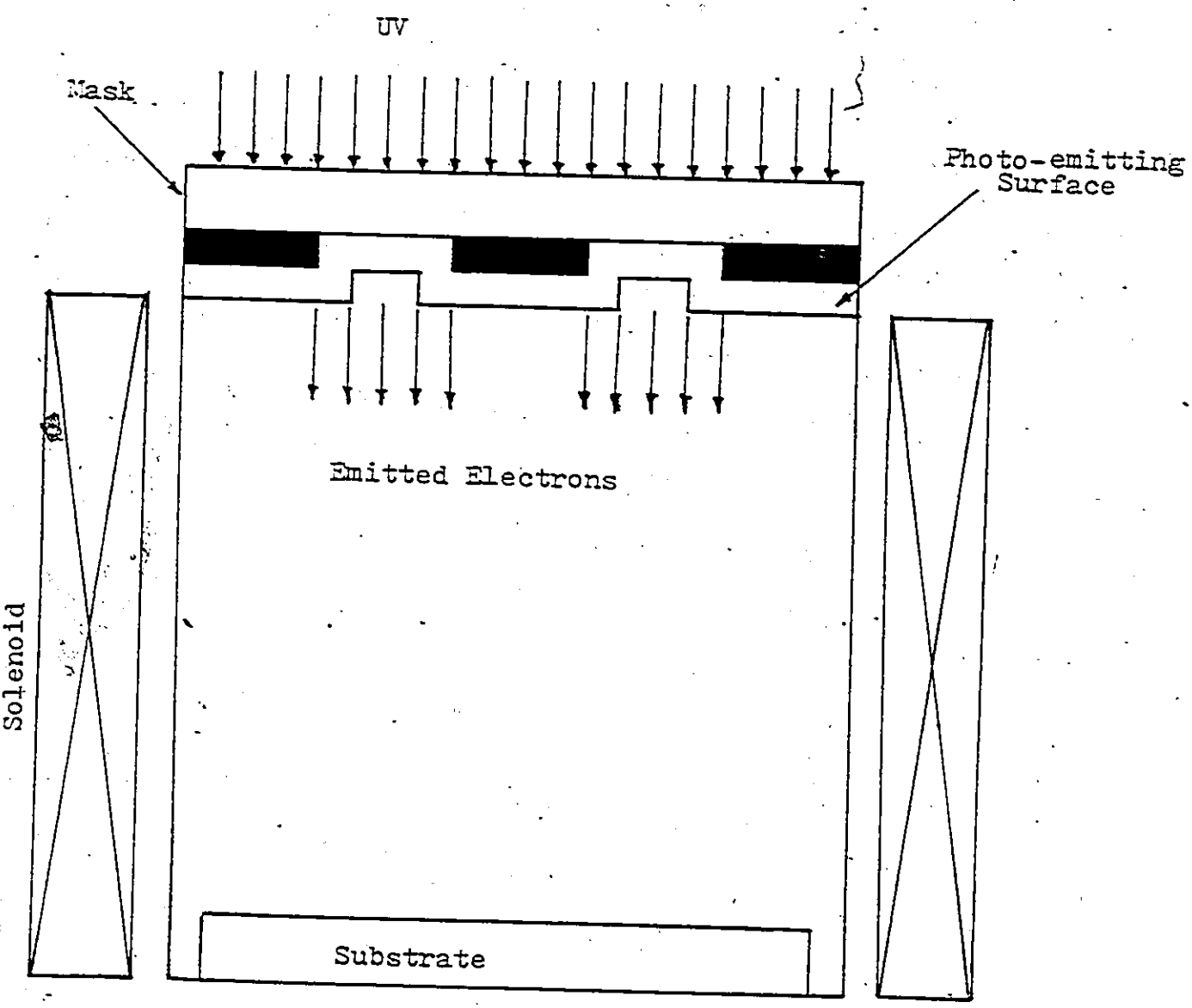


Figure 6.8 Basic Photo-emitted Electron Projection Lithographic System

straightforward. However, in practice there have been many problems in achieving the ideal conditions. Some of these problems are the distortion, orientation and precision of the image, and electrical and magnetic field uniformity and stability. Similar to transmission electron lithographic system, this system is in the research stage, and is yet too early to comment on its success.

6.4.6 X-Ray Lithography

Out of the three pattern replication techniques, x-ray lithography has proven to be the most successful to date. However, this relative success may very well be short lived, especially since the other two systems in principle can offer better exposure times. The basic principles of x-ray lithography are shown in Figure 6.9. Here the beam of electrons hit a metal target and emit soft x-rays. The wave lengths of the x-ray depends on the target used. Wave lengths of the x-rays used vary from 4 Å - 44 Å. These x-rays are then used to expose the substrate coated with polymer through a mask. One such polymer used for x-ray lithography is known as polymethyl methacrylate (PMMA). It is also one of the polymers used for electron beam lithography, and is a positive resist. There are other polymers [60] that can be used for x-ray and electron lithography. Also there are other negative resists that can

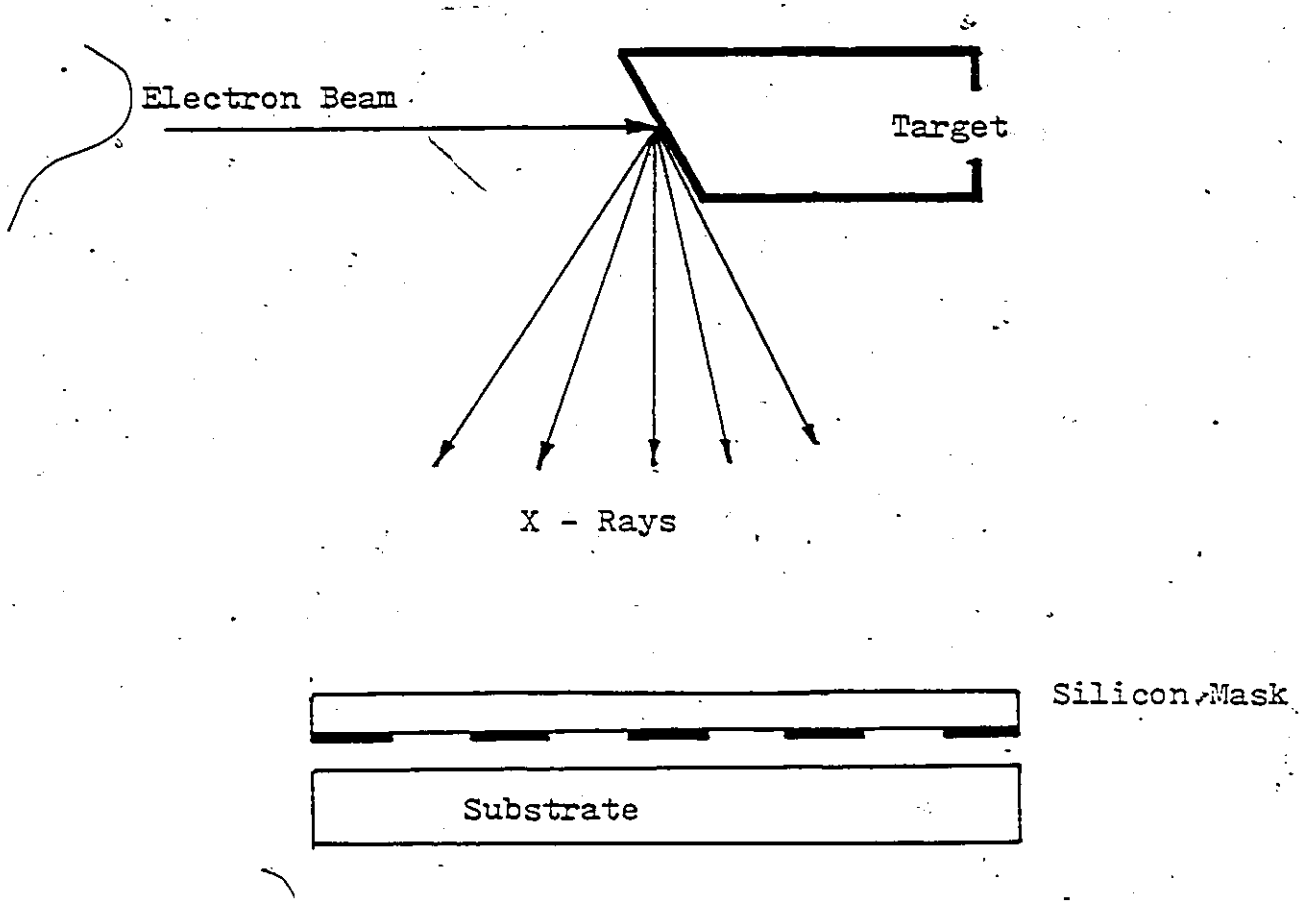


Figure 6.9 Basic Principle of X-Ray Lithography

be used for both types of lithography. Kodak micronegative resist (KMNR) is such an example.

One of the basic problems with x-ray lithograpy has been the exposure time. Initial experiments required exposure times of hours. The actual value of the exposure time depends on the x-ray source, the polymer, and the mask. The recent advances show that exposure time can be cut down to a few minutes (20 minutes for PMMA and 7 minutes for KMNR). This is a significant improvement, and there are good indications that this exposure time can be cut down further.

The mask used for x-ray lithography is a thin membrane mask. These thin membrane masks are necessary to reduce the absorption losses. A typical cross section of the mask is shown in Figure 6.10. The pattern is made with metal absorbers, and the material of the mask is silicon. The thickness of the membrane is 3-5 μm. It is possible to produce this silicon membrane [61] using crystallographic preferential etches. Such an etch is a mixture of ethylene diamine, pyrocatechol, and water [62]. This etch has a different etch rate for different silicon crystal planes. The least etch rate is found on <111> plane which is 3 μm/hour compared to 30 μm/hour for <110> plane and 50 μm/hour for <100> plane. Furthermore, it is an impurity preferential etch. If the p type (boron) impurity

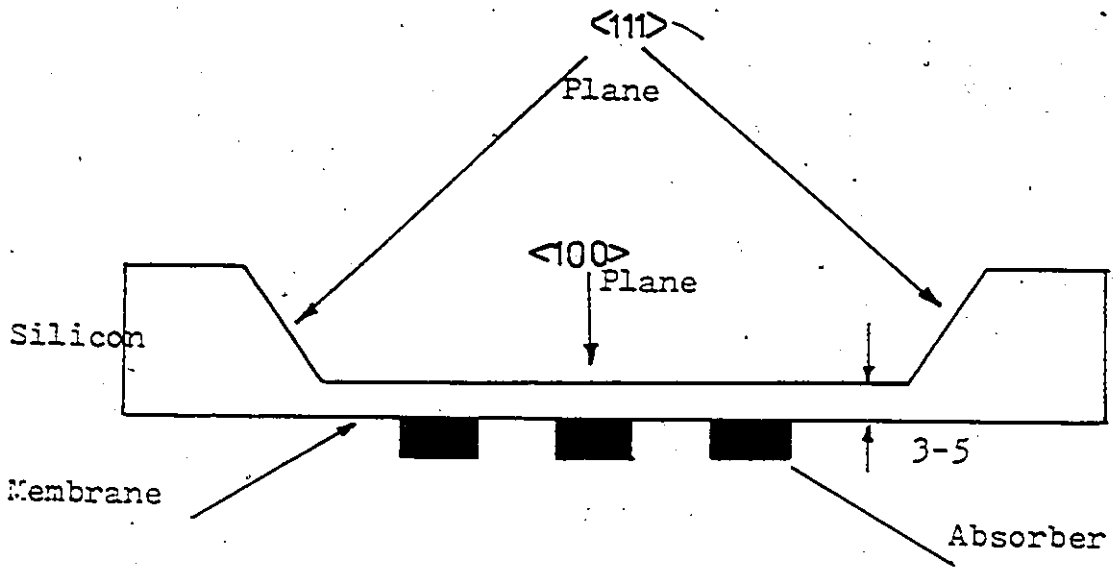


Figure 6.10 Cross Section of X-Ray Lithographic Mask

concentration is higher than 7×10^9 atom/cm^3, the etching does not occur. Therefore, if we choose the right plane of the silicon crystal and diffuse p-type impurities into the crystal to the depth equal to the thickness of the membrane, then we can etch this silicon wafer and make the membrane. The mask illustrated in Figure 6.10 is made with this technique on <100> plane silicon.

The x-ray lithographic technique is capable of producing resolution of about 500 Å, but is not free of diffraction. However, due to small the wavelengths used in x-ray techniques, these diffraction effects have a negligible degradation on resolution. Also, unlike electron lithography, x-ray lithography does not have any back scattering since the x-rays incident on the substrate will pass through the substrate. Therefore, there is no degradation of resolution due to back scattering.

6.5 ETCHING

When we discussed lithography, we mentioned etching in a very casual way. Here for the purpose of completeness, we shall review etching techniques very briefly. Within the last decade the science of etching has taken many steps toward better resolution, controllability, reproducibility, etc. Also, there have been some new additions to the state of the art. We shall start with chemical etching.

6.5.1 Chemical Etching

This form of etching has been and still is the most popular etching technique. This is because it is simple, economical, readily usable in mass production procedures, and one can usually be certain of the adhesion of the metal. The mechanism of chemical etching is a process of dissolution. However, there are not very many metals used in the electronic industry that dissolve in chemicals like sugar melts in water. Therefore, the most usual chemical etches usually use 2 or 3 different chemicals in their mixture. Sometimes the third component is a moderator. In general, this mechanism is one of oxidation and dissolution. One chemical may combine with the material and produce a soluble product (oxidation) and the other chemical will dissolve this product. This two step process continues while etching.

The rate of the chemical etching is a function of temperature, concentration of etchant, agitation of material properties of the substance to be etched and many other parameters which are not completely understood. It is not surprising to see with such a multi-variable process one finds random behaviours and difficulties in controlling the process accurately. Therefore accurate reproducibility is difficult to achieve. However, for most electronic devices which use the above-5 μm resolution the reproducibilities

that can be achieved have been sufficiently good. Below-5 μm circuits, chemical etching becomes a major cause for poor yield.

In general, most of the chemical etches used in photolithography are isotropic etches. By that we mean the rate of etching is independent of the direction. Therefore, we take the vertical etch rates to be the same as the horizontal etch rates. Now let us look at the situation where we want to fabricate a linewidth D. Here we have to take into account both the lateral etch rate as well as the vertical etch rate. In Figure 6.11 we pictorially show what happens. We see that the net result of lateral etch rate is a reduction in linewidth. Supposing the thickness of the metal we want to etch is 2000 Å, then our reduction of line width is 4000 Å. If the linewidth we want is 1 μm, then our linewidth after etching would be 6000 Å, which is a detrimental loss in linewidth resolution.

One major problem that we find in producing SAW transducers using chemical etches is the shorts that appear in the across the fingers. Any dust particle which settles on the IDT prior to etching acts as a mask. If the dust particle settles in between the fingers then the metal under this dust particle would be prevented from etching. The result is a short. In many surface wave devices each transducer may have a few hundred fingers. If there is one

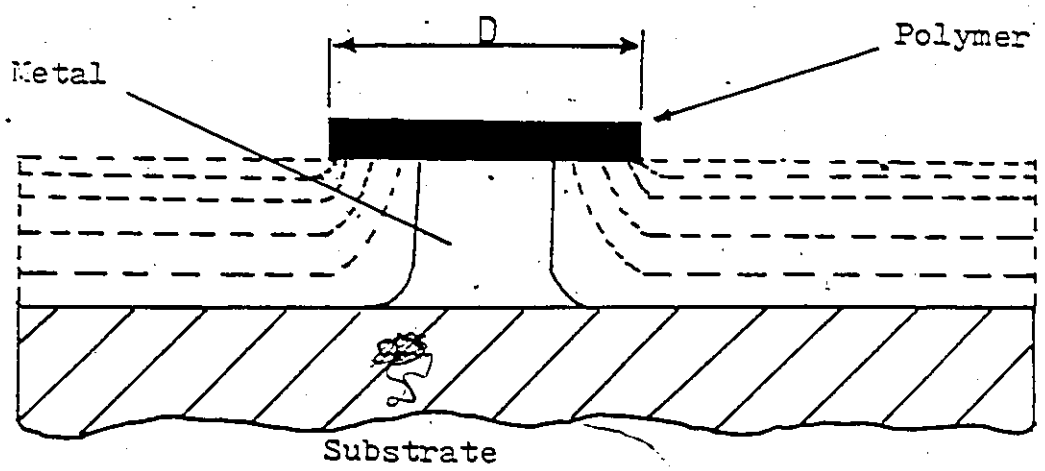


Figure 6.11 Chemical Etching Action

short between the two fingers the whole transducer is shorted and cannot be used to generate or receive a SAW. One advantage of using the lift-off technique is that any dust particle which acts as a mask to the evaporating metal will cause an open circuit instead of a short circuit. The result of an open circuit would be to lose a few fingers from the transducer, but the device is still functional. However, conventional photolithograpy is often used to produce SAW devices operating in the lower MHz region.

6.5.2 Sputter Etching

In sputtering, we place materials to be sputtered in the target position, and substrate in the collector position. In sputter etching, basically, we reverse these. In other words we use the substrate to be etched as a target. The basic mechanism of sputtering and sputter etching is the same. When ions or neutral atoms hit the surface of the target with sufficiently high velocity due to the change in momentum, atoms can be ejected from the surface. Now if the surface is the substrate surface then materials from the surface of the substrate can be removed. Therefore, if we have a pattern replicated in a polymer on the surface of the metal substrate, both polymer material and the metal will be removed. However, etch rates for these materials will be different. The etch rate (or

ejection rate) depends on the composition of the materials, the mass and the energy of the neutral ion and the angle of incidence. With the knowledge of the etch rate of the materials and proper control of the thicknesses of the materials, it is possible to etch patterns.

Etch rate of sputtering is a relatively well behaved function of energy. Etching threshold begins at about 10 ev and increases linearly up to about 500 ev. Above this level the etch rate tends to saturate. These etch rates are reproducible and undercutting can be avoided. By arranging the ions to strike at normal incidence to the substrate one can etch vertical walls. Therefore, this method of etching is ideally suited for submicron lithograpy.

The major disadvantage of the process is that there is no preferentiality in etching. Everything on the surface etches. Therefore, one needs to carefully control etch rates, and thickness of the metal and the polymers. Otherwise, one is faced with the danger of etching the substrate. However, this can be avoided by careful monitoring of the process of etching using a mass spectrometer. Ion beam sputter machines are commercially available. Adopting these machines for mass production is possible although they are not available at the present time. One surface wave device that uses ion beam technology is the REC (Reflective Array Compressor) filters [63]. The

grooves of these filters are etched using an ion beam source and also they are depth weighted very successfully.

Amplitude weighting has been discussed in Chapter 4 and in Chapter 6 we mentioned that we can vary the apodization function experimentally. We have developed a simple technique to modify the aperture of an IDT which can be used to determine the weighting function of a filter. We shall introduce this method next.

6.6 TECHNIQUE FOR MODIFYING THE APERTURE OF AN IDT [49]

In changing the weighting function of the aperture of an IDT, one usually has to go through the process of generating new masks starting from the initial stage of producing the art work and the subsequent reduction. Usually the reduction involves a double reduction. This process is very expensive and time consuming. Generally one may have to go through 5 or 6 alterations before an acceptable performance from the device can be achieved. This would mean 5 or 6 masks. The technique we are about to describe simplifies this process. Essentially, it involves producing a mask for the new apodization function and modifying the master pattern. Producing the mask for the new apodization function is much simpler than producing a completely modified device mask. This is because the new apodization function mask is a less complicated large area

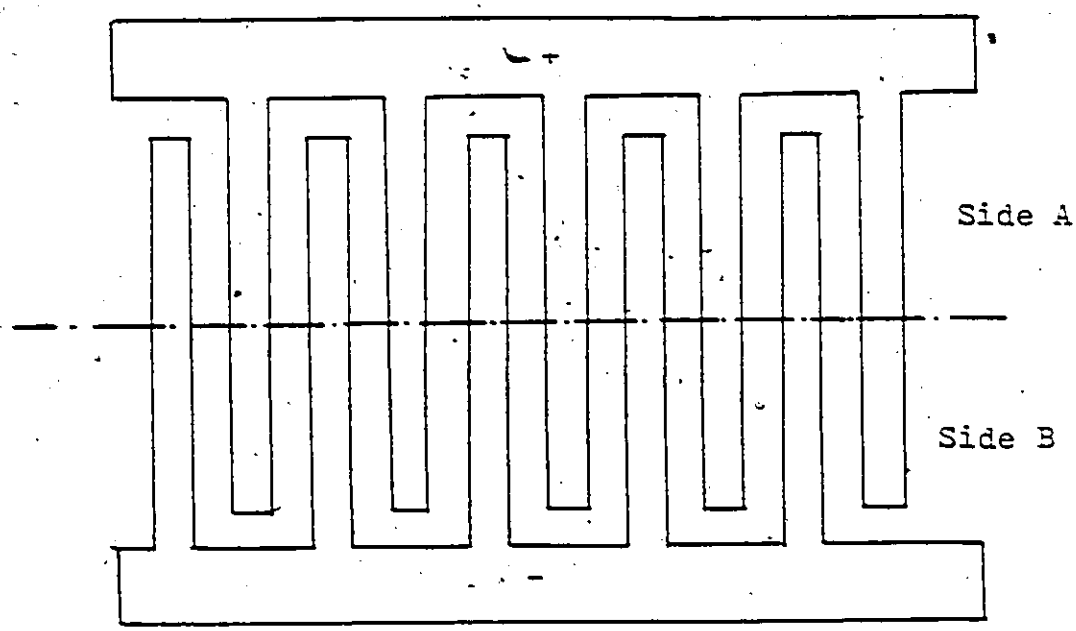
mask in comparison to the mask of the IDT. In fact modifying the mask may be made with a single reduction.

To illustrate the process let us consider a simple IDT pattern as shown in Figure 6.12(a). This is the pattern we find in our master mask which is produced on a glass plate as a positive photographic mask. We wish to modify this uniform aperture to a tapered aperture as shown in Figure 6.12(b). For illustrative purposes we designate our IDT fingers as positive and negative and the pattern is said to be symmetrical about the horizontal centre line which divides the pattern into two sides A and B. Let us select two adjacent fingers and show how one can remove one selectively. The step by step process is depicted in Figure 6.13.

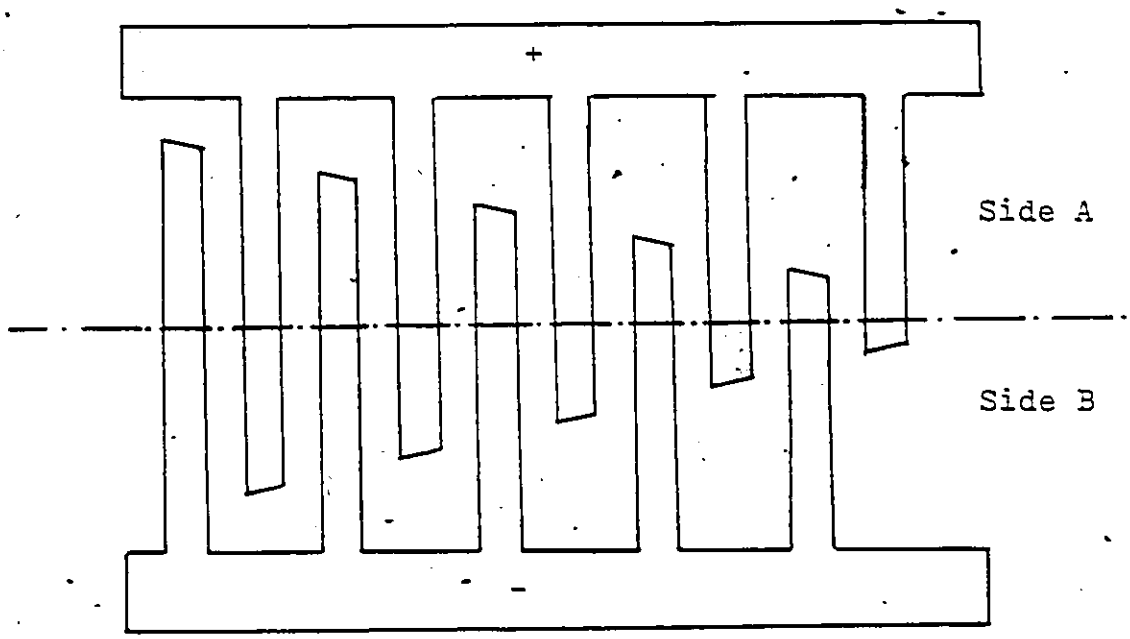
Step 1: The first step is to deposit four layers of chromium and gold. As illustrated in Figure 6.13(a). The first layer of chromium is 1000 Å - 1500 Å thick and it is deposited on the substrate.

Step 2: The second step is to fabricate the master IDT as shown in Figure 6.12(a) using standard photolithograpy and chemical etches. Figure 6.13(b) illustrates the cross section of the two adjacent fingers after completion of these two steps.

Step 3: In the third step we use a modifying mask to expose those areas that need to be modified, and remove



(a)



(b)

Figure 6.12 An Illustrative IDT Pattern
(a) Before Modification
(b) After Modification

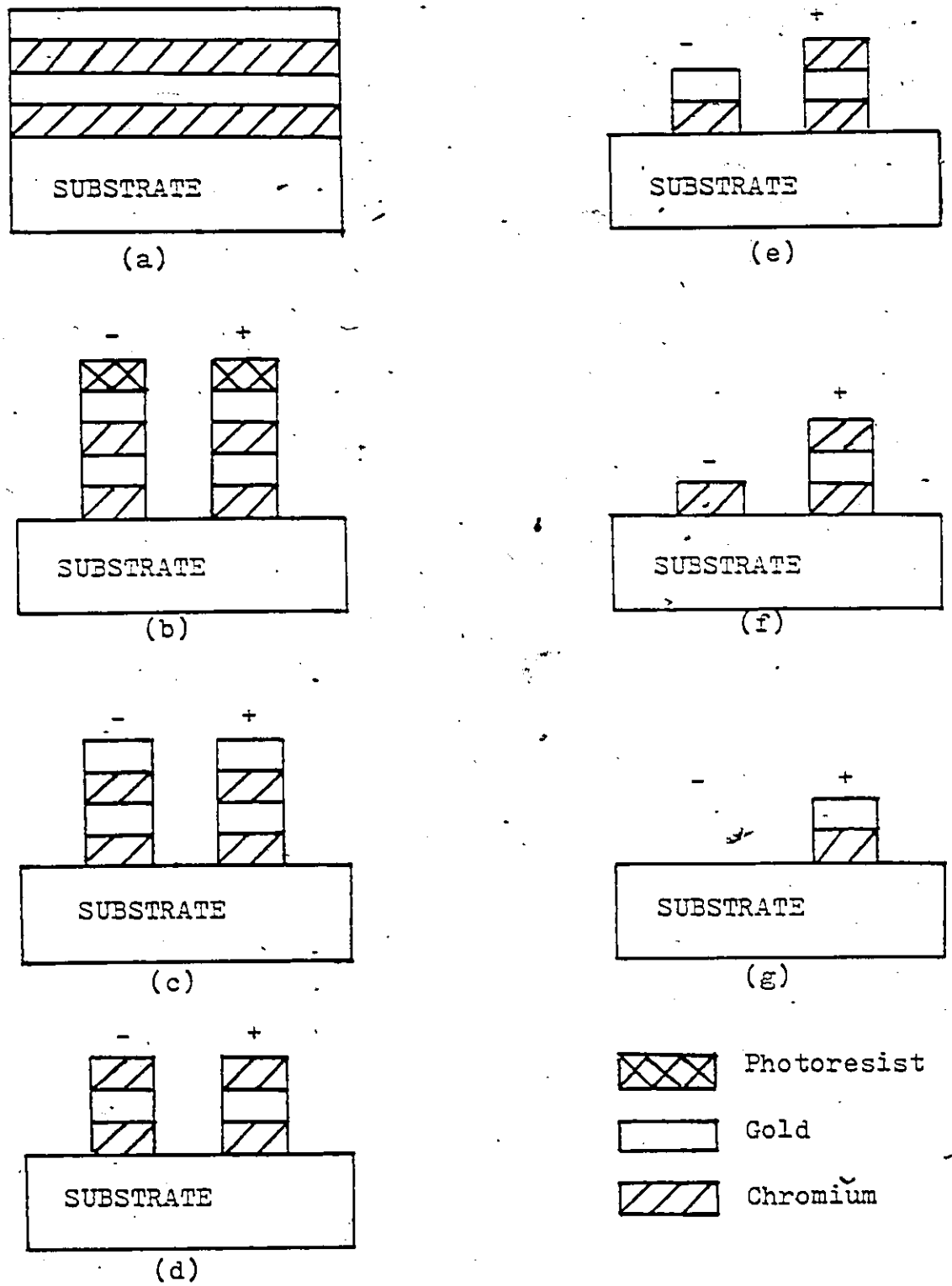


Figure 6.13 Aperture Modifying Technique

the photoresist to expose the metals for subsequent etching. As shown in Figure 6.13(c) we now have four identical metal layers in each finger, and we want to etch a portion of one finger without deleting the other. We designate the finger to be deleted as the negative finger and the other as the positive finger.

Step 4: At the fourth step we will remove the top layer of gold from both fingers leaving 3 layers of metal. Meanwhile the still-protected area has the original 4 metal layers surrounded by a layer of unexposed photoresist. In Figure 6.13(d) we show the 3 metal layers of the two adjacent fingers, and note that the top layer of metal is chromium.

Step 5: Our fifth step is to etch the chromium of the negative finger without etching the chromium of the positive finger. We do this by preferentially etching the negative finger using an electrolytic etch. The electrochemical etching bath contains a mixture of phosphoric acid and distilled water. The positive terminal of the electrolytic etching bath's power supply is connected to the pad of the IDT that contains the finger to be etched, while the negative terminal is connected to the electrode in the etching bath. In this way the upper layer

of chromium is etched away from the negative finger. The positive finger remains unetched because there is no electrical connection to it. In Figure 6.13(e) we illustrate the resultant structure after Step 5. There are two layers of metal on the negative finger, namely gold and chromium, while there are 3 layers of metal, namely chromium, gold and chromium respectively, on the positive finger. Using the same procedure we trim the necessary fingers on the opposite side.

Step 6: Now we may take off the photoresist from the protected area which contains 4 metal layers of which the top layer is gold. We will now etch the gold layer leaving a single chromium layer in the finger portions to be deleted while the rest of the transducer has 3 metal layers namely chromium, gold and chromium. The present status of the two fingers we considered so far is depicted in Figure 6.13(f).

Step 7: Finally, we etch the chromium away. Now we have deleted the only remaining chromium layer from the parts of the fingers that need to be trimmed, while the rest of the transducer has two layers of metal and chromium. The status of the two fingers we started with is shown in Figure 6.13(g) indicating

the deletion of the negative finger.

6.6.1 Experimental Verification of the Technique

To demonstrate the technique, we used an IDT structure with a uniform aperture. The value of the ratio of the finger width to spacing is made equal and the width d of each finger was set at 14.5 μm. For this example we wanted to reduce the aperture by 76 d. Therefore, we had to take out 38 d from each set of fingers. The modifying mask used to achieve this was made out of a sheet of rubylith material. The photoresist used was Shipley AZ1350 positive resist. Figure 6.14(a) illustrates a section of the IDT on four layers of chromium and gold, after exposing the area to be modified. The boundary between the area of the aperture to be modified and the area not to be removed is clearly defined by a vertical line on the right-hand side of the pattern. In Figure 6.14(b) we illustrate the same transducer after Step 5 which is the selective etching of fingers. The fingers with brighter lines are the fingers to be etched, and they are bright because of the exposed gold layer on them, while the others still have the chromium layer. Finally, in Figure 6.14(c) we illustrate one side of the transducer after completion of Step 7 which is the completion of the modification process.

In principle, we could use 3 layers omitting the top

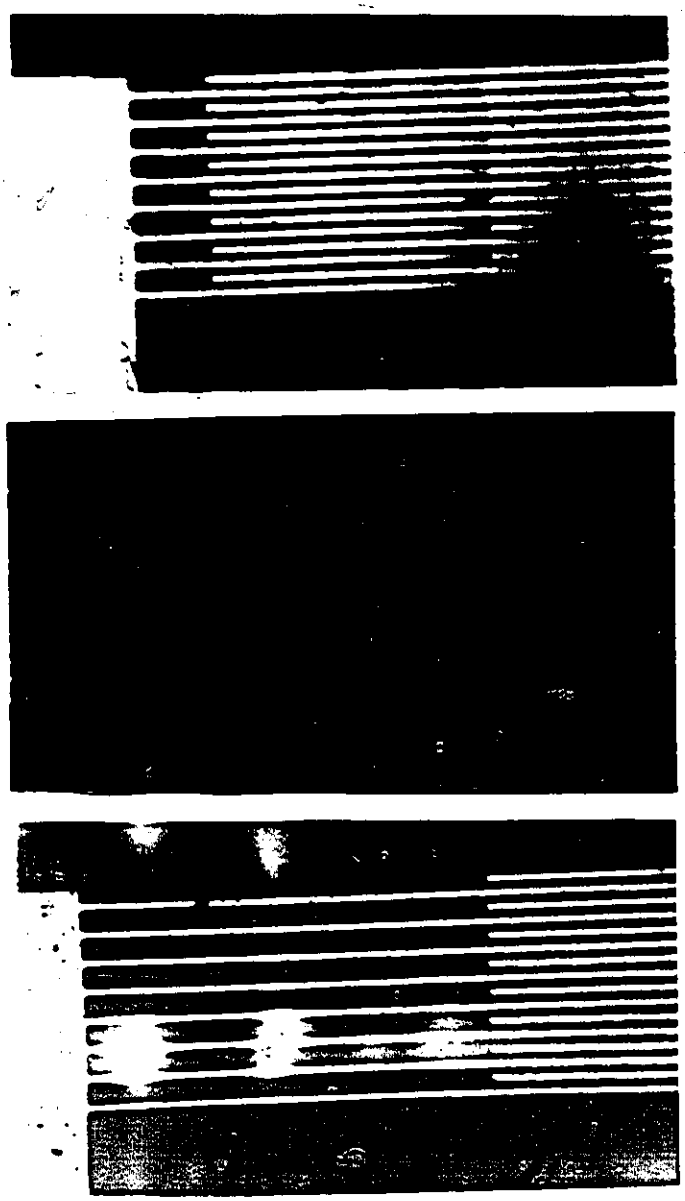


Figure 6.14

(a) A section of the IDT in 4 layers of chromium and gold after exposing the area to be modified.

(b) Same IDT after selective etching of chromium

(c) One side of the transducer after completion of the modification.

gold layer. However, we found that the interface properties between chromium and the photoresist to be poor, resulting in a high density of pin-holes. Therefore, we used the top gold layer. In addition, it acts as an additional mask during the process of preferential etching.

In principle, one might think the process could be cut down by using two layers. However, in practice we found it cannot be done with two layers, because during the electrolytic etching process the etch rate at the boundary of the photoresist is higher than elsewhere on the fingers. The result is to form an open circuit at the boundary long before the complete finger is etched away, and thereby leaving a part of unetched finger. Having a second gold layer under the finger insures the electrical continuity along the finger.

With this simple and economical method, it is possible to cut down the cost of design and development of SAW filters and other devices where apodizations are required. In this method we explained how to use this technique with conventional photolithography. However, the same method can be adapted to the lift-off technique as well. Therefore, one may also be able to modify submicron linewidth structures.

CHAPTER 7

EXPERIMENTAL RESULTS

In Chapter 6 we reviewed SAW device fabrication techniques briefly. In doing so we covered a range of technologies which can be used to produce devices which operate in the frequency range from 10 MHz - 1 GHz. Furthermore, we introduced a simple technique that can be used to modify apodization functions economically and effectively. We also demonstrated experimentally this thin film technique, and have shown that it offers the advantage of developing apodization functions experimentally. In this Chapter we will deal with the experimental verification of the new FM discriminator principle. It must be emphasized that the major objective of this work is to prove a new principle rather than producing a fully developed optimum discriminator. In fact, we could claim with some assurance that the performance of this discriminator reflects the worst case conditions, rather than the optimum capabilities. In evaluating the performance of the discriminator, we will investigate two main aspects. They are:

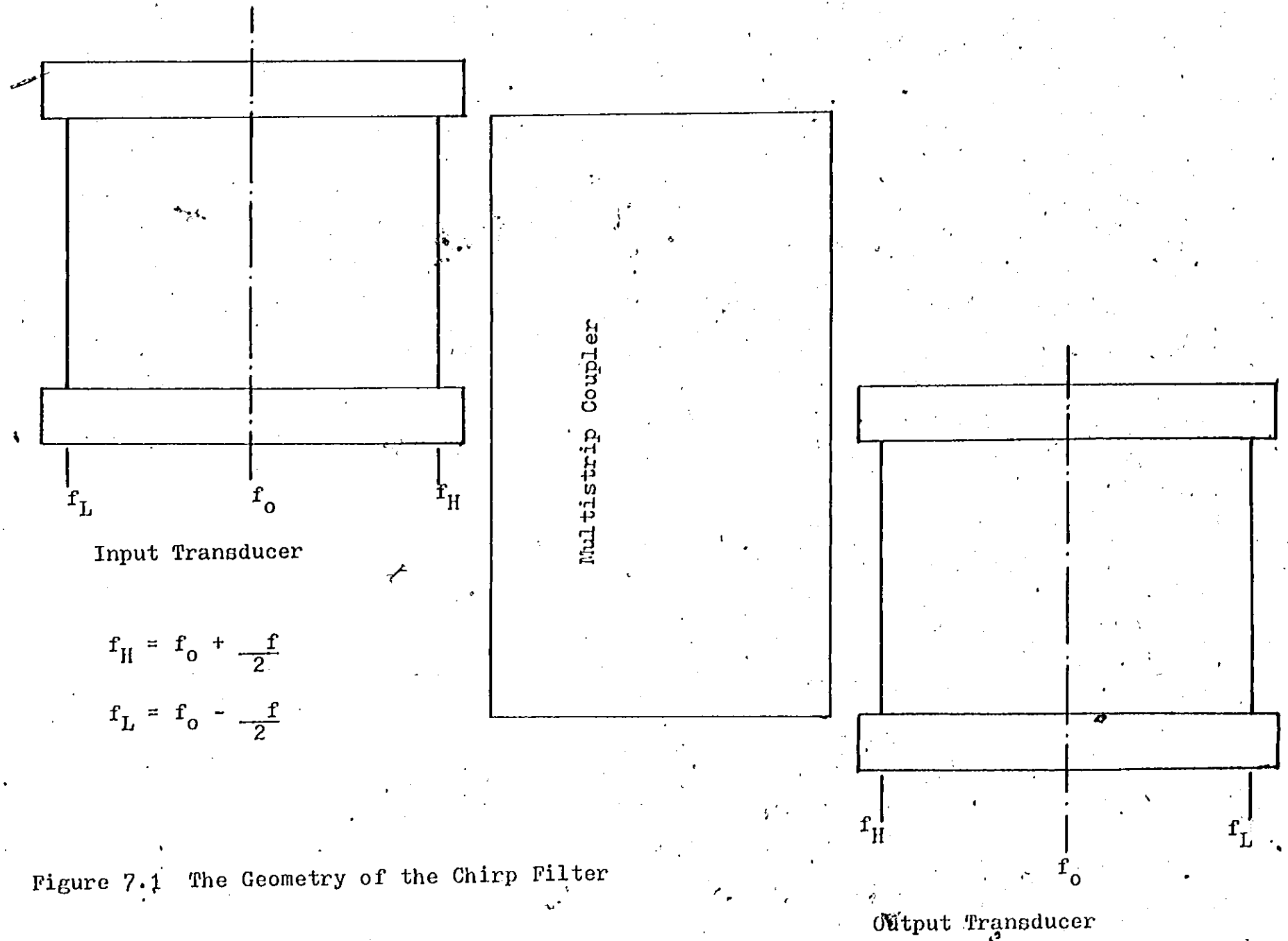
1. Signal performance of the discriminator
2. Noise performance of the discriminator

7.1 SIGNAL PERFORMANCE OF THE DISCRIMINATOR

In considering the signal performance of the discriminator we will initially look at the performance characteristic of the main signal processing device in the discriminator, and then later check the functional behaviour of the other electronic circuits.

7.1.1 Chirp Filter Performance Characteristics

The chirp filter was designed to operate at 70 MHz center frequency with ±10 MHz bandwidth on a yz cut LiN_bO_3 substrate. The time dispersion length of the transducer was to be 1.02 μs giving a time-bandwidth (TB) product of TB≈20. The inital geometry of the transducer configuration used was the in-line doubly-dispersive type with an apodization function as described in Chapter 5. We were forced to abandon this geometry due to the poor frequency response. The usable bandwidth of this device was only 4 MHz and the response was badly distorted due to second-order effects such as bulk wave generation, triple-transit echo and diffraction. To overcome these effects we modified the geometry by incorporating a track changing multistrip coupler. This geometry is shown in Figure 7.1.



$$f_H = f_o + \frac{f}{2}$$

$$f_L = f_o - \frac{f}{2}$$

Figure 7.1 The Geometry of the Chirp Filter

7.1.2 Frequency Response of the Chirp Filter

The incorporation of the multistrip coupler did show an enormous improvement in the frequency response of the device. Perhaps the most significant contribution of the multistrip coupler may have been the bulk wave suppression (or more correctly avoidance). The frequency response measurements were made on an HP 8407A network analyzer system. In Figure 7.2 we have illustrated the frequency response of the chirp filter. The 0 dB reference line is indicated by the top trace. The amplitude scale of this Figure is 10 dB per division. The maximum insertion loss in the absence of any matching is 37 dB and the maximum peak ripple is ≈6 dB and the average peak to peak ripple is ≈4 dB. Frequency values are shown on the horizontal axis to indicate the usable frequency range in the discriminator.

Compared to the state of art of chirp filters [38], this filter is extremely poor. We decided to use this device, however, because of the fact that this would be a good test vehicle to demonstrate the capability of the zero crossing detector techniques. In addition to this, it would verify the discriminator principle under the worst case situation.

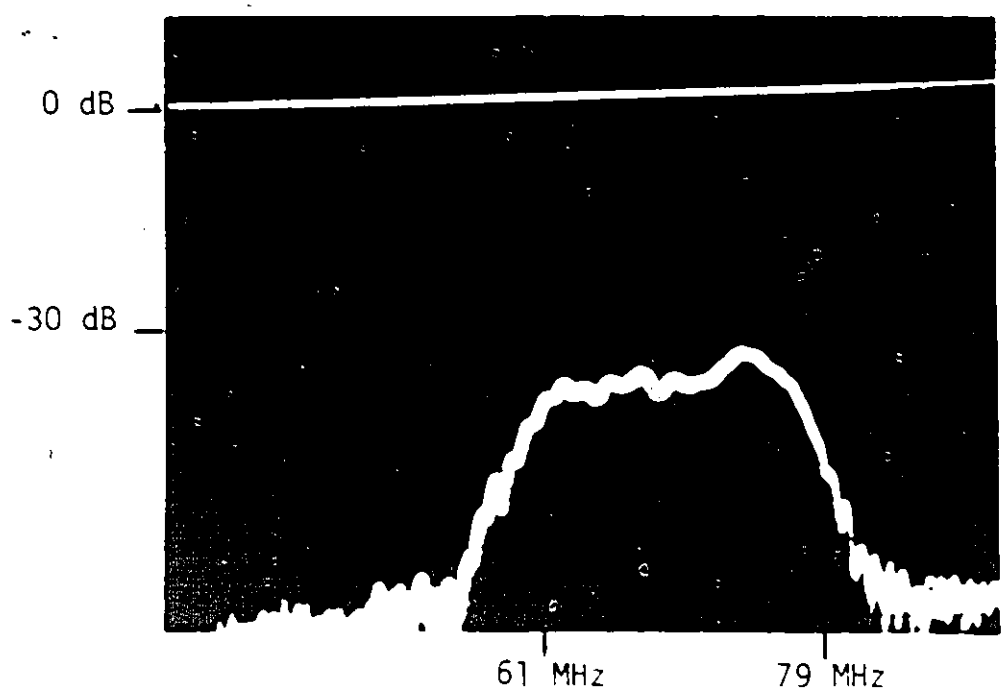


Figure 7.2 Frequency response of the chirp filter.

7.1.3 Time Responses of the Chirp Filter

Since we are to use the frequency vs. time delay characteristics of the SAW chirp filter as a measure of time scaling, we should verify the linearity of this characteristic. The arrangement of instruments for time delay measurements is shown in Figure 7.3. The output signal from the oscillator was pulse-modulated with the aid of a repetitive voltage pulse from the pulse generator and a mixer. This signal was then fed to the device and the output was amplified to overcome the insertion loss and displayed on the oscilloscope. Also, the input pulse modulated signal to the SAW chirp filter was tapped at the output of the mixer and displayed on the scope. The external trigger to the oscilloscope was supplied from the output of the pulse generator.

The displayed wave forms on the oscilloscope are shown in Figure 7.4(a). The upper trace illustrates the input pulse modulated input signal to the SAW chirp filter. The lower trace illustrates the delayed output correlated pulse from the device. The time delay between the input pulse and the output pulse must vary with frequency. This phenomenon is illustrated in Figure 7.4(b). The upper trace is the modulated output pulse whose frequency can be varied while the waveforms displayed at the centre and at the bottom of the photograph indicate clearly the time delay

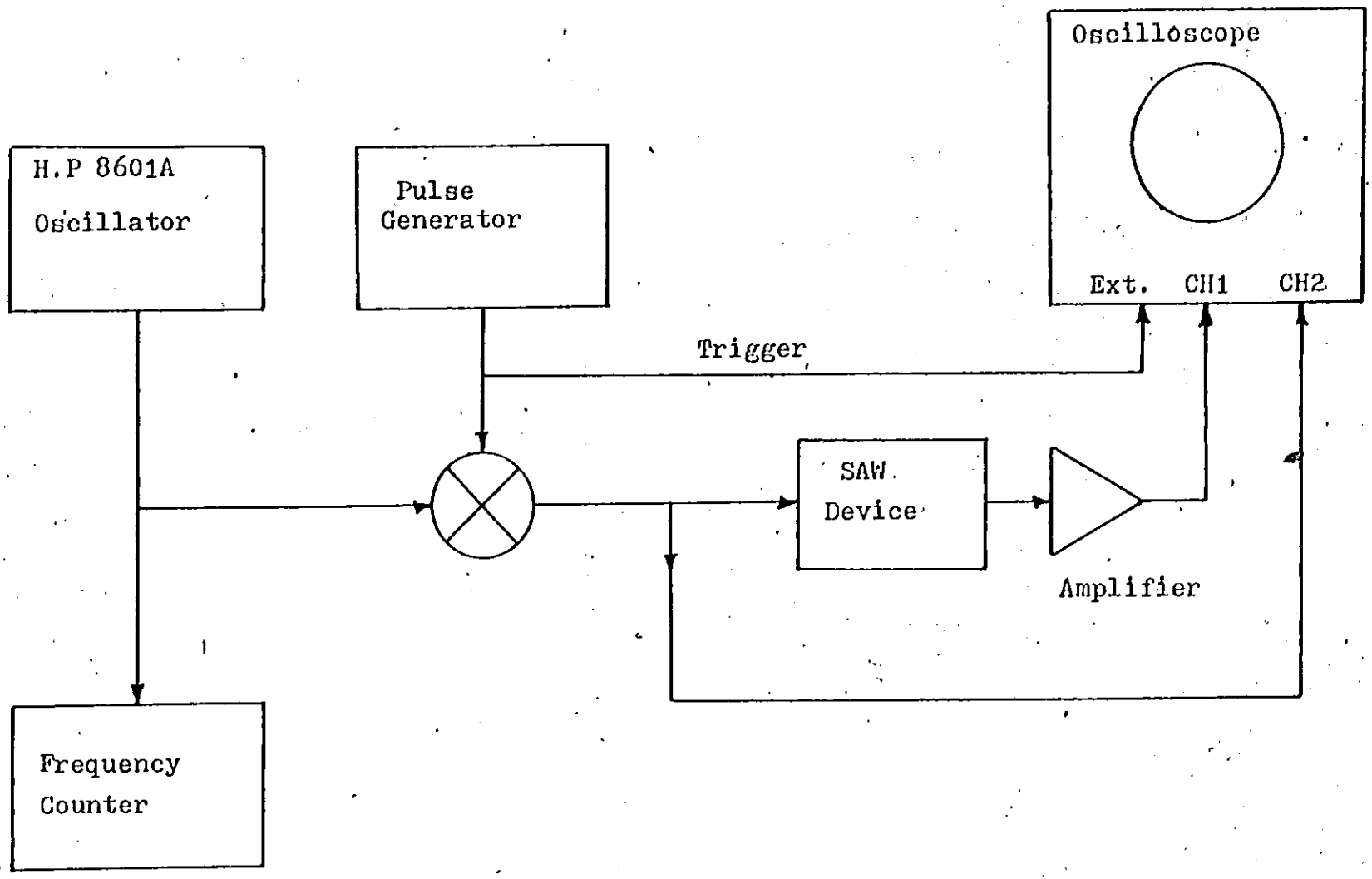


Figure 7.3 Arrangement of Instruments for Frequency vs Time Delay Measurement

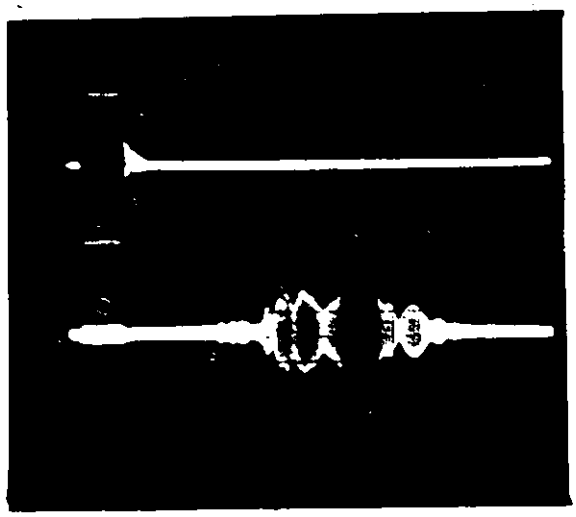


Figure 7.4(a) Upper trace: Sampled input signal at the SAW chirp filter.
Lower trace: Delayed output pulse from the SAW chirp filter.
Horizontal Scale: 0.5 μsec/div.

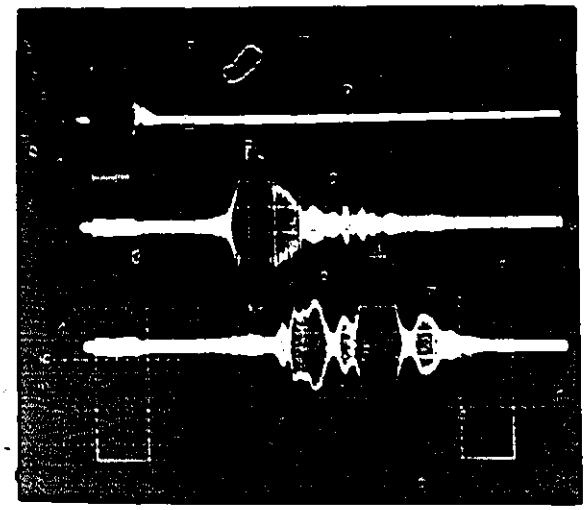


Figure 7.4(b) Upper trace: Sampled input signal at the SAW chirp filter.
Lower trace: Delayed output pulses for two frequencies
Horizontal scale: 0.5 μsec/div.

variation with frequency. The pulse width of the modulating or sampling pulse out of the pulse generator was set to 0.4 microsec. This pulse width was chosen experimentally to be the optimum pulse width so that the output of the SAW device produces a good "sinc" response throughout the bandwidth of the device. The frequency vs. time delay measurement characteristics are plotted on a graph and shown in Figure 7.5. As can be seen and expected, these characteristics are linear.

7.2 VERIFICATION OF SUBSYSTEMS' FUNCTION

Our next experimental objective is to verify the functions of various subsystems or components of the discriminator. We wish to carry out this task by monitoring waveforms at various test points. In Figure 7.6 we have illustrated the system diagram of the discriminator at various test points. The test instruments arrangement is similar to the set up used for time delay measurement, and the only difference is that the pulse generator and the mixer are now a part of the system. For the initial investigation, the input frequency was set to a frequency anywhere within the bandwidth of the device.

Starting at Test Point A we should see the sampled signal and at Point B we should see the correlated delayed output pulse from the SAW device. This delay should vary

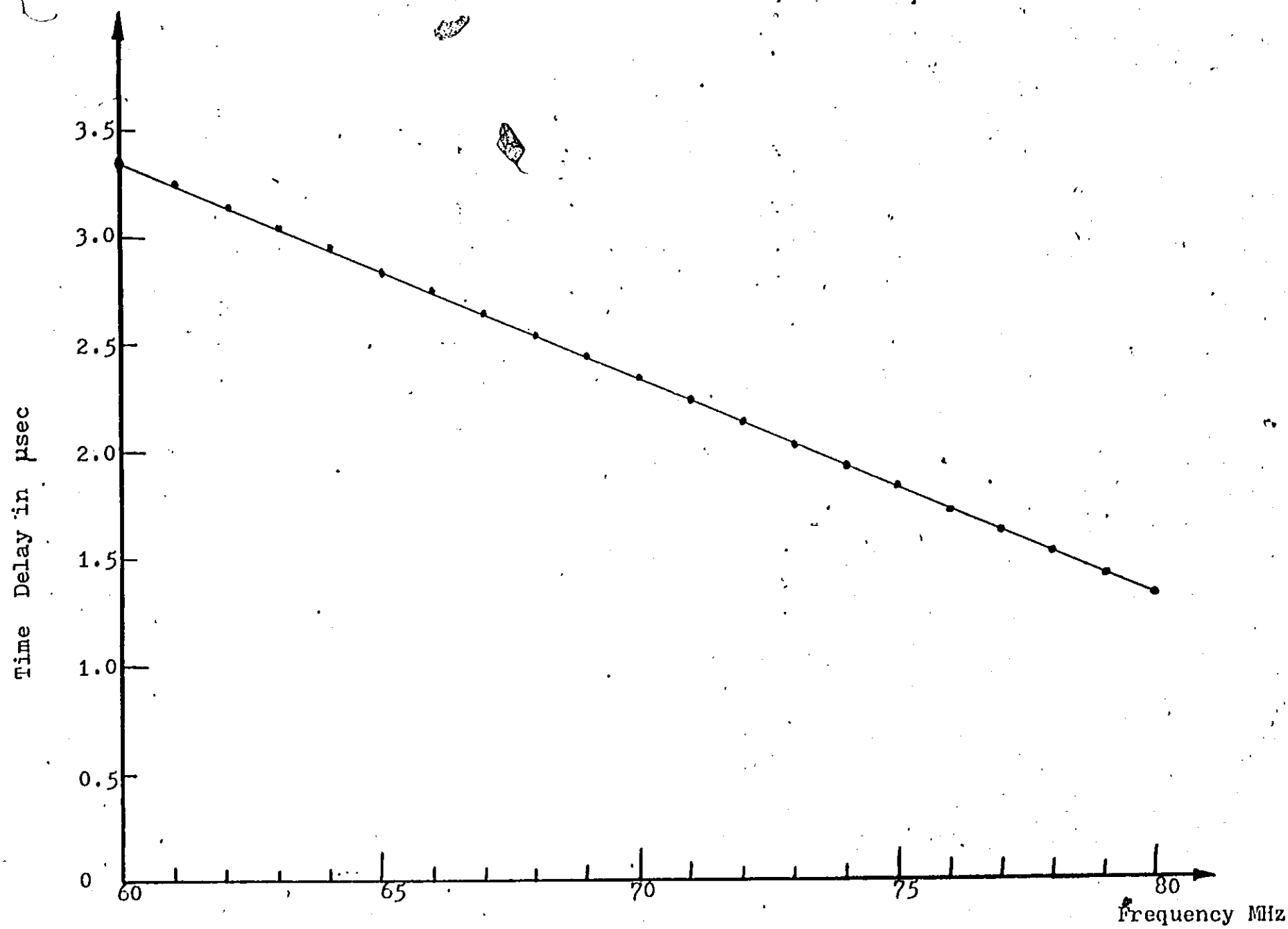


Figure 7.5 Frequency vs Time Delay Characteristics of the Chirp Filter

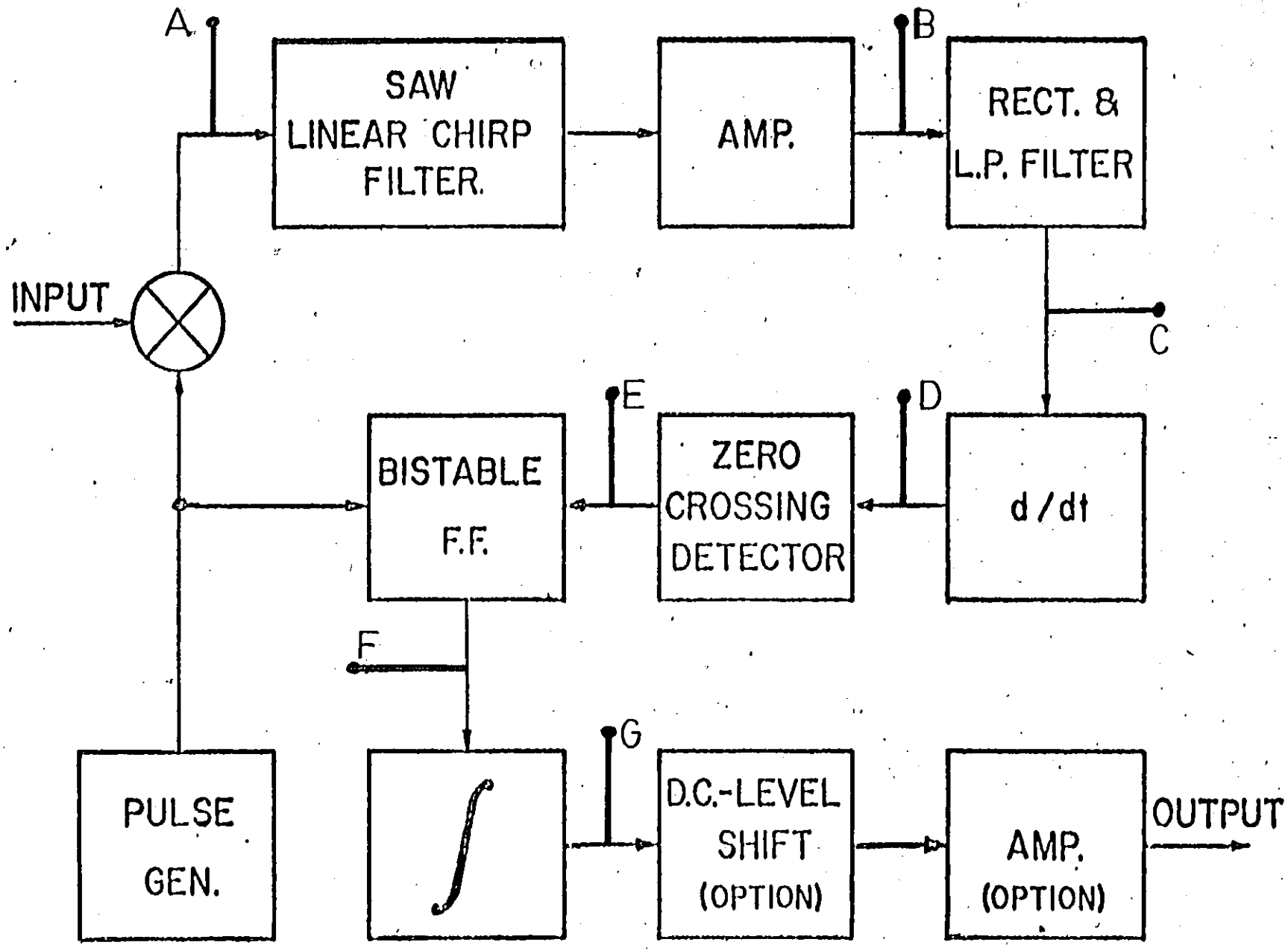


Figure 7.6 System Diagram with Test Points

with frequency. These functions were observed in the time response measurements and need not be carried out again. At Test Point C we should see rectified and lowpass filtered signals, and at point D we should see differentiated signals. In Figure 7.7(a) we have illustrated these observations. The upper trace indicates the output from the rectifier and lowpass filter circuit while the lower trace indicates the output waveform from the differentiator as monitored at Test Point D. This photograph was taken when the sampling frequency was 400 KHz to observe and to check any changes in the circuit responses when sampled at a higher frequency. As can be seen here, the zero-crossing point of the differentiator output signal corresponds to the time that peak of the envelope occurs. Any circuit delay present in the differentiator due to the time constant must be small and negligible. In Figure 7.7(b) we have illustrated the circuit delay and measured it to be nearly 50 ns. This delay was found to be constant throughout the bandwidth of the device.

Next we wanted to show the behaviour of the zero-crossing detector. In Figure 7.8(a), the upper trace indicates the output of the zero-crossing detector as observed at Test Point E, while the lower trace indicates the differentiated pulse. In Figure 7.8(b) the circuit time delay in the zero-crossing detector with an enlarged version

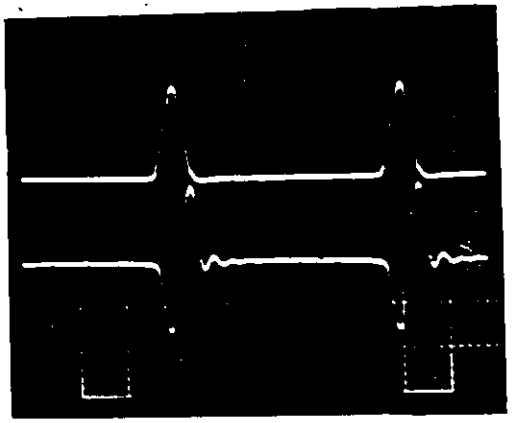


Figure 7.7(a) Upper trace: Output from the rectifier and lowpass filter circuit.
Lower trace: Output from the differentiator circuit
Horizontal scale: 0.5 μsec/div.

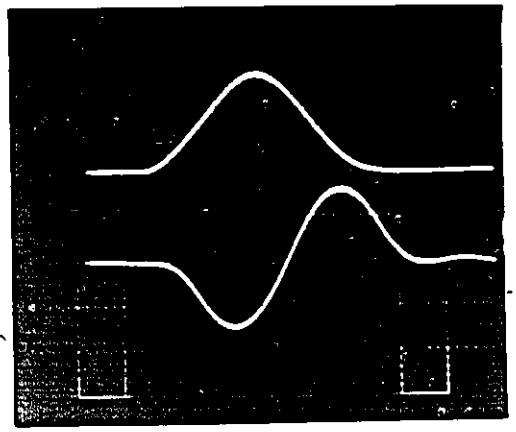


Figure 7.7(b) Enlarged version of 7.7(a) illustrating the circuit delay in the differentiator circuit
Horizontal Scale: 0.1 μsec/div.

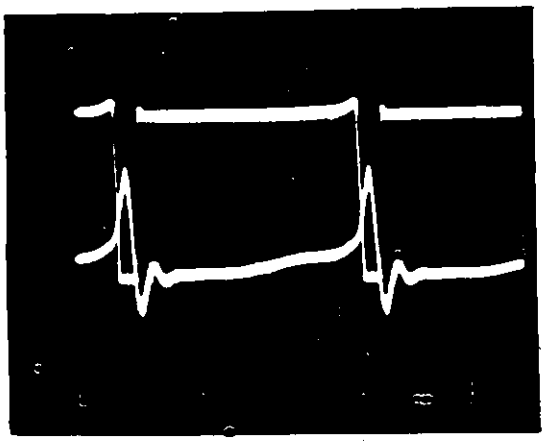


Figure 7.8(a) Upper trace: output from zero crossing detector circuit.
Lower trace: output from differentiator circuit.
Horizontal scale: 0.5 μsec/div.

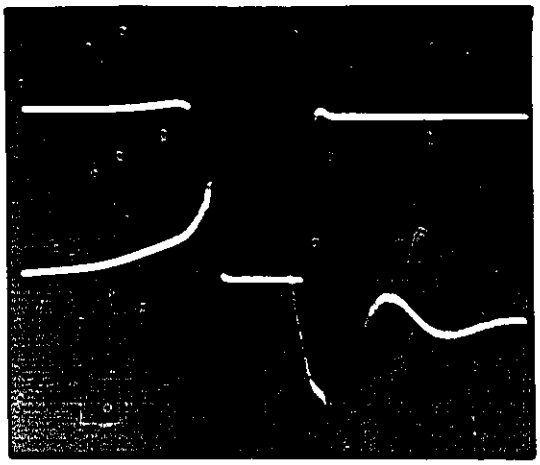


Figure 7.8(b) Enlarged version of 7.8(a) illustrating the zero crossing detector circuit delay
Horizontal Scale: 100 nsec/div.

of the Figure 7.8(a). The delay was measured to be 20 ns, and it is constant throughout the bandwidth of the device. This means the total circuit delay of the detector circuit is ≈70 ns, which was constant throughout the bandwidth of the device. This is a negligible delay and being constant throughout the bandwidth of the device indicates it will have hardly any effect on the linearity of the circuit. The fact that it operates at 400 kHz sampling frequency indicates this particular detector circuit can be used for discriminators which are capable of demodulating 200 KHz baseband signals. However, as we pointed out earlier the sampling frequency is determined by the maximum delay of the SAW chirp filter. In this particular chirp filter it happens to be 3.4 μsec. The corresponding sampling frequency is 294 kHz which means the maximum baseband signal it can demodulate is 147 kHz. Since the detector circuit works at 400 kHz, we can say with assurance that this circuit is more than adequate for this discriminator.

A zero-crossing-detector technique was designed to eliminate the errors that were caused by the variation of amplitude of the output pulse from the SAW chirp filter correlator. To demonstrate that this objective has been achieved, we present, in Figure 7.9, the output rectangular pulse from the bistable flipflop along with the output of the rectifier and lowpass filter circuit for 3 different

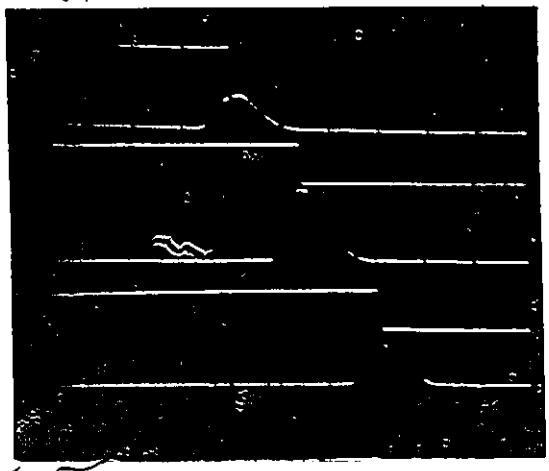


Figure 7.9 outputs from the bistable flipflop and the outputs from the rectifier and lowpass filter circuit for 3 different input frequencies. Horizontal Scale 0.5 μsec/div.

frequencies. As can be seen clearly, the bistable flipflop is reset to a low level right at the time in which the peak of the envelope occurs, irrespective of what the amplitude of the envelope is. This indeed is what we wanted out of this circuit technique.

In the final circuit function investigation, we want to show the behaviour of the baseband signal assignment circuit. The signal assignment circuit consists of the bistable flipflop and an integrator. The purpose of the bistable flipflop is to produce a constant amplitude rectangular pulse whose pulse width is proportional to the time delay of the SAW device. In fact, Figure 7.9 does demonstrate this behaviour. Effectively what we are doing here is converting the FM signal into a pulse-duration modulated (PDM) signal. To stress this point further, we varied the input signal and monitored the output of the bistable flipflop. In Figure 7.10(a) we illustrate the PDM signal as seen at the Test Point F. Pulse width modulation is illustrated in the right hand side of the picture. By integrating this pulse we can convert the pulse width variation to a voltage amplitude variation. In Figure 7.10(b) we have illustrated this point, at the right hand side of the picture which indicates the variation of the voltage amplitude with pulse width variation. The integrator used in this circuit is a bootstrap type

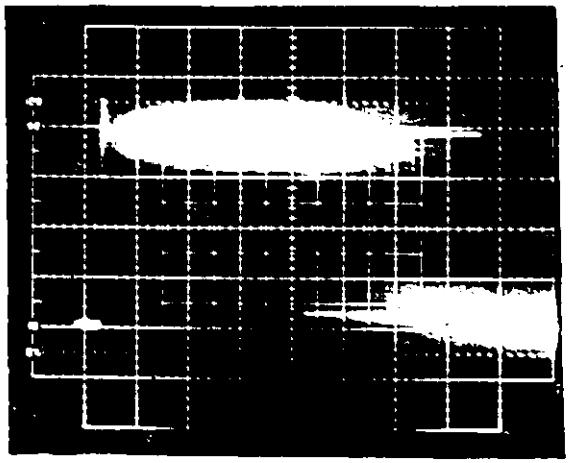


Figure 7.10(a) pulse width modulated signal at the output of the bistable flipflop.

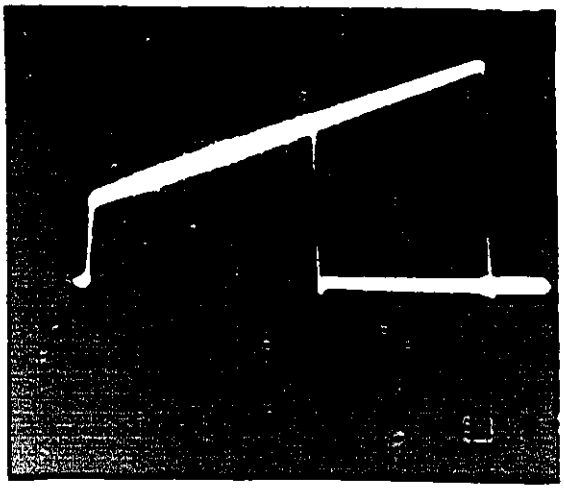


Figure 7.10(b) Output of the integrator indicating the variation of the voltage corresponding to the pulse width modulated signal.

integrator for which the circuit charging current could be adjusted over 20:1 range. This completes the waveform monitoring excercise. Now we have shown and verified the discriminator principle by verifying the systems component functions.

7.3 LINEARITY OF THE DISCRIMINATOR

Corresponding to the separation between the two transducers, the output of the bistable displays a constant pulse width which does not vary with frequency. In fact, to be more precise, constant pulse width is due to the constant time delay due to the separation between the two transducers and any circuit delays. This constant pulse width may be observed from the right hand side of the Figure 7.10(a) which displays the output of the bistable flipflop. Corresponding to this pulse there is a constant voltage at the output of the integrator. To measure the linearity of the discriminator we used a DC level shifter and an amplifier after the integrator in such a way that at 60 MHz the output of the whole system is zero. This DC level shifter and the amplifier are shown in the system diagram as options (see Figure 7.6). The measured frequency vs. voltage conversion characteristics are shown in Figure 7.11. In Figure 7.12 we have illustrated the measured non-linearity error of the conversion characteristics vs.

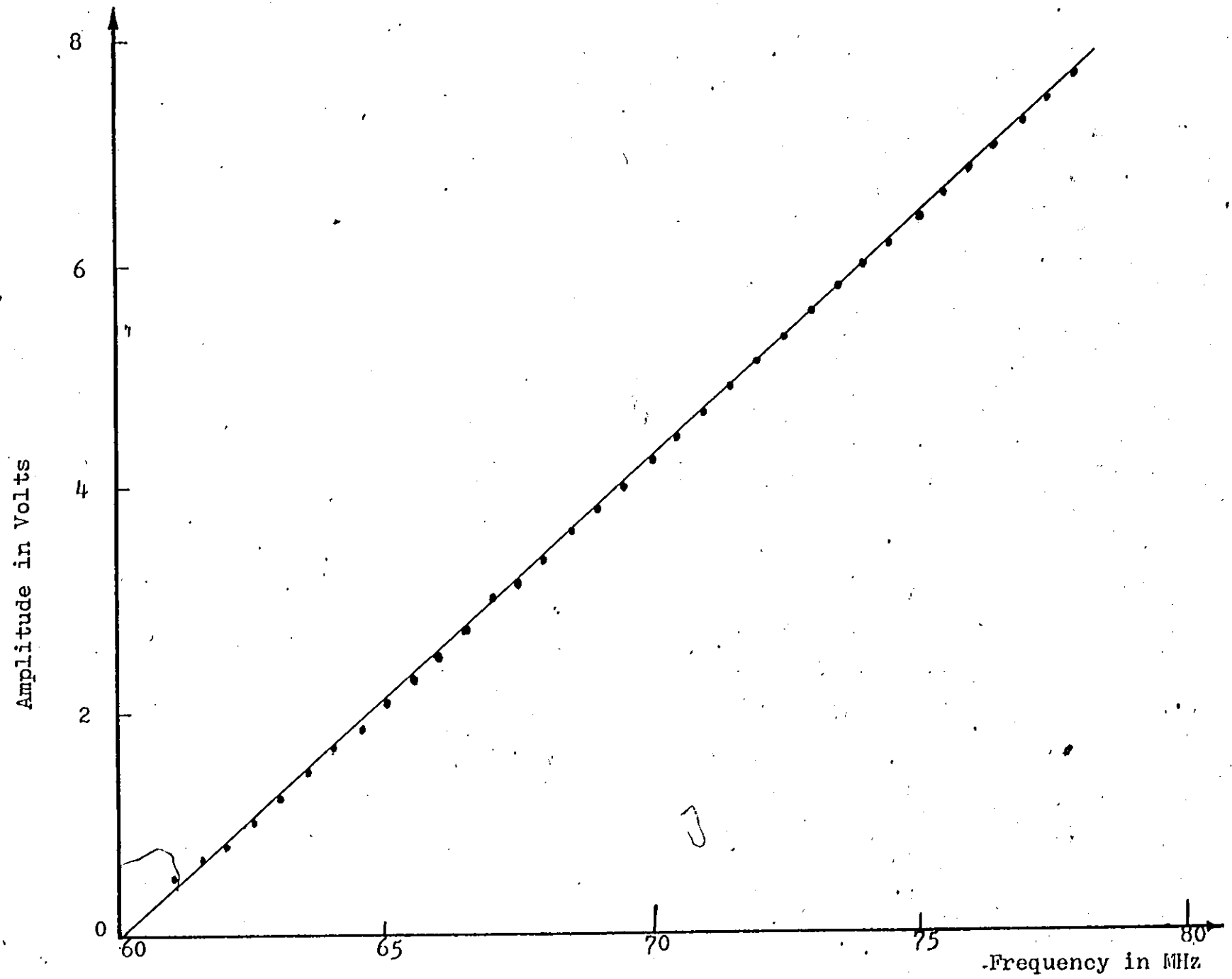


Figure 7.11 Frequency vs Voltage Conversion Characteristics

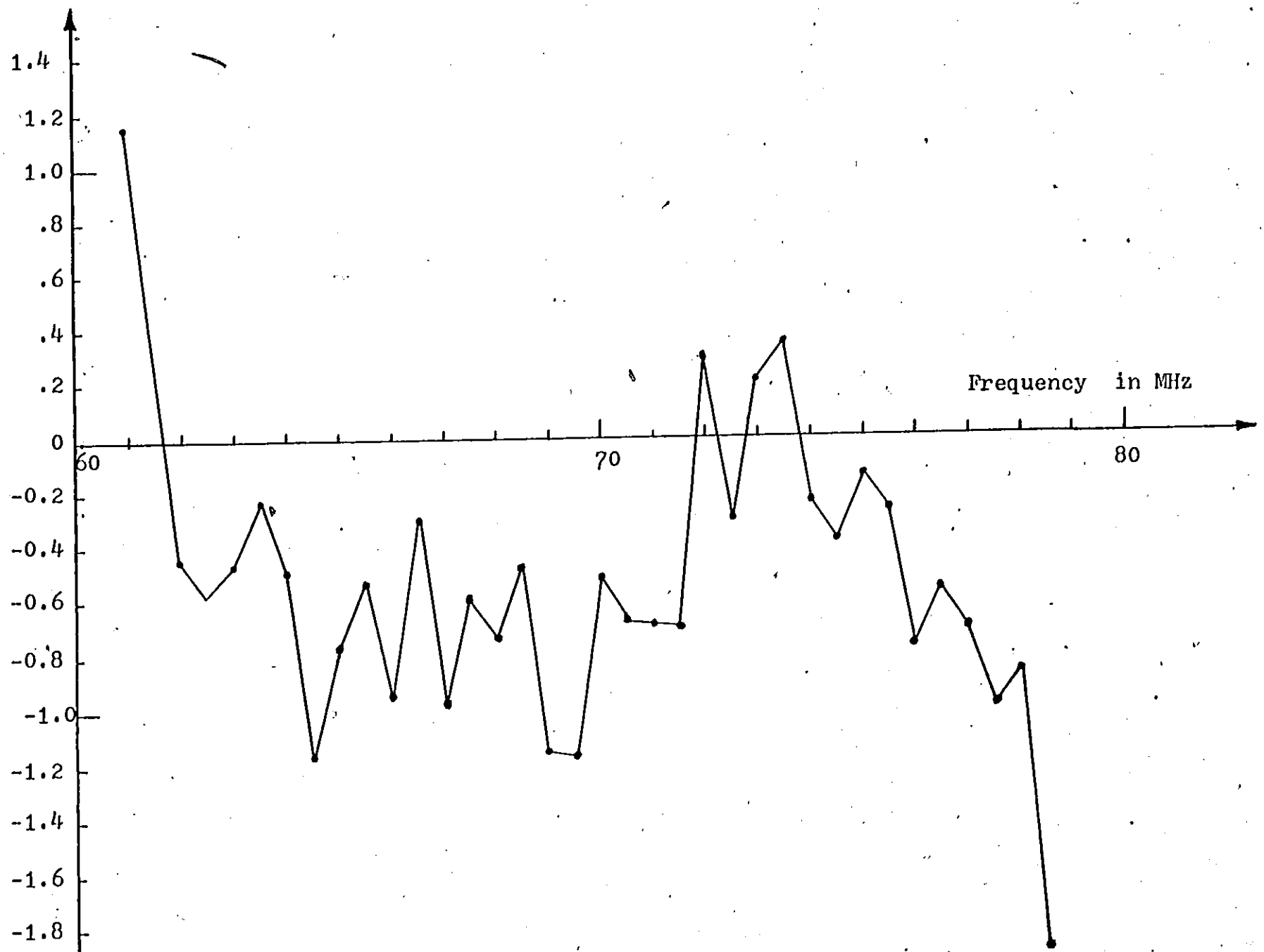


Figure 7.12 Non-linearity Error of the Conversion Characteristics vs Frequency

frequency. As can be seen, the maximum non-linearity as a percentage of the full scale was found to be 1.82%. This indeed is an excellent result considering the poor quality of the SAW device.

7.4 DEMODULATION CHARACTERISTICS

To demonstrate the demodulation characteristics of the device, we generated a single-tone frequency-modulated signal. The test set-up used for this experiment is shown in Figure 7.13. A sinusoidal waveform was generated by the baseband signal generator to externally FM modulate the carrier of HP 8601A generator/sweeper. The carrier frequency or centre frequency of the modulator was 70 MHz, and the frequency deviation was ±9 MHz. The sensitivity of the modulator was 5 MHz/volt. This FM signal was fed to the discriminator and the output of the discriminator was measured at two points. One at the output of the integrator, and the other at the output of the lowpass filter. These outputs were then compared with the original baseband signal. A lowpass filter was incorporated in the discriminator to suppress the sampling frequency present in the output of the integrator. We have not incorporated this lowpass filter as a part of the discriminator since the conventional FM receiver always has a low pass filter following it. As such, it is always regarded as a separate

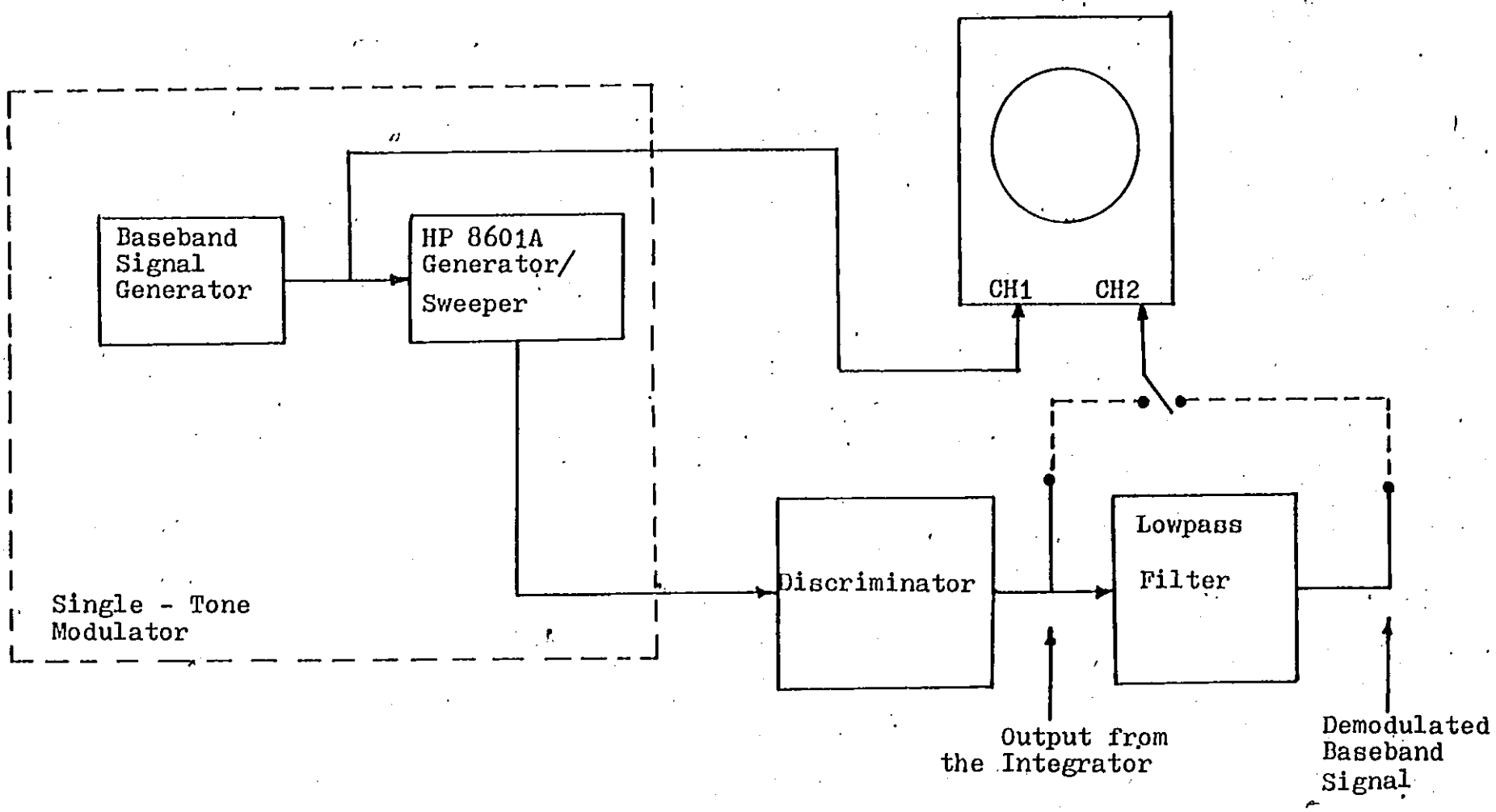


Figure 7.13 Arrangement used for Observing the Demodulation Characteristics

circuit. We have just followed this convention. In Figure 7.14(a) the upper trace illustrates the modulating baseband signal while the lower trace displays the output from the integrator. In Figure 7.14(b) the upper trace once again illustrates the baseband modulating signal while the lower trace indicates the demodulated signal seen at the output of the lowpass filter. Comparing these two signals, except for the delay found in the demodulated signal they are identical. In fact, this is what we expect from a good demodulator.

7.5 NOISE PERFORMANCE OF THE DISCRIMINATOR

In evaluating the noise performance of the discriminator we must pay attention to two important aspects. They are:

1. Noise performance of the SAW chirp filter correlator
2. Noise performance of the overall discriminator system

Usually we are only interested in the latter of the two. However, in the case of this discriminator, we are interested in both. Since this is a new principle of operation of FM discriminators, we will have to use new standards for measuring the noise performance. These standards must be compatible with conventional standards, in

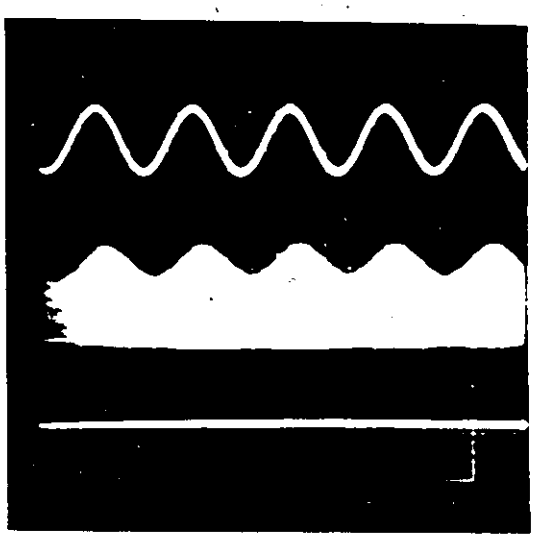


Figure 7.14(a) Upper trace: Input audio modulating signal.
Lower trace: Output from the integrator prior to lowpass filtering.
Vertical scale: 2 volts/div. Bandwidth 18 MHz
Horizontal scale: 5 msec/div.

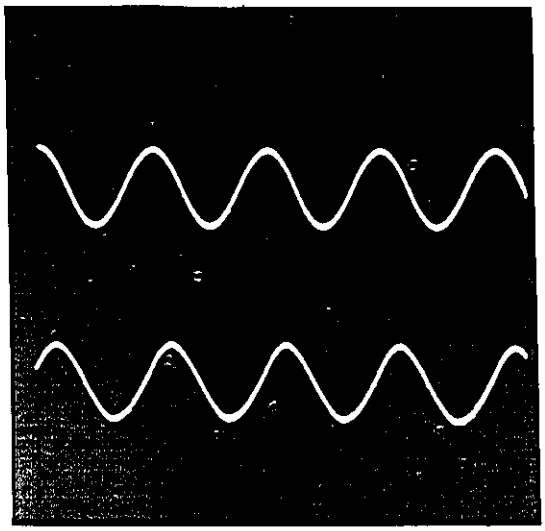


Figure 7.14(b) Upper trace: Input modulating signal.
Lower trace: Demodulated signal seen at the output of the lowpass filter.
Verical scale: 2 volts/div.
Horizontal scale: 5 msec/div.

order to make a sensible comparison between the noise performances of conventional discriminators and this discriminator. Therefore, when we define our noise measuring parameters, we should be particularly careful in our definitions, so that they do not contradict the conventional viewpoint as to the validity of the new principle of operation.

7.5.1 Noise Performance of SAW Chirp Filter Correlator

Since the major function of the SAW chirp filter correlator is to optimize the signal to noise ratio of the signal sample, we propose to look at the physical behaviour of the device in the presence of the noise. In doing so we hope to find a way to define the noise performance parameters of the device, and also to develop a method to measure these parameters.

The chirp filter is a bandpass filter. As such it will respond to the noise within the bandwidth of the device. The noise present is additive white Gaussian noise. If we take a sample of this noise and feed it to the chirp filter, the device will behave in the following way. All the noise spectral components outside the bandwidth of the device will be rejected by the filter. The spectral components of the noise within the bandwidth of the device will be found at the output of the filter, and the noise

voltage that we see at the output will be constant across the bandwidth assuming the filter passband characteristics are ideal. Suppose we have the situation where the input to the discriminator is signal plus additive Gaussian white noise. Then we sample this and feed it to the chirp filter correlator. Now the noise will be spread right through the bandwidth, but the signal energy would be compressed to a single point. The result is that the output signal to noise ratio is much better than the input signal to noise ratio. In Figure 7.15 we have illustrated this physical mechanism.

7.5.2 Detection Ratio and the Threshold of the SAW Correlator

The output of the correlator gives out the compressed signal whose maximum occurs at the time of perfect correlation. In the discriminator we detect this peak to measure the time of occurrence of the correlation. In other words, our information is in the peak of the pulse. Our detector will start making errors if the noise peaks are at the same level as the signal. Therefore, we may measure the ratio of the signal amplitude to noise amplitude to evaluate the performance of the SAW device in the presence of noise. We shall express the detection ratio D as

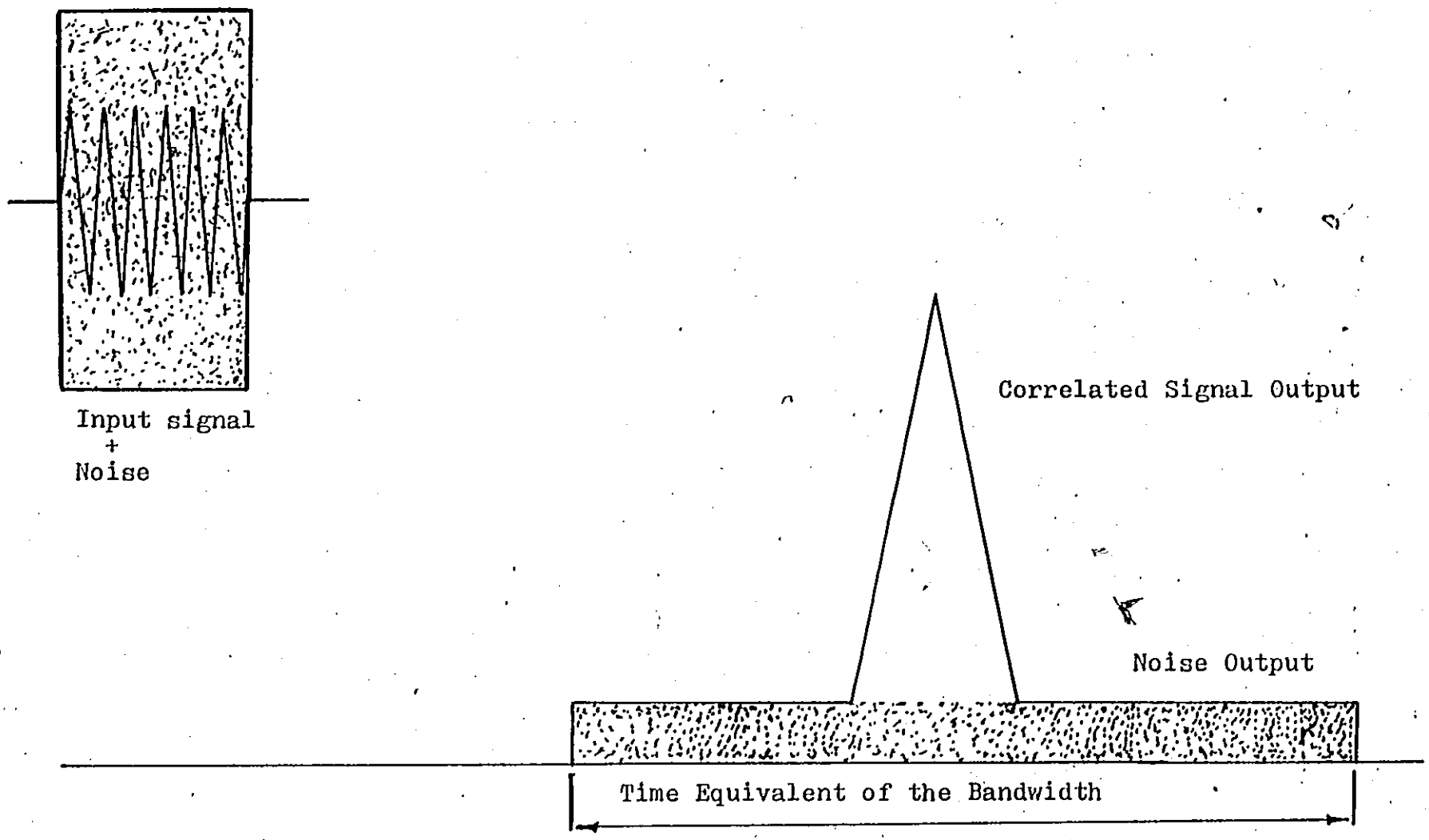


Figure 7.15 Physical Mechanism of Noise and Signal Compression

$$D = \frac{\text{Maximum Amplitude of the Signal at the output of SAW Device}}{\text{RMS Noise voltage amplitude at the input of the discriminator}}$$

We find this measurement parameter makes a sensible evaluation of the noise performance of the SAW device, and it is compatible with the principle of operation of the device.

7.5.2.1 SAW Device Threshold

If the detection ratio is greater than unity then the detector circuit will likely detect the signal reasonably accurately. If the detection ratio is less than unity then the detector will likely make errors.

Thus the SAW device threshold is said to have occurred at an input carrier to noise ratio such that the detection ratio has reached unity. It must be emphasized that this threshold is not the threshold of the discriminator. It is the threshold of the SAW device. It defines the device's capability and allows us to make two assessments. They are:

1. Assess the noise performance of the SAW device with regard to its performance quality
2. Assess the maximum threshold one can achieve with the present state of the art of the SAW chirp correlators.

We shall pursue these assessments after measuring the detection ratio.

7.5.3 The Detection Ratio Measurement Technique

In measuring the detection ratio, first we need to measure the input signal-to-noise ratio. The instrument arrangement for these measurements are shown in Figure 7.16. The FM signal from the modulator was connected to a resistive adder via a variable attenuator. Similarly, the amplified Gaussian white noise also passed through a variable attenuator and was connected to the resistive adder. Thus at the output of the resistive adder, we find the signal plus noise, and by varying the variable attenuator, we can vary the input signal to noise ratio. For the detection ratio measurement we use an unmodulated carrier.

Measuring the output signal amplitude is quite straightforward and can be done directly using an oscilloscope. However, the mean noise amplitude cannot be measured directly using the scope due to the random nature of the noise, although one could get a general idea of the behaviour of the device to noise using the scope. Thus we must use a power meter and calculate the mean noise amplitude.

The power meter reading indicates the mean noise measured during the sampling period. Suppose the sampling

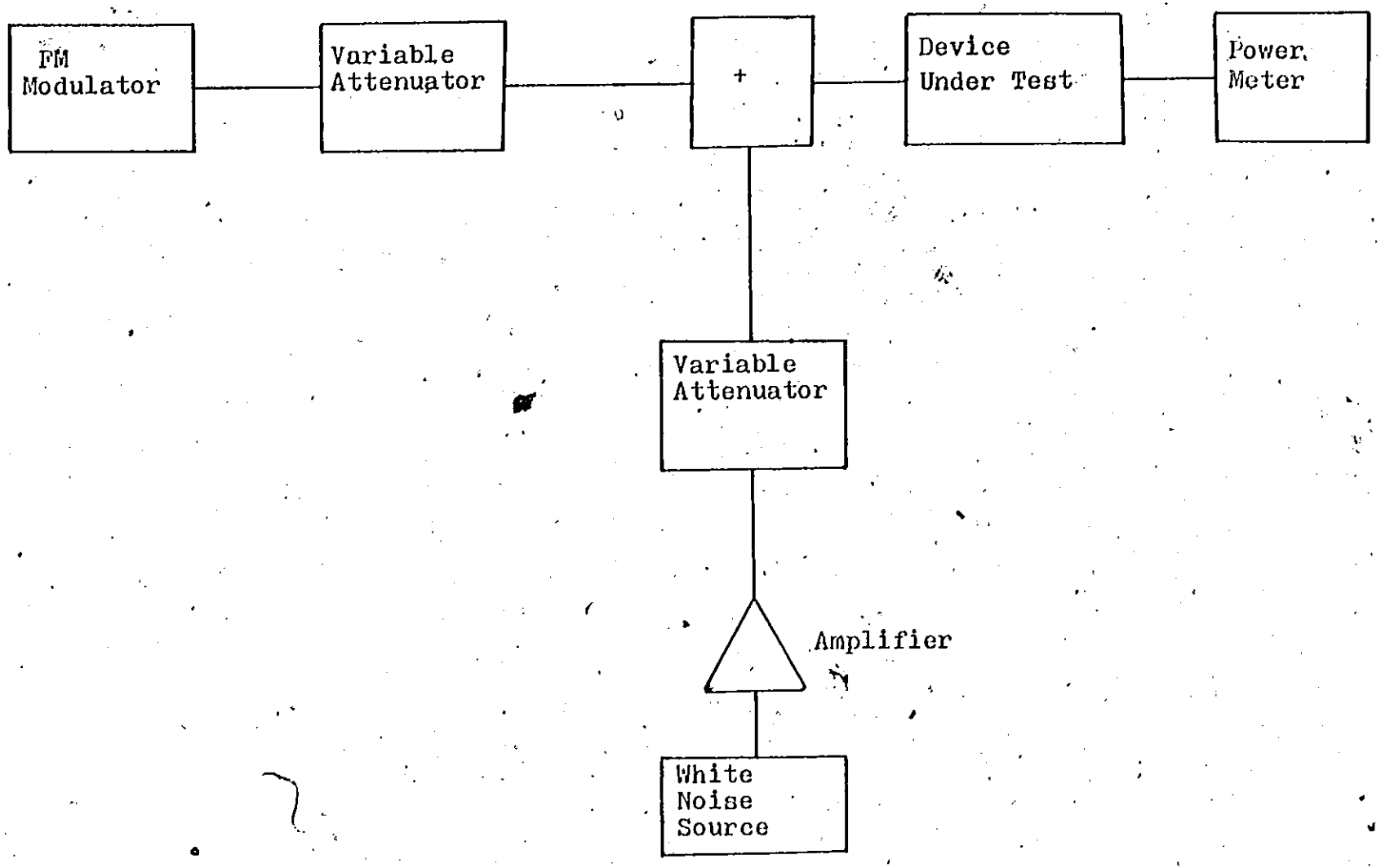


Figure 7.16 Instrument Set up for SNR Performance Measurements

period is T_S, and we know the noise is present at the output only a fraction of the time. Let this fraction be ΔT. This happens to be the time equivalent corresponding to the bandwidth of the device. With the knowledge of T_S, ΔT and P, we can calculate the mean noise voltage at the output. In Figure 7.17(a) we have indicated the mean noise amplitude V_n, in relation to the sampling period, and in Figure 7.17(b) we have illustrated the power measured by the power meter over the period. The area under the curve gives the noise energy. Let us calculate the mean noise voltage. The noise energy measured by the power meter is given by

$$E_n = PT_S \tag{7.1}$$

We know this energy is found only within ΔT.

The energy E_p within the ΔT is given by

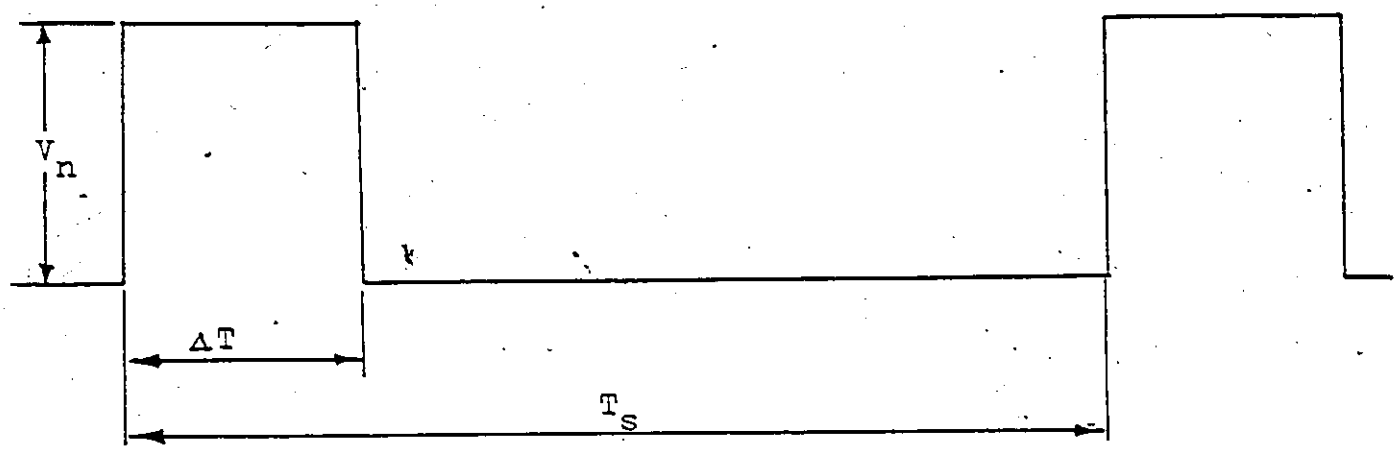
$$E_p = \frac{V_n^2}{R} \Delta T \tag{7.2}$$

where R = Input resistance of the power meter

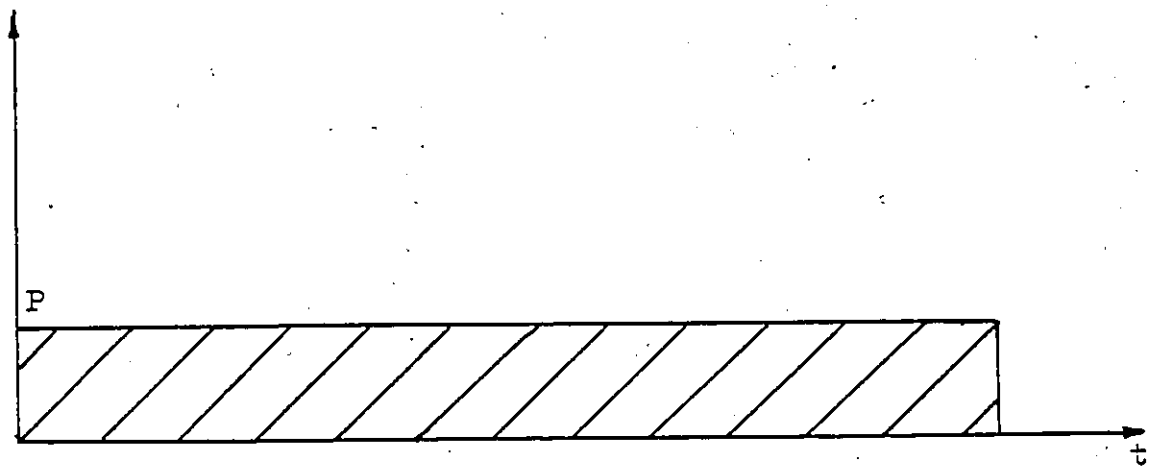
But $$E_n = E_p$$

therefore $$V_n = \frac{P\ T\ R}{\Delta T} \tag{7.3}$$

Now we know all the parameters that we need to calculate the mean noise voltage. Therefore, we can calculate the detection ratio.



(a)



(b)

Figure 7.17 (a) Sampled Noise Input
(b) Measure Power vs Time

7.5.4 SAW Device Noise Measurements

The objectives of the first set of measurements are:

1. To establish the linearity of the measuring system along with the SAW device.
2. To measure the main lobe level and the highest side lobe levels.

Both of these objectives can be met with one set of measurements. In Figure 7.18 we have illustrated the amplitude of the input signal vs. the amplitude of the output signal for the main lobe and highest side lobe. The linearity of the measuring system and the SAW device is self evident in these results. The highest side lobe level is ≈6 dB below the main lobe level. This information is valuable because we can DC shift the output signal from the SAW device at the rectifier lowpass filter stage by the same amount to avoid the side lobe level being detected. Just as a check to see that the linearity is still preserved in the presence of noise, we measured the input noise power vs. output noise power. These results are shown in Figure 7.19, and we find that the linearity is still preserved.

Our next set of measurements is to determine the detection ratio and the threshold of the SAW device. Although it is very difficult to measure any proper noise measurements with the oscilloscope, we have presented a few photographs of the noise outputs observed on the scope, so

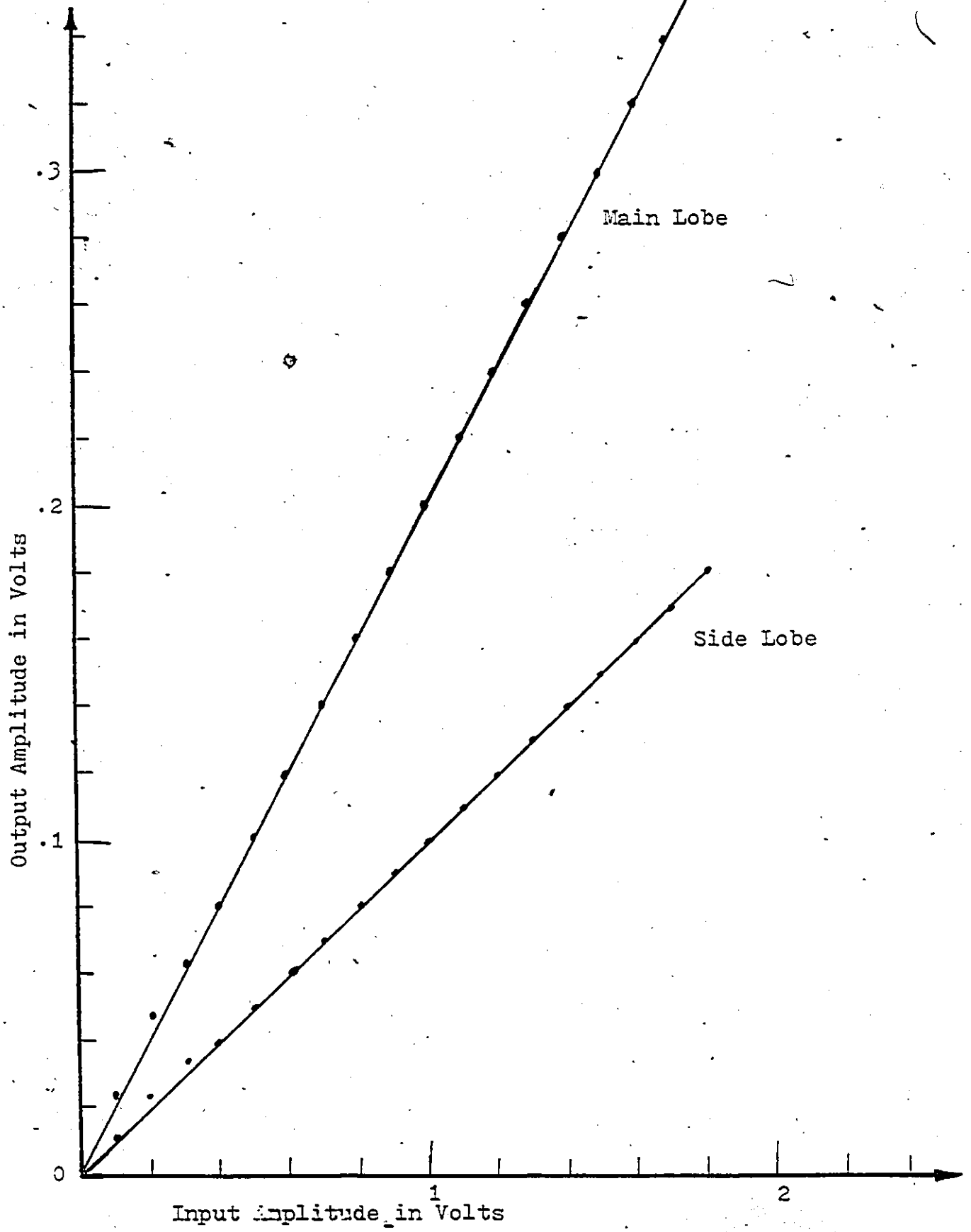


Figure 7.18 Input vs Output Signal Characteristics of the SAW Correlator

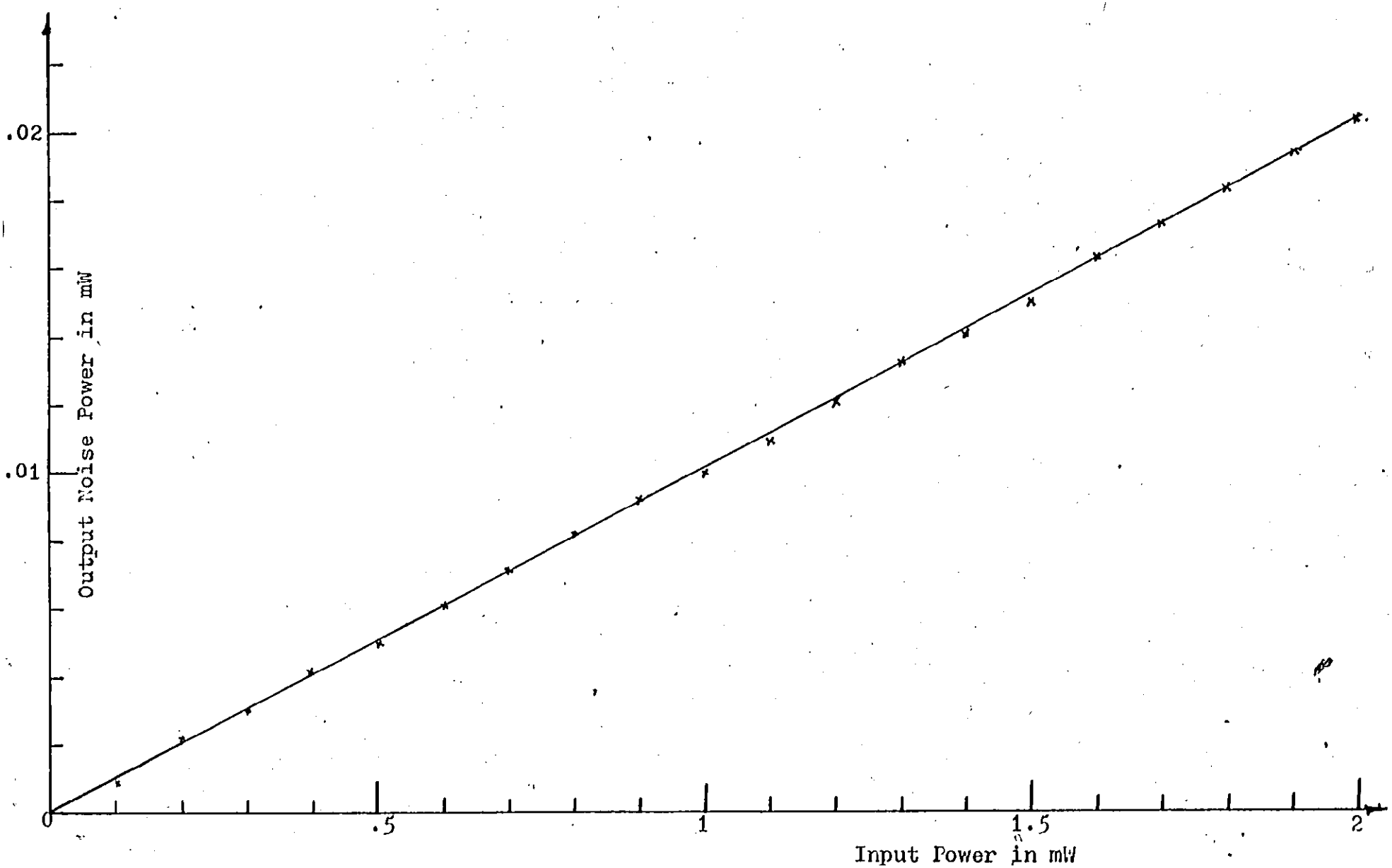
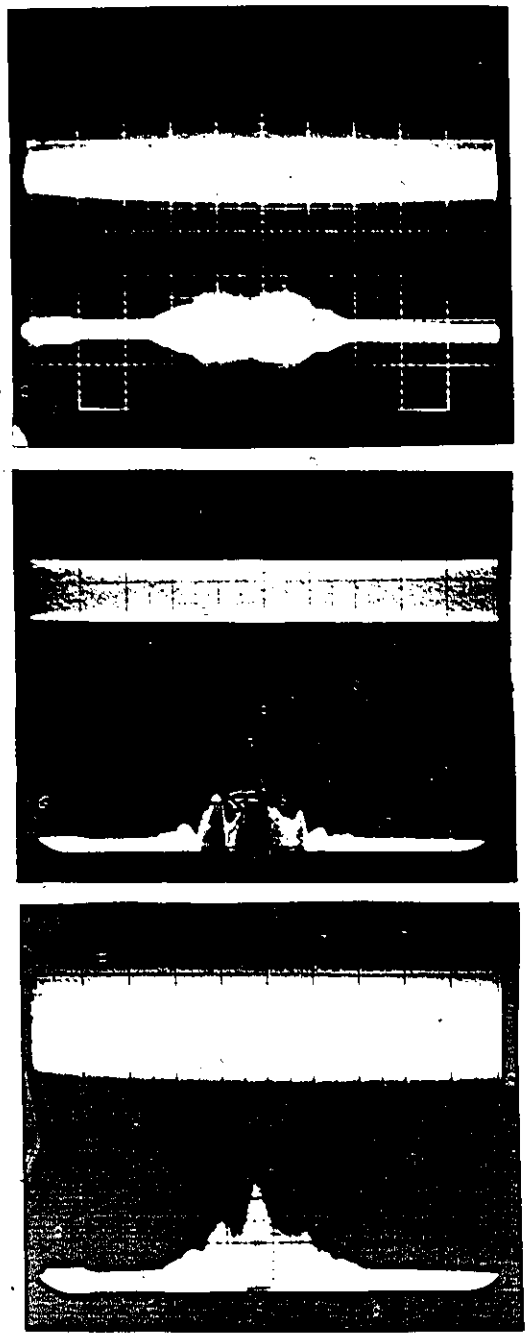


Figure 7.19 Input vs Output Noise Power Characteristics of the SAW Correlator

that one could get some feeling as to what is happening in the device. In Figure 7.20, photograph (a) indicates the input noise to the device in the upper trace while the lower trace indicates the output noise of the device. In (b) the upper trace indicates the signal input to the device and lower trace indicates the output signal. In (c) the upper trace indicates the signal plus noise. As may be observed, the signal is well buried in the noise. The lower trace indicates the output of the SAW correlator resulting from this input, and shows the noise, plus a detectable correlated signal. The detection ratio measurements were plotted on a graph and are shown in Figure 7.21. The horizontal dotted line shown in this figure is when detection ratio D is equal to unity. As we defined earlier, the threshold occurs when the detection ratio meets this line. As can be seen the SAW device threshold occurs when the carrier to noise ratio is ≈-3 dB. This result shows a definite correlation between the side lobe level and the threshold.

The result indicates the amount of white noise power required to bring the output noise level up to the main lobe level. This is, in our view, a very encouraging result. The encouraging aspect is that if we can increase the main lobe level with respect to the side lobe level or equivalently suppress the side lobe levels then we should be



(a)

(b)

(c)

Figure 7.20 (a) Upper trace: Input noise at the discriminator
Lower trace: Output noise of the SAW correlator

(b) Upper trace: Input signal at the discriminator
Lower trace: Output signal of the SAW correlator

(c) Upper trace: Signal plus noise prior to sampling
Lower trace: Output signal plus noise of the SAW device.

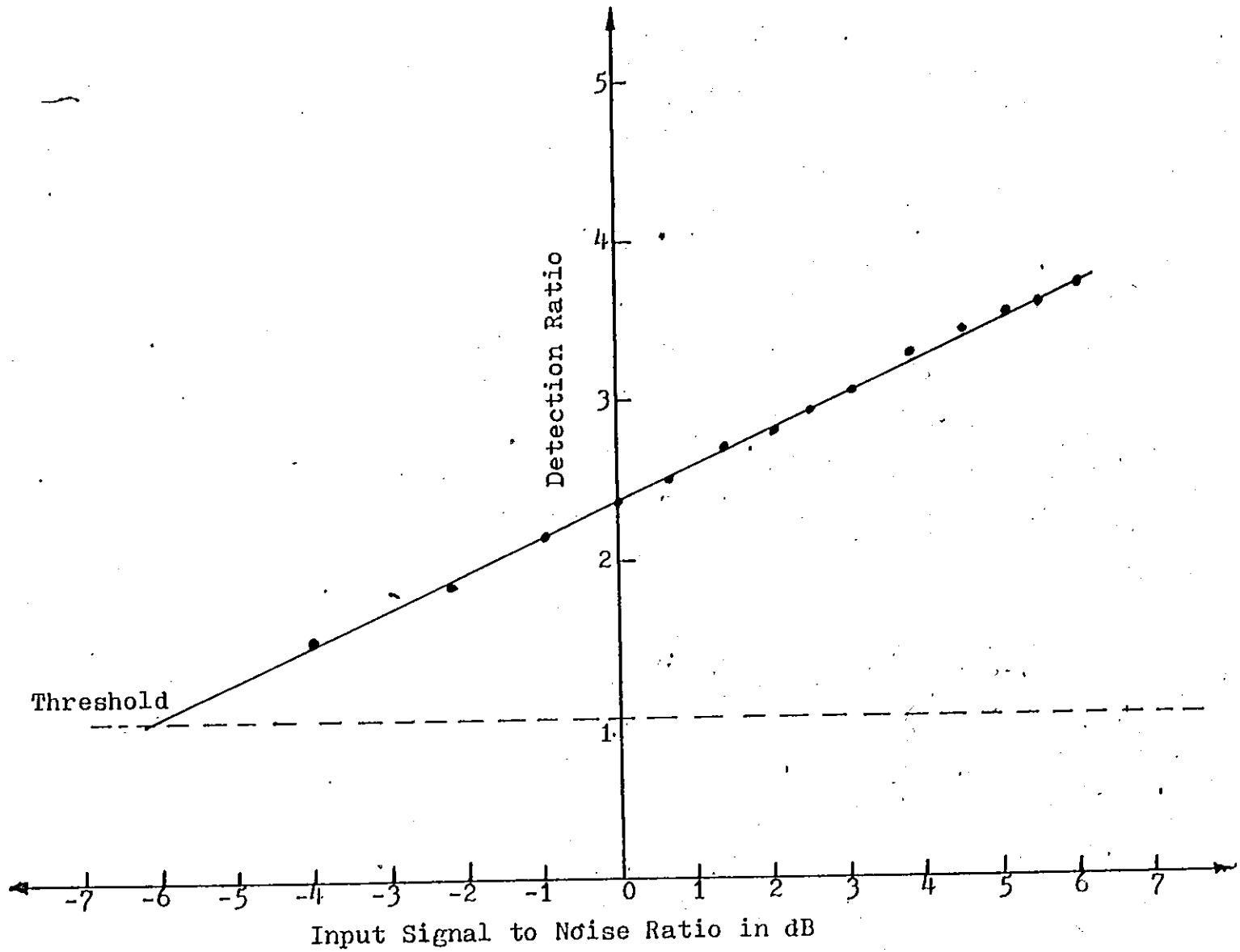


Figure 7.21 Detection Characteristics of the SAW Correlator

able to extend the device threshold. The side lobe level suppression of the chirp filters has been a popular branch of research ever since Sidney Darlington proposed the idea [64]. There are established design techniques available to suppress side lobe levels of SAW chirp filters, and the state of the art is 35 dB side lobe suppression of the chirp filters. This indicates device thresholds of about -35 dB input carrier to noise ratio could be achieved.

7.5.5 FM Threshold of the Discriminator

The threshold that we have discussed so far is a SAW device threshold. This threshold tells us the noise handling capabilities of the SAW device. However, the SAW device is just a part of the complete discriminator. Therefore, we must measure the FM threshold of the complete system. For this measurement we used a sine wave to modulate the carrier over the bandwidth of the device. In Figure 7.22 we have plotted the output signal-to-noise ratio as a function of the input carrier to noise ratio. If we define the FM threshold of this discriminator as the point at which the input carrier to noise ratio vs. output signal to noise ratio characteristics are no longer linear, then we find the threshold occurs at -3 dB of carrier to noise ratio. This definition is compatible with the conventional FM threshold definition.

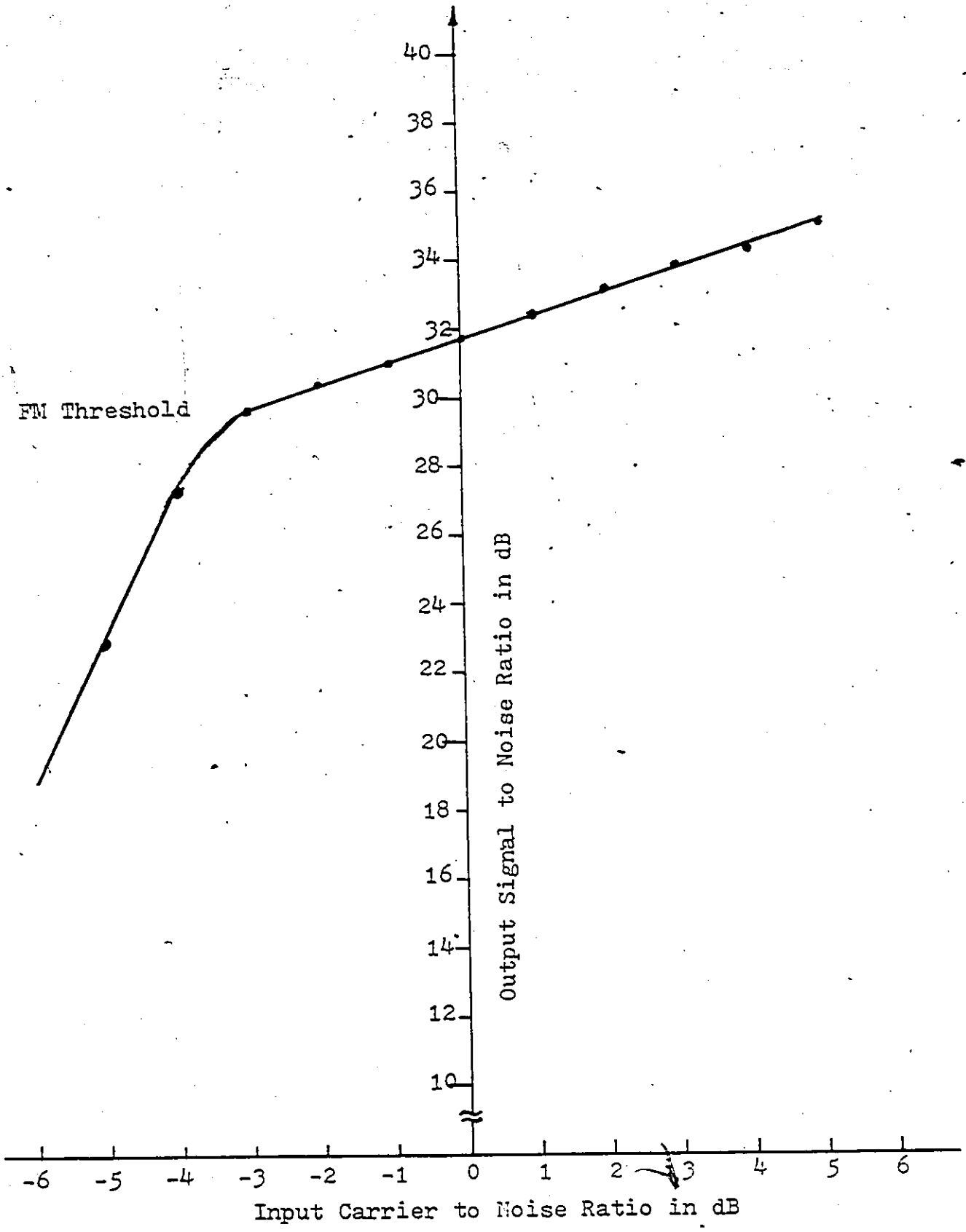


Figure 7.22 Output Signal to Noise Ratio vs Input Carrier to Noise Ratio

Comparing the FM threshold with the SAW device threshold we find a difference of 3 dB. This difference is due to the noise found in the electronic circuits. In producing these circuits we paid little attention toward selecting good quality components. In fact they are being built with cheap, standard, off-the-shelf ICs and other components. With good components and better control over the circuit design it could be possible to eliminate this difference. Since our objective is to prove the principle rather than to produce an optimum discriminator, we felt that choosing these components is justified. What is important to notice here is even with these worst case conditions this discriminator has been capable of extending the FM threshold to -3 dB which is an achievement that portrays the superior signal processing capabilities of the SAW device.

CHAPTER 8

APPLICATIONS

In Chapter 7 we presented the experimental results of the discriminator designed to prove the principle of operation. We have shown the linearity of the device over 18 MHz bandwidth to be within 1.8%. We measured the detection ratio and the threshold of the SAW device which gives an indication of the signal processing capability of the device. Finally, we measured the threshold of the FM discriminator and found it to be around -3 dB. This discriminator by no means is the optimum discriminator. In fact, a much better threshold can be achieved by using this principle.

In communication and in other electronic instrumentation, one requires wideband FM discriminators. One particular field of communications in which one uses wideband FM discriminators is Spread Spectrum communications. In this Chapter we shall propose a modem in which this device can be used. We shall only outline the basic principle of the modem and the implementation of the modem rather than giving details. We shall begin the discussion by briefly introducing the Spread Spectrum concepts.

8.1 SPREAD SPECTRUM COMMUNICATION

By definition, Spread Spectrum communication systems are systems in which a transmitted signal spreads over a wide frequency band. In fact, the bandwidth used is much wider than the bandwidth required to transmit the signal. For example, a voice channel of a few kHz is spread over many MHz. There are 4 main modulation schemes used in Spread Spectrum. These are:

1. "Direct Sequence Modulation" - one in which each bit to be transmitted is transmitted at a much higher rate. Pseudo-noise biphase modulated signals are a good example of this type of modulation.
2. "Frequency Hopping" - another name for this type of modulation scheme is "multiple frequency code selected frequency shift keying". This is somewhat similar to the frequency shift keying except for the fact that here many frequencies are used - unlike the conventional FSK system which use only two frequencies - one for "mark" and another for "space".
3. "Chirp modulation" - this is pulsed FM modulation familiar to radar engineers except one may use more than one chirp signal to transmit the information.
4. "Time-hopping" - here the time of transmission is governed by the code sequence.

There is no restriction on how one is to use these

modulation schemes. One may combine one or two modulation schemes. Usually, in Spread Spectrum systems one uses at least two modulation schemes. Such hybrid systems are known as "lock and key" systems.

All modems except for modems which use chirp modulation have two major problems. They are, the synchronization problem and the hardware implementation problem. This has resulted in complex and cumbersome systems. In spite of these drawbacks, Spread Spectrum offers some advantages that are attractive for secure communication. For this reason and others, this system has been widely used in military communication. Needless to say because of the security blanket over this technology, the state of art is not well divulged to the general public. However, this technology does not need to be used only for military applications. There are sections of the technology which are released for public adoption. They can be used for civil systems where the benefits of Spread Spectrum techniques can be exploited. Some of these benefits as Dixon [65] points out are:

1. Selective addressing capabilities
2. Code division multiplexing
3. Low density power spectra for signal hiding
4. Message screening from eavesdroppers
5. High resolution ranging

6. Interference rejection

To explain how these benefits come about would require a complete review of Spread Spectrum technology which is not the scope of this thesis. As such we shall avoid undertaking this task.

8.2 ROLE OF SAW DEVICES IN SPREAD SPECTRUM COMMUNICATION [66]

In general, SAW devices can help to solve two problems associated with Spread Spectrum systems. They are:

1. Synchronization problem
2. Hardware implementation problem

The synchronization problem can be solved by SAW devices [67]. Correlation is an action easily realized by SAW devices. As such it is possible to make self-synchronizing systems using these devices. The type of SAW devices and how they are to be used depends entirely on the type of modulation being used and how the system is organized.

The hardware implementation problem is solved on two accounts. First, the SAW device itself carries most of that burden. Secondly, SAW devices are small and compact, reproducible, reliable, and economical. An equivalent signal processing function made with conventional circuitry may include a multitude of integrated circuit (IC) and other

components which make the circuitry complex, bulky, unreliable; in addition to these, they have poor reproducibility and are less economical [68] [69].

8.3 MULTI-FREQUENCY POSITION MODULATION (MFPM) MODEM

We propose a multi-frequency position modulation modem as an application for the FM discriminator. We call it by this name because the transmitted signal occupies a large number of frequency positions in the spectrum. Conceptually, this is similar to a frequency hopping system except for the fact that between frequencies there is a guard frequency; that can be used as a reference to synchronize the system. These frequencies are normally referred to as "chips". In normal frequency hop systems there are no guard or reference chips, and an error in synchronization leads to chip interference. As a result, a great deal of effort is being spent on producing an accurate synchronization method. In MFPM modem we overcome this problem by the synchronization of the system prior to processing the chip. The signal set of this modulation scheme is shown in Figure 8.1(a). Let us see how to use our discriminator to generate this signal set.

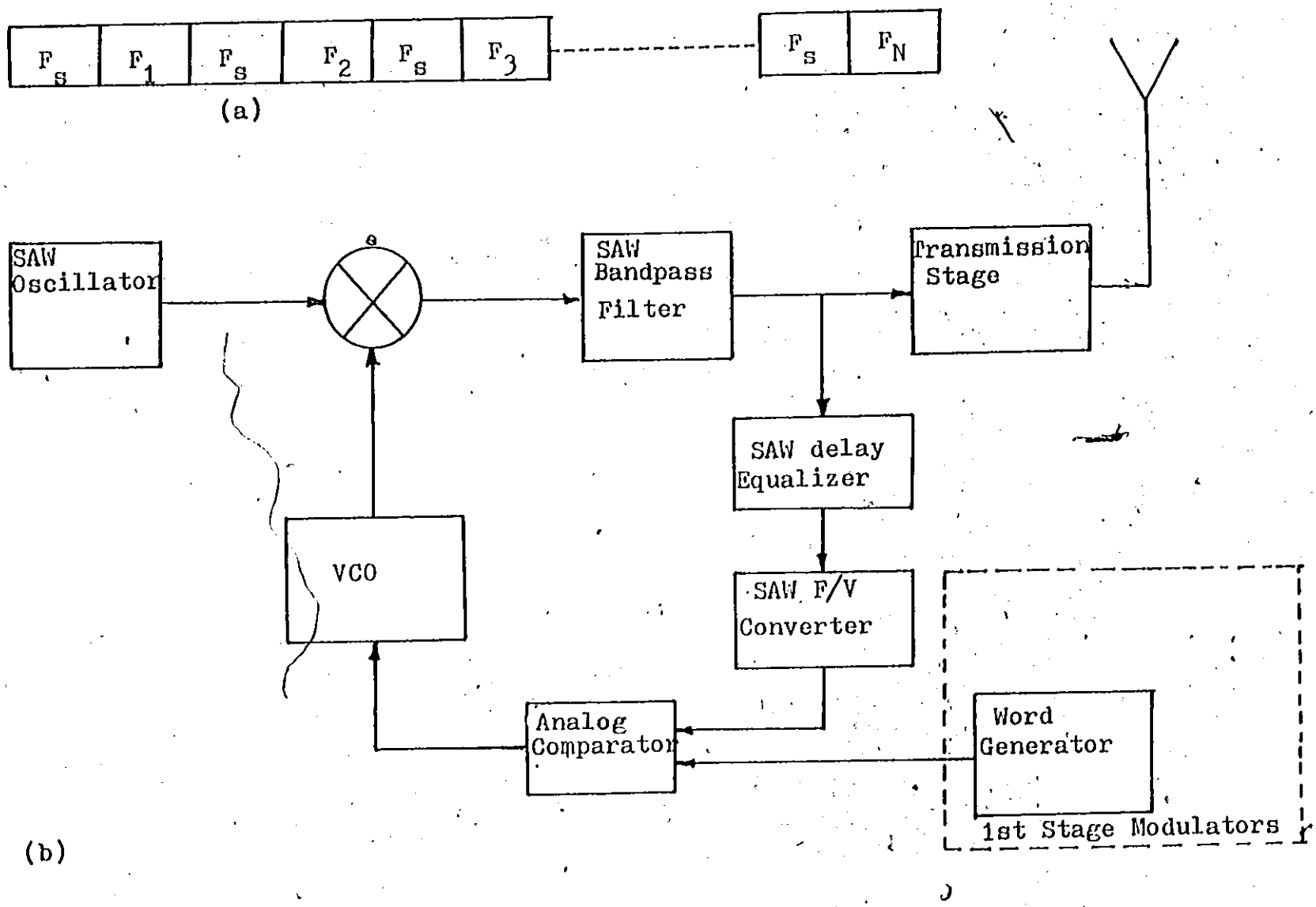


Figure 8.1 (a) MFPM Signal Set
(b) Multi-Frequency Position Modulator and Transmitter

8.3.1 Transmitter

In Figure 8.1(b) we illustrate the transmitter. In this diagram MFPM is the final stage of the multi-modulation scheme. We shall not deal with other modulation schemes prior to MFPM except to say that we require a discrete set of voltage levels out of the word generator. For the purpose of explanation, let us bypass the delay equalizer and assume that the frequency out of the VCO is f_o when the input to it is zero volts. The SAW oscillator output is f_r and is constant and stable. Therefore, the output of the mixer is

$$f_s = f_o + f_r \tag{8.1}$$

We shall ignore all the other frequencies that the mixer generates. This is reasonable because we can choose the bandwidth of the bandpass filter following the mixer to do this. Thus at the input of the transmitter stage and at the input of the discriminator the frequency f_s is present. We have designed the discriminator (F/V converter) in such a way that the output of the device is zero volts when f_s is present. When there is no information from the word generator, the output of it is also zero volts. Now the comparator inputs are zero volts. Therefore, the comparator output is also zero volts. This is the steady state condition in the absence of any information.

Now the word generator gives a voltage level V_1.

This will cause the comparator output of the VCO input to be V_1. Correspondingly, the VCO output will be f_i. Consequently, the mixer output will give

$$f_1 = f_i + f_r \tag{8.2}$$

Once again, we assume the bandpass filter allows us to omit all the other frequencies generated at the mixer output. Now the discriminator sees f_1 and gives out a voltage equal to V_1. This leads the output of the comparator to be zero again and a frequency f_s will be generated. Now we have been successful in generating two chips - one (f_1) corresponding to the word V_1 and the other at f_s (the guard chip). The discriminator generates different delays depending on the frequency. This means the loop delay will vary with frequency, and different chips will have different time lengths. This is an undesirable situation. We need some uniformity in chip lengths. We solve the variable chip length problems by introducing a SAW delay equalizer which has opposite frequency vs. delay characteristics to the one found in the discriminator. Having done that, all we have left are scaling problems; such as selecting the time length of the chips, how many chip pairs for a word, etc. For the purpose of discussion, let us assume the time length of the word is adjusted in such a way that the modulator generates one chip pair per word. By "a pair" we mean an information chip at f_N and

guard chip at f_s. These frequencies are then up converted and transmitted.

8.3.2 Receiver Stage

We can also build a receiver stage using a SAW discriminator. Such a scheme is illustrated in Figure 8.2. Let us assume two chips have been up converted and transmitted. These two chips are a guard chip and an information chip. Thus, at the receiver, we have an up converted version of these two chips. We need to recover the original word. First we down convert the received signals such a way that we obtain the original two chips f_s and f_n. We send these two signals through two SAW bandpass filters. The first bandpass filter is designed in so that it will pass a band of frequencies containing f_s. The second bandpass filter bandwidth occupies all the other chip frequencies. Thus the guard frequency will be correlated by the SAW correlator and produce a reference pulse to the discriminator. In other words, the sampling pulse generator of the SAW discriminator is replaced by the SAW correlator and other associated electronics with it. Following f_s we have an f_n, and this will be passed through the bandpass filter and fed to the discriminator. The discriminator demodulates f_n and gives out the original voltage level corresponding to the word transmitted. This output from the

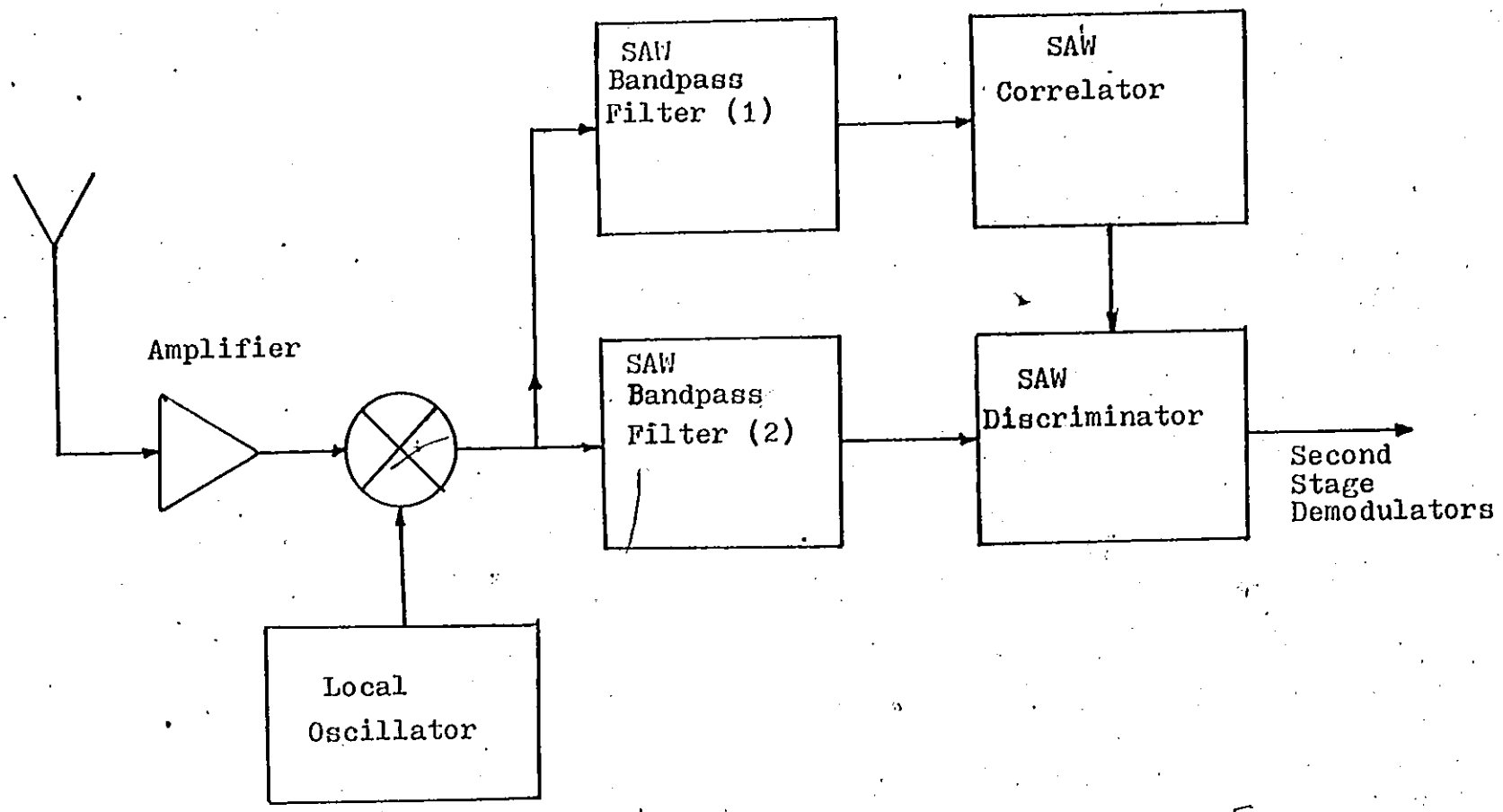


Figure 8.2 Receiver for MFPM Signals

demodulator is then fed to other demodulation sections of the receiver to recover the baseband signal. This is the very basic principle of operation of the receiver.

Here we have outlined the philosophy of producing a multi-frequency position modem. We have also outlined the very basic principles of implementation using SAW discriminators. We have avoided details such as timing, amplification and number of chips per bit design, etc. They are standard systems engineering problems to which solutions depend on the systems specification. A detailed performance analysis of the system is also out of the scope of this thesis. However, this scheme in principle will solve two major problems which plague the frequency hop system. They are chip interference due to synchronization errors and hardware implementation. Notice the number of vital circuits in the modem that can be built with SAW devices. For example, 50% of the circuit functions of the transmitter can be realized with SAW devices, and in the receiver stage all the circuit functions of the demodulator stage are realized with SAW devices.

8.4 PULSE AMPLITUDE MODULATION AND FREQUENCY MODULATION (PAM-FM)

A modem similar to MFPM is PAM-FM which can also be built by using the same philosophy. In this case the first

stage modulator is a PAM modulator. The spacing between the PAM samples can be used as the synchronization frequency or the guard frequency. The receiver would employ the same idea.

Apart from these two communication applications, there are other hybrid modulation schemes [67] where FM is used. It is possible to develop principles of operation for those using this discriminator. We shall not deal with them here.

8.5 APPLICATION TO INSTRUMENTATION

Quite apart from the application to systems, there are many measuring systems in which one requires F/V converters. As we have mentioned earlier it is normally difficult to build F/V converters in the frequency range of 10 MHz - 1 GHz. Having shown a way to overcome this problem we would like to point out that it is possible to cascade chirp filters to produce F/V converters to function over a very wide frequency range. Thus employing such F/V converters in these measuring systems, it is possible to improve the dynamic range and the measurement capabilities of these systems. We shall show two applications that are generally used in these systems. They are

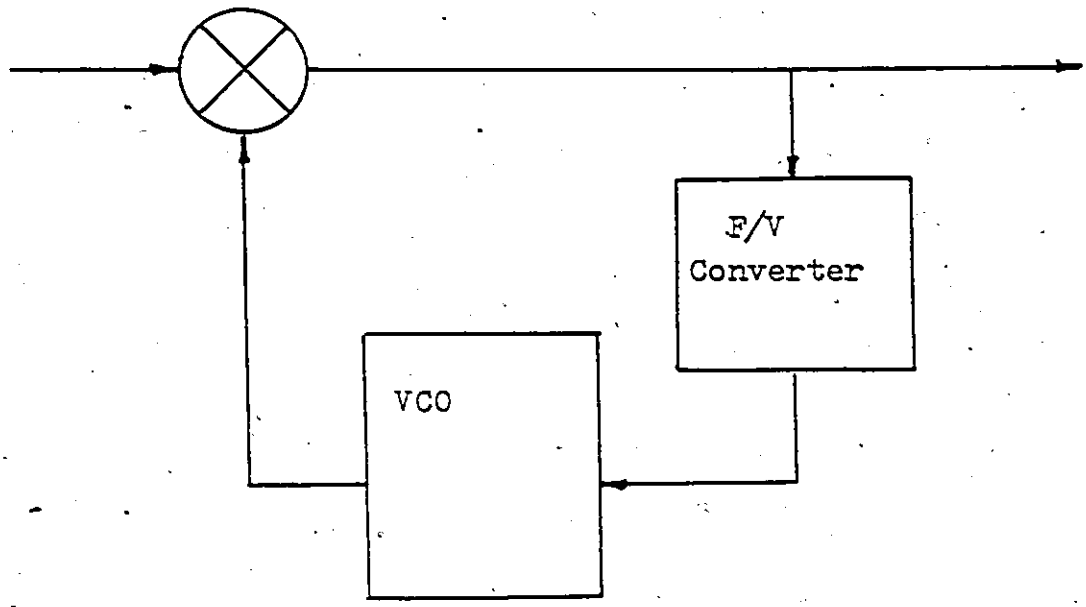
1. Automatic Frequency Control System (AFC)
2. Frequency Measuring System

8.5.1 Automatic Frequency Control (AFC) System

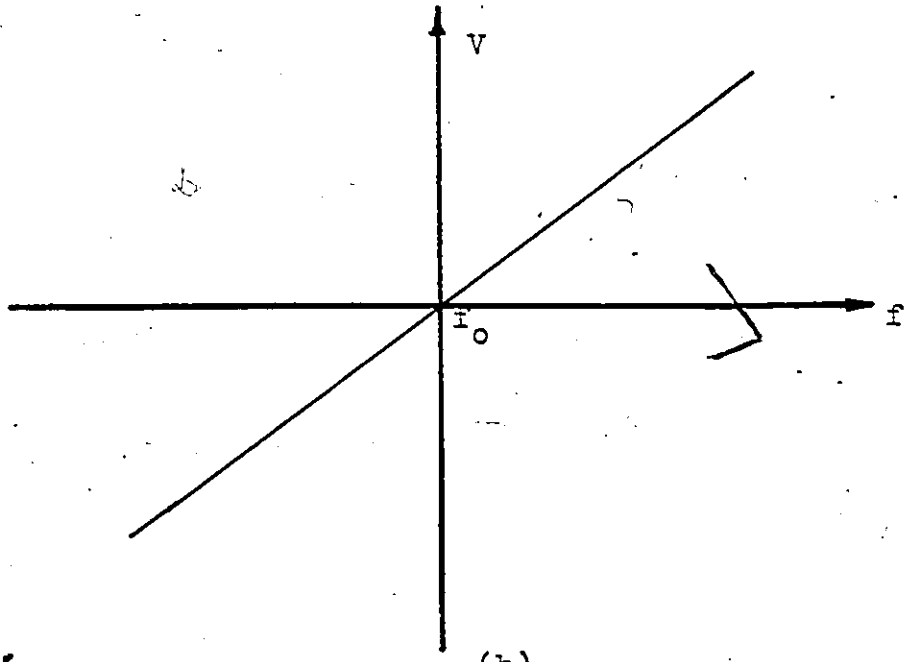
The basic block diagram of the AFC system is shown in Figure 8.3(a). The transfer function of the F/V converter is shown in Figure 8.3(b), in which the object is to produce a stable frequency at the output. Any deviation from the required frequency is detected by the F/V converter and will drive the VCO to generate a correction frequency. The correction is made at the mixer. This is a conventional AFC except that the SAW F/V converter is incorporated to improve the performance.

8.5.2 Frequency Measuring System

In Figure 8.4 we have illustrated the frequency measuring system. Here, the SAW discriminator is being used without the integrator. The input frequency to be measured is fed to the discriminator. At the time the frequency enters the discriminator the bistable output sits high, thus opening the gate between the pulse generator and the counter. The bistable output goes low after a delay time which is proportional to the input frequency. At that time the gate closes, and no more pulses are given to the counter to count. Now the counter output is proportional to the delay which is proportional to the input frequency. Therefore, the frequency is proportional to the counter output. Using proper scaling and calibration of the counter output and pulse repetition frequency of the pulse



(a)



(b)

Figure 8.3 (a) Automatic Frequency Control System
(b) Transfer Characteristics of the F/V Converter

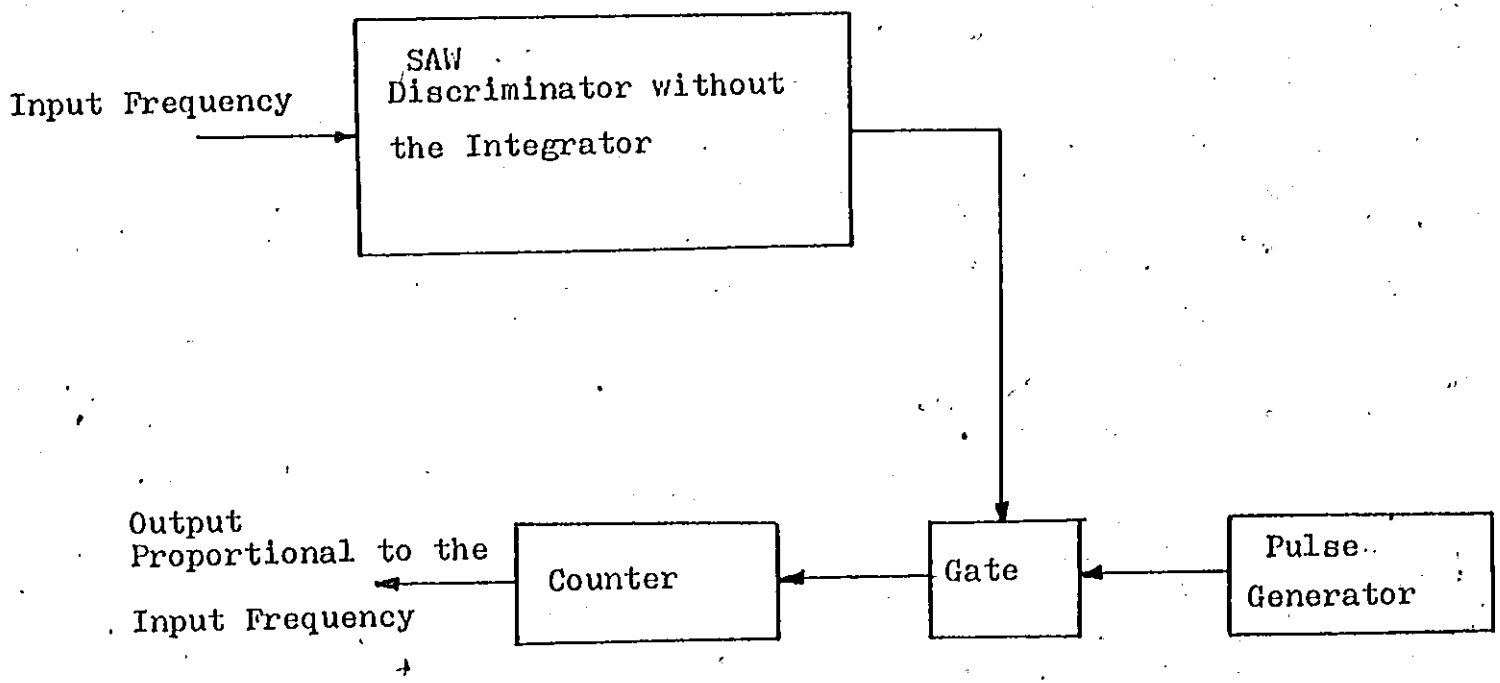


Figure 8.4 Frequency Measuring System

generator, we can make a frequency measuring system.

These are some of the applications. Some of the possible applications of this new discriminator have been mentioned. However, we expect it is possible to find many other applications for this new device. Thus we can optimistically say that the new SAW discriminator may be used for producing better systems. With that optimistic note we conclude this chapter on applications.

CHAPTER 9

CONCLUSIONS

In Chapter 2 we reviewed the existing SAW discriminator types. In doing this, we pointed out the possibility of using SAW devices for narrowband discriminators. Two such specific discriminators proposed are the SAW resonator discriminator and the differential delay line discriminator. However, the general principle of operation of the SAW resonator filter discriminator is identical to the conventional filter discriminator. Therefore, there is no originality as far as the principle is concerned. What would be novel in this is the recognition of the SAW resonator application which has its own advantages. Some of these advantages were discussed in Chapter 2. As far as the differential delay line discriminators are concerned, the principle of using delays and mixing the outputs is known. However, the technique of using differential delay, and mixing the outputs, seems to have a biasing effect on the discriminator, and the operation of this discriminator resembles the second-order phase-locked loop operation. As such it may have the same threshold properties as the PLL demodulator.

Both of these discriminators have certain advantages over the conventional discriminators. They can be designed to operate in the 10 MHz - 1 GHz frequency range. Also, they solve the traditional problems such as reproducibility, reliability and economics. It is more than likely they will have better noise performance purely on the basis of the noise rejection properties of the SAW devices. As we mentioned earlier the basic SAW transduction mechanism involves correlation. This mechanism helps to reject out-of-band noise very effectively. The major disadvantages of these discriminators would be the bandwidth. They may prove to be excellent narrowband FM discriminators, but they cannot be made to operate as very wideband discriminators. This is due to the fundamental limitation of the principle of operation.

At the end of Chapter 2 we summarized the main problems associated with the existing discriminators, and undertook to solve them. In carrying out this task we found two major causes of these problems. They are:

1. limitations in the operational principle
2. limitations in the hardware used to implement these discriminators

In Chapter 3 we formulated a functional diagram to overcome the limitations found due to the principle of operation. What may be unique in this approach is that the

FM signal is processed sample by sample and assembles as the baseband signal at the output. Also, in the processing we have optimized the signal-to-noise ratio. Using this approach we have developed the principle of operation of the discriminator. We found that by using this principle we would solve three specific problems. These are:

1. Bandwidth limitation
2. Linearity over wide bandwidth
3. Noise performance

The discriminator principle was akin to the M-ary principle. The key to signal detection comes about from time scaling. Increased bandwidth of operation can be achieved by increasing M. Since we are processing the signal sample by sample and assembling the baseband signal, we control the assembling process. Therefore, we can control the linearity of the discriminator. Poor noise performance of the conventional discriminator leads to the presence of FM threshold. In FMFB demodulators and in PLL demodulators, extension of the threshold was achieved by using the tracking filter principle. In the new discriminator the threshold extension is achieved by SNR optimization of the signal sample. The reasoning behind this scheme is that if we can suppress the noise and feed the optimum signal to demodulator, then we should extend the threshold. This is because the existence of the threshold

is caused by the inability of the discriminator to separate noise from the signal. In other words, conventional discriminators respond to signal as well as noise and produce a term at the output corresponding to the noise. When the expected power of this noise component becomes larger than the signal component, the threshold occurs. So if we can optimize the signal prior to demodulation, we should be able to extend the threshold.

Having developed the principle of operation, we defined the hardware requirements and pointed out that the conventional devices cannot fulfill the requirements. We claimed that SAW devices, on the other hand, could fulfill these requirements. In Chapter 4 we reviewed SAW device principles. We explained the fundamentals of the SAW devices, generation and detection mechanism and modelling of transducers. Then we went on to explain the SAW filters and recognized them as a class of tranvsversal filters. We explained various advantages of these filters and showed the design freedom they offer. Next we explained the various higher order effects that affect the performance of the device, and how to overcome these effects. One particular device which is useful in suppressing second-order effects is the multistrip coupler, and we explained the operation of this device.

Having introduced the SAW devices and explained the

pertinent materials to this particular work, we developed the hardware. In Chapter 5, we explained how to build a correlator with a pair of transducers, and went on to explain how to extend this principle to build an M-ary correlator with time scaling. We decided not to use that device because of its a cumbersome nature. The reason why we mentioned this is to show that this device can meet the demanded requirements for the design of the M-ary correlator. Having shown that, we wanted to show that a chirp matched filter is an approximate equivalent to this M-ary correlator. Furthermore, we pointed out that for this particular application the chirp matched filter is better suited because its correlation mechanism allows us to relax the restriction placed on sampling; namely that each sample should have only one frequency. This is no longer necessary because the device operates as a plurality of chirp matched filters. Therefore, a slight deviation of frequency would not cause any errors. In fact, it will help to get a better correlated pulse. Having explained these points, we explained how to go about designing a chirp matched filter. Furthermore, we explained a rule of thumb method to design a track changing multistrip coupler to overcome some of the second-order effects. Unfortunately, it must be emphasised that the multistrip coupler is useful only on high coupling-constant materials. The material we used was $L_iN_bO_3$ which

is a high K^2 material. However, for low coupling materials one could use inclined transducers and obtain very good responses [72]. Next we explained how to implement the complete discriminator system. In doing so, we presented a zero-crossing detector technique developed to detect the peak output correlated pulse of the SAW device. This detector principle eliminates errors due to the amplitude variations of the output pulse, thus minimizing any non-linearity errors due to inband ripples in the SAW chirp filter response.

Since SAW devices are being used here, we needed to explain how these devices are made. Therefore, in Chapter 6 we reviewed SAW fabrication techniques. The techniques included in that Chapter indicate how to fabricate SAW devices to operate in the frequency region from 10 MHz to 1 GHz. It must be said that, as SAW technology has borrowed from the already existing integrated circuit fabrication technology, it has contributed just as much new technology back to the semiconductor field. The lift-off technique and x-ray lithography are two such examples. In reviewing the fabrication method, we have made an attempt to highlight the problem areas and show how some of these problems have been solved, and what is being done at the present time to solve other problems. One particular problem of SAW device development work has been in the area of modification of

transducer geometry. More often than not one has to modify the initial transducer geometry to obtain the required response. This means it is necessary to generate new masks from the inital stages of the art work, reduce them, etc. This is expensive and time consuming. In that Chapter we have introduced a simple thin film technique to modify the apertures of interdigital transducers. Furthermore, we have experimentally verified the technique and the results were presented. This technique offers the facility to develop apodization functions experimentally.

In Chapter 7 we presented the experimental verification of the principle of the discriminator. In particular the experimental results reflect the performance under the worst-case condition. The time-bandwidth product of the particular chirp filter is only 20.2 per transducer, and consequently does not reflect the state of the art of chirp matched filters. Devices with time-bandwidth product of 10,000 are being produced [73]. Passband ripples (0.2 dB) and side lobe levels of -35 dB with respect to the main lobe levels are now being attained [38]. Furthermore, these devices are being made on a semi-routine basis in some laboratories. Their reliability and reproducibility are proven facts. Needless to say, there is no comparison between our chirp device and these devices. However, our device has been a good vehicle to test the rest of the

circuits. The detector circuit in this discriminator employs a zero-crossing detector technique for peak detection. It was developed to minimize any detection errors that are caused by amplitude variation of the received pulse from the SAW device. Therefore, a chirp filter with poor passband ripple characteristics would be a good vehicle to test the operation of these circuits. We proved the amplitude insensitivity of this detector circuit in the results. Even with this poor chirp filter, this discriminator linearity was measured to be better than 1.8% over 18 MHz bandwidth with a centre frequency of 70 MHz. This of course was more than we expected with this device. Needless to say, with better SAW devices, one could obtain extremely linear characteristics. Furthermore, bandwidths in the order of 100% could be realized using chirp filters since the only bandwidth dependent components in this circuit are the chirp filter and the amplifier immediately following the SAW device.

There are no problems in finding amplifiers which operate in the frequency range of interest (10 MHz - 1 GHz). In fact, they are off-the-shelf components these days. In regard to the SAW chirp matched filter, bandwidth is not a problem. Wide-band chirp matched filters have been produced [74]. The frequency of operation of this discriminator is also primarily governed by the SAW device. Chirp matched

filters operating at 1.3 GHz with 500 MHz bandwidth and side lobe levels of -30 dB below the main lobe level have been reported [75]. Therefore, we can confidently say that the frequency of operation is not a problem in this type of discriminator.

In the evaluation of the signal performance of this discriminator, we have presented the results of the device operating as a demodulator. For this purpose we used a lowpass filter after the integrator to suppress the high frequency components that are present due to sampling. To clear up any confusion as to why it has not been incorporated in the discriminator system, conventional FM receivers usually have lowpass filters after the discriminator. As such, they do not need to be incorporated into the discriminator system. The reason why there is a high frequency component in the output of the integrator can be understood by recollecting the principle of operation of this discriminator. The discriminator processes the input FM signal sample by sample. In other words, in the discriminator process one sample, assigns the baseband value to the output, and then waits for the next sample. Thus, at the output, we have a frequency component corresponding to the sampling frequency that needs to be suppressed to recover the baseband signal.

In measuring the noise performance of the device, we

need to evaluate the SNR optimization capability of the SAW matched filter. The yardstick we defined for this purpose was the so-called detection ratio. Also, we defined another measuring parameter called the SAW device threshold. This parameter must not be confused with the conventional FM threshold. It is a parameter used to measure the noise performance capability of the SAW device. We have also explained the measurement techniques developed to measure these parameters. The most important result we found in these measurements was the correlation between side lobe levels of the SAW chirp filter and the SAW device threshold. We found them to be identical. This is a far reaching result, and it means that if we can suppress the side lobe levels relative to the main lobe, we should be able to extend the SAW device threshold. Translating this result to the state of the art of chirp filters, it is possible to extend the SAW device threshold up to -35 dB. According to the definition, the threshold occurs when the mean noise amplitude is equal to the signal amplitude, in which case the detection ratio is one. The SAW device threshold occurs when the mean noise amplitude is equal to the signal amplitude, in which case the detection ratio is one. The FM threshold occurs when the detector starts making errors, assuming the thermal noise in the other electronics can be neglected. Since detector errors are more likely to occur

when the detection ratio is greater than unity, the FM threshold can be made equal to the SAW device threshold. This means it should be possible to extend the FM threshold to -40 dB. However, there is going to be some thermal noise in the electronic circuits and the FM threshold measurement of the discriminator indicates a difference of about 3 dB. That is to say, the FM threshold occurs 3 dB below the input carrier to noise ratio. It must be emphasized that the electronic components used in this circuit are not high quality. We believe it is possible to improve these circuits further so that the difference between the thresholds can be made even smaller. It is clear to us that by using better SAW chirp matched filters and better electronic design, we can extend the threshold further.

The FM threshold of conventional discriminators occurs at 10 dB of input carrier to noise ratio. FMFB and PLL demodulators show a threshold reduction of about 5-7 dB over this figure. Comparing the new discriminator FM threshold with these, we find a reduction improvement of 13 dB compared to the conventional discriminator and 8-10 dB compared to the FMFB and PLL demodulators. This represents a significant improvement in the design of the minimum power FM systems. What is more important is that this is the worst case results and even then it portrays superior results.

In Chapter 8 we proposed some applications of this discriminator. One area of communications to which this discriminator and other SAW devices can contribute, is Spread Spectrum systems. In fact, this is a well recognized fact, and a great deal of research and development are presently being carried out in this field. Until recently, Spread Specrum had been reserved for military communication. It has unique advantages some of which have been discussed in Chapter 8. One could exploit these advantages in civil communication. For example, they could be used in the areas of space communication, civil mobile communication, and air traffic control communication and in remote data transmission etc. One wideband system we proposed is a multi-frequency position modem. We have explained the general principle of operation of this modem. It overcomes the problem of synchronization and simplifies the hardware found in the conventional frequency hop Spread Spectrum system. A similar system can be used for PAM-FM modems. Finally, we have shown how the device can be used in measurement systems, namely, as an AFC unit and a frequency counter.

As a final check we shall look at the summary of the problems that we found in conventional discriminators, and see how many of them have been solved. We have summarized them at the end of Chapter 2 in Table 2.1. The first

problem listed in the Table is the narrow bandwidth found in the conventional discriminators. We have solved this problem. The second one is the poor linearity problem which also has been solved. Entry 3(a) is the limitation in frequency of operation and we have shown that the present discriminator does not face that up to 1 GHz. Entry 3(b) is the limitation in bandwidth at higher frequencies. This too we have overcome. The fourth problem was the limitation in FM threshold. We have experimentally shown how to extend the FM threshold, and have shown that this threshold can be further extended. The fifth and sixth listed problems, namely, reproducibility and reliability problems are overcome by using SAW devices. These are inherent qualities of SAW devices. Finally, the seventh problem is economy. The SAW devices face the same problem that the transistor faced at the early stages of its introduction. They were expensive, but now the same device which once cost many dollars can be bought for a few cents. What brought the cost of these devices down was mass production capability of the technology. The production techniques for SAW devices are much simpler than for transistors. Therefore, the price of SAW devices will also go down when produced on a mass scale. In fact, they already have. It must be emphasized that in principle all the components which go into this discriminator can be produced on a single substrate using

integrated SAW device techniques. Thus the cost reduction is possible if they are produced in large quantities. In Table 9.1 we have summarized the properties of the new discriminator.

Table 9.1 Properties of the Sampling FM Discriminator

1. Wide bandwidth operation (up to 100%)
2. Excellent linearity over the device bandwidth
3. Frequency of operation in the range from 10 MHz - 1 GHz
4. No limitation in bandwidth at higher frequency of operation
5. Better FM threshold (presently -3 dB and with the possibility of extension)
6. Good reproducibility
7. Simplicity
8. Good reliability
9. Compactness
10. Economy

In summary, we have explained a novel principle of operation of an FM discriminator using a chirp matched filter as the central signal processing element [48]. We have shown how to design this discriminator and have explained the circuit techniques developed to overcome any nonlinearity error. We have shown a new simple thin film

technique that can be used to develop SAW devices and the technique was experimentally verified [49]. The discriminator principle was experimentally verified, and the technique used to measure signal and noise performance have been explained [71]. This new principle overcomes many of the problems that were associated with conventional discriminators. Finally, we have proposed some applications for this device. As a final concluding remark, we can say that this SAW device based discriminator allows us to exploit the potential advantages of FM technology which were not possible previously.

APPENDIX A

CLEANING PROCEDURE

1. Immerse the substrate into trichlorethylene and slowly bring it to a boil (use a beaker and hot plate).
2. Transfer this beaker into the ultrasonic cleaner and agitate for 10 minutes.
3. Repeat step 1 with fresh trichlorethylene.
4. Repeat step 2 for 10 minutes.
5. Transfer the substrate into a beaker of methanol.
6. Repeat step 2 for 10 minutes.
7. Repeat step 5.
8. Repeat step 6.
9. Transfer the substrate into a beaker of distilled water.
10. Repeat step 2 for 5 minutes.
11. Rinse off the substrate twice with distilled water.
12. Transfer the substrate into a beaker of detergent and water (5% detergent). Bring to a boil slowly.
13. Repeat step 2.

14. Rinse off the detergent with distilled water twice.
15. Immerse the substrate into fresh distilled water and warm it up.
16. Repeat step 2.
17. Rinse off twice with fresh distilled water and remove all the water from the beaker.
18. Add hydrogen peroxide and sulfuric acid (ratio 1:1) and let it stand for 20 minutes.
19. Rinse off 3 times with double distilled water.
20. Load the substrate on a spinner and spin at 3000 RPM while wetting the surface with double distilled water for 15 seconds and continue to spin it for another 45 seconds without wetting. This should be dry by now for metallization.

This cleaning procedure is generally adapted in this laboratory for recycling crystals. Possibly, one could avoid a few of the steps if we were certain that the substrates were precleaned. For example, fresh substrates from the manufacturer are precleaned. For those we could start from step 12 and continue. There are other methods which are used in semiconductor industries that could be adapted for large scale production purposes, and of course, there are methods used in other research and development laboratories that are equally good, but may use much better

equipment such as vapour degreasers, recirculating de-ionized water baths etc. However, with our limited facilities, the above mentioned process produced a reasonably high yield for most of our needs.

REFERENCES

[1] H.L. Kirice, "Frequency modulation - BBC field trials", BBC Quarterly, Vol. 1, July 1946, p. 62.

[2] D. Maurice, G.F. Newell, and J.G. Spencer, "Proposed test procedure for FM broadcast receiver", Electronic Engineering, Vol. 24, March 1952, p. 106.

[3] Lord Rayleigh, "On waves propagated along the plane surface of an elastic solid", Proc. London Math Soc., Vol. 17, 1885, pp. 4-11.

[4] E.A. Ash, "Realization of microwave circuit functions using acoustic waves", presented at IEEE GMTT International Microwave Symp., Boston, MA, May 1967.

[5] R.M. White and F.W. Voltmar, "Direct piezo-electric coupling to surface elastic waves", App. Phys. Lett., Vol. 7, 1965, pp. 314-316.

[6] L.B. Arguimbau. Vacuum Tube Circuits. New York: Wiley 1948, pp. 486-494.

[7] Simon Haykin. Communication Systems. New York: Wiley 1978.

[8] C.S. Hartmann, "Surface wave discriminator", U.S. Patent 3,750027, July 31, 1973.

[9] P. Hartemann and O. Menager, "Rayleigh wave frequency discriminator", Elect. Lett., Vol. 8, No. 8, April 1972, p. 214.

[10] P. Hartemann, "Surface acoustic wave phase and frequency discriminators", in Proc. International Specialist Seminar on Component Performance and Systems Application of S.A.W. Devices, (Aviemore, Scotland) Sept. 1973, pp. 152-166.

[11] P. Cross, W.H. Haydl and R.S. Smith, "Design and application of two port SAW resonators on Y-Z LiN_bO_3", Proc. IEEE, Vol. 64, No. 5, May 1976, pp. 682-685.

[12] A.D. Robertson, G.W. Hurley and J.R. Yeager, "Discriminator design uses acoustic delay lines", Microwave, July 1971, pp. 27-30.

[13] H. Taub and D.L. Schilling. Principles of Communication Systems. New York: McGraw Hill, 1971, p. 352.

[14] J.M. Wozecraft and I.M. Jacobs. Principles of Communication Engineering. New York: Wiley, 1965.

[15] M. Schwartz, W.R. Bennett and S. Stein. Communication Systems and Techniques. New York: McGraw-Hill, 1966.

[16] B. Arthurs and H. Dym, "On the optimum detection of digital signals in the presence of white Gaussian noise", I.R.E. Trans. Commun. Systems, CS-10, Dec. 1962, pp. 336-372.

[17] A.H. Nuttall, "Error probabilities for equi-correlated M-ary signals under phase coherent and phase incoherent reception", I.R.E. Trans. Inform. Theory, Vol. IT-8, July 1962, p. 308.

[18] J.D. Maines and E.G.S. Paige, "Surface acoustic wave device for signal processing applications", IEEG Proceedings, Vol. 64, No. 5, May 1976, pp. 639-653.

[19] J.D. Maines and E.G.S. Paige, "Surface acoustic components devices and applications", Proc. IEE Reviews, 120 (1973) pp. 1078-1110.

[20] E.A. Ash, R.M. da la Rue and R.F. Humphreys, "Microsound surface wave guides", IEEE Trans. Microwave Theory Tech., Vol. MTF17, Nov. 1969, pp. 882-892.

[21] A.A. Oliner, "Wave guides for acoustic surface waves", IEEE Proc., Vol. 64, No. 5, pp. 615-627.

[22] S. Nanayakkara and E.A. Ash, "Indirect coupling to topographic acoustic wave guides", Waves Electronics, Vol. 1, No. 4, Feb. 1976, pp. 247-263.

[23] M. Munasingha and G.W. Farnell, "Acoustic surface wave scattering on a homogeneous three quarter space", J. Appl. Phys., 44, 1973.

[24] H. Engan, "Excitation of elastic surface waves by spatial harmonics of interdigital transducer", IEEE Trans., ED-16, 1969, p. 1014.

[25] K.A. Ingebrigtsen, "Surface wave in piezo-electrics", J. App. Physics, 40, 1969, pp. 2681-2686.

[26] R.F. Milsom and M. Redwood, "Piezo-electric generation of surface waves by interdigital array; variational method analysis", Proc. IEE, 118, 1971, pp. 831-840.

[27] W.R. Smith, H.M. Gerard, J.H. Collins, T.M. Reader and H.J. Shaw, "Analysis of interdigital surface wave transducer by use of an equivalent circuit model", IEEE Trans. Microwave Theory Tech., MTT 17, 1969, pp. 856-864.

[28] W.P. Mason. Electromechanical Transducers and Wave Filters. 2nd Ed., Van Nostrand, New York, 1948, pp. 200-399.

[29] G.W. Farnell, I.A. Cermak, P. Silvester, and S.K. Wong, "Capacitance and field distributions for interdigital surface wave transducers", IEEE Trans. Sonics Ultrasonics, SU-17, 1970, pp. 188-195.

[30] H.E. Kallman, "Transversal filters", Proc. IRE 28, 302 (1940).

[31] R. Tancrell and M. Holland, "Acoustic surface wave filters", Proc. IEEE, 59, 1971, p. 393.

[32] S. Nanayakkara and C.K. Campbell, "A Technique for modifying the aperture of an interdigital transducer finger pattern for a surface acoustic wave device", Thin Solid Films, 46, 1977, pp. 75-80.

[33] R.H. Tancrell, "Analysis design of band pass filters", IEEE Trans. on Sonics and Ultra Sonics, Su-21, No. 1, Jan. 1974, pp. 12-22.

[34] C.S. Hartmann, "Weighting interdigital surface wave transducers by selective finger withdrawal of electrodes", in Proc. IEEE Ultrasonics Symposium, 1973, pp. 423-426.

[35] J.P. Reilly, C.K. Campbell and M. Suthers, "The design of SAW bandpass filters exhibiting arbitrary phase and amplitude response characteristics", IEE Trans. Sonics and Ultrasonics, Vol. SU-24, No. 5, Sept. 1977.

[36] R.L. Miller and A.J. DeVries, "A simple building block method for design of SAW filters having nonlinear phase response", in IEEE Ultrasonics Symposium Proceedings, 1979, IEEE Cat. #76CH1120-SSU.

[37] H. Engan, "Series-Weighting of surface acoustic wave transducers", IEEE Ultrasonic Symposium Proceedings, 1974, IEEE Cat. #74CH0896-ISU.

[38] P.B. Snow in H. Matthews, eds. Surface Wave Filters. Chapter 5, New York: Wiley 1977.

[39] T.W. Bristol, W.R. Jones, P.B. Snow and W.R. Smith, "Application of double electrodes in surface surface wave device design", in Proc. IEEE Ultrasonics Symposium, 1972, pp. 343-345.

[40] V.M. Ristic, "Bulk mode generation in surface wave devices", in Proc. IEEE Ultra Sonic Symposium, 1972, pp. 424-428.

[41] R. La Ross and C.F. Vasile, "Broad band cancellation in acoustic surface wave devices", Electron Lett., 8, 1972, p. 478.

[42a] F.G. Marshall and E.G. Paige, "Novel acoustic surface wave directional coupler with diverse applications", Elect. Lett., 7, 1971. pp. 460-462.

[42b] F.G. Marshall, C.O. Newton, and E.G.S. Paige, "Surface acoustic wave multistrip components and their applications", IEEE Trans. Sonics Ultrasonics, SU-20, 1973, pp. 134-142.

[43] I.M. Mason and E.A. Ash, "Acoustic surface wave beam diffraction on isotropic substrate", J. Appl. Physics, 42, 1971, pp. 5343-5351.

[44] G.W. Farnell and E.L. Adler, "An overview of acoustic surface wave technology", Final Report to Communication Research Center on DSS Contract 36001-3-4406 (1974).

[45] W.R. Smith, H.M. Gerard and W.R. Jones, "Analysis and design of dispersive interdigital surface wave transducers", IEEE Trans. Microwave Theory Tech., MTT-20, 1972, pp. 485-471.

[46] A.J. Slobodnik, Jr., "UHF and microwave delay lines: Design", Air Force Cambridge Research Lab. Report No. AFCRL-TR-73-0538 (1973).

[47] F.G. Marshall, C.O. Newton, and E.G.S. Paige, "Theory and design of the surface wave multistrip coupler", IEEE Trans. Microwave Theory Tech., MTT-21, 1973, pp. 206-215.

[48] S. Nanayakkara and C.K. Campbell, "Method for frequency to voltage conversion using a surface wave device", IEEE Trans. on Sonics and Ultrasonics, (in press).

[49] S. Nanayakkara and C.K. Campbell, "A technique for modifying the aperture of an interdigital transducer finger pattern for a surface acoustic wave device", Thin Solid Films, 46, 1977, pp. 75-80.

[50] A.J. Slobodnik, "The effect of surface quality on acoustic surface wave attenuation on Yz $LiNbO_3$" in Proc. IEEE Ultrasonic Symposium, 1973, pp. 369-371.

[51] R.W. Dietz and J.M. Bennett, "Bowl feed technique for producing super smooth optical surfaces", Appl. Opt., 5, 1961, p. 881.

[52] H.I. Smith, F.J. Bachner and N. Efremow, "A high yield photolithographic technique for surface wave devices", J. Electrochem. Soc., 118, 1971, p. 821.

[53] H. Seidel and D.L. White, "Ultra sonic surface waveguides", U.S. Patent 3406358, Oct. 15, 1968.

[54] F.S. Hickernell, "The role of layered structures in surface acoustic wave technology", in Proc. Conf. Component Performance and Systems Applications of Surface Acoustic Wave Devices (Aviemore, Scotland). Sept. 1973, pp. 11-21.

[55] T. Sniosaki, "High-speed fabrication of high quality sputtered Z_nO thin-films for bulk and surface wave applications", in Ultrasonic Symposium Proceedings, 1978, IEEE Cat. No. 78CH1344-ISU.

[56] F.S. Hickernell, "Zinc oxide thin film-surface wave transducers", Proc. IEEE, Vol. 64, No. 5, May 1976.

[57] A.J. Shuskus, T.M. Reeder and E.L. Paradis, "R.F.-sputtered aluminium nitride films on sapphire", Appl.

Phy. Lett., 24, 1974, pp. 155-156.

[58] H. Koops, "On electron projection systems", J. Vac. Sci. Technol., 10, 1973, p. 909.

[59] W.R. Livesay, "Integrated circuit production with electron beam", J. Vac. Sci. Technol., 10, 1973, pp. 1028-1032.

[60] L.F. Thompson, E.D. Feit, M.J. Bowden, P.V. Lenzo and E.G. Spencer, "Polymeric resists for x-ray lithography", J. Electro. Chem. Soc., 121, 1974, p. 1500.

[61] S. Nanayakkara, "A method of fabricating silicon mask for x-ray lithography using anisotropic chemical etches", U.C.L. Technology Report, E.P.I., 1973.

[62] S. Nanayakkara. Fabrication of Acoustic Surface Wave Guides Using Chemical Etching. Post Grad. Dip. Thesis, University College, London, 1973.

[63] H.I. Smith, R.C. Williamson and W.T. Brogan, "Ion beam etching of reflective array filters", in Proc. IEEE Ultrasonic Symposium, 1972.

[64] J.R. Klauder, A.C. Price, S. Darlington and W.J. Albersheim, "The theory and design of chirp radars", Bell Sys. J., July 1960, pp. 745-820.

[65] R.C. Dixon. Spread Spectrum Systems. New York: Wiley 1976.

[66] D.T. Bell, Jr., J.D. Holmes, and R.V. Riddings, "Application of surface wave technology of spread spectrum communication", IEEE Trans. Microwave Theory Tech., MTT-21, 263 1973.

[67] W.C. Fifer, et al., "Switchable acoustic matched filter", Final Report F30602-71-C-0199, Rome Air Defence Command, Rome, New York, 1972.

[68] J.H. Collins and P.M. Branth, "The role of surface acoustic wave technology in communication", Ultrasonics 10, 1972, p. 57.

[69] G.S. Kino and H. Matthews, "Signal processing in acoustic surface wave devices", IEEE Spectrum Vol. 8, No. 22, 1971.

[70] C.B. Feldman and W.R. Bennett, "Band width and transmission performance", B.S.T.J., 28, p. 490.

[71] S. Nanayakkara and C.K. Campbell, "Evaluation of signal and noise performance of a sampling wideband SAW FM discriminators", in Proc. Ultra Sonic Symp., 1979.

[72] B.R. Potter and C.S. Hartmann, "Surface acoustic wave slanted correlators for linear FM pulse compressors", in Ultrasonic Symposium Proceedings, 1977, IEEE Cat. No. 77CH1264-ISU, pp. 607-610.

[73] R.C. Williamson, "Properties and application of reflective array devices", Proc. IEEE, Vol. 64, No. 5, May 1976, pp. 702-710.

[74] E.G.S. Paige, "Surface acoustic wave devices: A transatlantic view", in Ultrasonics Symposium Proceedings, 1977, IEEE Cat. No. 77 CH1264-ISU, pp. 563-567.

[75] R.D. Weglein, "SAW chirp filter performance above 1 GHz", Proc. IEEE, Vol. 64, No. 5, May 1976, pp. 695-698.