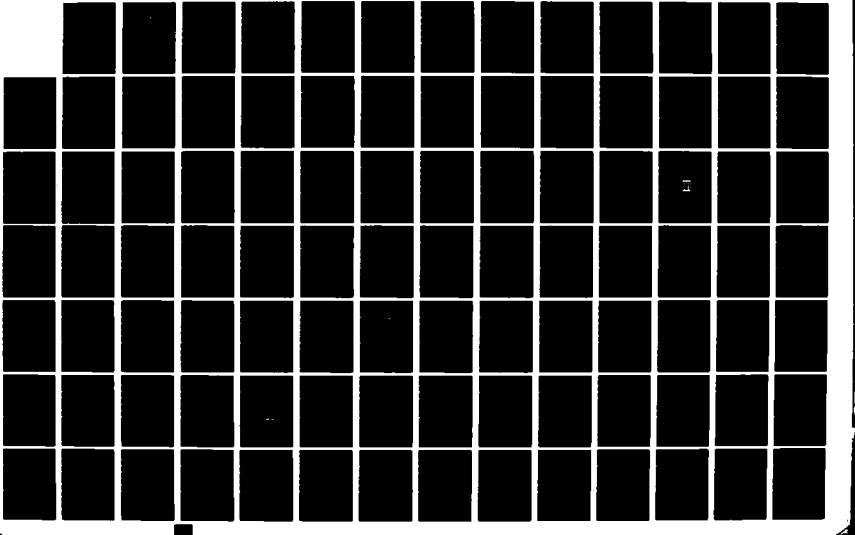
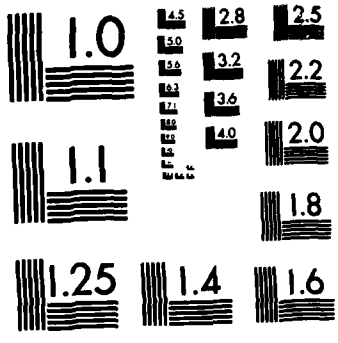
AD-A127 845 MULTIFUNCTION SPREAD SPECTRUM TECHNOLOGY(U) NAVAL AIR 1/2
DEVELOPMENT CENTER WARMINSTER PA COMMUNICATION
NAVIGATION TECHNOLOGY DIRECTORATE Y LUI 04 OCT 82
UNCLASSIFIED NADC-82188-40 F/G 17/2 NL





12

REPORT NO. NADC-82188-40

MULTIFUNCTION SPREAD SPECTRUM TECHNOLOGY

Yau-Foon Lui
Communication Navigation Technology Directorate
NAVAL AIR DEVELOPMENT CENTER
Warminster, Pennsylvania 18974

FINAL REPORT

Airtask No. F66212001
Work Unit No. ED212

Approved for Public Release; Distribution Unlimited

Prepared for
NAVAL AIR SYSTEMS COMMAND
Department of the Navy
Washington, D.C. 20361

NOTICES

REPORT NUMBERING SYSTEM - The numbering of technical project reports issued by the Naval Air Development Center is arranged for specific identification purposes. Each number consists of the Center acronym, the calendar year in which the number was assigned, the sequence number of the report within the specific calendar year, and the official 2-digit correspondence code of the Command Office or the Functional Directorate responsible for the report. For example: Report No. NADC-78015-20 indicates the fifteenth Center report for the year 1978, and prepared by the Systems Directorate. The numerical codes are as follows:

CODE	OFFICE OR DIRECTORATE
00	Commander, Naval Air Development Center
01	Technical Director, Naval Air Development Center
02	Comptroller
10	Directorate Command Projects
20	Systems Directorate
30	Sensors & Avionics Technology Directorate
40	Communication & Navigation Technology Directorate
50	Software Computer Directorate
60	Aircraft & Crew Systems Technology Directorate
70	Planning Assessment Resources
80	Engineering Support Group

PRODUCT ENDORSEMENT - The discussion or instructions concerning commercial products herein do not constitute an endorsement by the Government nor do they convey or imply the license or right to use such products.

APPROVED BY: William F. Lyons Jr. DATE: 4 October 1982

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NADC-82188-40	2. GOVT ACCESSION NO. AD-A122 045	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* Multifunction Spread Spectrum Technology		5. TYPE OF REPORT & PERIOD COVERED Final Report 6. PERFORMING ORG. REPORT NUMBER –
7. AUTHOR(s) Yau-Foon Lui		8. CONTRACT OR GRANT NUMBER(s) –
9. PERFORMING ORGANIZATION NAME AND ADDRESS Communication Navigation Technology Directorate Code 40, Naval Air Development Center Warminster, PA 18974		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Airtask Number 2. F66212001 Work Unit No. ED 212
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Air Systems Command Department of The Navy Washington, DC 20361		12. REPORT DATE – 4 OCTOBER 1982 13. NUMBER OF PAGES –
14. MONITORING AGENCY NAME & ADDRESS *(if different from Controlling Office)* –		15. SECURITY CLASS. *(of this report)* UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE –

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for Public Release; Distribution Unlimited

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

–

18. SUPPLEMENTARY NOTES

–

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

Communications
Spread Spectrum
Correlation
Signal Processing

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

The techniques for developing a multifunction modulator/demodulator for processing spread spectrum waveforms are investigated. Based on a trade-off analysis of techniques such as CCDs, SAWs, digital and optical correlators, the best technical approach is determined.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF TABLES

SUMMARY

With the widespread use of spread spectrum techniques in the military, the need to determine the feasibility of using a common demodulator (correlator) for processing various signals to reduce the logistics, size, and weight of using different receivers becomes inevitable. Signals such as GPS, JTIDS, HFIC, UHF ECCM, and SINCGARS are analyzed to determine their commonalities in waveform structure for processing.

Consideration of technologies such as surface acoustic wave devices, acousto-optical devices, charge-coupled devices, and digital devices indicated that it is possible to process most of these waveforms using a programmable correlator. The convolver approach is ideal in the sense that it can be operated at RF/IF with no restriction to the kind of waveform used. The technology is relatively new and presents a risky approach.

Like digital circuits, charge coupled devices are baseband devices. The technology is more mature than that of the convolver and is therefore less risky. It has an advantage over a digital correlator because no quantization or multichannel correlations are required. In any case, digital technology is the most advanced technology and as such it is the least risky approach.

BACKGROUND

The optimum communication system for military application is one that is secure, has low probability of intercept or high interference rejection. The systems that satisfy these requirements are well known to be spread spectrum systems. Due to technology limitations, spread spectrum communication systems were not widely used. With the advance in signal processing techniques in recent years, wideband spectrum systems are becoming more widespread and reached a degree of sophistication that was never achieved before. A look at the military developmental communication systems definitely indicated this trend. Systems such as the Single Channel Ground and Airborne Radio System (SINCGARS), Time Division Multiple Access (TDMA) and Distributed Time Division Multiple Access (DTDMA) Joint Tactical Information Exchange System (JTIDS), Global Positioning System (GPS), Ultra-High-Frequency Electronic Counter Counter Measure (UHF ECCM) system, High Frequency Intra-Task-Force Communication (HFIC) system, all utilize some form of spread spectrum techniques for signalling.

Table I shows the operational requirements of the naval aircraft for these developmental communication systems. The addition of each function to an aircraft requires the addition of a modem

Table I. Operational Requirements for Naval Air Platforms

Platforms	Operational Requirements	Platforms	Operational Requirements
E-2C	UHF ECCM JTIDS GPS HFIC SINCGARS	A-6E	JTIDS UHF ECCM SINCGARS GPS
F/A-18	JTIDS UHF ECCM SINCGARS		

because these are independent developments that are not interoperable. The addition of several functions to an aircraft represents a significant increase in the weight and space requirements of the existing aircraft which is undesirable since these aircraft already reached their weight and volume limits.

The alleviation of this problem is to design a modem that is capable of processing all these waveforms. The objective of this IED is, therefore, to investigate the techniques used in signal processing for the development of a multifunction modulator/demodulator for spread spectrum communication for JTIDS, GPS, SINCGARS, UHF ECCM, and HFIC.

SPREAD SPECTRUM SYSTEMS

A spread spectrum system can be defined to be a system in which the transmitted signal is spread over a wide frequency band using some form of modulation. The bandwidth is much wider, in fact, than the minimum bandwidth required to transmit the information being sent using the same modulation. An amplitude modulated (AM) voice signal, for example, can be transmitted with twice the information bandwidth using double sideband modulation. The distribution of this same information over an extremely large bandwidth (e.g. megahertz) by encoding this signal with another wideband signal before modulating the carrier produces a spread spectrum. It should be emphasized that the encoding of the baseband signal with a wideband signal is a requirement for producing the spread spectrum. Frequency modulation (FM) systems with large deviation ratios, or phase shift keying (PSK) systems are not wideband systems according to this definition although the bandwidths are relatively large.

In general, the signal level has to be higher than the noise level at the receiving antenna for good reception of the transmitted signal unless somehow, an increase in signal-to-noise ratio can be obtained. This is precisely what a spread spectrum provides. In the course of processing the wideband spread spectrum, the signal-to-noise ratio at the receiver output is increased depending on the degree of spreading of the signal. The difference in output and input signal-to-noise ratio in any processor is its "processing gain". A processing element that produces an output signal-to-noise ratio of 20 db with an input signal-to-noise ratio of 10 db is said to have a processing gain of 10 db. In a spread spectrum system, the processing gain can be estimated by the ratio of the transmitted RF bandwidth and the baseband information bandwidth. Therefore, the wider the RF bandwidth, for a given baseband information bandwidth, the higher the processing gain. With no jamming power, the RF signal level can be reduced by the processing gain to end up with the same output signal-to-noise ratio as in the conventional approach. If the same signal is transmitted using relatively high power, interference rejection can be readily accomplished.

Since the wideband spectrum are obtained by encoding the baseband signal with a wideband signal, an enemy cannot usually listen to messages being sent. The system is by no means secure since the wideband signal may not be secure. By the same token, efforts have to be made to decode the message if the wideband signal is not known.

Because of the ease of generation of the spread spectrum by using digital data, it is almost universal that spread spectrum systems are digital systems. Being digital systems, more than one code can be transmitted on the same frequency provided the codes have low correlation between one another. Therefore, the same bandwidth can be used to accommodate many users by means of code division multiplexing techniques. This feature of the spread spectrum systems may compensate for the larger bandwidths used.

WAVEFORM STRUCTURE

Tables II through VI show the structures of the military waveforms under development as of March, 1980 and may have slightly changed since then. Table VII is a summary for these waveforms. SINCGARS, which is an analog FM system and UHF ECCM are not wideband waveforms according to the present definition since the baseband information are not encoded in any way. With the exception of these two waveforms, it can be seen that only two types of modulations are used – Minimum Shift Keying (MSK) and Binary Phase Shift Keying (BPSK). A brief description of these two waveforms are necessary if we were to understand the signal processing aspect of these signals.

Table II. JTIDS Signal Structure

Minimum Shift Keying (MSK) Modulation
Frequency hopping
Time hopping (DTDMA)
Chip rate: 5 MHz
Code length: 32 bit cyclically shift keying
Frequency band: Lx band (960 to 1215 MHz)
Number of channels: 51
Channel spacing: 3 MHz

Table III. GPS Signal Structure

Composite Binary Phase Shift Keying (BPSK) Modulation
No frequency hopping
No time hopping
Chip rates: C/A code – 1.023 MHz
P code – 10.23 MHz
Code lengths: C/A code – 1 msec
P code – 1 week
Frequency band: Lx band (960 – 1215 MHz)
Number of channels: 2 – 1575.42 MHz, 1227.60 MHz

Table IV. HFIC Signal Structure

Minimum Shift Keying (MSK) Modulation
Frequency hopping
No time hopping
Chip rate: ~ 70 kHz
Code lengths: 3 variable code lengths less than 70 bits each
Frequency band: HF band (2 to 30 MHz)
Number of channels: 1000
Channel spacing: 1 KHz

Table V. SINCGARS Signal Structure

Frequency Modulation (FM)
Frequency hopping
No time hopping
Chip rate: no digital data
Code length: not applicable
Frequency band: VHF band (30 to 88 MHz)
Channel spacing: 25 KHz
Number of channels: 2320

Table VI. UHF ECCM Signal Structure

Minimum Shift Keying (MSK) Modulation
Frequency hopping
No time hopping
Data rate: ~ 60 kHz
Code length: no PN spreading
Frequency bands: UHF (225 to 299.975 MHz) UVHF (118 to 173.975 MHz) LVHF (30 to 87.975 MHz)
Number of channels: UHF – 3000 UVHF – 2240 LVHF – 2320
Channel spacing: 25 KHz

Table VII. Waveform Summaries

Waveform	Modulation	PN Spreading	Frequency Hopping	Time Hopping	Chip Rate
DTDMA JTIDS	MSK	yes	yes	yes	5 MHz
TDMA JTIDS	MSK	yes	yes	no	5 MHz
GPS	BPSK	yes	no	no	1.023 MHz 10.23 MHz
SINCGARS	FM	no	yes	no	Not Applicable
HFIC	MSK	yes	yes	no	~ 70 kHz
UHF ECCM	MSK	no	yes	no	~ 60 kHz

BINARY PHASE SHIFT KEYING (BPSK)

If we let ω_c to be the transmitted carrier frequency, then a binary phase shift keying signal can be represented as follows:

$$f(t) = \cos(\omega_c t + \theta + \alpha)$$

where θ is an arbritrary phase angle of the carrier and $\alpha = 0°$ or $180°$ corresponding to a data of value 1 or 0 respectively.

Since $\cos(\omega_c t + \theta + 0°) = \cos(\omega_c t + \theta)$

and $\cos(\omega_c t + \theta + 180°) = -\cos(\omega_c t + \theta)$

f (t) can be rewritten as follows:

$$f(t) = \pm 1 \cos(\omega_c t + \theta)$$

If a doubly balanced mixer were to be used as a modulator, the digital data which has voltage values of V_h and V_ℓ volts respectively have to be adjusted to bipolar voltages symmetrical about 0 volt to cause a change in the carrier phase and produce a constant amplitude output. Assuming that the data are normalized to a ±1 level, f(t) can then be written as, $f(t) = D \cos(\omega_c t + \theta)$ where D is the data at a ±1 level. This waveform can be generated simply by using a mixer as shown in figure 1. The BPSK signal has a power spectral density as shown in figure 2. Figure 3 shows the fractional out of band power for BPSK and is an indication of the bandwidth efficiency of the BPSK waveform. The horizontal axis is a bandwidth scale normalized to the data rate with the carrier frequency at the origin. A bandwidth of two based on this normalization is equivalent to 4 MHz if the data rate is assumed to be 2 MHz. The bandwidth, being symmetrical to the carrier frequency, is 2 MHz wide on both sides of the carrier frequency. The out of band power is indicated by the vertical scale in dB. Assuming the same data rate of 2 MHz, it can be seen that the out of band power with a bandwidth of 2 (4 MHz) is approximately −10 dB which is computed to be 0.1. Since the main lobe of the BPSK waveform has a normalized bandwidth of 2, the main lobe therefore contains 0.9 of the total energy. The probability of error for a BPSK signal is shown in figure 4.

MINIMUM SHIFT KEYING (MSK)

The MSK waveform is a particular case of frequency shift keying (FSK) in which a logical one is represented by one frequency F_H and a logical zero is represented by another frequency F_L. The MSK waveform, in addition to the requirements of a FSK signal, has the following properties:

1. The frequency deviation Δf is exactly $\pm(1/4)\, f_d$ from the rf carrier frequency where f_d is the data rate.

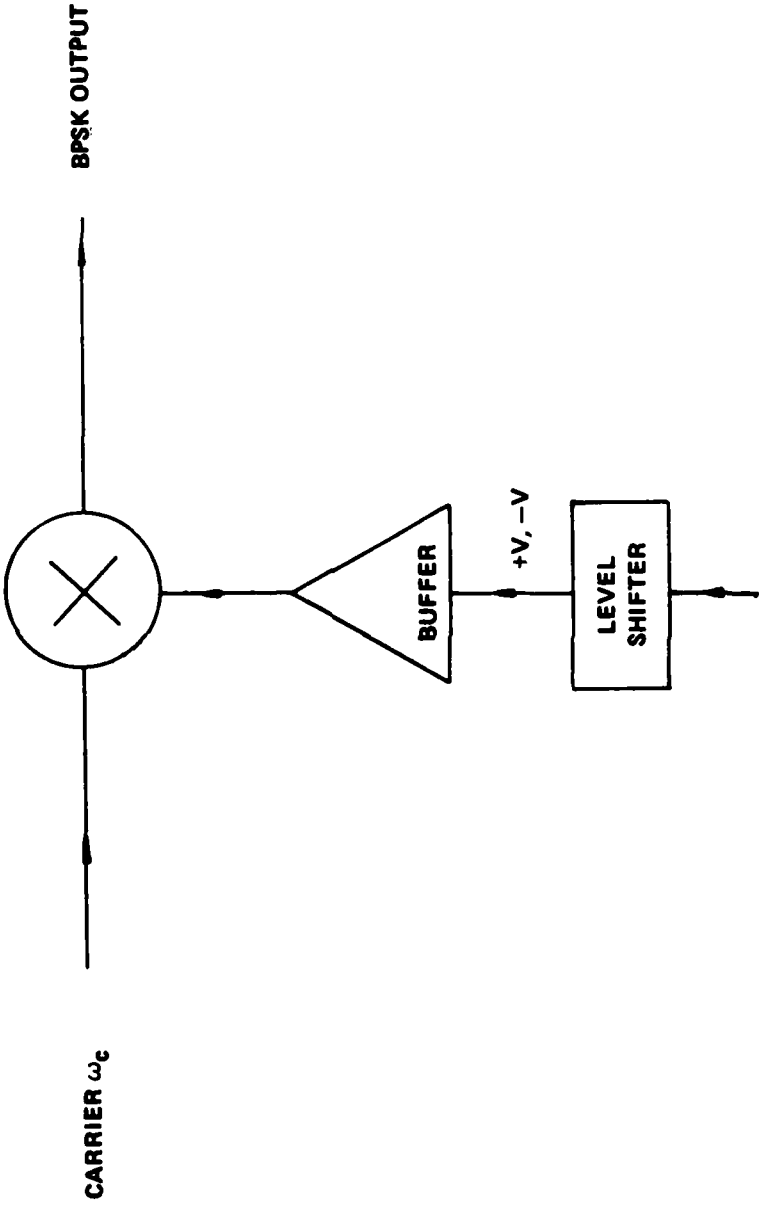


Figure 1. A BPSK Modulator

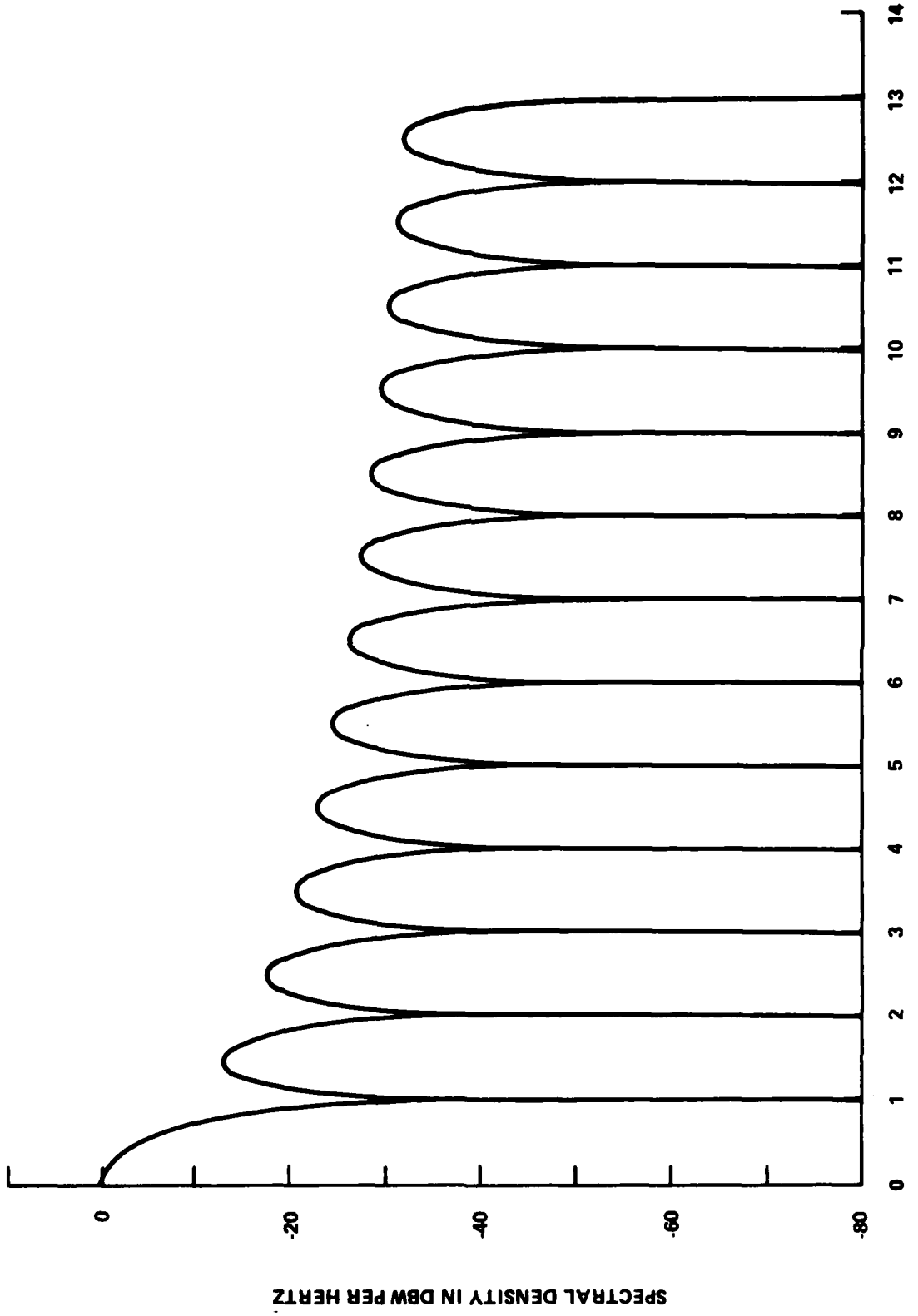


Figure 2. Spectral Density of BPSK

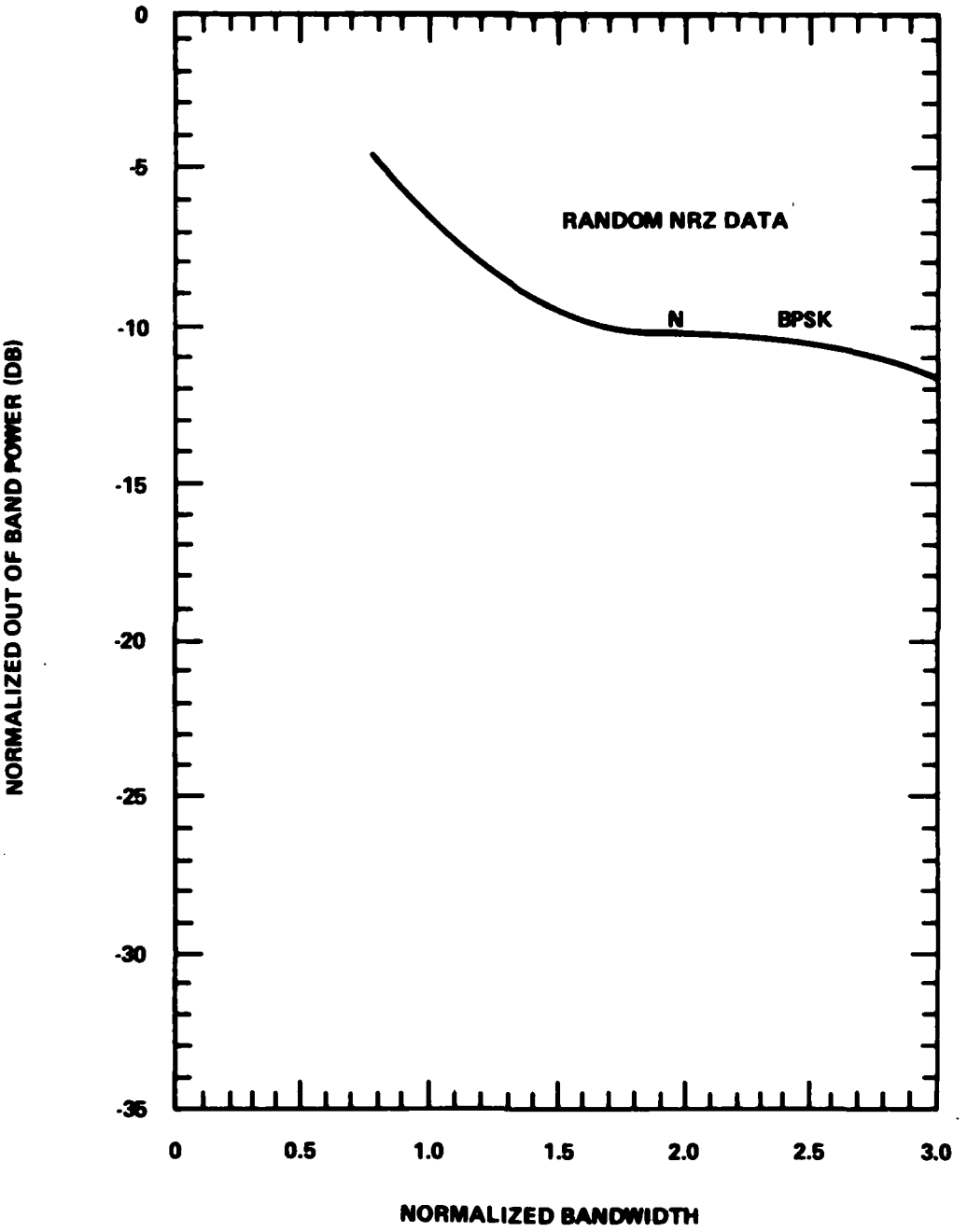


Figure 3. Power Outside of Normalized Channel Bandwidth for BPSK

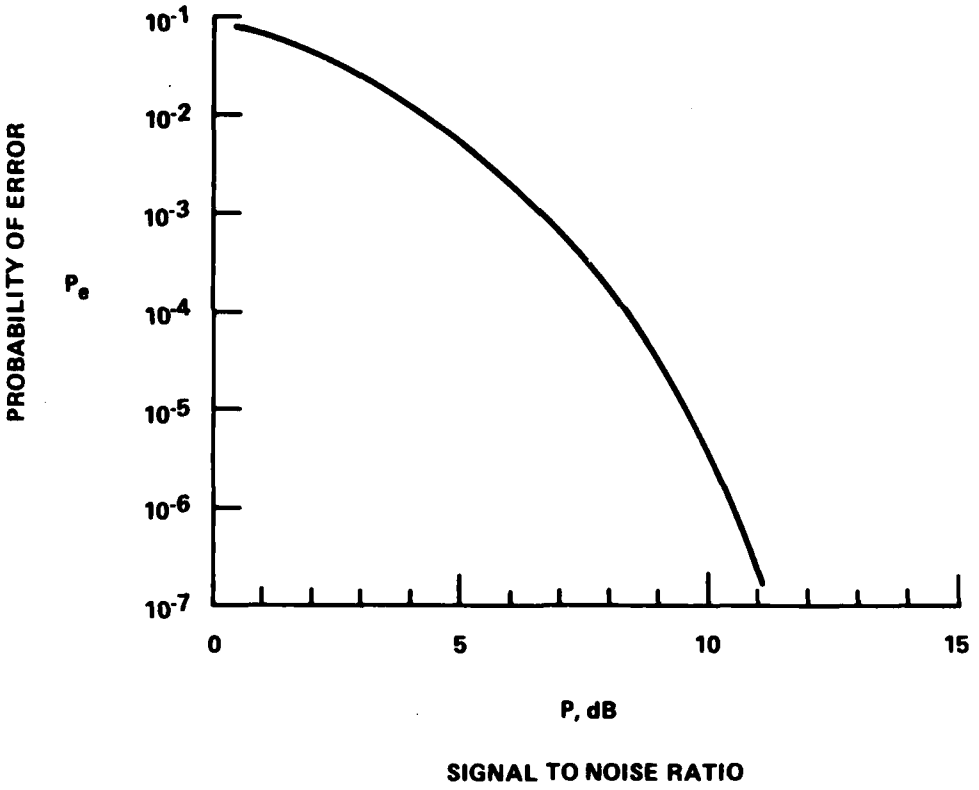


Figure 4. Error Probability for Coherently Detected BPSK Signal

2. The RF phase varies linearly exactly ±90° with respect to the carrier during each data period $T = \frac{1}{f_d}$.

3. There is phase continuity of the modulated rf carrier at the interbit switching instant and consequently phase is continuous at all times.

4. The envelope of the signal is constant at all times.

5. The data are differentially encoded so that phase ambiguity in the receiver can be resolved without the necessity of transmitting a known block of data within the messages.

The waveform for a MSK signal in the time domain is shown in figure 5.

Mathematically, a FSK signal can be written as $\pm \cos (A \pm B)$ where $A = \omega_c t + \theta$, $B = \omega_m t + \alpha$. ω_c, ω_m are respectively the carrier and modulating frequencies with phase angles θ and α.

The requirements for a FSK signal to be a MSK signal is that of phase continuity and the differential encoding of baseband data. If the cosine function is expanded, the following set of equations are obtained.

$$+\cos (A-B) = +\cos A \cos B + \sin A \sin B$$

$$+\cos (A+B) = +\cos A \cos B - \sin A \sin B$$

$$-\cos (A-B) = -\cos A \cos B - \sin A \sin B$$

$$-\cos (A+B) = -\cos A \cos B + \sin A \sin B$$

The waveform can be seen to be composed of two signals. They are cos A cos B and sin A sin B. Each of these two signals can be looked upon as an amplitude modulated signal with carrier A and modulating signal B. These two signals can be obtained by using doubly balanced mixers. These are conventional AM signals with zero DC bias and 100 percent modulation. Notice that the second term has both the carrier frequency as well as the modulating frequency 90 degrees out of phase of the respective components in the first term. Notice also from the four expanded equations that if the signs for the two terms are the same, the difference frequency (A-B) is obtained while different signs for the two terms produce the sum frequency (A+B). Since the baseband data provides the information for differential encoding and differentially encoded data in turn determines the frequency that has to be transmitted, the signs for the two terms have to be determined by the data. The rules for differentially encoding the data are as follows:

1. If $X(n) = X(n-1)$, the lower frequency is transmitted.

2. If $X(n) \neq X(n-1)$, the higher frequency is transmitted.

The following seven bit Barker code can be used as an example:

Barker Code: 1110010

Differentially encoding the Barker code according to the above rules yields the following code:

001011

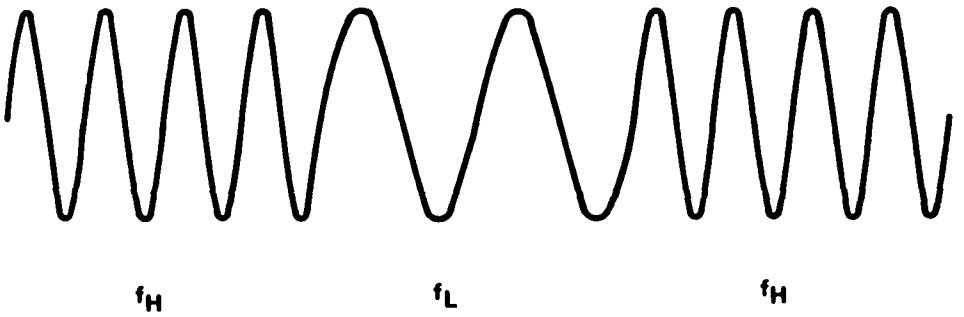


Figure 5. An MSK Signal in the Time Domain.

Notice that a seven bit code becomes a six bit code after encoding. The upper and lower frequencies will be transmitted accordingly as follows:

$$f_L, f_L, f_H, f_L, f_H, f_H$$

where f_L is the lower frequency and f_H is the higher frequency.

As in the BPSK case, if the logical levels for the data are normalized to a +1 or -1 level, the data can be applied alternately to the two quadrature components that make up the signal. The first, third, fifth data bits are multiplied with the cos A cos B term while the second, fourth, sixth data bits are multiplied with the sin A sin B term. As can be seen from the equations, two adjacent bits of the same value will produce the lower frequency while the higher frequency is produced if the bits are different.

It should be noted that each of the transmitted frequencies, f_H, f_L are determined by two consecutive bits — the present bit and a previous bit. Therefore, the duration of the data of each channel is twice as long as the period for the individual data. This can be seen more clearly by using the same Barker code as an example:

Barker Code: 1110010

As shift in reference level yields the following:

+1 +1 +1 -1 -1 +1 -1

The first bit is now applied to the cos A cos B term while the second bit is applied to the sin A sin B term yielding the first of the set of four equations. This is the lower frequency since: + cos (A - B) = + cos A cos B + sin A sin B. The third bit, which is also +1, is now applied to the first term as stated earlier. The data to the second term remains unchanged during this time interval since this information is required to determine the frequency to be transmitted. The output in this case is also f_L. The fourth bit is -1 and is applied to the second term with the first term now remaining fixed. This will provide the fourth equation which is the high frequency. The fifth bit is -1 and with this bit applied to the first term, the third equation is obtained. This is the low frequency. Similarly, with the application of +1, -1, the second equation was used twice providing the last two periods with the high frequency. This is shown as follows:

Real	Quad
cos A cos B (+1) + sin A sin B (+1)	$= + f_L$
cos A cos B (+1) + sin A sin B (+1)	$= + f_L$
cos A cos B (+1) - sin A sin B (-1)	$= + f_H$
- cos A cos B (-1) - sin A sin B (-1)	$= - f_L$
- cos A cos B (-1) + sin A sin B (+1)	$= - f_H$
- cos A cos B (-1) + sin A sin B (+1)	$= - f_H$

Assuming that both terms of the MSK signal have opposite signs, the MSK signal can be shown in the frequency plane in figure 6.

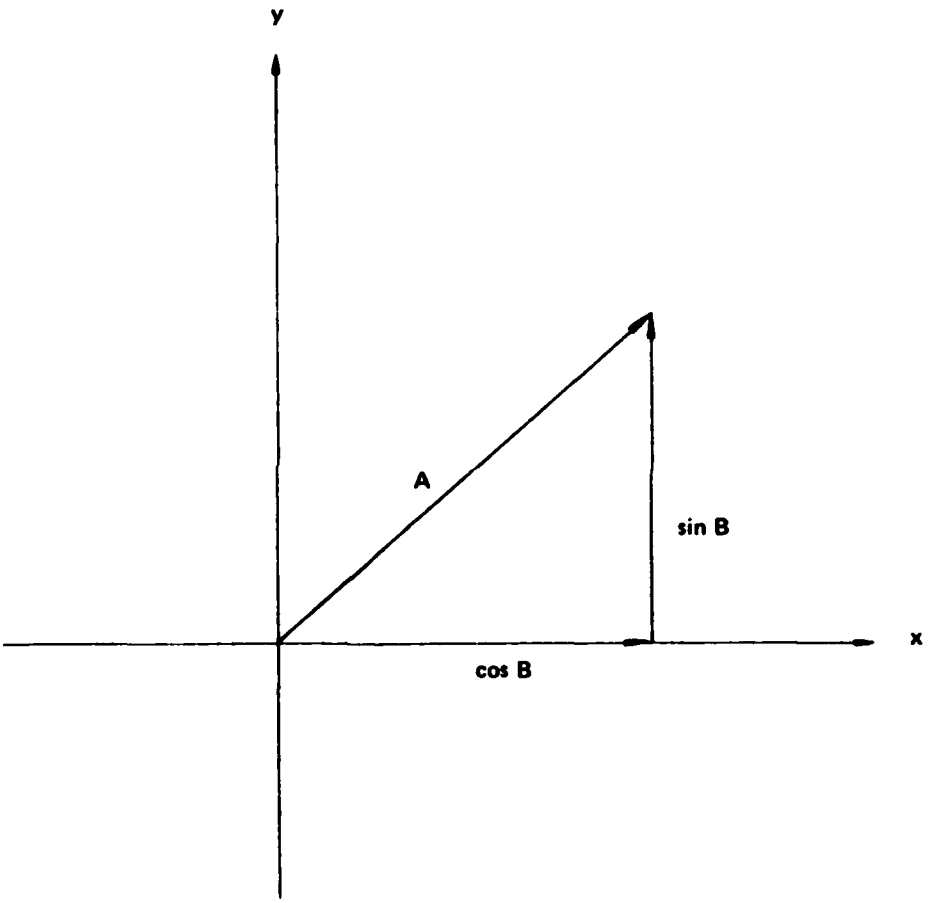


Figure 6. Vector Diagram for an MSK Signal in the Frequency Plane

The amplitude of the resultant vector is of constant amplitude because $|Z| = \sqrt{\cos^2 B + \sin^2 B} = 1$. Since sin B is at its maximum when cos B is at its minimum (and vice versa), switching can be done to the channel that reaches its minimum. The resultant phase is continuous since a vector of zero amplitude makes no contribution to the phase. To synchronize the switching of the data and the zero crossing of the modulating frequency, the modulating sinusoid is usually obtained from the timing circuit that clocks the data out. Figure 7 shows a block diagram for a MSK modulator with the timing circuit.

The 90° phase shift per period requirement can be satisfied by choosing the correct modulating frequency as follows:

Let f_C be the carrier frequency,

f_M be the modulating frequency,

f_L be the low frequency,

f_H be the high frequency.

For a 90° phase shift at each chip duration, the following equations have to be satisfied:

$$(f_C - f_L)\frac{1}{f_M} = \frac{1}{4}$$

$$(f_H - f_C)\frac{1}{f_M} = \frac{1}{4}$$

where $|f_C - f_L| = |f_H - f_C|$ is the frequency deviation between the carrier frequency and the high and low frequencies; $\frac{1}{f_M}$ is the period of the data. Therefore,

$$f_C - f_L = f_M/4$$

and $$f_H - f_C = f_M/4$$

If the modulating frequency is chosen, then the lower and higher frequencies can be determined. There is no requirement that a particular number of cycles be transmitted per chip. However, the relationship between the deviation and the chip width requires the number of cycles in the higher frequency chip to be exactly 1/2 more than the number in the lower chip. The difference between the two frequencies is $\frac{1}{2T}$, which is 2.5 MHz for 200 nanosecond chips.

Figures 8 and 9 respectively show the power spectral density and the power outside of normalized channel bandwidth for a MSK signal with random data.

BIT ERROR RATE AND BANDWIDTH COMPARISONS FOR BPSK AND MSK

In any data transmission system, the one most important element in its evaluation is the bit error rate. It was shown in various texts that coherently detected binary phase shift keying has the best bit error rate for a given signal-to-noise ratio of the receiver input.

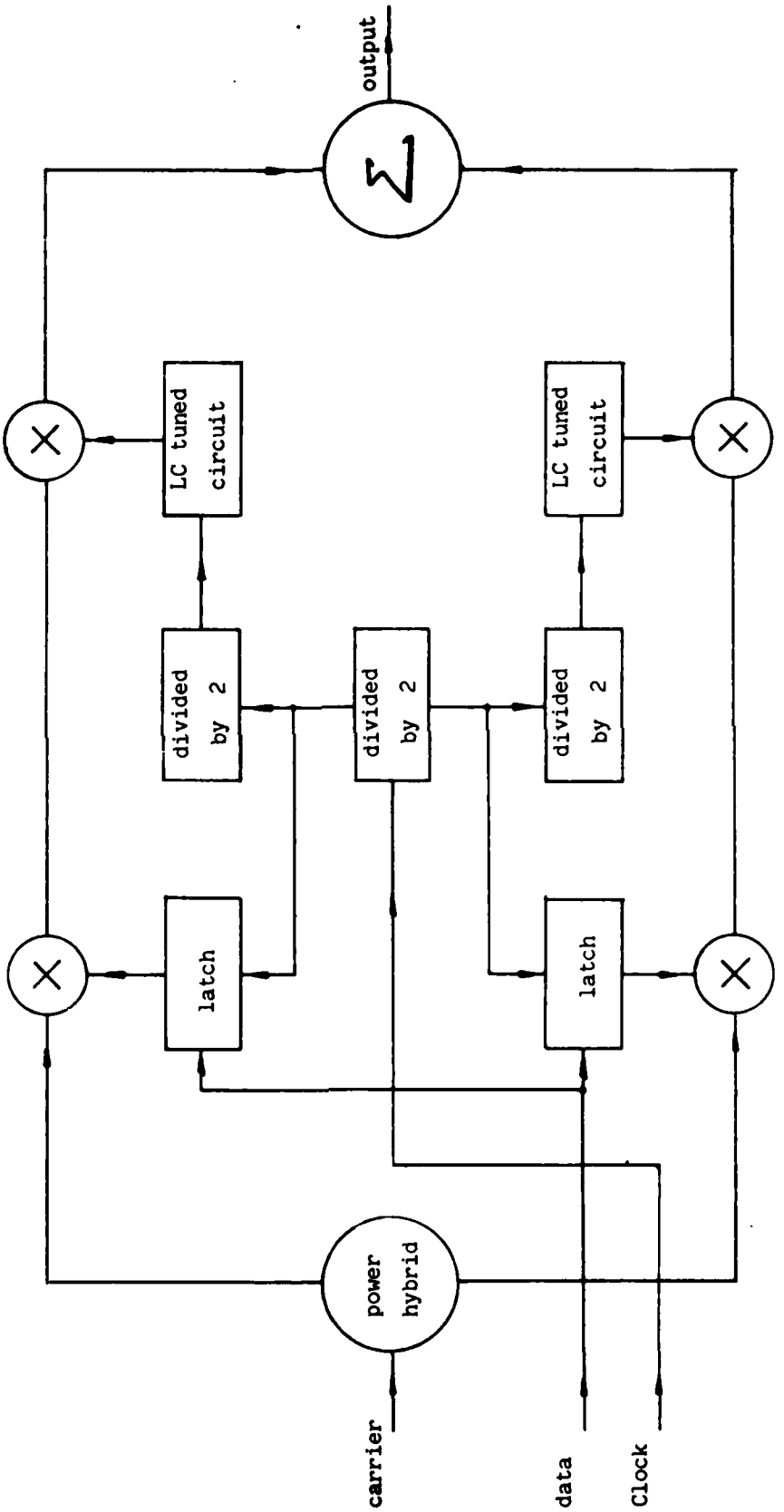


Figure 7. An MSK Modulator.

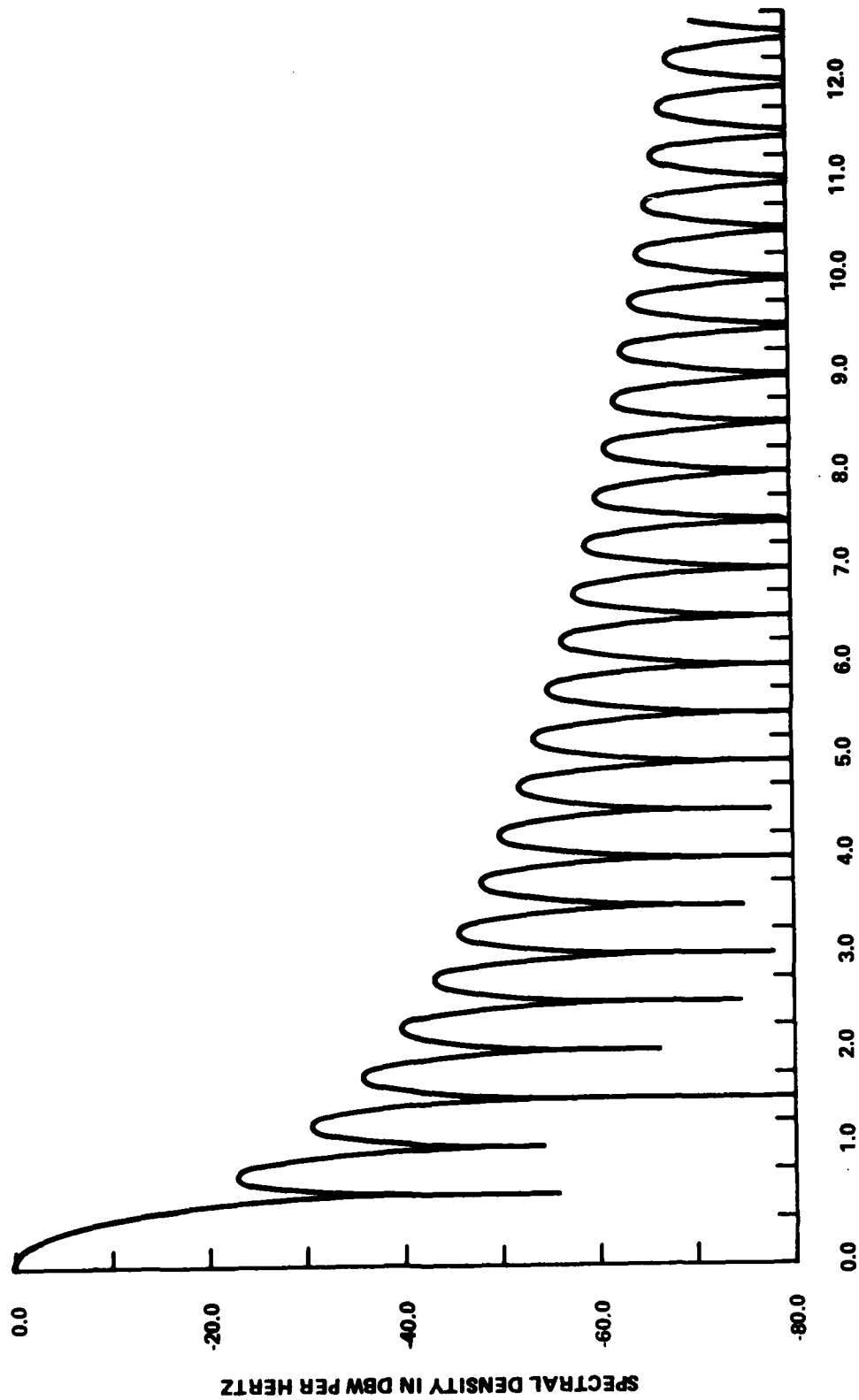


Figure 8. Spectral Density of MSK.

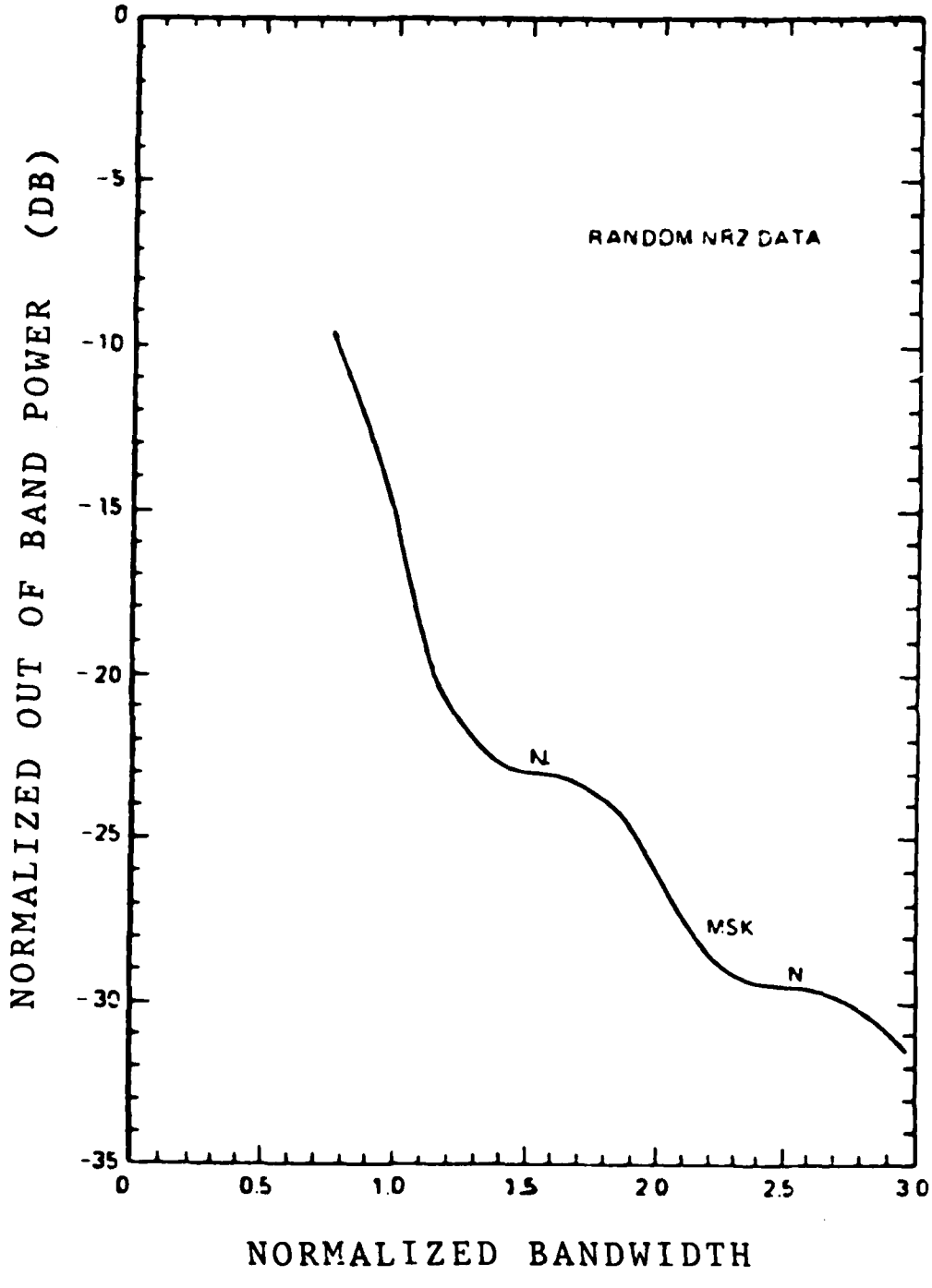


Figure 9. Power Outside of Normalized Channel Bandwidth for MSK.

The probability of bit error, p_e, of a coherently detected BPSK signal is shown in figure 5 and is given by

$$p_e = Q\left[\sqrt{2\left(\frac{E_b}{N_o}\right)}\right]$$

where E_b/N_o is a signal to noise power ratio.

$$Q(k) = \frac{1}{\sqrt{2\pi}} \int_k^\infty e^{-\lambda^2/2}\, d\lambda$$

and is shown in figure 10. For a probability of error of 10^{-5}, the $\frac{E_b}{N_o}$ value is 9.6 db.

The MSK signal, with baseband data which are not differentially encoded, has a probability of error that is identical to that of the coherently detected BPSK signal as given in the previous equation. Differential encoding of the baseband data changes the error probability. This is given as follows:

$$P_e = 2\,p_e - p_e^2$$

where p_e and P_e are the error probabilities before and after differential encoding of the baseband respectively.

For an error probability of 10^{-5} before differential encoding of baseband data, the probability of error after differential encoding is:

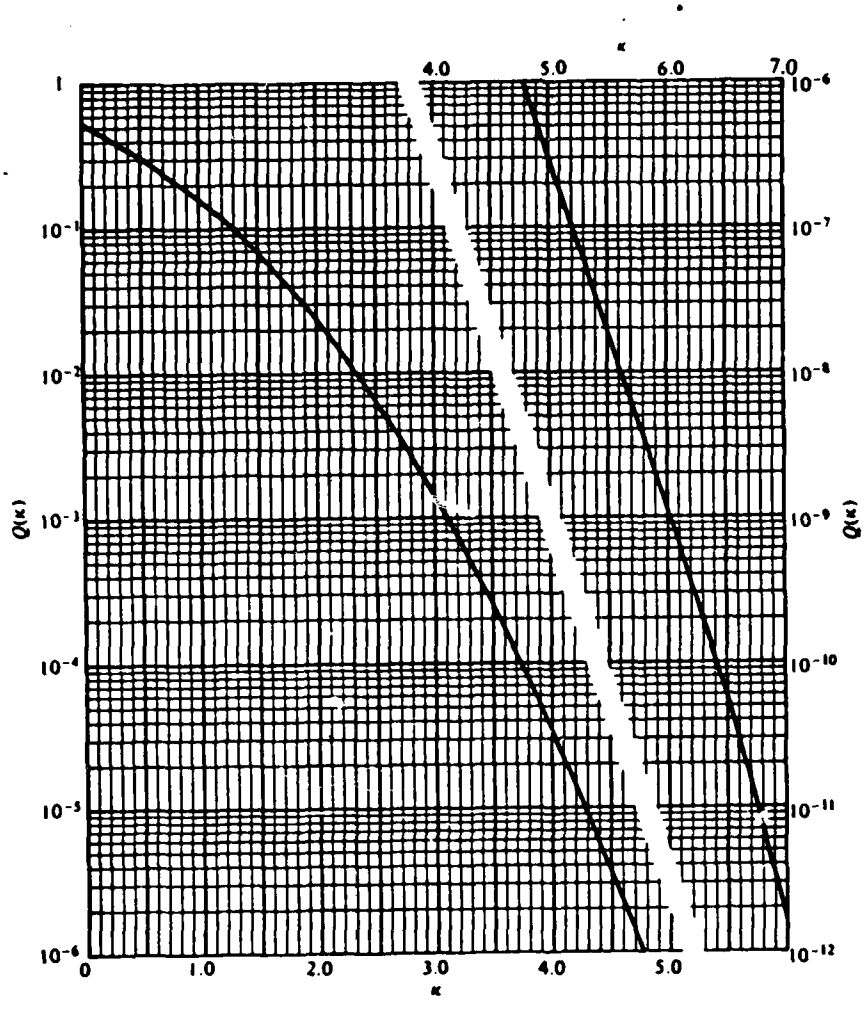
$$(P_e) = (2)\,(10^{-5}) - (10^{-5})^2$$

$$= 1.99999 \times 10^{-5}$$

$$\approx 2 \times 10^{-5}$$

The probability of error is increased due to differential encoding of baseband data for a given signal-to-noise ratio. To achieve the same error probability as in the BPSK case of 10^{-5}, a higher signal-to-noise ratio is required. This is computed to be 9.9 db, an increase of 0.3 db.

From a bandwidth stand point, the MSK signal has better bandwidth utilization than the BPSK signal. Figure 11 shows the power outside of channel bandwidth for BPSK and MSK signals. It can be seen that if the definition of bandwidth is taken as specified by the FCC as that which contains 99 percent of the energy, the bandwidth of the MSK signal is about 1.1 times the data rate. With the BPSK signal, only 82.2 percent of the energy is contained in the same bandwidth. It is expected in the future that MSK or BPSK will be the modulation scheme to use for digital signals. With the ever increasing traffic in the RF band allocated for transmission, bandwidth conservation is becoming more important. It seems that MSK definitely have an edge over BPSK in this regard.



Q(k) may be approximated by $e^{-k^2/2}/(2\pi)^{1/2}(k)$ for $k \gg 1$

Figure 10. Graph of Q (k) vs k.

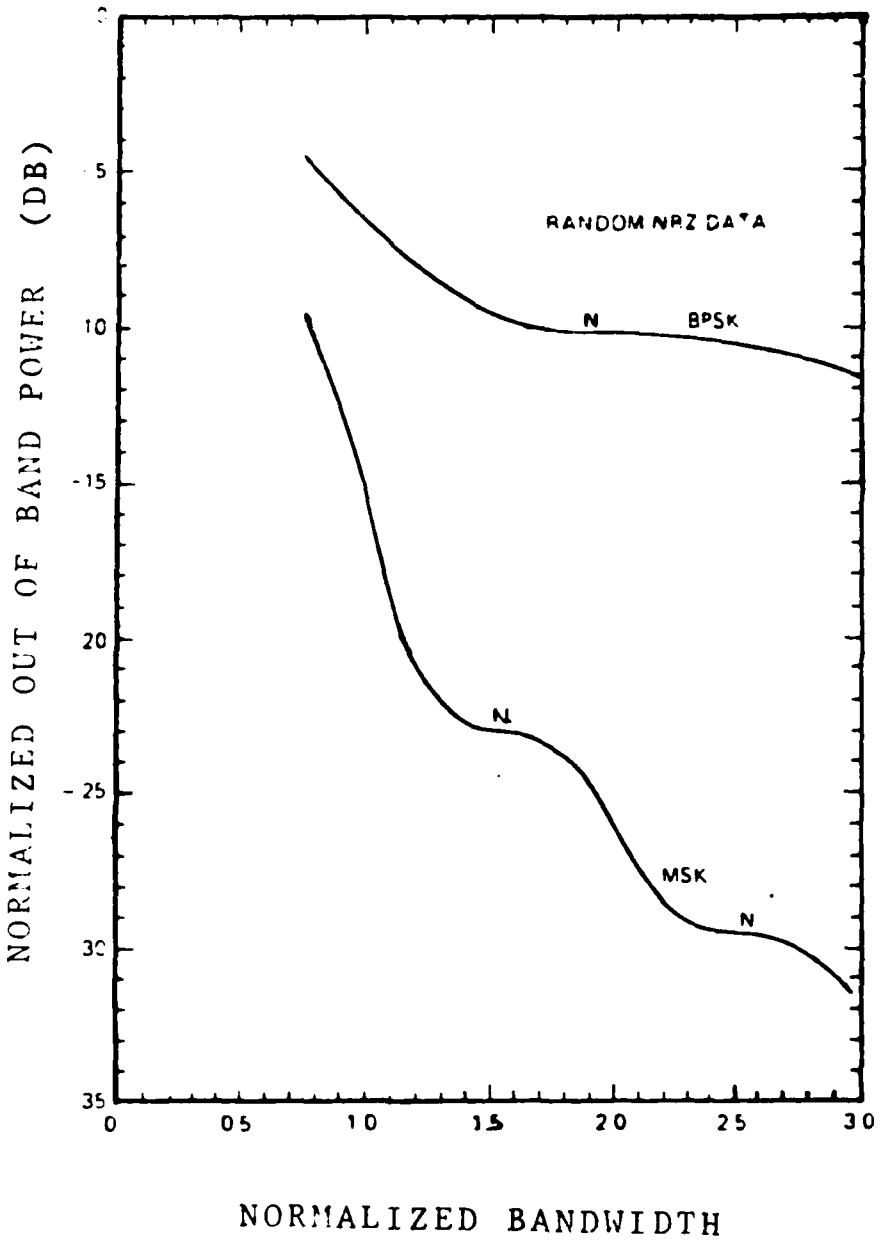


Figure 11. Power Outside of Normalized Channel Bandwidths for MSK and BPSK.

MATCHED FILTER RECEIVERS

In a wideband system, the detection of the presence of signal is of utmost importance. This is especially true in the case of GPS in which the signal level at the receiver is extremely low (-160 dbm). To maximize the detection probability requires the maximization of the receiver output signal-to-noise ratio. A decision will be made based on whether the output of the receiver exceeds a certain threshold. The matched filter receiver turns out to be the optimum approach for this application.

MATCHED FILTERING PROCESS

A matched filter is a filter that is "matched" to the incoming signal. In the frequency domain the matched filter has a frequency response that is the complex conjugate of the incoming signal. In the time domain, the matched filter has an impulse response that is identical to the incoming waveform except inverted in time. To satisfy the causality requirement that there would not be an output if there is no input, the impulse response is delayed in time as shown in figure 12.

A signal, entering any device with impulse response h(t) will convolve with the impulse response of the device. This is also true with the matched filter. The convolution process is described by the following equation

$$y(t) = \int_{-\infty}^{+\infty} g(x)\, h\,(t - x)\, dx$$

$$= g(t) * h(t)$$

where y(t) is the output, h(t) being the impulse response of the device and g(t) being the input signal.

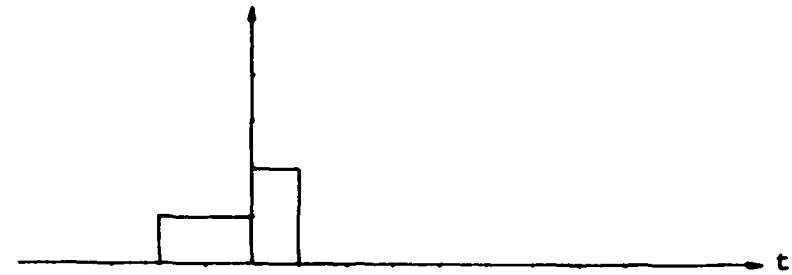
The convolution process can best be understood graphically. This is shown in figure 13. As can be seen, in the process of convolution, the impulse response is inverted to the other side of the t axis. The output is obtained by sliding the inverted waveform over the stationary waveform and integrating. If we had an impulse response that is a time inverted replica of the incoming signal, the inversion process in convolution will produce a replica of the incoming signal. The convolution between the incoming signal with the impulse response of the matched filter is equivalent to the correlation between the signal and a reference which is the inverted impulse response of the matched filter.

The equation that described the correlation process is as follows:

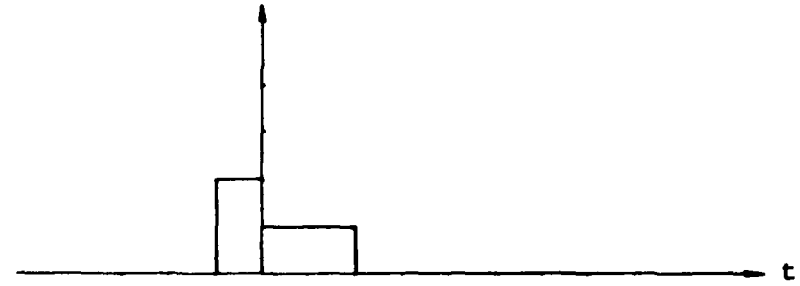
$$y(t) = \int_{-\infty}^{+\infty} g(x)\, h\,(t + x)\, dx$$

The difference between the correlation equation and the convolution equation is the change in sign in the second term representing the inversion process. A correlation of two waveforms is shown in figure 14. It can be seen that if one of the functions is even, the correlation function is identical to the convolution function.

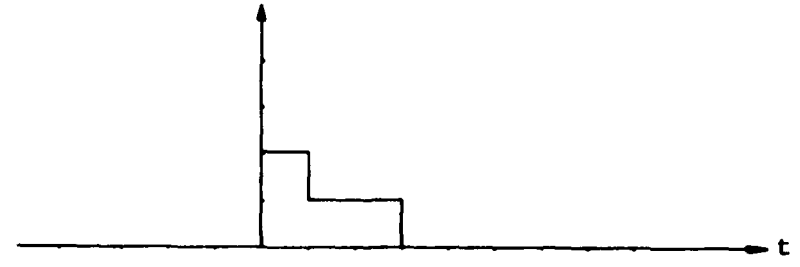
The choice of the code for signaling is therefore of great importance. It has to have good auto correlation properties so that the receiver, in determining the presence of signal, make very little mistakes.



(a)



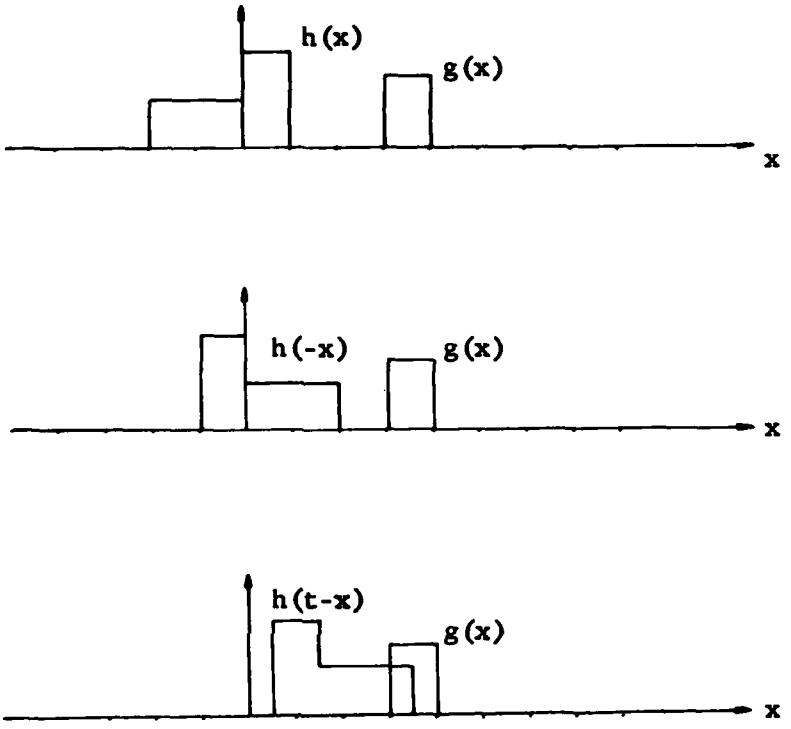
(b)



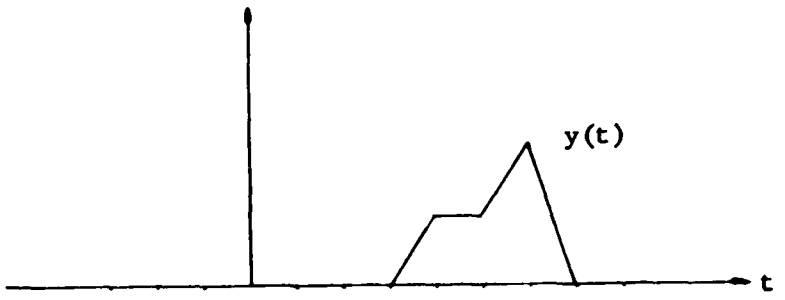
(c)

Figure 12. Impulse Response of a Matched Filter

(a) Incoming Signal
(b) Incoming Signal Inverted in Time
(c) Impulse Response of Matched Filter

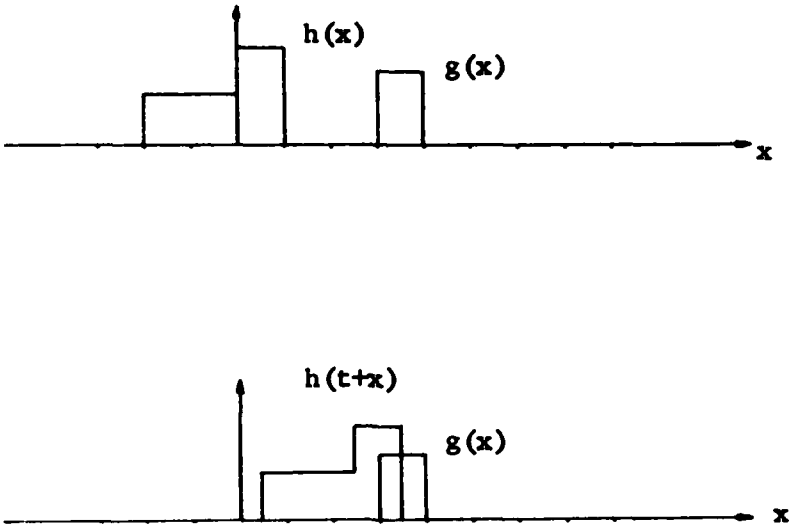


(a) Steps employed in graphical convolution

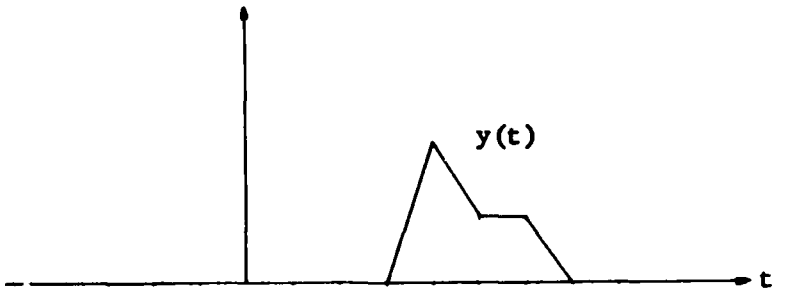


(b) Resulting convolution solution

Figure 13. Graphical Convolution of 2 Signals



(a) Steps employed in correlation



(b) Resulting correlation output

Figure 14. Cross Correlation of 2 Signals

IMPLEMENTATION OF THE CORRELATION FUNCTION

The correlation function can be implemented in two different ways. They are,

1. Multiplication with the incoming signal with a reference waveform and integrate, and

2. Fabrication of a device that has an impulse response that is identical to the incoming signal but inverted in time. The causality requirement is satisfied automatically for being a physical device. By the process of convolution, the correlation function can be obtained.

ACTIVE CORRELATION

In the first approach, a reference waveform is multiplied with the incoming signal and the result is then integrated to obtain the correlation function for threshold detection. This process requires a multiplier and an integrator. The timing of the reference is crucial for the proper detection of the signal. If the reference signal is offset by one bit from the incoming signal, the proper correlation will not occur.

Failure for the correlation peak to exceed the threshold,the reference waveform can be advanced by one bit at the end of the period and the correlation is performed again in the next cycle. This process is repeated until the threshold is exceeded. It is the active seeking of correlation that the term active correlation is used. One disadvantage is obviously the time required for the synchronization process. Depending on how misaligned the two codes are, synchronization can take a relatively long time. Another requirement being that the same code has to be transmitted repetitively until the threshold is exceeded. For codes that do not repeat, a bank of correlators will have to be used in parallel. The reference waveform to each correlator has to be offset by one bit.

The advantage of active correlation is that it is simple to implement. The main disadvantage is that of synchronization. In the first approach for synchronization, it is a time consuming process. In the parallel approach, a large amount of hardware are required for long codes. In either case, the control software/hardware may be very complex.

PASSIVE CORRELATION

A second approach in obtaining the correlation function is to design a device with the impulse response matched to the incoming signal. The device may be an active device. It is passive in the sense that it does not have to be actively updated for correlation. Once the device is fabricated or programmed, no updating control is necessary until the threshold is exceeded. At this time, the device, if programmable, may be reprogrammed for reception of another code. It can be seen that even if the receiver and transmitter are not in synchronization, correlation peak can still occur.

The main disadvantage being that of fabricating a device with the appropriate impulse response. This may be hard or impossible to implement. Furthermore, the amount of hardware required may be tremendous.

THE OPTIMUM CORRELATOR

The following is a set of criteria that a correlator should have for processing the different wideband spread spectrum waveforms.

1. Programmable Chip Rate

Different data rates are used for the different waveforms. Therefore, for a correlator to be applicable to these waveforms, this is the one most important requirement.

2. Programmable Reference Waveforms

The waveforms under consideration are BPSK and MSK. MSK can be viewed as FSK with more stringent requirements as described earlier or as an offset keyed quaternary phase shift keying signal. Nevertheless, the correlator should be capable of processing these waveforms.

3. Passive Correlation

The use of a passive correlator eliminates the timing control circuit used for synchronization. In addition, the signal can be acquired in a shorter duration than an active correlator. For non-repeating burst signals that a receiver must acquire within the duration of the burst, this seems to be the logical approach.

4. RF Processing

A receiver is the best if it has the simplest design and meets performance requirements. It is a goal of all systems. To this end, a correlator receiver that operates at RF is the best correlator receiver since all the signal conditioning functions that are required for down converting the signal into intermediate frequencies or baseband can be eliminated. An extreme case would be for a correlator to operate at the output of the antenna.

5. Real Time Output

The correlator should have real time output for real time signal processing.

6. Large Bandwidth

The correlator should have as large a bandwidth as physically feasible to accommodate the anticipated wideband signals.

7. Large Scale Integration

In view of the present thrust in the very high speed integrated circuit (VHSIC) development by the military, it should be a requirement that a correlator can be large scale integrated. Doing so will reduce the size, power, and weight requirement and increase the reliability of the device.

8. Low Insertion Loss and Low Temperature Sensitivity

The insertion loss of the correlator is to be minimized so that minimal external amplifications are required. The temperature sensitivity has to be low so that the characteristics of the device will not change with temperature variation, thereby eliminating the need of a temperature controlling device.

CORRELATOR TECHNOLOGY

An examination of the present technologies in the fabrication of a correlator indicates the following:

1. Surface Acoustic Wave (SAW) Devices.
2. Acousto-Optical Devices.
3. Optical Fibers.
4. Charge Coupled Devices.
5. Digital Devices.

These devices cover the part of the communication spectrum from dc to rf in the hundreds of megahertz range. Specifically, the rf devices are optical fibers, the rf/if devices are surface acoustic devices, acousto-optical devices, and the baseband devices are charged coupled devices and digital devices. The various technologies are covered in the following paragraphs.

SURFACE ACOUSTIC WAVE (SAW)

A SAW is an elastic wave that travels along the surface of a solid and is confined to the vicinity of that surface. This is in contrast to bulk waves which occupy the entire cross section of the medium. While SAW technology makes use of all waves which can propagate through a solid and which are confined to be near surfaces of that solid, the most common of these are known as Rayleigh wave which is the wave propagation that is associated with earth. The ordinary sound wave that the human ear can detect (at frequencies less than 30 KHz) are bulk waves that travels in air. For waves that are above 30 KHz, they are known as ultrasound waves. A SAW is simply an ultrasound wave that is confined to the surface of a solid.

Figure 15 illustrates the propagation of the bulk sound wave in a solid represented by grid points. The surfaces of the solid is the plane $z = 0$, and an increasing value of z measures increasing depth of the solid from the surface. Each of the grid points on the solid is uniformly spaced before the wave propagates. While the wave is propagating (in the y direction) one observes compression and elongation along the y axis and uniformity along the x axis.

Figure 16 shows the SAW propagation. Notice that for large value of z, the wave barely exists.

To launch the surface acoustic wave onto a piece of piezoelectric material requires the use of an interdigital transducer (IDT). An interdigital transducer is simply a set of metal strips placed on the piezoelectric substrate on which the surface acoustic wave is to be generated. A typical IDT is shown in figure 17. Alternate strips, commonly called fingers, are connected to each other by two other metal strips which form electrical inputs. The usual width of each metal finger is one quarter of a wavelength ($\lambda/4$), with the fingers also separated by $\lambda/4$. When an rf voltage is applied to the two input terminals of this interdigital transducer, an electric field is set up between all the adjacent fingers simultaneously. Thus, the piezoelectric material alternately gets compressed and elongated between the fingers and generates a surface acoustic wave which starts traveling with a velocity v, the surface wave velocity. Since the surface acoustic wave is generated simultaneously by each pair of adjacent fingers, a situation arises in which the sources are distributed. To obtain efficient surface wave generation, all these excitations must generate waves which add coherently.

If the period of the RF voltage is to be T seconds, to propagate a distance of $\lambda/2$ requires T/2 seconds. Since this is also the time it takes for the electric field between two adjacent fingers to change polarity, all the generated waves will add constructively. For detection, the reverse is true. As the elastic deformation passes under the fingers, it induces in phase voltages between each finger pair.

Surface acoustic wave devices are inherently lossy devices because the basic transducer is bidirectional. That is, half the power is radiated toward the output transducer while the other half radiates towards the end of the crystal substrate and is lost (figure 18a). By reciprocity, half the power is lost at the output transducer, and thus, a bidirectional SAW device is limited to a minimum insertion loss of 6 db. This difficulty can be alleviated by using the three transducer approach (figure 19b). The center transducer launches the surface acoustic wave in each direction with 50 percent of the total available energy for a loss of 3 db. Each receiving transducer in turn receives 50 percent of this energy for a total of 50 percent of the energy. The loss is therefore 3 db.

By using an unidirectional approach in which the energy is transmitted in one direction only, the insertion loss can be reduced further (figure 18c).

The above mentioned schemes will cut down on the loss of the energy after it has been coupled onto the substrate. The insertion loss due to the electromechanical coupling coefficient of the substrate material is a characteristic of the material itself and there is no way to change it. All commonly used substrate materials have low coupling coefficients, but some are much lower than others. That part of the input electrical signal that does not excite surface acoustic wave is dissipated, partly in the form of bulk waves through the body of the material. They are different phenomenon from surface waves and in a surface acoustic wave device, they are undesirable because they can be reflected from all the surfaces and when they arrive at the surface where the transducers are mounted, they interact with the surface acoustic waves in undesirable ways. The largest source of such reflections is the lower surface of the substrate, it can be slightly ground so that the bulk wave striking it will be dispersed instead of reflected.

Table VIII lists the coupling coefficients of some common SAW device substrate materials, together with their temperature coefficient. There is clearly a trade off to be made between better coupling coefficient and better temperature stability, depending on which is more important in a given application. This figure also shows wave velocity of these materials. Wave velocity is of interest because the line width of a transducer pattern for a given frequency is wider for materials with higher wave velocity, making the fabrication process less exacting.

A SAW device can be made into a filter by varying the lengths of the fingers which can be obtained by specifying a frequency response in the frequency domain. The normalized samples of the impulse response in the time domain are then the lengths of the fingers.

Since the finger width is $\lambda/4$, SAW filters are limited by the resolution of the lithographic process in the high end of the frequency spectrum. The low end of the frequency spectrum is limited by the obtainable size of the crystal substrate.

SAW Tapped Delay Line (TDL)

To use the SAW device as a correlator, the impulse response of the SAW device has to be the time inversion of the incoming signal. This can be easily accomplished by using the SAW device as a tapped delay line with a length that will accommodate the signal. If taps are deposited on the

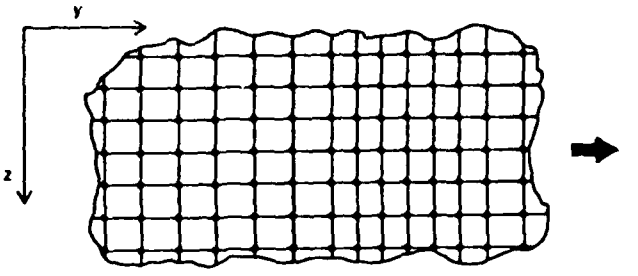


Figure 15. Propagation of Bulk Sound Wave

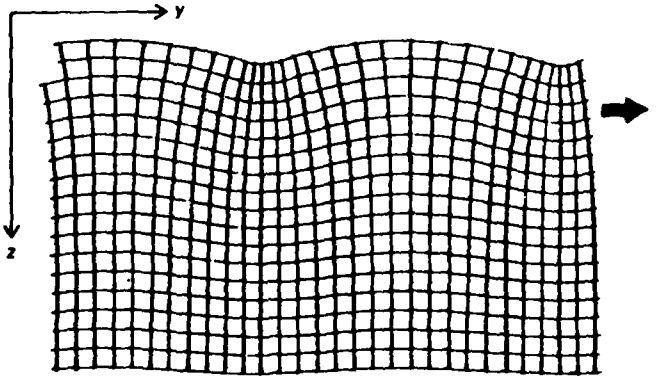


Figure 16. Propagation of Surface Acoustic Wave

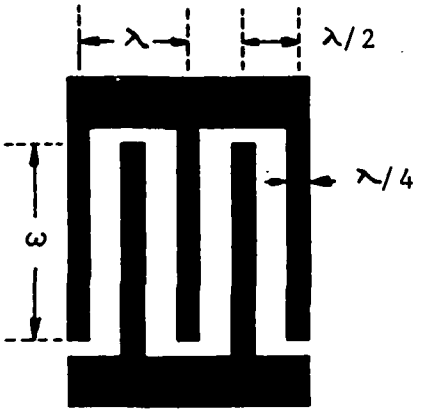


Figure 17. A Typical Interdigital Transducer

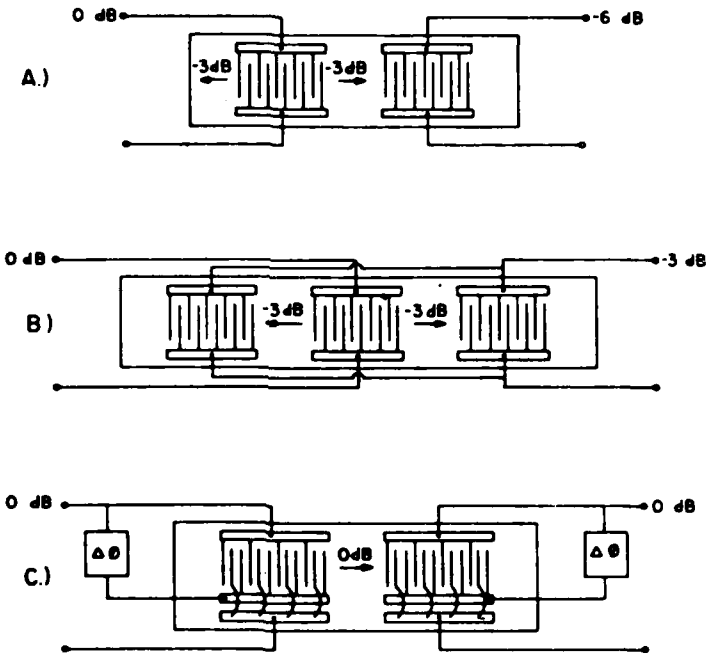


Figure 18. SAW Interdigital Transducer Structures

a. Two Transducers
b. Three Tranducers
c. Unidirectional Transducers

Table VIII. Properties of Three Commonly Used SAW Materials

	Wave Velocity meters/sec	Coupling Coefficient	Temperature Coefficient ppm/deg C
Lithium Niobate (Y,Z cut)	3488	4.82	94
Lithium Tantalate (Y,Z cut)	3230	0.66	35
Quartz (ST cut)	3158	0.116	Negligible

Essential characteristics of commonly used substrate materials for SAW devices are shown in above table.

Most dramatic differences are reversed from the point of view of desirability; coupling coefficient of lithium niobate is well over an order of magnitude better than quartz, while its temperature coefficient is very poor and that of quartz is nearly perfect. Wave velocity is of interest because the linewidth of a transducer pattern for a given frequency is wider for material with higher wave velocity making the fabrication process less exacting. Lithium tantalate is seen to be intermediate in all characteristics.

piezoelectric substrate with spacings that are equal to the bit period of the data T, the application of a pulse of short duration at the incoming carrier frequency will result in an output that lasts a duration NT (where N is the number of taps on the substrate) if the tapped outputs are brought to a summing circuit. This is a matched filter for a carrier input. This is shown in figure 19. To obtain the impulse response for a PSK input signal of a fixed code, the output of the taps corresponding to the phase shifts can be inverted by external inverting amplifiers with unity gain. The impulse response of such a structure will be identical to the incoming signal except inverted in time. This is shown in figure 20 and is the desired matched filter. To make such a filter programmable, inverting and non-inverting amplifiers can be put at the output of each tap. Logic circuits, e.g. shift registers, can then be used to determine which amplifiers are to be turned on according to the reference code in such a way that the impulse response is the time inversion of the incoming signal. This is shown in figure 21. The SAW tapped delayed line is now a programmable matched filter/correlator. The correlator is programmable in the sense that the code can be changed. However, it can be seen that once the taps are made, the data rate remained fixed and there is no way of changing it unless a different SAW tapped delay line is used. The advantages of a SAW tapped delay line are as follows:

1. A passive correlator.

2. A mature technology that has a low risk factor.

3. A passive device that has high reliability.

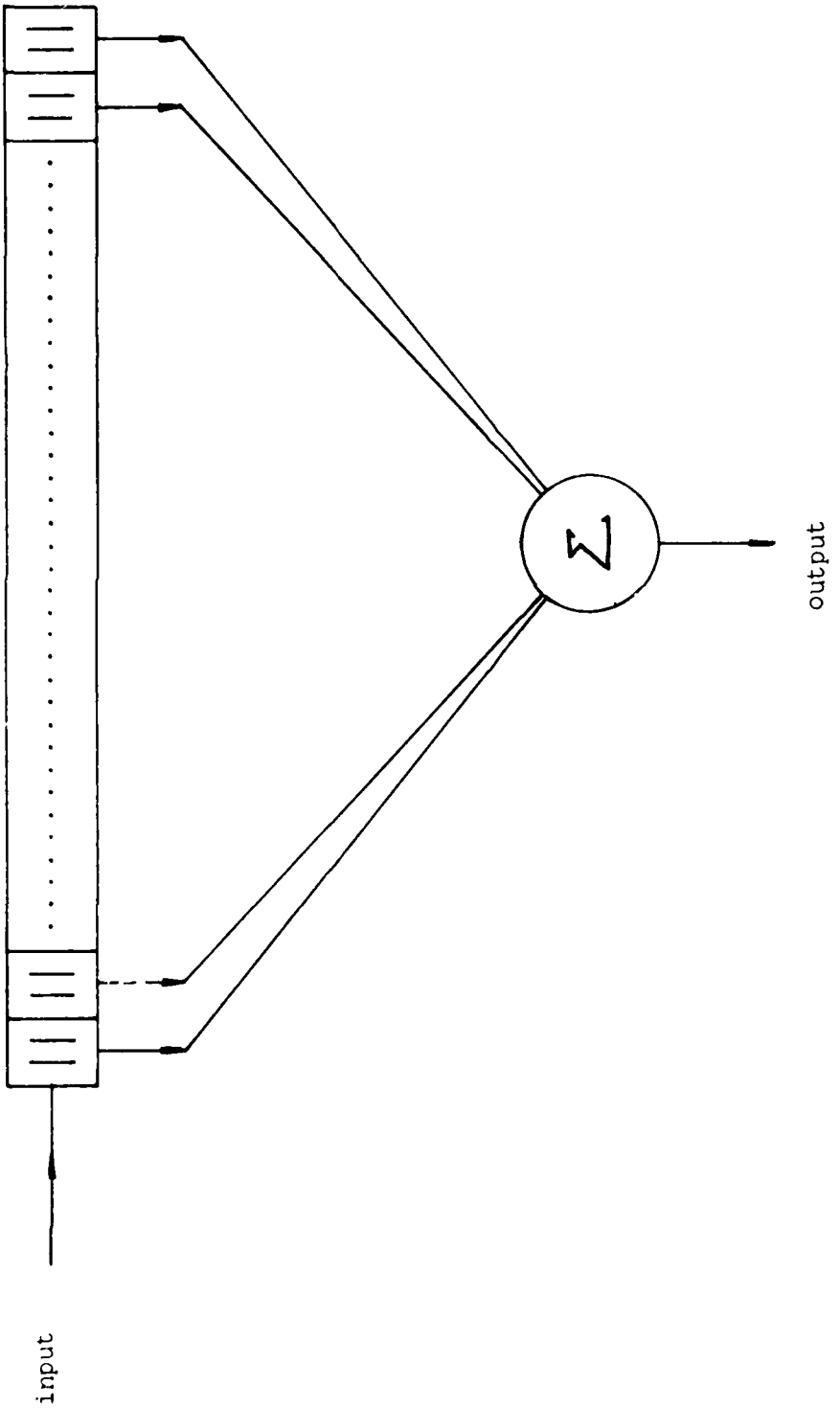


Figure 19. A SAW TDL Correlator for a Carrier Wave Input

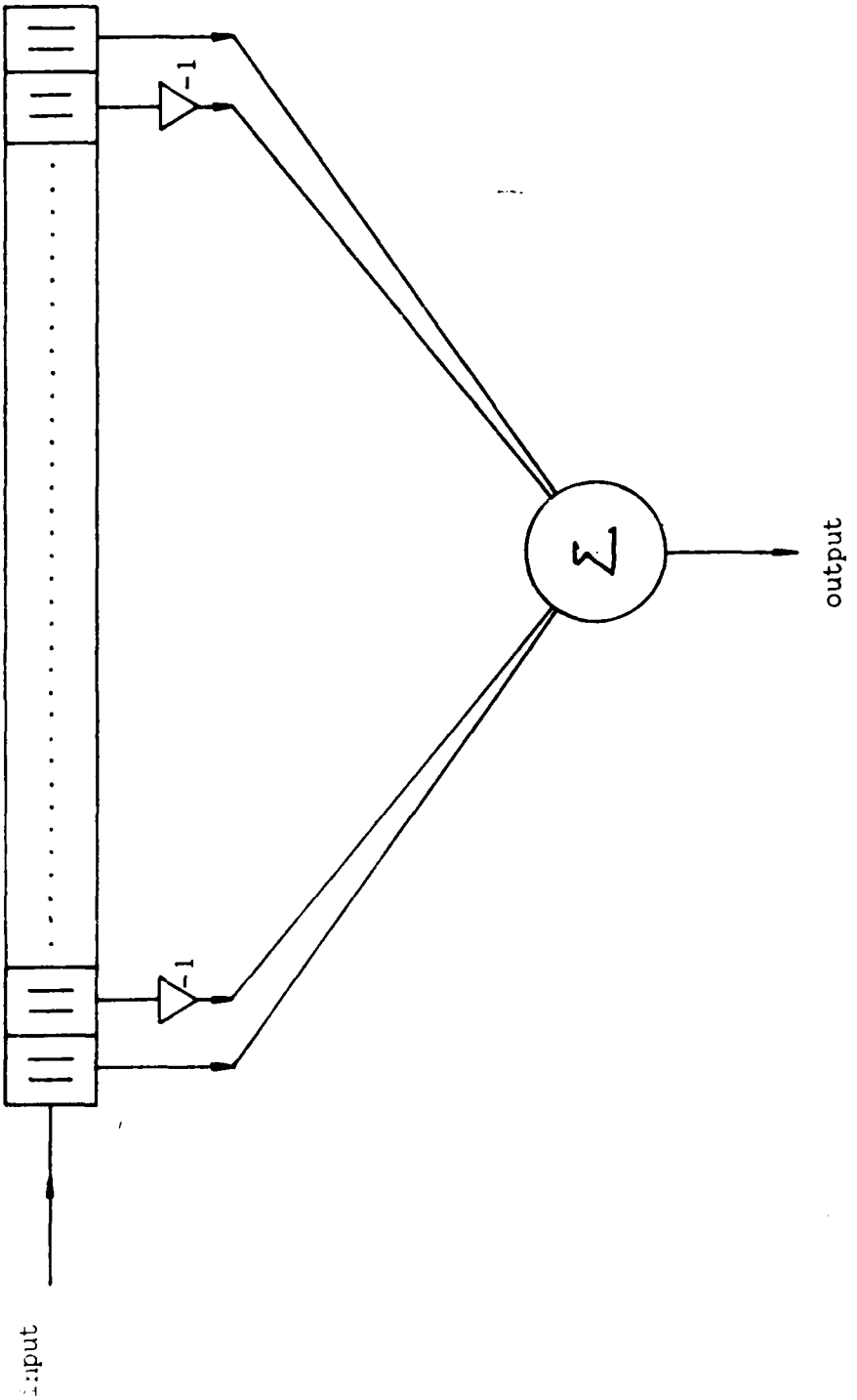


Figure 20. A Fixed Code SAW TDL Correlator

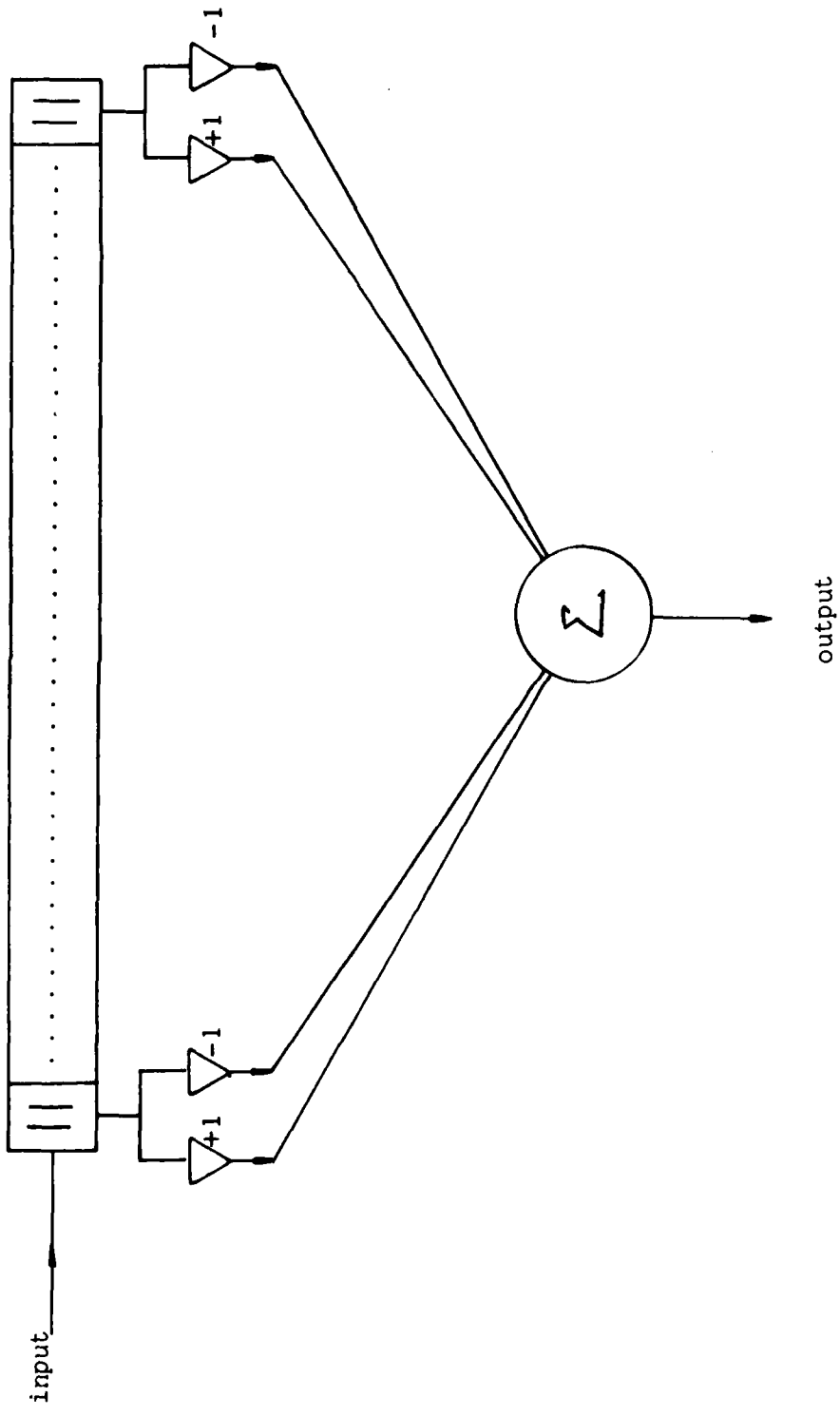


Figure 21. A Programmable SAW TDL Correlator

4. Can be used at IF frequency for signal processing.

The disadvantages of a surface acoustic wave tapped delay line are,

1. Large signal loss for ST cut quartz for low temperature coefficient.

2. Large temperature coefficient for Lithium Niobate for large coupling efficiency.

3. A hybrid approach that requires both digital and analog circuitry.

4. Length of the SAW device is dependent on the signal frequency and, for low frequency signals, the SAW device may be bulky.

5. SAW tapped delay lines do not have variable chip rate capability.

6. SAW's do not lend itself to large scale integration.

The fixed chip rate of SAW TDL is the one factor that limited its application as a multiwaveform correlator.

SURFACE ACOUSTIC WAVE REFLECTION

As can be seen in a SAW tapped delay line, reflection of the acoustic wave in the piezoelectric material causes distortions in device response and often cause unattractive tradeoffs between device performance and insertion loss. In recent years, devices have been developed that take advantage of the reflective wave by depositing reflective grating on the substrate. Reflection grating devices have proven to be remarkably free of spurious signals and second order effects. As a consequence, devices of this type have better performance and time bandwidth products greater than those demonstrated with devices based on distributed tapping in interdigital transducers.

The surface wave solution to the equation of motion for a semi-infinite homogeneous medium differ from bulk wave propagation because the surface wave must satisfy a set of boundary conditions on a plane surface. Any change in the boundary that couple to the surface wave will reflect the surface wave and possibly scatter some energy into bulk waves. The surface perturbations which cause reflection can be an overlay layer or an alteration of material properties by ion implantation or diffusion as shown in figure 22. The different geometries employed in reflective devices are shown in figure 23.

Reflective Array Compressor (RAC)

A reflective array compressor has the basic structure of figure 23(c) except the grating in this device have a spacing that increases (or decreases) as a function of distance from the input transducers. The device is made based on the fact that the surface wave is strongly reflected at a right angle at a point in the reflective array where the spacing between two consecutive lines in the propagation direction matches the wavelength of the surface wave. A second grating which is the mirror image of the first grating is symmetrically placed to send the waves to the output transducer. In the time domain, if a signal of different frequencies is launched onto the device, the different frequencies will come out at different times at the output transducer since the distances the different frequency components have to travel before reflection occurs are different. This is shown in figure 24. SAW based compressive receiver can be designed within the range of parameters approximately bounded by maximum bandwidth 500 MHz, resolution 3 MHz and at the other extreme bandwidth 2 MHz, resolution 20 kHz. The way the reflective array compressor can be used is based on the chirp Z transform.

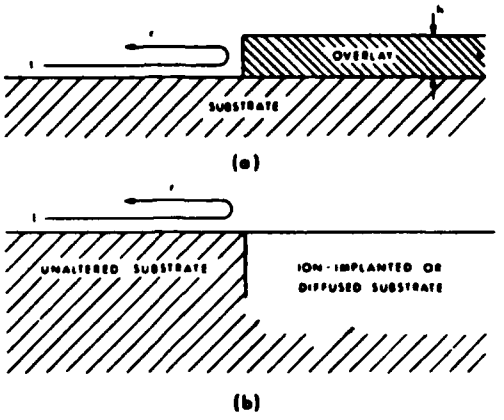


Figure 22. Surface Perturbations Which Cause Surface Wave Reflection

a. An Overlay Layer
b. Alteration of Material Properties by Ion-Implantation or Diffusion

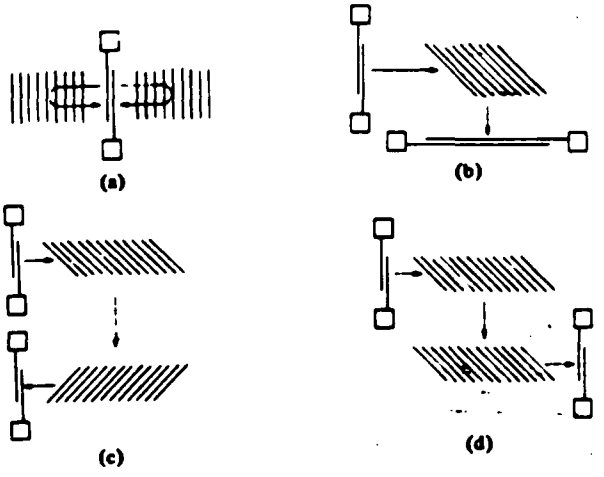


Figure 23. Geometries Employed in Reflective-Array Devices

a. Normal-Incidence Resonator
b. One-Bounce at Oblique Incidence
c. Two-Bounce U Path
d. Two-Bounce Z Path

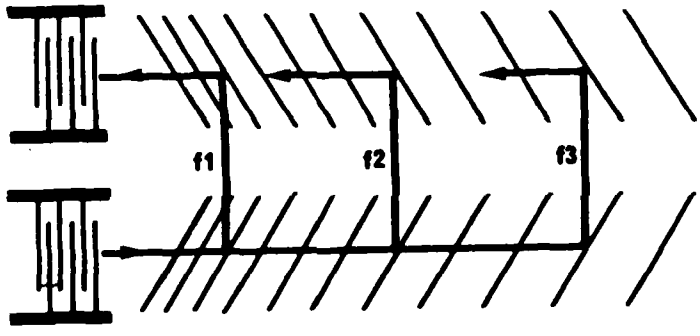


Figure 24. A Reflective Array Compressor

Chirp Z Transform

The chirp Z transform can be realized by a chirp filter, two chirp generators and two multipliers as shown in figure 25. The output function F(t) can be written as.

$$F(t) = \left\{[S(t)\, C_1(t)] * I(t)\right\} C_2(t)$$

where * represents convolution.

To see that F(t) is the Fourier transform of the input signal, complex notation is used.

Let

$$S(t) = S'(t)\, e^{j2\pi f_o t}$$

$$C_1(t) = \begin{cases} e^{j2\pi\left(f_1 t - \frac{\Delta F t^2}{2\Delta T}\right)} & |t| < \frac{\Delta T}{2} \\ 0 & |t| > \frac{\Delta T}{2} \end{cases}$$

$$I(t) = \begin{cases} e^{j2\pi\left[(f_1 + f_o)\, t - \frac{\Delta F t^2}{2\Delta T}\right]} & |t| < \Delta T \\ 0 & |t| > \Delta T \end{cases}$$

$$C_2(t) = C_1(t)$$

Substitution of these equations into the expression for F (t) yields the following:

$$F(t) = C_2(t) \left\{[S(t)\, C_1(t)] * I(t)\right\}$$

$$= C_2(t) \left\{\left[S'(t)\, e^{j2\pi f_o t}\, e^{j2\pi\left(f_1 t - \frac{\Delta F t^2}{2\Delta T}\right)}\right] * I(t)\right\}$$

$$= C_2(t) \int_{-\frac{\Delta T}{2}}^{\frac{\Delta T}{2}} S'(\tau) \left\{e^{j2\pi\left[(f_1+f_o)\tau - \frac{\Delta F \tau^2}{2\Delta T}\right]}\right\} \left\{e^{j2\pi\left[(f_1+f_o)(t-\tau) + \frac{\Delta F (t-\tau)^2}{2\Delta T}\right]}\right\} d\tau$$

$$= C_2(t)\, e^{j2\pi\left[(f_1+f_o)\, t + \frac{\Delta F t^2}{2\Delta T}\right]} \int_{-\frac{\Delta T}{2}}^{\frac{\Delta T}{2}} S'(\tau)\, e^{-j\left[2\pi\left(\frac{\Delta F t \tau}{\Delta T}\right)\right]} d\tau$$

The integral represents the Fourier transform of the incoming signal with

$$\omega = \frac{2\pi \Delta F t}{\Delta T}$$

This function, however, is weighted by a linear FM function. Multiplication by $C_2(t)$ will eliminate the FM sweep as shown below.

$$F(t) = e^{j2\pi\left(f_1 t - \frac{\Delta F t^2}{2\Delta T}\right)} \quad e^{j2\pi\left[(f_1+f_o)t + \frac{\Delta F t^2}{2\Delta T}\right]} \quad S(\omega)$$

$$= e^{j2\pi(2f_1 + f_o)\,t}\ S(\omega)$$

The term $e^{j2\pi(2f_o + f_1)\,t}$ represents a frequency translation process.

A qualitative understanding of the operation of the transformer can be obtained by examining the frequency time diagram shown in figure 26. The signal amplitude is not shown. Frequency is displayed on the vertical axis and time on the horizontal axis. The bandwidth of the signal corresponds to its height and the duration corresponds to its length.

As shown in the figure, input signal S(t), which is assumed to have bandwidth from 125 to 175 MHz is mixed with a linear frequency down chirp C(t) centered at 275 MHz. This mixing gates the input in time and superimposes a chirp on the signal. The sum frequency product term A(t) centered at 425 MHz is fed into the chirp filter shown in the block diagram of figure 25. The chirp filter output B(t) covers the same frequency range as the input A(t), but it is delayed in time, and the dispersive up chirp characteristic of the filter has introduced an up chirp on the output. Since the linear FM filter I(t) must handle the sum of the bandwidths of S(t) and $C_1(t)$ and also possesses the same chirp slope as $C_1(t)$, then I(t) is twice the time length of $C_1(t)$. The duration of the convolution of A(t) and I(t) is therefore three times the length of the the signal A(t). The only portion of the signal that corresponds to a true Fourier transform of the input is the center part that corresponds to the time when the entire signal is in the chirp filter. The slope of the chirp $C_1(t)$ is matched to the filter chirp slope so that the filter output consists of compressed pulses corresponding to the input frequency components. The chirp filter does not change the frequencies present in the input signal A(t).

The filter output B(t) is mixed with another down chirp $C_2(t)$. This mixing gates out the useful portion of the output and cancels the chirp characteristics of the signal. The product term F(t) at 700 MHz corresponds to the true Fourier transform of the input signal S(t) over the duration (ΔT) of the chirp signal. The high frequency can be avoided by mixing with an up chirp instead of down chirp $C_2(t)$ and using the difference product term only.

Correlation in the Frequency Domain

Let y(t) be the correlation function of two functions h(t) and x (t). y(t) can be written as follows:

$$y(t) = \int_{-\infty}^{\infty} h(\lambda)\, x(t+\lambda)\, d\lambda$$

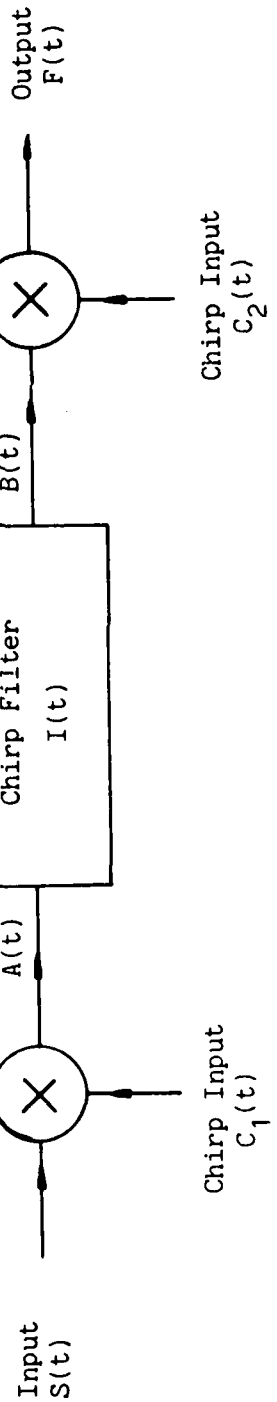


Figure 25. Basic Chirp-Z Transform Unit

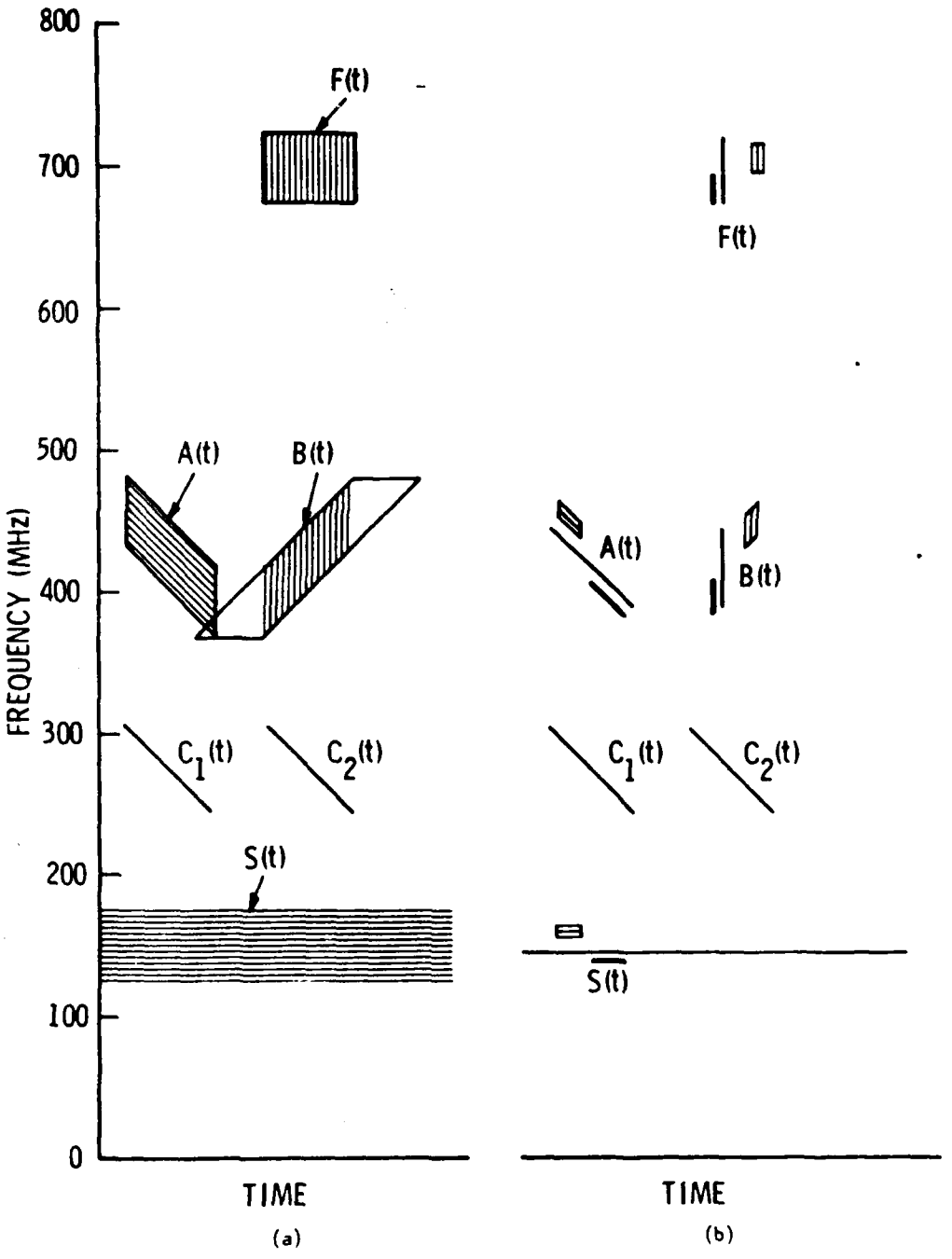


Figure 26. Frequency-Time Diagram of a Chirp-Z Transform Unit

a. Full Band Coverage

b. A Three Tone Example

Taking the Fourier transform of the above equation yields,

$$y(\omega) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} h(\lambda)\, x(t+\lambda)\, d\lambda \right] e^{-j\omega t}\, dt$$

$$y(\omega) = \int_{-\infty}^{\infty} h(\lambda) \left[\int_{-\infty}^{\infty} x(t+\lambda) e^{-j\omega t} dt \right] d\lambda$$

Let

$$t + \lambda = \xi$$

then,

$$t = \xi - \lambda$$

$$\frac{dt}{d\xi} = \frac{d}{d\xi}(\xi - \lambda)$$

$$= 1$$

and, $dt = d\xi$

Therefore,

$$y(\omega) = \int_{-\infty}^{\infty} h(\lambda) \left[\int_{-\infty}^{\infty} x(\xi)\, e^{-j\omega(\xi - \lambda)}\, d\xi \right] d\lambda$$

$$= \int_{-\infty}^{\infty} h(\lambda)\, e^{j\omega\lambda}\, X(\omega)\, d\lambda$$

$$= H(-\omega)\, X(\omega)$$

$$= H^*(\omega)\, X(\omega)$$

where * stands for the complex conjugate of the function.

It can be seen that in the frequency domain, the correlation of two functions is the product of the Fourier transform of one function and the complex conjugate of the Fourier transform of the other function. The time domain response will be the inverse Fourier transform of this function as follows,

$$y(t) = F^{-1}[H^*(W)\ X\ (W)]$$

where * denotes the complex conjugate of a function.

It was shown earlier that the Fourier transform of a signal can be obtained by a reflective array compressor in conjunction with two chirp signals.

The difference between normal and inverse Fourier transform is that the slopes of the chirp filter and the two chirp generator must be reversed i.e. change up chirp to down chirp and vice versa.

Therefore, the correlation can be done in the frequency domain and output can be inverse transformed back to the time domain. Figure 27 shows the configuration of such an approach.

Such a receiver provides fast spectral analysis of the input waveform. The transforms are in real time and at speeds which are far in excess of that available in digital fast Fourier transform. The number of resolvable frequencies in a compressive receiver or discrete frequencies in a Chirp Z transform is less than or equal to $T\Delta f/4$ where $T\Delta f$ is the time bandwidth product of the filter. The availability of RACs with time bandwidth products of several thousand make the applications potentially attractive. The chirp Z transform performs the Fourier transform of a finite segment of duration T of the incoming signal, where T is the duration of the chirp with which the incoming signal is mixed.

In using RAC's as a correlator, the main problem being one of matching. Since there are three reflective array compressors that have to be used for performing the Fourier and inverse Fourier transforms, these devices have to match exactly to the slope of the chirp function for proper operation. The complex conjugate operation can be eliminated if the time inverted reference is available.

These devices can be used for different waveforms with different chip rates provided that the duration of the incoming waveform is within the delay and frequency range of the device. For longer codes, the signal has to be divided up into intervals for processing.

The correlation is an active one and so synchronization may be more time consuming.

The following lists some of the advantages and disadvantages of reflective array compressor. The advantages are,

1. Real time processing.
2. Programmable chip rate.
3. Programmable waveform.
4. IF/RF processing.

The disadvantages of reflective array compressor are,

1. Active correlation.
2. The two Fourier transformers have to be very closely matched for proper operation.
3. The inverse transformer, which is also a reflective array compressor, has to match the slopes of the two Fourier transformers.
4. Three devices are used instead of one, therefore, less reliable.

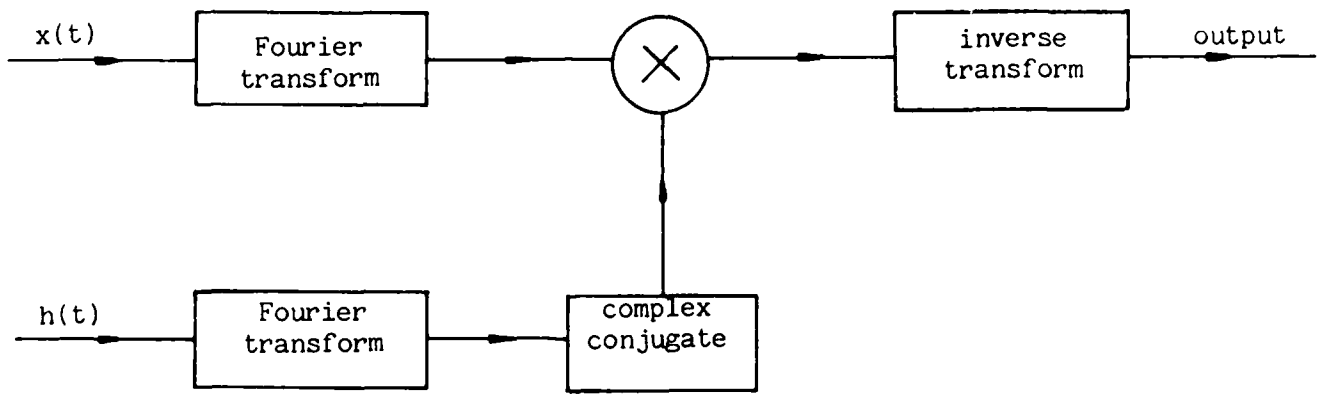


Figure 27. Correlation Using Dispersive Delay Line

5. These devices are not cascadable and therefore, does not allow for expansion in the future if the need arises for new waveforms. To make a device that covers the frequency band of interest requires the use of a very large substrate.

6. Temperature sensitive.

7. Large insertion loss.

8. Cannot be large scale integrated.

ACOUSTO ELECTRIC INTERACTIONS

When a semiconductor or metallic conductor is placed close to the surface of a piezoelectric material on which a surface acoustic wave is propagating, current will flow in the semiconductor. Because of the highly nonlinear relation between the current and the electric field in a semiconductor, nonlinear acousto-electric interaction between an acoustic wave and the semiconductor can be relatively strong. This makes it possible to devise various types of devices. An important class of such devices are the so called convolvers and correlators, these take the product of two signals and form the correlation or correlation integral of the signals.

Principle of Operation of the Acoustic Convolver

To describe the principles of operation of the convolver, a piezoelectric material in which a nonlinear interaction between two acoustic surface waves propagating along the surface of the substrate is assumed. Let these two signals be two carrier frequencies at ω_1 and ω_2 and are launched onto the piezoelectric substrate at each end, respectively. If the substrate is of length L, the signals at any point z along the device will be of the forms $e^{j\omega_1 (t-z/v)}$ and $e^{j\omega_2(t+z/v)}$ respectively, where v is the velocity of the acoustic wave. The change in sign is due to the fact that the two waves are traveling in opposite directions. Due to the nonlinear effect of the piezoelectric material, second order potentials will be generated at frequencies $\omega_1 \pm \omega_2$, $2\omega_1$ and $2\omega_2$ as well as dc potentials proportional respectively to the squares of the two input signals. At the z plane, the signal at the sum frequency will have an associated potential at the surface of the substrate of the form.

$$\phi\,(t, z) = A\, e^{j[(\omega_1 + \omega_2)\, t - (\omega_1 - \omega_2)\, z/v]}$$

where A is a constant.

The sum frequency can therefore be detected by an interdigital transducer of proper finger spacing. For the case $\omega_1 = \omega_2$, the potential of the nonlinearly generated signal $\phi\,(t, z)$ does not vary with z and can be detected between metal films laid down on the top surface and lower surface of the piezoelectric substrate.

Let $F(t)\, e^{j\omega t}$ and $G(t)e^{j\omega t}$ respectively be two signals with modulation. The convolver will yield an output of the form.

$$V = C\, e^{2j\omega t} \int F\,(t - z/v)\, G\left[t - \frac{(L - z)}{v}\right] \quad dz$$

where C is a constant which is related to the strength of the nonlinear interaction, the distance between the input transducer is taken to be L and the integration is carried out over the length of the output transducer L_o.

Let

$$t - z/v = \tau$$

Then,

$$V = D\, e^{2j\omega t} \int F(\tau)\, G(2t - \tau - T)\, d\tau$$

where D is a constant and T = L/v is the delay time of an acoustic surface wave passing along the delay line. This equation can be easily recognized as being closely related to the convolution of the two input signals. The difference being that the output signal is compressed by a factor of two in time. This is due to the fact the two acoustic waves are each traveling with speed v and so passes each other with speed 2v.

If one of the waveforms is made the time inverted reference waveform, the process of convolution becomes that of correlation. For an arbitrary input, it was shown that the time inverted signal could be obtained from the device by inserting a signal of frequency ω_1 into one end of the device, and a narrow pulse of frequency ω_3 at the center transducer to obtain a time inverted output signal at the input transducer of frequency $\omega_2 = \omega_3 - \omega_1$. So, by using one device as the inverter and one device as the convolver, the correlation can be obtained.

Semiconductor Convolvers

An alternative approach is to increase the strength of the nonlinear interaction by making use of the nonlinear response of a semiconductor coupled to the RF electric fields of the acoustic waves propagated along the piezoelectric delay line. Two types of interactions with semiconductors are possible: longitudinal interaction in which the nonlinearity is caused by the electric field parallel to the direction of propagation and transverse field interactions caused by an electric field perpendicular to the direction of propagation and to the surface of the semiconductor.

The configuration that receive the most attention for use as a semiconductor convolver is of the transverse field interaction kind. The interaction is essentially between the electric field E_y normal to the surface of the semiconductor and the carriers in the semiconductor. Typically, a relatively thick semiconductor layer is employed so that the parallel field component E_z tends to be shorted out. Semiconductor depletion layer theory leads to the conclusion that a depletion layer of thickness ℓ will be formed in an n-type semiconductor of donar density N_d such that

$$E = q\, N_d\, \ell/\epsilon$$

where E is the field normal to the surface and ϵ is the permitivity of the semiconductor. In turn, this implies that the potential across the depletion layer is,

$$\phi = q\, N_d\, \ell^2/2\epsilon$$

$$= \epsilon\, E^2/2q\, N_d$$

It is thus apparent that the potential formed across the depletion layer is proportional to the square of the field and varies inversely with the donor density. It is as if the semiconductor behaves as a distributed varactor, with a considerably stronger nonlinearity then can be obtained in the piezoelectric material itself. By employing this principle, a convolver can be constructed.

In this case, the potential generated across the depletion layer at any point is proportional to the product of the two input signals. The output is detected between an electrode on the lower surface of the piezoelectric material, capacitively coupled to the surface of the depletion layer, and an electrode on the top surface of the semiconductor. The output potential will be proportional to the integral of the induced depletion layer potential along the length of the semiconductor.

The frequency response of the convolver, like other SAW devices, is dependent on the photolithographic techniques. Since the output frequency is the sum of the input frequencies which are usually the same, the output frequency will be twice that of the input frequency. This represents the maximum frequency that can be generated by the two inputs. Therefore, if the state-of-the-art in lithographic technique can produce finger spacing of 1000 MHz, 500 MHz is the limit on the two input frequencies.

Unlike the surface acoustic wave tapped delay line and the reflective array compressor, which have fixed tap distance or linear grating on the substrate, the convolver has interdigital transducers for launching and receiving the signals only. The fingers do not perform any filtering or weighing function. Therefore, as long as the input waveforms do not have frequencies outside the bandwidth of the transducers, convolution can be obtained.

The requirement in using a SAW convolver is that the two signals have to exist within the device processing time window ΔT. The input waveform as well as the chip rate is immaterial. Therefore, if a convolver can be made to accommodate the longest desired waveform, the convolver can then be used as a multi waveform correlator since the reference waveform can be changed externally. The coexistence requirement of a convolver offers significant advantage over active correlation in that proper correlation is not dependent on the critical alignment of the received signal with the locally generated reference. The dynamic range of the correlator is a maximum when $\tau = \Delta T$ and for $\tau < \Delta T$, it is reduced by $20 \log \tau/\Delta T$.

It can be seen that a SAW convolver is a very powerful tool for processing wideband signals. The maximum delay for a convolver that manufacturers are capable of fabricating is on the order of 50 μsec. The device is relatively new and therefore poses considerable risk.

The advantages of SAW Convolvers are as follows:

1. Real time processing.
2. Programmable chip rate.
3. Programmable waveform.
4. IF/RF processing.

The disadvantages of SAW Convolver are;

1. Active correlation.
2. Large insertion loss.

3. Not cascadable.
4. Relatively large and bulky device.
5. Temperature sensitive.
6. Cannot be large scale integrated.
7. Relatively new technology.
8. Length of device is dependent on the duration of the pulsed signal.
9. Huge supporting circuitry.

ACOUSTO-OPTICAL EFFECTS

A beam of light, directed at an angle $\pm \alpha_\beta$ at a high frequency sound wave traveling in an acoustic substrate generates a new one whose direction differs by $\pm 2\alpha_\beta$ from the incident light with frequency offset by $\omega_\ell \pm \omega$, ω_ℓ, ω are respectively the frequencies of light and sound. The angle α_β is called the Bragg angle and is given by

$$\sin \alpha_\beta = \lambda/2\Lambda$$

where λ is the wavelength of light in the acoustic medium and Λ is the acoustic wavelength. The value $\sin \alpha_\beta$ refers to the angle observed inside the medium of sound propagation. Outside the medium, Snell's law changes the value of α_β to α_β' so that

$$\sin \alpha_{\beta'} = \lambda_v/\Lambda$$

where λ_v denotes the vacuum wavelength.

This is shown in figure 28. The diffracted beam of the difference frequency is called the -1 order corresponding to an incidence angle of $-\alpha_\beta$. Similarly an incidence angles of $+\alpha_\beta$ will produce the sum frequency and is called the +1 order.

The essential properties of Bragg diffraction can be thought of as an interaction between photons and phonons. The momenta of the interacting particles are given by $\hbar\bar{k}$ and $\hbar\bar{K}$ for the photon and phonon respectively, where $\hbar = h/2\pi$, h is Planck's constant, $\bar{k}$ and $\bar{K}$ are the propagation vectors of light and sound with magnitudes $1/\lambda$ and $1/\Lambda$ respectively. A vector diagram can be drawn for the interaction between the two waves and is shown in figure 29. The Bragg angle can be derived easily as follows:

$$\sin \alpha_\beta = \frac{1}{2}\frac{K}{k}$$

$$= \frac{1}{2}\frac{\lambda}{\Lambda}$$

Frequency up or downshift follow from the same diagram by considering energy conservation in terms of photon and phonon energies $\hbar\omega_\ell$ and $\hbar\omega$.

If the width of the transducer L is decreased, as shown in figure 30, the column of sound in the medium will look less and less like a plane wave. If L is further decreased, more and more orders are generated, as shown in figure 31. (Note that the n^{th} order is shifted in frequency by an amount $n\omega$). The extent to which this amplitude scattering process can operate is determined by the ratio of the angular width Λ/L of the sound radiation pattern to the separation λ/Λ between orders. A parameter Q, inversely proportional to this ratio is used and is defined as follows:

$$Q = \frac{K^2 L}{k} = \frac{2\pi\lambda L}{\Lambda 2}$$

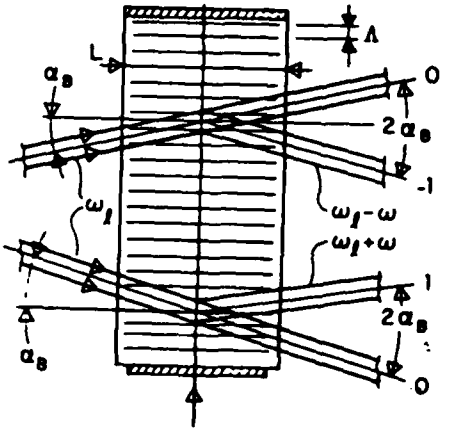


Figure 28. Bragg Diffraction Showing the Downshifted (Top) and Upshifted (Bottom) Interaction

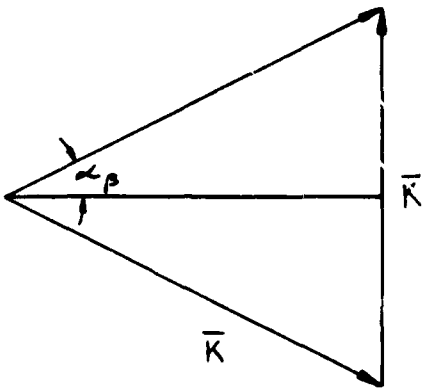


Figure 29. Wave Vector Diagram Illustrating Diffraction in the Bragg Region

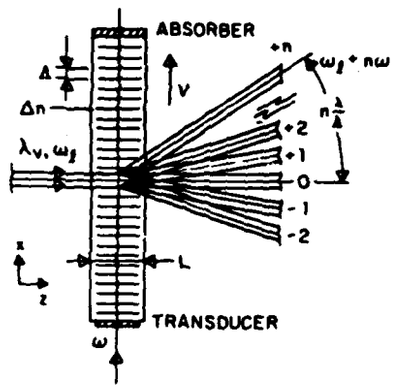


Figure 30. Diffraction in the Debye-Sears Region

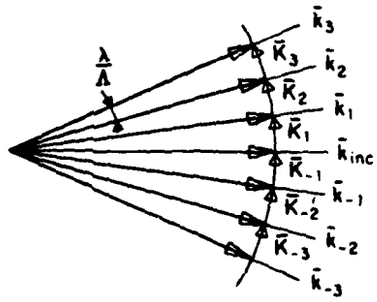


Figure 31. Multiple Scattering in the Debye-Sears Region

The condition for Bragg diffraction (the Bragg Region) is

$$Q = \frac{K^2 L}{k} >> 1$$

While the opposite condition of which $Q<<1$ denotes the multiple scattering regime where many orders may be generated. This latter regime is called the Raman-Nath or Debye-Sear region after early investigators. The amplitude of the diffracted order is given by

$$A_n = (-j)^n E_{inc} J_n(\hat{\alpha})$$

where E_{inc} is the amplitude of the incident light and $\hat{\alpha}$ denotes the peak phase drift of the light due to the peak refractive index variation. $J_n(\hat{\alpha})$ is the Bessel function and is shown in figure 32.

Quite often a modest sound pressure is used ($\hat{\alpha}<<1$) and effectively only the +1 and -1 orders are generated. In this case,

$$A_0 \approx E_{inc}$$

and

$$A_{\pm 1} = -j E_{inc} \cdot \frac{1}{2} \hat{\alpha}$$

In the case for Bragg Refraction

$$E_0 = E_{inc} \cos \frac{\hat{\alpha}}{2}$$

$$E_{\pm 1} = -j E_{inc} \sin \frac{\hat{\alpha}}{2}$$

This is shown in figure 33. For weak interaction ($\hat{\alpha}<<1$), these expressions being identical to the ones for diffraction in the Debye-Sears regime.

For maximum efficiency in Bragg diffraction, it is necessary that $\hat{\alpha} = \pi$.

Beam Deflection

One of the most important application of Bragg diffraction is beam deflection. This is shown in figure 34 where the diffracted beam of light is shown in three successive positions, corresponding to sound frequencies f_1, f_2, and f_3. Because the relation between deflection angle and frequency sweep is linear ($\alpha_{defl} = \lambda/\Lambda = \lambda f/v$) a simple frequency sweep is all that is needed to operate a beam deflector. In such a device, the number of resolvable angles N is determined by the ratio of the range of deflection angle $\Delta\alpha_{defl}$ to the angular spread (λ/d) of the light beam. N is given by the following equation,

$$N = \frac{\Delta\alpha_{defl}}{\lambda/d}$$

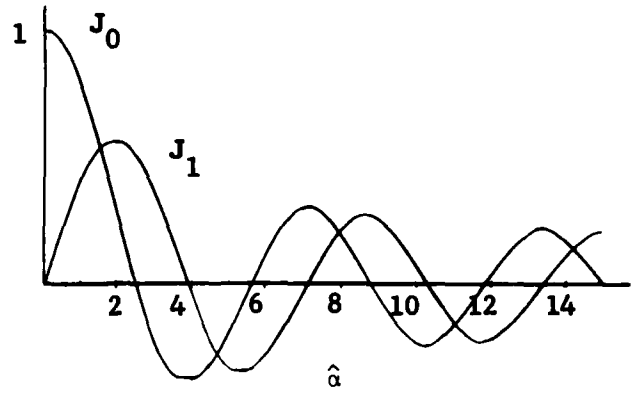


Figure 32. Amplitude of Diffracted Order Versus Peak Phase Shift $\hat{\alpha}$ in the Debye-Sears Region

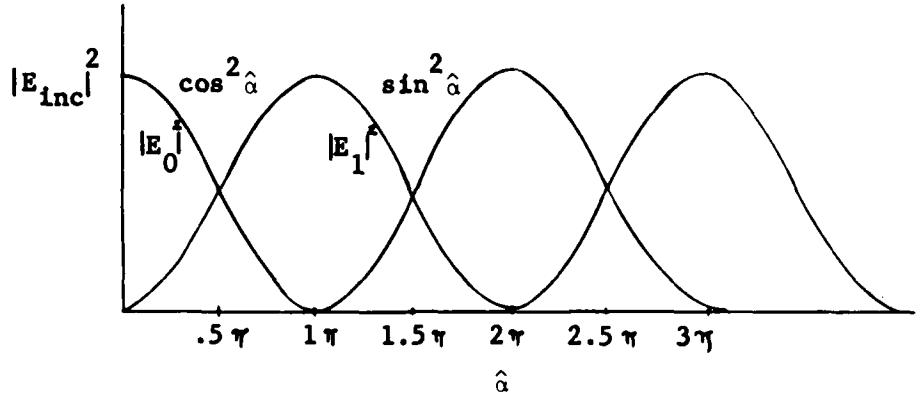


Figure 33. Intensity of Diffracted Order Versus Peak Phase Shift $\hat{\alpha}$ in the Bragg Region

As $\Delta\alpha$ is determined by the maximum frequency deviation Δf,

$$\Delta\alpha_{defl} = \frac{\lambda}{v} \Delta f$$

Combining these two equations,

$$N = \tau\Delta f$$

If the deflector is used to randomly access N positions by using N different frequencies within Δf, τ also denotes the access time of the system. Likewise, for general signal processing the parameter $\tau\Delta f$ represents the cell's time bandwidth product. Consideration of maximum optical aperture and finite acoustic loss limit this product at present to values of 1000 – 2000, with τ ranging from 1 μsec to 10 μsec and Δf from 10 MHz to 1 GHz.

In figure 34, three frequencies are fed into the sound cell simultaneously. Each frequency generates a beam at a specific angle. A lens can be used to converge these deflected light onto the focal plane as shown in figure 35. This plane will then display the frequency spectrum or Fourier transform of the sound signal. Notice that both the Bragg and Debye-Sear diffraction is capable of light reflection. However, if the cell is used in the Debye-Sears regime, filtering of the higher order components are required. The spatial resolution in the Fourier transform plane corresponds to a frequency resolution of $\delta_f = 1/\tau$.

Acousto Optical Correlator

The correlation function, in terms of analytic signal representation, can be written as
Re $\int \underline{S}_1^* (x-\tau)\, \underline{S}_2 (x)\, dx$.

where S_1, S_2 are the analytic signals to be correlated and * denote the complex conjugate. Figure 36 is an arrangement in which this correlation function can be implemented.

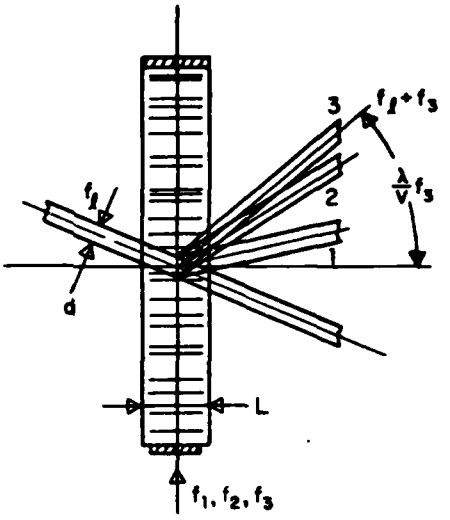


Figure 34. Bragg Diffraction Sound Cell Used As Beam Deflector

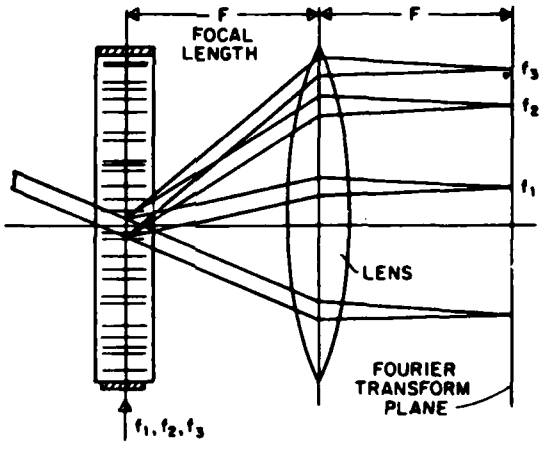


Figure 35. Bragg Diffraction Sound Cell Used As a Spectrum Analyzer

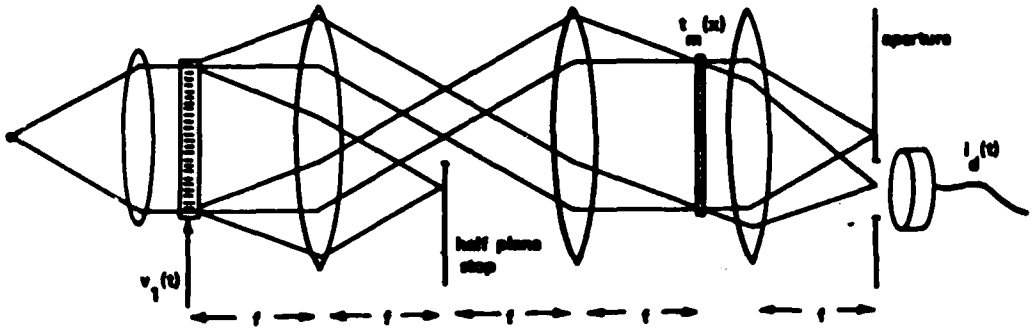


Figure 36. Acousto Optical Correlator With Zero and +1 Diffraction Components Imaged on Reference Mask $t_m(x)$

The light, originating from a point source, passes through a lens L_1 which is used as a collimator and forms, at the output of the lens, plane waves with the direction of motion perpendicular to the direction of motion of the acoustic wave. In going through the acousto-optical cell, the light is deflected into the ± orders with an undeflected zeroth order. The -1 order can be filtered out by first passing all of the components through a lens which converges the different parallel lights onto different points on the focal plane of the lens. The two frequencies can then be filtered out by blocking the unwanted -1 order component at the focal plane of the lens (which in a way, acts like a transformer) by means of a dark phototransparency or any opaque object. There will only be the +1 and 0 orders in the other side of the focal plane. A third lens, with the identical focal length, is used to convert these components into plane waves again before passing through the phototransparency, which has the reference signal on it. The transmittance function for the mask is given as follows:

$$t_m(x) = 1 + S_2(x)$$

$$= 1 + \frac{1}{2}\underline{S}_2(x) + \frac{1}{2}\underline{S}_2^*(x)$$

The wave amplitude incident on the mass can be written as,

$$U_{inc}(x, t) = 1 + \frac{1}{2}\underline{S}_1^*(x - vt)$$

where 1 is the zeroth order component and $\frac{1}{2}\underline{S}_1^*(x - vt)$ is the +1 order component.

The output of the mask is:

$$U_{trans}(x,t) = \left[1 + \frac{1}{2}\underline{S}_1^*(x - vt)\right] \left[1 + \frac{1}{2}\underline{S}_2(x) + \frac{1}{2}\underline{S}_2^*(x)\right]$$

$$= 1 + \frac{1}{2}\underline{S}_1^*(x - vt) + \frac{1}{2}\underline{S}_2(x) + \frac{1}{4}\underline{S}_1^*(x - vt)\,\underline{S}_2(x)$$

$$+ \frac{1}{2}\underline{S}_2^*(x) + \frac{1}{4}\underline{S}_1^*(x - vt)\,\underline{S}_2^*(x)$$

Of the above terms, only the second and fifth terms are retained. The rest of the terms can be filtered out by using a fourth lens and an opaque filter. The wave that enter the detector can then be written as follows,

$$U_d(x,t) = \frac{1}{2}\underline{S}_1^*(x - vt) + \frac{1}{2}\underline{S}_2^*(x)$$

The detector output is,

$$i_d(t) = \int \left| u_d(x,t) \right|^2 dx$$

$$= \frac{1}{4} \int \left| \underline{S}_1^*(x - vt) \right|^2 dx + \frac{1}{4} \int \left| \underline{S}_2^*(x) \right|^2 dx$$

$$+ \frac{1}{2} R_e \int \underline{S}_1^*(x - vt)\, \underline{S}_2(x)\, dx$$

The last term in the above equation is the desired correlation function. The term can be separated by filtering the other two terms.

It should be noted that another acousto-optical cell can be used for the reference signal instead of the phototransparency. The reference signal in this case, has to be injected into the acousto-optical cell from the opposite side of the first cell.

The advantages of acousto-optical correlator are,

1. Real time processing.
2. RF processing.
3. Programmable chip rate.
4. Programmable waveform.

The disadvantages of acousto-optical correlator are,

1. Active correlation (can be passive, but then, the waveform can no longer be programmable).
2. Not cascadable.
3. Requires the use of a collimated light source (which can be a laser).
4. This is a hybrid approach that involves light, acoustics and electronics.
5. Introduction of optical deficiency into system.
6. Relatively new technology for integrated optics.
7. Risky approach.
8. High insertion loss (10 dB per Bragg cell).
9. Relatively temperature sensitive.

FIBER OPTICAL TAPPED DELAY LINE

An optical fiber can be used as a delay line since the refraction index is larger than one. With the large bandwidth associated with the fiber optical cable, this may be the ideal approach for the processing of certain waveforms. The propagation attenuation for several typical delay material is shown in table IX.

It can be seen that among the different materials that can be used as delay lines, the surface acoustic wave devices are the best under 1 GHz. For application above 1 GHz, optical fiber seems to be the optimum choice for low loss applications.

Table IX

Propagation Attenuations for Several Typical Delay Materials in Db/μs

	Frequency in GHz			
	0.01	0.1	1.0	1.0
Acoustic Waves	10^{-4}	10^{-2}	1	100
Coax RG-19/U	–	5	28	125
Coax RG-58/U	–	14	35	>500
X-Band Waveguide	–	–	–	30
Optical Fiber	0.4	0.4	0.4	0.4

Internal Reflection

An optical fiber consists of a core fiber having a refractive index n_1 surrounded by a cladding having an index n_2 where $n_1 > n_2$ as shown in figure 37. A light ray, striking the interface between the core and cladding at an angle greater than the critical angle θ_c, given by $\sin \theta_c = n_2/n_1$ will be totally internally reflected back into the core. Consequently, the ray will be confined to and propagate down the optical fiber. The maximum angle with which a ray can enter the end of the fiber and still be totally internally reflected is given as follows,

$$\sin \phi_{max} = \left(n_1^2 - n_2^2\right)^{1/2}$$

which is also known as the numerical aperature (N.A.) of the fiber. The angle is measured from the normal to the end of the optical fiber.

Since the fiber is acting as an electromagnetic waveguide, only certain angles or nodes satisfy the boundary conditions. The total number of nodes possible for such a guide is given approximately as,

$$N \simeq 4.9 \left(\frac{d\,N.A.}{\lambda}\right)^2$$

where

λ is the optical wavelength and

d is the core diameter.

It can be seen that as the core diameter is decreased, the angle within which total internal reflection is reduced. There is a point at which only a single mode will propagate. The condition for single mode operation is,

$$d \leq (0.76\ \lambda/N.A.)$$

A step index guide is the name given to optical fiber that is made up of a core of uniform refractive index n_1, and a cladding of uniform refractive index n_2. However, guiding can be achieved in a fiber in which the refractive index follow any number of profiles with a maximum at the center. One such useful symmetric profile is the parabolically graded guide, in which the number of modes is half the value for step index fiber having the same core radius and index difference.

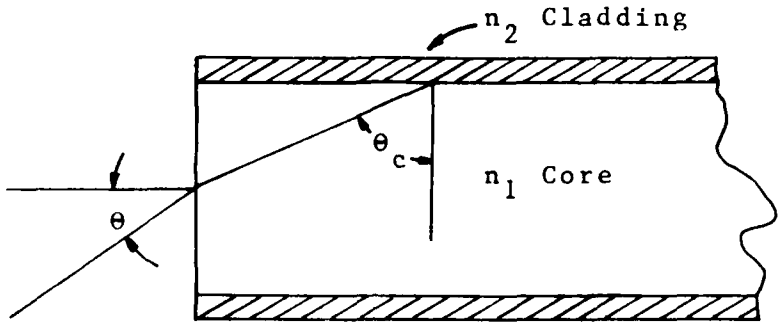


Figure 37. A Typical Multi-Mode Optical Fiber Configuration

To determine if optical fiber are suitable for signal processing application (as a tapped delay line), two aspects has to be considered. The first is the time delay and the second is the loss. The bandwidth does not seem to be a problem as can be seen in table VIII.

Time Delay for Optical Fiber

If a refractive index of 1.5 is used, the propagation delay will be 5 μs/km. For a short code such as the one used in JTIDS, the duration of the code is 6.4 μsec. Therefore the required length is 1.28 km which is unusually long to say the least. For the GPS C/A code, the required length is computed to be 200 km which is not at all practical to implement.

Bending Loss

To use any optical fiber in a practical manner, the long fiber lengths involved for even moderate time delay require the fiber to be wound into a coil. However, as the radius of curvature is reduced, the number of modes which the fiber will support is reduced and the higher order modes are radiated. The number of modes that a curved guide will support is given as follows,

$$N(R) = N \left[1 - \left(\frac{d}{2R\,\Delta n}\right)\right]$$

and

$$N_\rho(R) = N_\rho \left[1 - \frac{d}{R\,\Delta n}\right]$$

for step index and parabolic index fibers respectively, where R is the radius of curvature.

If it is assumed that a step index fiber with a core diameter of 0.1 mm and index difference Δn = 0.01, half the modes (3 dB) would be lost for a radius of curvature of 1 cm.

Coupling Loss

1. Input Coupling Loss

Input coupling losses occur at the source/fiber interface. Usually, a short length of fiber (called a pigtail) is permanently attached to the source's emitting area in single fiber communications. Any mismatch between this emitting area and the pigtail's core area results in the first input coupling loss factor. Additional loss occurs if the core area is smaller than the source's emitting area. A rough value of this fractional loss is simply the ratio of the core area to the emitting area.

A second input coupling loss factor relates to the light gathering ability of the fibers themselves. This was expressed as the numerical aperture previously. Fibers with a large numerical aperture will be able to couple more light then fibers with smaller numerical aperture.

The last, and least important, input coupling loss comes from light reflection from the input end of the fiber. A relatively small loss, it usually equals about 0.2 dB.

2. Output Coupling Loss

At the receiving end of the link, output coupling losses are generally not as severe as input losses. Total output losses is estimated to be at approximately 1 dB.

Access Coupler Loss

To use the fiber optical cable as a tapped delay line, the signal has to be tapped off at different points. There are two ways in which this can be accomplished.

The generic T coupler design is shown in figure 38. Signals can be injected or tapped off onto the main optical fiber at each coupler. In one approach, the tap fiber is fused in the main trunk fiber, bring the cores closely enough together to provide transfer of energy between the fibers. The amount of light coupling can be adjusted by changing the core-to-core spacing and interaction length.

The star configuration which is shown in figure 39 offers an alternative coupling solution for multiterminal applications. The operation is as follows.

A signal is transmitted on a dedicated fiber towards the coupler. Entering the coupler at port 1, the light spreads out and is reflected by the dielectric mirror into all ports.

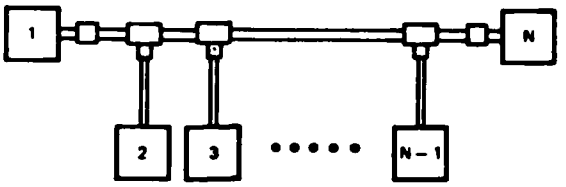
Between the two couplers, the T approach suffers loss at every place it is tapped. Worst case system loss (in decibels) increase linearly with the number of terminals. On the other hand, the star configuration introduces only one coupler loss. Furthermore, as the number of port increases, the star coupler's power splitting loss increases only as 10 log n. The results of a loss comparison analysis is shown in figure 40 in which worst case system losses are plotted against the number of terminals for both in-line and star-coupler configurations. It is assumed that a T coupler has constant 10 dB tap loss, plus 2 dB insertion loss and a 7 dB insertion loss for the star design. Values of 1 and 3 dB for connector and I/O splitting losses are also assumed. It can be seen that the star design has lower loss for systems with more than about 4 terminals.

Since to use the optical fiber as a tapped delay line requires the coupler to be placed at predetermined fixed distances away and that the tapped signal only has to feed one multiplier, there exist no advantage for the star configuration.

Figure 41 shows the configuration of a correlator using optical fiber as a tapped delay line. The correlator cannot be made programmable to accommodate signals with various chip rates.

Notice in the diagrams also that a receiver that converts the light energy into electrical signal is required for each coupler before the multiplication process. This means that more hardware is required in addition to the tremendous length of the optical fiber itself. The advantages of fiber optical delay line are,

1. Passive correlation.
2. Real time RF signal processing.
3. Very wide bandwidth.
4. Immune to electromagnetic interference.
5. Relatively constant insertion loss per unit length up to 10 GHz.
6. Relatively temperature independent (10 ppm/°c)



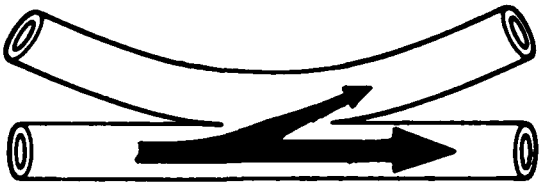


Figure 38. The T Optical Coupler Configuration

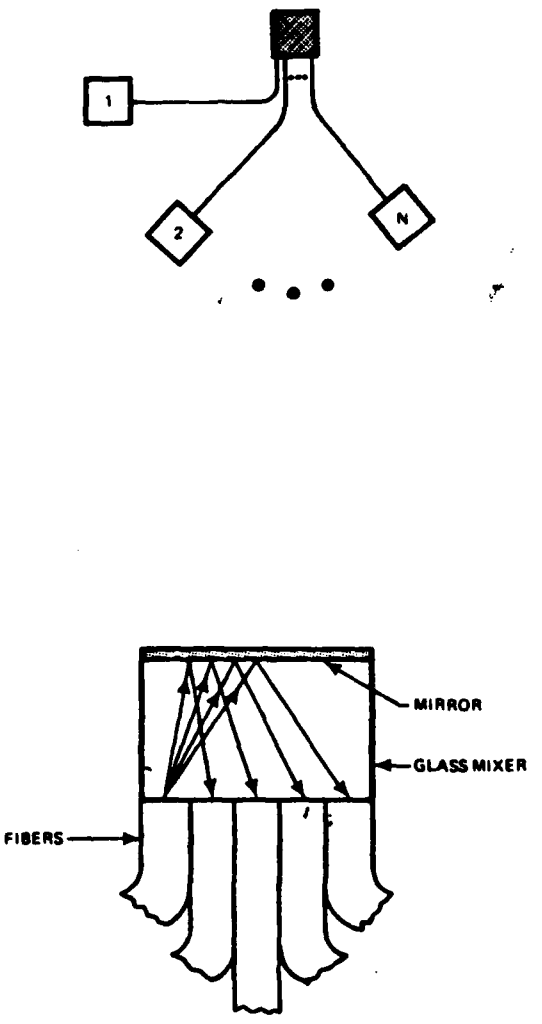


Figure 39. The Star Optical Coupler Configuration

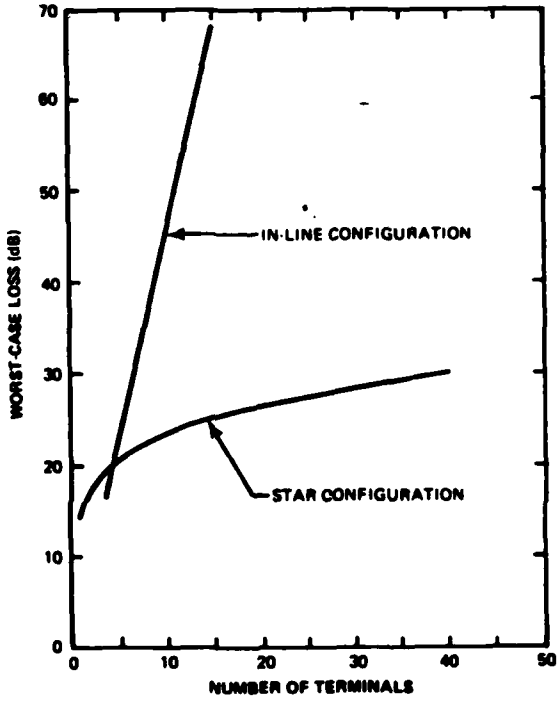


Figure 40. Worst Case Loss for the Star and T Optical Coupler Configurations

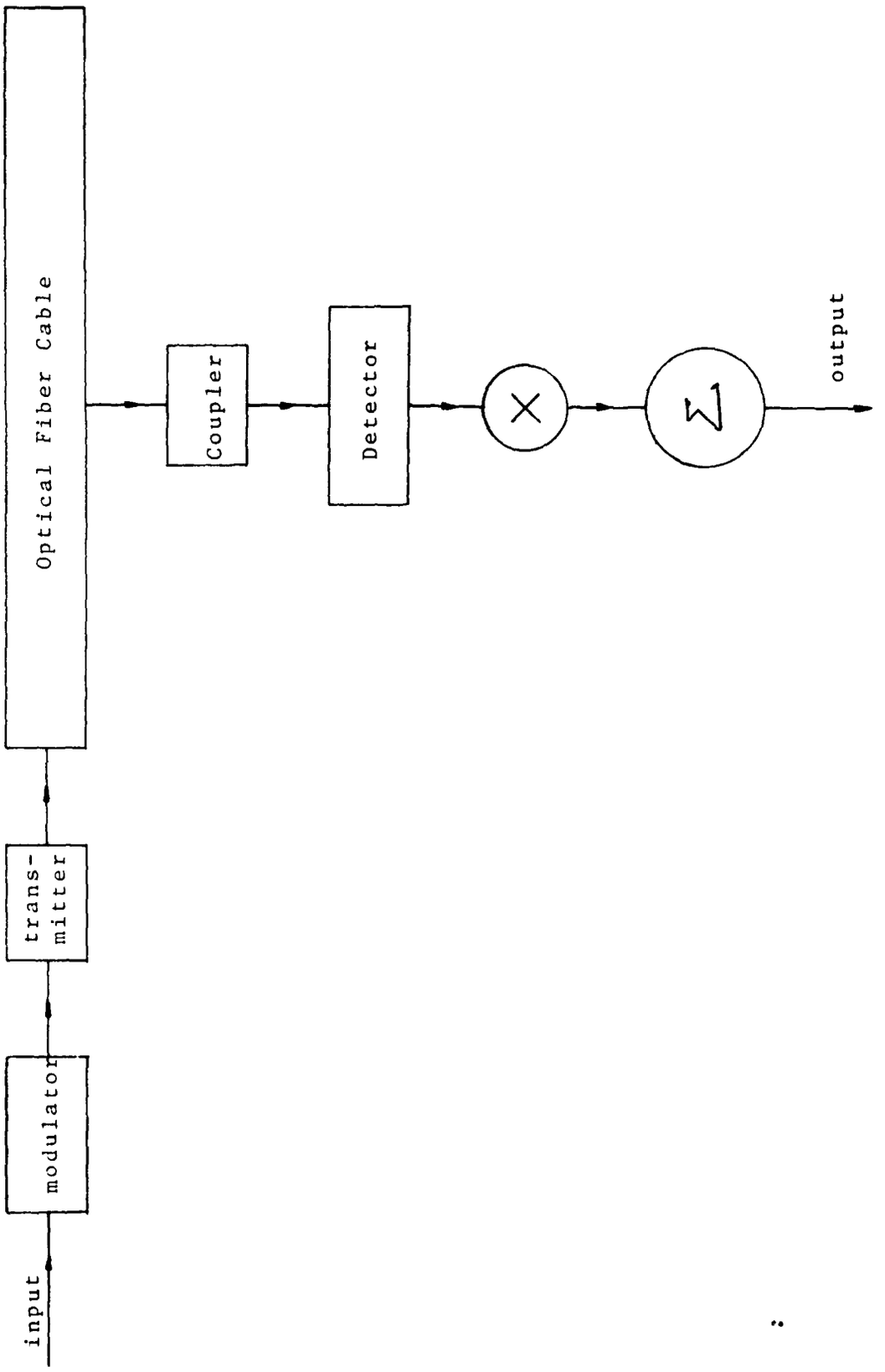


Figure 41. An Optical Fiber Tapped Delay Line Correlator

The disadvantages of fiber optical delay line are,

1. Extremely long fiber optical cable for the required delay (200 km/1 μsec delay).
2. Large and bulky device when wound into a coil.
3. High insertion loss for the required delay (1 dB/km).
4. Required optical receiver at each coupling point.
5. Programmability can be acquired at great loss.
6. Cannot be large scale integrated.

DIGITAL CIRCUITS

Digital circuits are primarily baseband devices. Like any conventional digital signal processor, the input signal has to be down converted into a baseband signal and digitized into binary numbers for processing. A diagram that shows the general procedure for signal processing using digital hardware is shown in figure 42. Notice that unlike the previous methods discussed, the digital approach requires a frequency translator, a sample and hold circuit as well as an analog to digital converter. A digital to analog converter is usually needed since the output has to be in analog form eventually.

As can be seen, two channel quadrature processing are required in general since the phase of the incoming signal is unknown. Each of these two signals are digitized into the corresponding binary equivalents of the analog samples. In the case of 1-bit quantization, the A/D converters are in reality comparators which effectively perform hard limiting of the signal amplitude at the sample times. Thus, the signal available to the remainder of the unit is a binary sequence representing a 1-bit quantization of the input signal at the sample times. It is well known that such hard limiting of the input signal in the presence of a strong interfering function such as a CW jammer will allow the strong function to capture the limiter and cause a loss in signal detectability. To prevent this, multi-level quantization is required. Since two channels are used, the input signal can be sampled at half the Nyquist rate at each channel. Figure 43 shows the configuration of a digital correlator for one of the channels.

Depending on the number of levels of quantization, the appropriate sets of shift registers are used. A sample-by-sample multiplication in each channel with a digital reference sequence are made. For each product the correlators generate a unit output current for each pair of coincident states. These output currents for each digit are summed separately and the current sum is then weighted according to the significance of the digit it represents. The weighted components of each channel are then summed by a resistor network. The correlation function is given by the following equation,

$$A = \sqrt{I^2 + Q^2}$$

where

A is the correlation amplitude.

I is the inphase correlation function.

Q is the quadarature correlation function.

There are many places where noise may get into the digital system. These are discussed in appendix A.

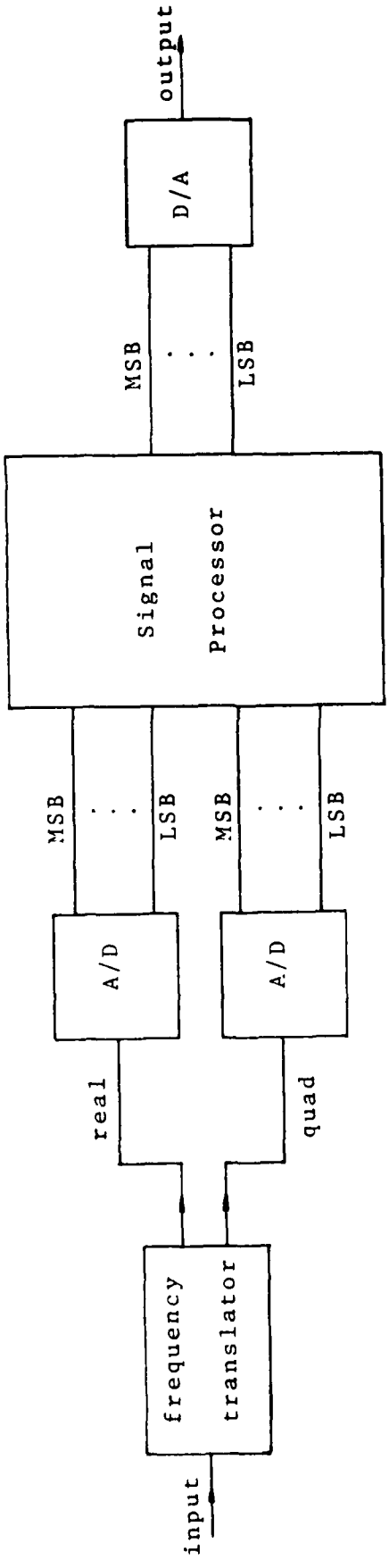


Figure 42. A Digital Signal Processor

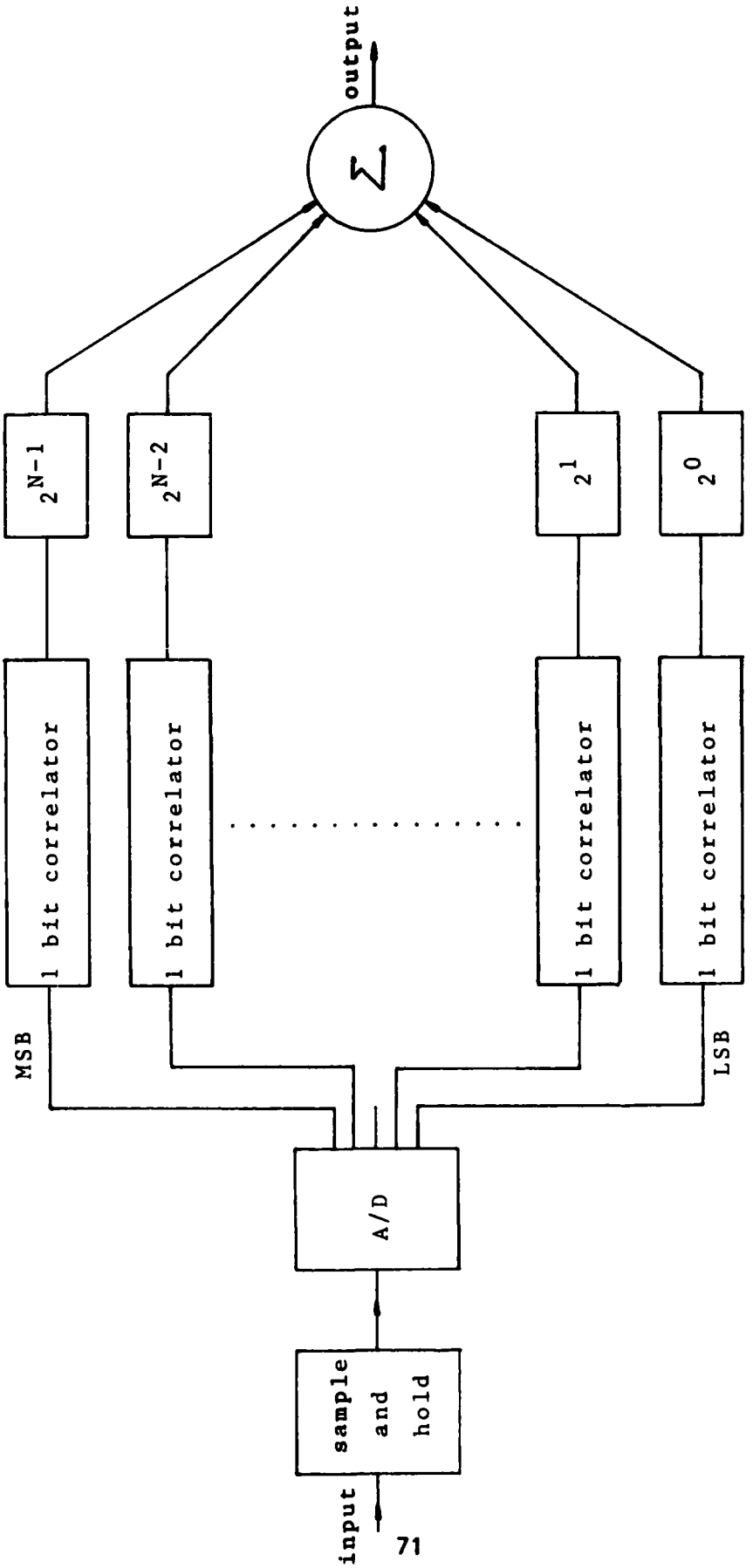


Figure 43. A Digital Correlator

Digital technology is a fast advancing, mature technology that is extremely reliable. The major components for the digital approach are sets of shift registers and multipliers. However, simplications can be made usually so that the multiplier can be eliminated. The major drawback for the digital approach is power consumption and the amount of external hardware required for the support of the correlator. However, the main advantage being that the digital correlator is fully programmable as far as waveform and data rate are concerned. Furthermore, the digital memories can store the reference and input sample indefinitely and it is a well known fact that digital devices can be large scale integrated. The advantages of digital circuits are,

1. Passive correlation.
2. A mature technology that poses a low risk factor.
3. No insertion loss.
4. Programmable chip rate and waveform.
5. Can be large scale integrated.
6. Temperature insensitive.
7. Cascadable correlator.

The disadvantages of digital circuits are,

1. Requires samples and hold, analog to digital conversion circuits.
2. Introduction of noise due to the A/D conversion process.
3. High power consumption.
4. Relatively small bandwidth.
5. Baseband processing.
6. Multiple channel required for multiple bit quantization.

CHARGE COUPLED DEVICES

Theory of Operation

Charge coupled devices (CCD) operate by transferring packets of minority charge, which represent analog signals, from one potential well to another. These potential wells are formed by a linear array of deeply depleted MOS capacitors, either on a uniformly doped substrate (surface channel, SCCD) or on a substrate with a thin, depleted layer of opposite conductivity at the surface (buried channel, BCCD). In either case, the minimum potential energy of these wells is determined by the voltage applied to the gate electrode, so that the appropriate manipulation of the phase voltages causes charge packets to transfer along the line, since charge always moves to the local potential minimum. This is shown in figure 44.

It can be seen that the CCD can be used as an analog delay line in which the incoming signal is sampled and transferred down the line in an analog fashion like a digital shift register by applying a potential difference to the cells in the charge coupled device. This voltage difference is the clock to the analog delay line. Notice that unlike a digital circuit in which the clock can be a single phase square wave, a multiphase clock is required to propagate the charge down the delay line. This is due to the fact that a single phase clock applied to all CCD cells represents no voltage difference between the cells at the instant of application and hence no charge transfer.

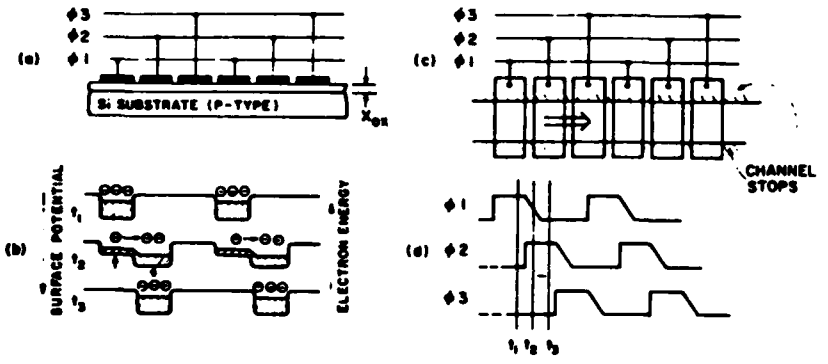


Figure 44. Basic Charge Transfer Action With a Three-Phase, n-Surface-Channel CCD.
a. Cross-Sectional View
b. Surface Potential Profile at Three Different Times
c. Top View of the CCD
d. The Clock Waveform

The most important figure of merit for a charge transfer device is charge transfer efficiency denoted by η, the fraction of the original charge packet that is transferred from one storage site to the next. The fraction that is not transferred is denoted by ϵ, the charge transfer inefficiency, or charge transfer loss. Since the potential well used to store and transfer minority carriers also serve to repel majority carriers, recombination is negligible. Thus, minority carriers are lost from the original charge packet only by being left behind; i.e. charge is either transferred or it is not. Thus,

$$\eta + \epsilon = 1$$

The effect of imperfect transfer efficiency is to decrease the amplitude of the signal packet, so that after η transfers, the ratio of the signal level A_n to the original level A_o is given by,

$$A_n/A_o = \eta^n$$

$$= (1 - \epsilon)^n$$

$$\approx e^{-\epsilon n}$$

for small ϵ.

Since many CCD applications requires in excess of 1000 transfers, ϵ must be very small to prevent excessive signal degradation.

CCD's perform many of the same functions as surface wave devices (SWD's) such as time delay, transversal filtering, etc. There are two important differences, however: (1) CCD's operate at much lower frequency and can achieve a much longer time delay than surface wave delay lines; and (2) CCD's are sampled data, which means that the input analog signal is sampled and charge packets representing the analog samples are processed in the device.

Figure 45 shows a way in which a charge coupled device can be used as a correlator. Notice that this is identical to the way digital devices are used as a correlator in which a frequency translator as well as a sampler are required. However, there is no need for an A/D converter and since the sampled signal is in analog form, one analog delay line per channel is sufficient and consequently, fewer number of devices can be used. The sampling circuit, unlike a sample and hold circuit for the digital approach, is usually incorporated into the CCD delay line when the cells are made on a substrate.

Unlike digital circuits, which has states determined by whether the output transistor is being turned on or off, thereby providing infinite storage (provided that the power supply is on), the charge coupled device begins to lose the charge through thermal leakage the moment the signal is sampled. The storage time therefore, is dependent on how fast the capacitor is discharging.

By the same reasoning, the maximum clock rate for a charge coupled device is dependent on the rate that the capacitor charges. Commercially available CCD delay lines can be clocked from 10 KHz to 20 MHz. The advantages of Charge Coupled Devices are,

1. Does not require quantization.
2. One delay line per channel operation.
3. Cascadable.
4. Programmable chip rates.
5. Programmable reference waveform.

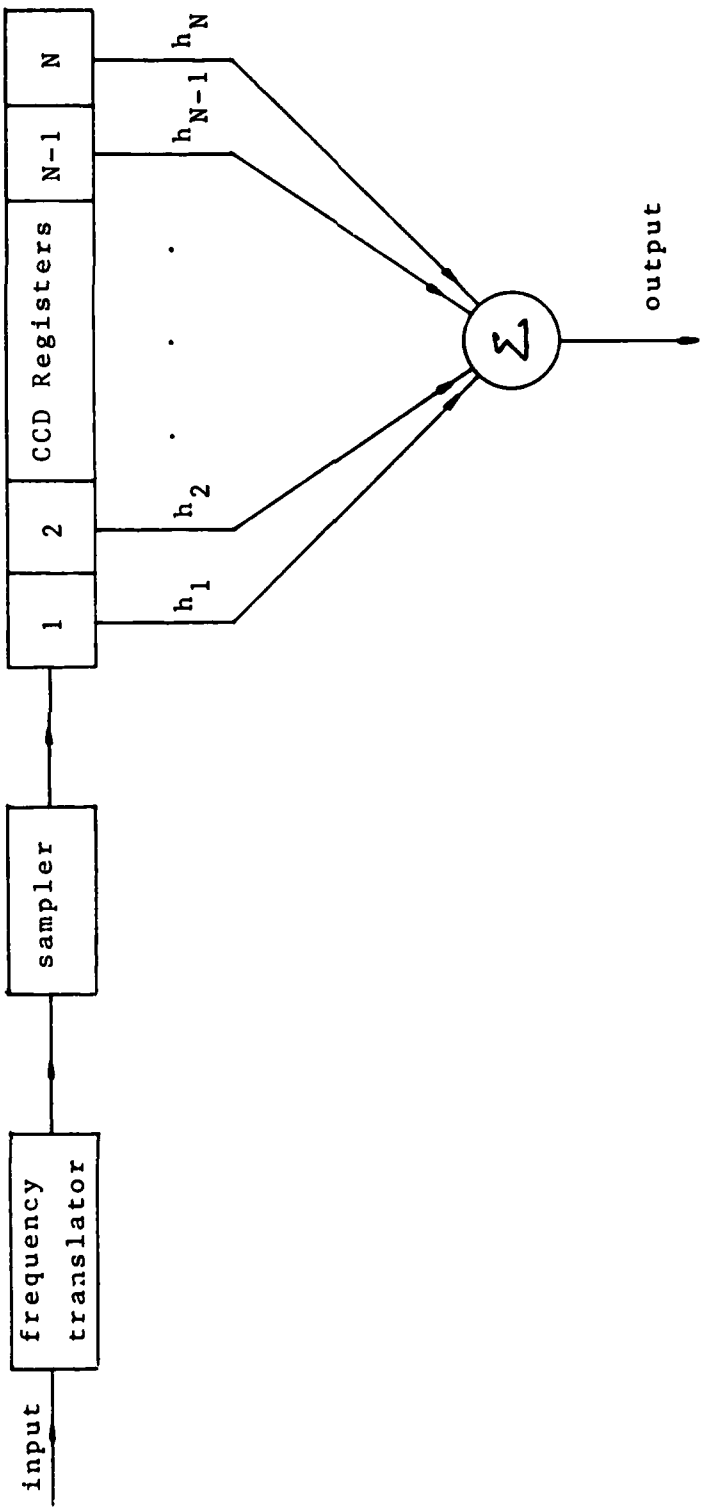


Figure 45. A CCD Correlator

6. Passive correlation.
7. Small size.
8. Low power consumption.
9. Can be large scale integrated.

The disadvantages of charge coupled devices are,

1. Baseband process that required a frequency translator and sampling circuits.
2. Relatively small bandwidth.
3. Transfer efficiency needs improvement.
4. Temperature sensitive.

TRADE-OFF ANALYSIS

Table X is a comparison between the various technologies that can be used for correlation together with the criteria for an ideal correlator. One most important criteria for a universal correlator, as was shown earlier, is that of programmability since the various waveforms have different data rates and modulations.

Surface acoustic wave delay line and the fiber optical cable can be eliminated since they are not programmable.

Among the RF/IF processing devices, the convolver has an advantage over the acousto-optical devices since no collimated light source or optical elements are involved.

Among the baseband processing devices, the charge coupled device has an edge over digital circuit in that no quantization is required and consequently the number of devices required is reduced.

Since a convolver has a larger bandwidth than either of the baseband devices and the fact that it can be used at IF/RF makes it the ideal choice for the correlation function. However, to use the convolver for initial synchronization purpose will require a relatively long time as discussed earlier. The present state-of-the-art for convolver design has a maximum processing time of 50 μsec which is not quite adequate for JTIDS application nor is appropriate for GPS unless partial correlation is considered.

Since these devices are not cascadable, the time delay cannot be increased by using two or more convolvers. The same problems are faced by the acousto-optical approach with the aperture length of the Bragg cell to be 10 μsec. Therefore, it seems that the baseband devices are the ones to use for the correlation function for short term purpose. However, it should be kept in mind that with the advance in technology in the production of SAW devices, it would not be long for researchers to come up with a convolver that meets the delay requirements for the various waveforms of interest. Notice that the requirement in using a convolver is that the two incoming waveforms has to coexist during the correlation interval. Consequently, the timing requirement is greatly reduced. Correlation will occur even the reference signal does not coincide exactly with the incoming signal. It seems difficult, if not impossible, to push the processing speed of either digital circuit or CCD's to IF (RF) processing.

Table X. Correlator Technology Comparisons

	Ideal Correlator	SAW TDL	Acousto-Optics	Fiber Optics	Digital Circuits	Charge Coupled Device	Convolver
Active or Passive	Passive	Passive	Active	Passive	Passive	Passive	Active
Power Consumption	Low	Medium	Medium	Medium	Medium	Low	Medium
Ease of Implementation	Easy	Medium	Difficult	Medium	Easy	Easy	Difficult
Bandwidth	Large	Medium	Medium	Large	Small	Small	Medium
Risk	Low	Medium	High	Medium	Low	Medium	High
Size	Small	Medium	Large	Large	Medium	Small	Large
Weight	Small	Small	Large	Large	Medium	Medium	Medium
Cost	Low	Medium	High	Medium	Low	Low	High
Program-mability	Yes	No	Yes	No	Yes	Yes	Yes

Digital technology is a much more mature technology than CCD's which needs improvements in transfer efficiency, dynamic range, as well as temperature sensitivity. Based on technology trade offs, it seems that the digital approach is the present solution that can be used for the implementation of the correlator. The CCD approach will be the followup candidate while the SAW convolver, which is a relatively new technology, is most risky and costly. The use of acousto-optical Bragg cells seems to be an impractical approach. The situation can be changed only if integrated optic which is at its infancy proved to be feasible.

SIGNAL GENERATION

A bandpass signal can be represented as follows,

$$f(t) = x(t) \cos \omega_c t - y(t) \sin \omega_c t$$

where x(t) and y(t) are known as the inphase modulation component and quadrature modulation component respectively. A simple scheme of implementing the above equation is shown in figure 46. It can be seen that to generate a bandpass function requires two multipliers and one adder.

The GPS signal, which is a composite binary phase shift keying signal, can be written as follows.

$$f(t) = (D \oplus PN_I) \cos \omega_c t + (D \oplus PN_Q) \sin \omega_c t$$

where $\oplus$ is the "exclusive OR" operation defined as follows,

$$0 \oplus 0 = 0$$

$$0 \oplus 1 = 1$$

$$1 \oplus 1 = 0$$

$$1 \oplus 0 = 1$$

and D is the baseband data. PN_I and PN_Q are the inphase and quadrature pseudo sequences respectively. A more generalized expression can be obtained by using different data for the real and quadrature channels as follows.

$$f(t) = (D_I \oplus PN_I) \cos \omega_c t + (D_Q \oplus PN_Q) \sin \omega_c t$$

A minimum shift keying signal can be expressed as follows.

$$f(t) = (PN_I) \cos \omega_m t \cos \omega_c t + (PN_Q) \sin \omega_m t \sin \omega_c t$$

where PN_I and PN_Q are again the inphase and quadrature pseudo random sequences and $\cos \omega_m t$ and $\sin \omega_m t$ are the inphase and quadrature phase amplitude modulating signals. Notice that the cyclical shifting of the PN sequences in each burst of the signal represents a fixed number of bits of baseband data. Therefore, the data does not appear in the expression for the MSK signal.

A comparison between the two equations indicates that due to the $\cos \omega_m t$ and $\sin \omega_m t$ terms in the MSK expression, one amplitude modulator is required on each channel of the MSK modulator. A modulator that is capable of generating both the MSK and composite binary phase shift keying signals is shown in figure 47.

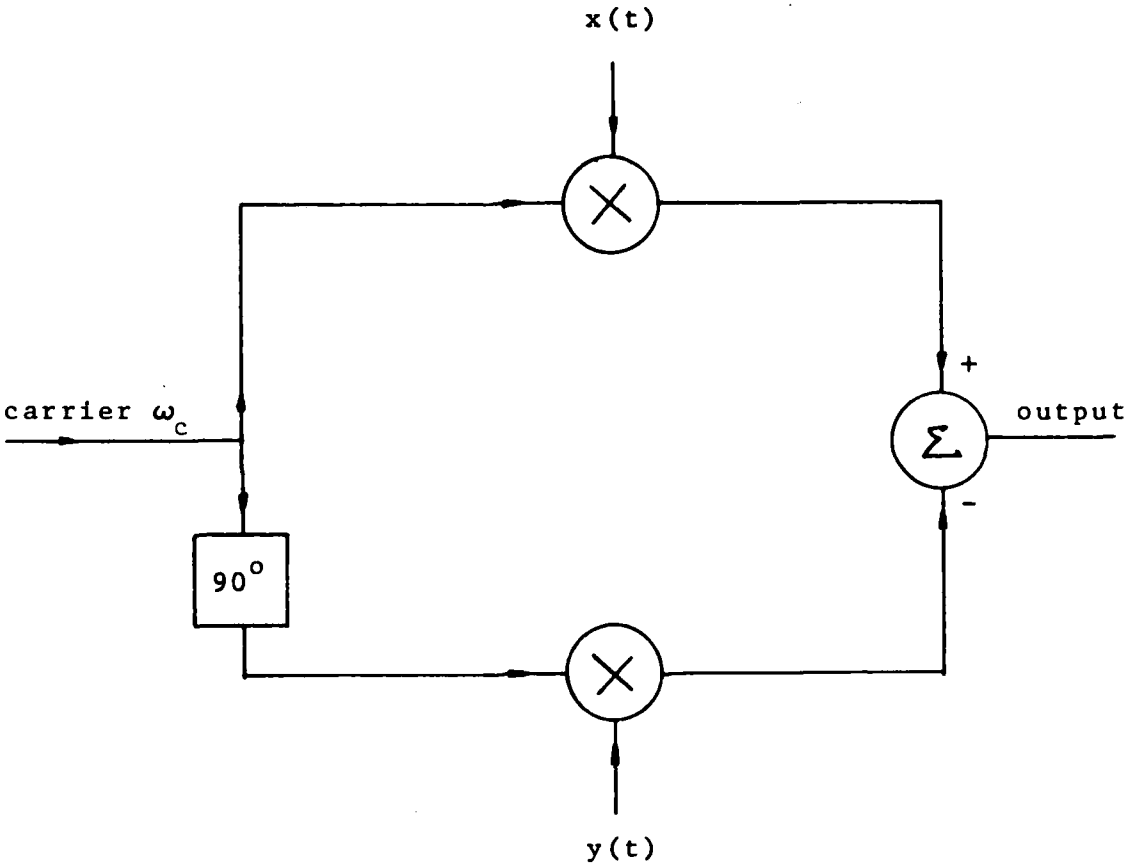


Figure 46. Generation of a Bandpass Signal

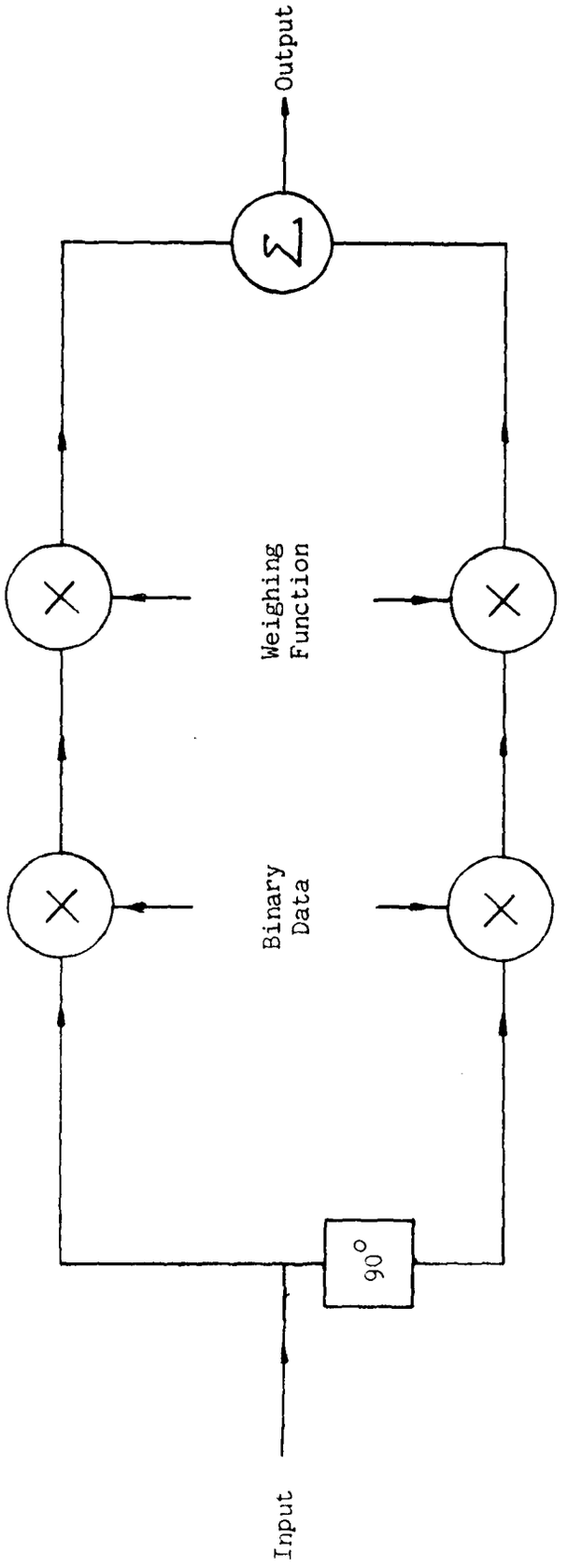


Figure 47. An MSK/BPSK Signal Generator

The incoming carrier is split into an inphase and quadrature phase components (cos $\omega_c t$ and sin $\omega_c t$. The first multiplier (a mixer in reality) in each channel is used to provide biphase modulated signals which represent (D $\oplus$ PN_I) cos $\omega_c t$ (D $\oplus$ PN_Q) sin $\omega_c t$ or PN_I cos $\omega_c t$ PN_Q sin $\omega_c t$ depending on the desired waveform. For the GPS case, the input to the second mixer can be set to a specific level so that the biphase signal will pass through the mixer unchanged. For the MSK signal, an amplitude modulating function can be used. The outputs are then summed together by an adder to produce the required signal. Since timing is very critical in the MSK case, extra circuitries are required to ensure the correct timing.

CORRELATION ANALYSIS FOR BASEBAND PROCESSING

As seen in the previous discussion, the incoming signal can be represented as,

$$f(t) = x(t) \cos(\omega_c t + \theta) + y(t) \sin(\omega_c t + \theta) \tag{1}$$

Multiplication by cos $\omega_c t$ of the above expression yields the following,

$$\begin{aligned} f(t) \cos \omega_c t &= x(t) \cos(\omega_c t + \theta) \cos \omega_c t + y(t) \sin(\omega_c t + \theta) \cos \omega_c t \\ &= \frac{x(t)}{2} \left[\cos(\omega_c t + \theta - \omega_c t) + \cos(2\omega_c t + \theta)\right] + \\ &\quad \frac{y(t)}{2} \left[\sin(\omega_c t + \theta - \omega_c t) + \sin(2\omega_c t + \theta)\right] \\ &= \frac{x(t)}{2} \left[\cos\theta + \cos(2\omega_c t + \theta)\right] + \frac{y(t)}{2} \left[\sin\theta + \sin(2\omega_c t + \theta)\right] \end{aligned} \tag{2}$$

Low pass filtering of the above expression yields,

$$I(t) = \frac{1}{2} x(t) \cos\theta + \frac{1}{2} y(t) \sin\theta \tag{3}$$

Similarly, multiplication of the incoming signal by sin $\omega_c t$ and low pass filtering the result yields,

$$Q(t) = \frac{1}{2} x(t) (-\sin\theta) + \frac{1}{2} y(t) \cos\theta \tag{4}$$

The following correlations can then be made,

1. The inphase channel samples with the inphase PN reference sequence.
2. The inphase channel samples with the quadrature PN reference sequence.
3. The quadrature channel samples with the inphase PN reference sequence.
4. The quadrature channel samples with the quadrature PN reference sequence.

The results are listed below:

1. $$I_{II} = \frac{1}{2} x(t) \cos\theta R_I + \frac{1}{2} y(t) \sin\theta R_I \tag{5}$$

2. $$I_{IQ} = \frac{1}{2} x(t) \cos\theta R_Q + \frac{1}{2} y(t) \sin\theta R_Q \tag{6}$$

3. $Q_{QI} = -\frac{1}{2} x(t) \sin\theta R_I + \frac{1}{2} y(t) \cos\theta R_I$ (7)

4. $Q_{QQ} = -\frac{1}{2} x(t) \sin\theta R_Q + \frac{1}{2} y(t) \cos\theta R_Q$ (8)

An assumption has to be made here that the cross correlation between the inphase and quadrature PN references are small. With this assumption, the equations can be reduced as follows,

1. $I_{II} = \frac{1}{2} x(t) \cos\theta R_I$ (9)

2. $I_{IQ} = \frac{1}{2} y(t) \sin\theta R_Q$ (10)

3. $Q_{QI} = -\frac{1}{2} x(t) \sin\theta R_I$ (11)

4. $Q_{QQ} = \frac{1}{2} y(t) \cos\theta R_Q$ (12)

The following results can be observed:

1. I_{II} is the projection of a vector of magnitude $\frac{1}{2} x(t) R_I$ in the positive x direction.
2. I_{IQ} is the projection of a vector of magnitude $\frac{1}{2} y(t) R_Q$ in the positive y direction.
3. Q_{QI} in the projection of a vector of magnitude $\frac{1}{2} x(t) R_I$ in the negative y direction.
4. Q_{QQ} is the projection of a vector of magnitude $\frac{1}{2} y(t) R_Q$ in the positive x direction.

If the sum of the two inphase components are formed ((9) + (12)), the following expression can be obtained:

$$\begin{aligned} R_I &= \frac{1}{2} x(t) \cos\theta R_I + \frac{1}{2} y(t) \cos\theta R_Q \\ &= \frac{1}{2} [x(t) R_I + y(t) R_Q] \cos\theta \end{aligned} \quad (13)$$

Taking the difference for the quadrature channels ((10) - (11)) yields the following,

$$\begin{aligned} R_Q &= \frac{1}{2} y(t) \sin\theta R_Q - \left[-\frac{1}{2} x(t) \sin\theta R_I\right] \\ &= \frac{1}{2} [y(t) R_Q + x(t) R_I] \sin\theta \end{aligned} \quad (14)$$

A vector diagram can again be drawn with R_I and R_Q to be the two components. For phase modulating signals, R_I and R_Q represent the correlation function for the two channels each with a different PN sequence. Two simultaneous outputs can be obtained from R_I and R_Q. For a MSK signal, the situation is slightly different. The correlation output is the sum of the two components. The magnitude can be seen easily to be 1/2 [y (t) R_Q + x (t) R_I].

This is the desired correlation function, which can be obtained by taking the square root of the sum of the sequences of the individual components to eliminate the sin θ and cos θ scalers as follows;

$$
\begin{aligned}
R &= \sqrt{[x(t)\,R_I \cos\theta + y(t)\,R_Q \cos\theta]^2 + [y(t)\,R_Q \sin\theta + x(t)\,R_I \sin\theta]} \\
&= \sqrt{[x(t)\,R_I + y(t)\,R_Q]^2 \cos^2\theta + [y(t)\,R_Q + x(t)\,R_I]^2 \sin^2\theta} \\
&= \sqrt{[x(t)\,R_I + y(t)\,RQ]^2 (\cos^2\theta + \sin^2\theta)} \\
&= x(t)\,R_I + y(t)\,R_Q
\end{aligned}
\qquad (15)
$$

Notice that the first expression for R is the one that has to be implemented.

Figure 48 shows a diagram for the baseband approach for the correlation function. The correlator is applicable for the digital approach also by adding the A/D converters. (The analog shift registers has to be changed to digital shift registers.)

One very important point to remember in this analysis is that it is assumed that the PN sequence used for the real channel has to have low correlation with the quadrature PN sequence. Otherwise the correlation between the two PN sequences will not be small and simplication cannot be made for equations (5), (6), (7), and (8).

However, the assumption may not be a valid one since the PN sequences used for the MSK or bi-phase modulation does not necessarily have this property. PN sequences in general are chosen because they have low cross correlation properties as one complete sequence and not as composite sequences.

CONCLUSIONS

The optimum approach towards a universal correlator that can be operated in the RF region is the convolver. By feeding the time inverted reference into one port of the convolver, the device can be used as a correlator. The requirement in using a convolver is for the two signals to coexist during the time of correlation. Thus, the total delay of the convolver has to be twice the duration of the signal. The state-of-the-art convolver has a delay of 50 μsec which is marginal for JTIDS (for total correlation) not to mention the GPS or HFIC signals which have much longer durations. The GPS P code, which is reset each week, represents the extreme case which is impractical and impossible to implement using a single convolver. It seems that in the near future, implementation of a convolver for the C/A code may prove to be a difficult task since the C/A code has a duration of 1 msec. Implementation of the correlator using the convolver will mean that the convolver will have a delay of 2 msec which is an increase of 40 times the state-of-the-art convolver delay. Since the convolvers are not cascadable, one way of getting around this problem will be to break up the incoming signals and reference sequence into groups and perform a piecewise correlation. The outputs will be added together to determine the correlation output. However, doing so will imply more complicated circuitry. Since the convolver is more of a delay line nature, the device does not have

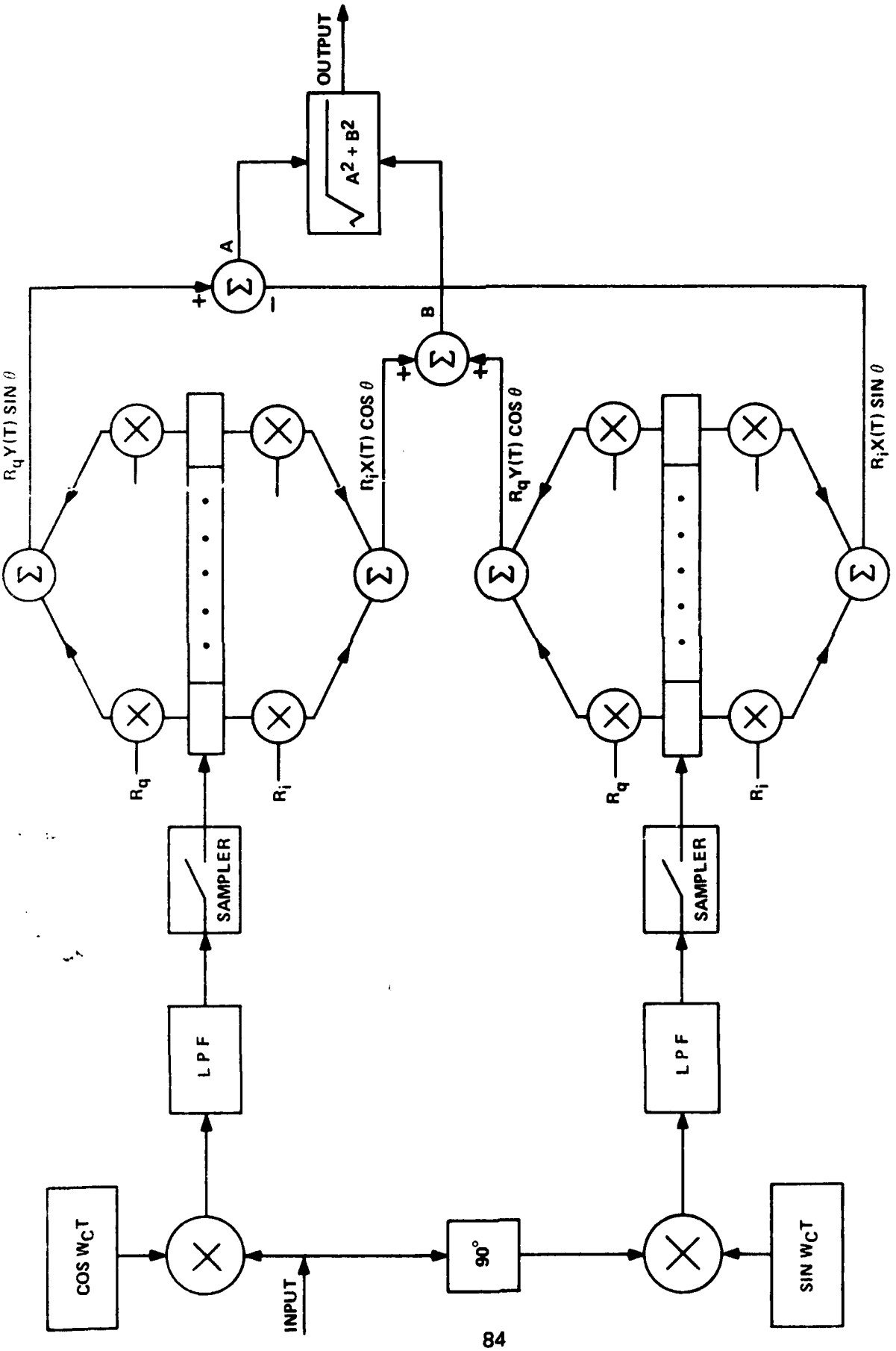


Figure 48. A Bandpass Correlator Structure for CCD or Digital Circuits

any capability to store the reference and therefore, active correlation will have to be used. The implication is that either a long time is required for synchronization or many convolvers are required with the reference being shifted by one successive bit going into each convolver.

The short term solution for a universal correlator seems to be either the CCD approach or the digital approach. Both are baseband devices that require a frequency translator. The digital approach is attractive in that it is a very mature technology and poses very little risk. This is the technology that one can fall back on. CCD's represents a more ambitious approach that has a lot of merits if the transfer efficiency can be increased. The technology is not as mature as digital techniques and is more risky. One important point to remember in either approach is that if the cross correlations between the inphase and quadrature PN sequences are not small, the analysis will not hold. The convolver, of course, does not have such a problem. Provided that convolvers can be made to satisfy the signal processing requirements and therefore is the optimum, a dilemma will be faced because of the fact that the convolver which is manufactured based on piezoelectric surface acoustic wave technology, cannot be large scale integrated and the device is large and bulky.

Currently, commercially available CCD delay lines are available that can be clocked to 20 MHz which is marginally adequate for GPS processing for the P code. Experimental devices are fabricated that can be run at hundreds of megahertz. The limitation to speed, it seems are the clock drive requirements and the difficulty of sensing the output charge at high rate.

Based on the risk factor involved, the following is recommended: Digital – the present solution; CCD – the near term solution; and Convolver – the long term solution for the implementation of a multi-waveform correlator in the overall order of difficulty. A universal correlator at RF is very attractive. The convolver is the optimum in RF processing if size does not present any problem in an application. The approach is ideal if it can be made to be a passive process.

APPENDIX A

SIGNAL REPRESENTATION

Definition

Roughly speaking, a signal is an embodiment of a message, whether the message is intentional or nonintentional. It is a physical quantity (e.g. pressure, displacement, voltage, current, velocity) which may vary with time (or some other real variable, such as distance).

Lowpass and Bandpass Signals

The classification of a signal as a low pass function or bandpass function depends to a certain extent upon the application one has in mind. In the spectrum of a signal shown in figure A1-(a), one would have no hesitation in classifying it as lowpass. On the other hand, the spectrum of figure A1-(b) might be classified either as lowpass or bandpass. It would be considered lowpass if it is to be passed through a filter whose passband extends from zero frequency to a frequency greater than f_o. On the other hand, if the filter has a narrow passband near f_o, it would be more convenient to classify the signal as bandpass.

A bandpass signal can be represented as follows:

$$x(t) = x_c(t) \cos \omega_c t - x_s(t) \sin \omega_c t \tag{A-1}$$

where $x_c(t)$ and $x_s(t)$ are lowpass functions called, respectively, the inphase and quadrature modulation components, and where ω_c is the angular velocity of a reference frequency called the carrier frequency.

Any signal can be put into this form, and $x_c(t)$ and $x_s(t)$ will have spectrum concentrated around f_c, the carrier frequency. However, a complex notation turned out to be more convenient.

Complex Representation of Bandpass Signal

The real signal $x(t)$ can be represented as the real part of a complex signal $\underline{x}(t)$, called the analytic signal or pre-envelope where

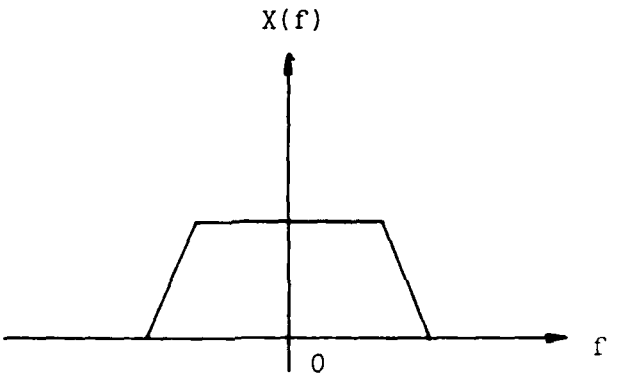
$$\underline{x}(t) = x(t) + j\hat{x}(t) \tag{A-2}$$

The imaginary part of $\underline{x}(t)$ is taken to be the Hilbert transform of $x(t)$. This is defined by the following integral:

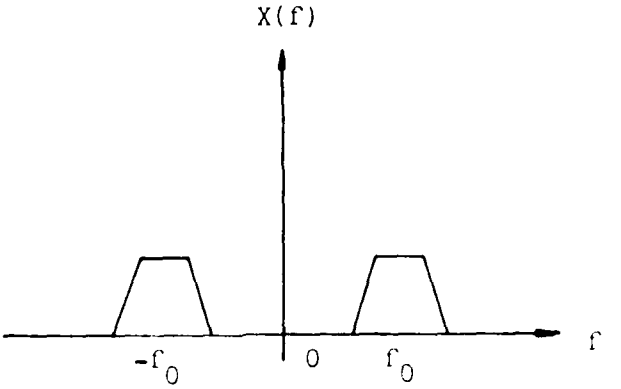
$$\hat{x} = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(\tau)}{t - \tau} \, d\tau \tag{A-3}$$

Since $\underline{x}(t)$ is complex, we may express it as follows:

$$\begin{aligned} \underline{x}(t) &= \tilde{x}(t) \; e^{j\omega_c t} \\ &= \tilde{x}(t) \; e^{j2\pi f_c t} \end{aligned} \tag{A-4}$$



(a)



(b)

Figure A1 - Signal Spectra of band limited signals

where f_c is again the reference (carrier) frequency.

$\tilde{x}(t)$ is known as the complex envelope of $\underline{x}(t)$.

The Fourier transform of $\hat{x}(t)$ and of $\underline{x}(t)$ may be found by noting that equation A-3 is a convolution of $x(t)$ with $1/\pi t$.

That is;

$$\hat{x}(t) = x(t) * \frac{1}{\pi t} \tag{A-5}$$

Therefore, the Fourier transform $\hat{X}(\omega)$ of $\hat{x}(t)$ is given by

$$\hat{X}(\omega) = X(\omega)\ F\ [1/\pi t] \tag{A-6}$$

The Fourier transform of $\frac{1}{\pi t}$ is,

$$F\left(\frac{1}{\pi t}\right) = -j\ \text{sgn}\ \omega$$

$$= -j \qquad \omega > 0$$

$$0 \qquad \omega = 0$$

$$j \qquad \omega < 0 \tag{A-7}$$

Therefore,

$$\hat{X}(\omega) = -j\,X(\omega) \qquad \omega > 0$$

$$0, \qquad \omega = 0$$

$$j\,X(\omega) \qquad \omega < 0 \tag{A-8}$$

From equation A-2, the Fourier transform of $\underline{x}(t)$ is,

$$\underline{X}(\omega) = X(\omega) + j\,\hat{X}(\omega)$$

$$= 2\,X(\omega) \qquad \omega > 0$$

$$X(\omega) \qquad \omega = 0$$

$$0 \qquad \omega < 0 \tag{A-9}$$

This result shows that the pre-envelope has a one-sided spectrum. That is, it vanishes from $\omega < o$.

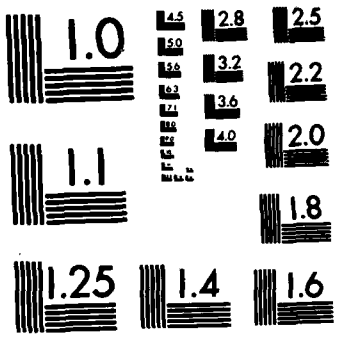
Conversely, any time function which has a one-sided spectrum (to the right) is decomposable into a complex time function whose imaginary part is the Hilbert transform of the real part.

Notice that equation A-4 can be expressed as

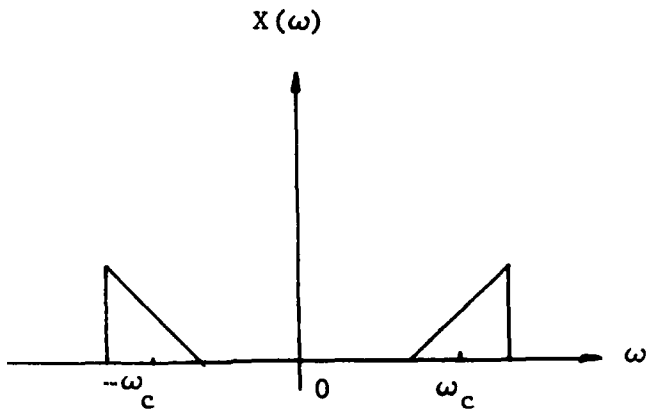
$$\tilde{x}(t) = \underline{x}(t)\, e^{-j\omega_c t} \tag{A-10}$$

if we let $x(t)$ to be a real bandpass signal whose Fourier transform $X(\omega)$ is indicated in figure A-2(a). Note that there is symmetry in the spectrum. In fact $|X(\omega)|$ is even in ω while the phase of $X(\omega)$ is odd in ω. Figure A-2(b) shows the spectrum $X(\omega)$ of the pre-envelope and figure A-2(c) shows the

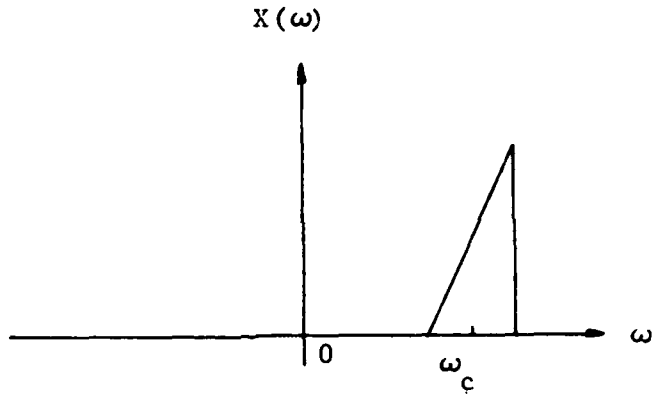
AD-A127 845 MULTIFUNCTION SPREAD SPECTRUM TECHNOLOGY(U) NAVAL AIR 2/2
DEVELOPMENT CENTER WARMINSTER PA COMMUNICATION
NAVIGATION TECHNOLOGY DIRECTORATE Y LUI 04 OCT 82
UNCLASSIFIED NADC-82188-40 . F/G 17/2 NL



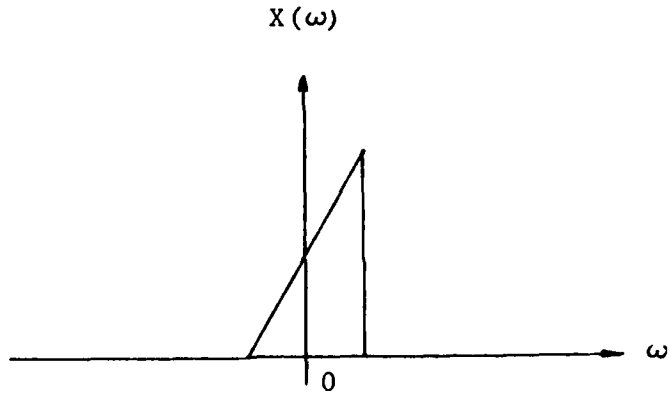
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A



a. Spectrum of x(t)



b. Spectrum of $\underline{x}(t)$, the Pre-Envelope



c. Spectrum of $\tilde{x}(t)$, the Complex Envelope

Figure A2 - Spectra of a Bandpass Signal

spectrum $X(\omega)$ of the complex envelope. Note that if $X(\omega)$ is confined to a band in the vicinity of ω_c then $\tilde{X}(\omega)$ will be confined to a similar band in the vicinity of $\omega = 0$.

It can be seen that all of the knowledge of $x(t)$ is contained in its complex envelope $\tilde{x}(t)$. Therefore, we wish to examine more closely the properties of $\tilde{x}(t)$. It can be shown that

$$H(\cos \omega_c t) = \text{sgn}(\omega_c) \sin \omega_c t \tag{A-11}$$

and

$$H(\sin \omega_c t) = -\text{sgn}(\omega_c) \cos \omega_c t \tag{A-12}$$

If we let $x_c(t)$ and $x_s(t)$ of equation A-1 to be lowpass waveforms whose total nonzero frequency (including positive and negative frequencies) extent is less than $2 f_c$. This is the same as saying that $x(t)$ has a Fourier transform which is zero outside an interval B less than $2 f_c$ for $f > o$. It may be shown that

$$H[x_c(t) \cos \omega_c t] = x_c(t) \sin(\omega_c t) \text{sgn}(\omega_c) \tag{A-13}$$

$$H[x_s(t) \sin \omega_c t] - \text{sgn}(\omega_c)\, x_s(t) \cos(\omega_c t) \tag{A-14}$$

Thus,

$$\begin{aligned} \underline{x}(t) &= x(t) + j\hat{x}(t) \\ &= [x_c(t) + j x_s(t)]\ [\cos \omega_c t + j \sin \omega_c t] \\ &= [x_c(t) + j x_s(t)]\ e^{j\omega_c t} \end{aligned} \tag{A-15}$$

A comparison between equations A-15 and A-4 shows that

$$\tilde{x}(t) = x_c(t) + j\, x_s(t) \tag{A-16}$$

If $x(t)$ is a band limited signal (bandwidth $< 2 f_c$), it makes sense to write equation (A-1) as

$$x(t) = a(t) \cos[\omega_c t + \phi(t)] \tag{A-17}$$

where

$$a(t) = [x_c(t)^2 + x_s(t)^2]^{1/2} \tag{A-18}$$

$$\phi(t) = \tan^{-1}[x_s(t)/x_c(t)] \tag{A-19}$$

$a(t)$ is called the envelope or the amplitude modulation function, and $\phi(t)$ is called the phase modulation function. Comparing equations A-17, A-18, A-19 with A-15, A-16 shows that

$$a(t) = |\tilde{x}(t)| = |\underline{x}(t)| \tag{A-20}$$

$$\begin{aligned} \phi(t) &= \arg \tilde{x}(t) \\ &= \arg x(t) - \omega_c t \end{aligned} \tag{A-21}$$

$|\tilde{x}(t)|$ is sometimes called the "instantaneous" envelope and $\phi(t)$ is called the "instantaneous" phase modulation.

APPENDIX B

NOISE SOURCES IN DIGITAL CORRELATION

In the course of digital correlation, there are many places that noise may be introduced that may affect the quality of the signal. In the following paragraphs, some of these noise sources are investigated.

Noise Sources from a Sampling Process

The quality of the signal begins to deteriorate the moment it is sampled – assuming sampling is high enough in frequency so that there is no aliasing effect. The first group of noise sources come from the sample-and-hold circuit which freezes the fast moving signal for the A/D converter to make the conversion. These noise sources are listed below:

1. Offset – This is the extent to which the output deviates from zero with zero input and is usually a function of time and temperature. The offset can often be adjusted by means of an external (or built-in) potentialmeter.

2. Non linearity – This is the amount by which the output versus input deviates from a straight line.

3. Scale Factor Error – This is the amount of which the output deviates from specified gain (usually unity). This parameter can also be adjusted.

4. Aperture Uncertainty – This is the time elapsed between the command to hold and the actual opening of the hold switch.

5. Droop – This is the drift of the output caused by the discharge of the storage capacitor.

Among the above noise sources from the sample-and-hold, the most important ones are the aperture uncertainty and the droop.

The aperture uncertainty has two components – a nominal time delay, and an uncertainty caused by jitter or variation from time to time and unit to unit and is shown in Figure B-1.

The aperture uncertainty of a sample-and-hold unit dictates the rate that the input signal can vary given that the signal is to be resolved to a certain desired minimum value as in the case of an A/D converter. Assuming that a signal of the form Asin (ωt) is sampled by a sample-and-hold circuit the rate at which the signal varies with time will be the time derivative of the signal and is given by the following equation.

$$\frac{d}{dt} A \sin \omega t = \omega A \cos \omega t \text{ volts/second}$$

If an A/D converter is to be used to put the sampled value into 10 binary bits, the aperture should be designed so that the output of sample and hold circuit will not vary more than half the weight of the least significant bit (LSB) of an A/D converter. Exceeding such a value will cause the A/D converter to change the value of the A/D converter. Assuming the A/D converter has 1 volt as its full scale, therefore 1/2 LSB = 2^{-11} volt. The maximum rate at which the input signal can vary is ωA volts/sec and therefore, the maximum aperture error is $2^{-11}/\omega$A seconds. Therefore, to specify the aperture error, one must know the weight of the LSB of the A/D converter and the amplitude of the incoming signal. Notice that if ωA becomes larger indicating that the frequency is larger (A is a constant), a smaller aperture uncertainty is required.

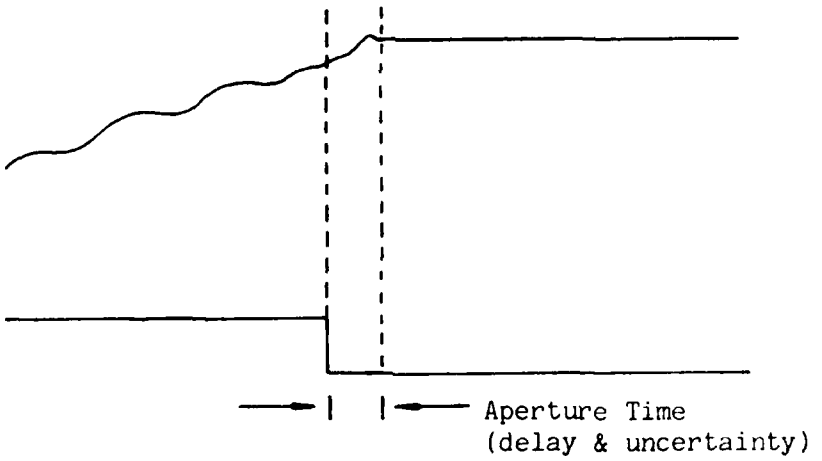


Figure B1 - Aperture Error in a Sample-and-Hold

As aperture time is important to fast moving signals, the droop is important to slow moving waveforms. A typical diagram for the droop is shown in figure B2 and is the effect caused by the discharge of the capacitor in the "hold" circuit. Since slow moving signals imply a slower sampling rate, the capacitor will have more time to discharge than in higher rate sampling circuits. Therefore, given the LSB for the A/D converter using in conjunction with the sample and hold with a certain droop, the minimum sampling rate can be determined. For example, assume a sample and hold has a droop rate of 50 μv/μsec and a 1 volt full scale 10 bit A/D converter has 2^{-11} volt as the weight of half of its LSB. The minimum sampling rate should be:

$$f = [(2^{-11} \text{ volt})\ (1 \text{ sec}/50 \text{ volts})]^{-1}$$

$$= 102.4 \text{ khz}$$

Sampling less than 102.4 KHz may yield an erroneous result at the output of the A/D converter.

Quantization Noise

The source of error coming from the A/D converter is known as quantization noise. The noise arises because of the fact that only a finite number of bits are used to represent an analog voltage. This is known as truncation error, or round off error depending on the scheme used to obtain the fixed binary number.

For two's complement representation, the error E_T for truncation was found to be $0 \geqslant E_T > -2^{-b}$. Rounding off a two's complement number will produce an error given by the following expression;

$$-(\tfrac{1}{2})\ (2^{-b}) < E_R \leqslant (\tfrac{1}{2})\ (2^{-b})$$

The mean and variance of the quantization noise are shown to be

$$m_e = o$$

$$\sigma_e^2 = \frac{\Delta^2}{12} = 2^{-2b}/12$$

for rounding and,

$$m_e = -2^{-b}/12$$

$$\sigma_e^2 = 2^{-2b}/12$$

for two's complement truncation.

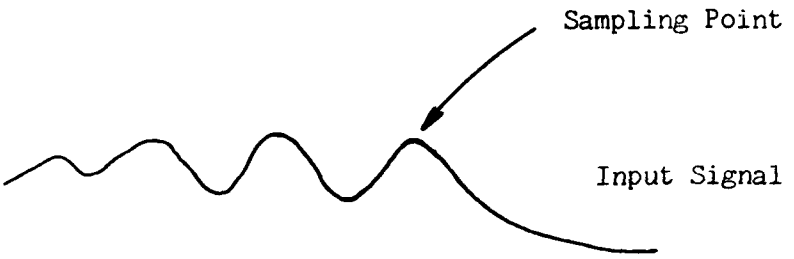
If it is assumed that the signal power to the A/D converter is σ_x^2, then the signal-to-noise ratio produced by the quantization will be,

$$\frac{\sigma_x^2}{\sigma_e^2} = \left(\frac{12}{2^{-2b}}\right)\ \sigma_x^2$$

Expressing this in terms of db, we have the following;

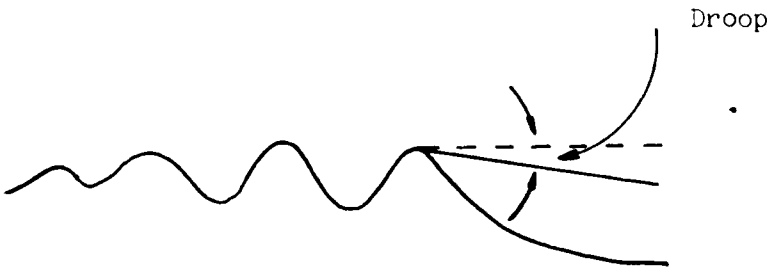
$$SNR = 10 \log_{10} (\sigma_x^2/\sigma_e^2)$$

$$= 6b + 10.8 + 10 \log_{10} (\sigma_x^2)$$



(a) Sampling of Input Signal

(b) Ideal Response



(c) Actual Response

Figure B2 - Droop in a Sample-and-Hold

It can be seen that the signal-to-noise ratio increases 6 db for each bit added to the register length. To prevent the signal from exceeding the dynamic range of the A/D converter, the incoming signal can be scaled. Let A be the scaling factor, the resulting signal-to-noise ratio will then be

$$10 \log_{10}\left(\frac{A^2 \sigma_x^2}{\sigma_e^2}\right) = 6b + 10.8 + 10 \log_{10} (\sigma_x^2) + 20 \log_{10} (A)$$

Since A is less than one, $20 \log_{10}(A)$ is negative and so by reducing the input amplitude, the signal-to-noise ratio is reduced. If A is assumed to be (1/4) σ_x the signal-to-noise ratio is (6 b – 1.24) db. To obtain a signal-to-noise ratio of 60 db, 10 bits are required.

For systems in which it is desired that the signal waveform remain undistorted at the output of the digital filter, it is necessary to use very fine quantization step in order to provide a large signal-to-quantizing noise ratio.

Since the spectrum of the quantizing noise is much wider than the signal bandwidth, sampling tends to alias the quantizing noise back into the signal bandwidth further reducing the ratio of signal-to-quantizing noise.

For detection systems where the preservation of the waveform is unimportant, relatively coarse quantization can be used with only a slight reduction in performance. A simple explanation of why it suffices to use fewer levels is based on the fact that in match filtering operation the various components are delayed and summed in order to maximize the ratio of signal peak to the rms noise. In the summing process, the quantizing noise components tend to cancel and thus, have a small effect on the output value. The principle can also be explained from the point of view that the energy in the essentially "white" quantizing noise is reduced by the correlator processing gain.

Theoretically, the number of quantizing levels needed to produce the signal statistics can be determined from the characteristic function of the signal probability density. The effective loss in signal-to-noise ratio as a function of the number of quantizing levels is shown in figure B3.

Assuming that the input signal-to-noise ratio is low, the figure shows that with eight quantizing levels, the output signal-to-noise ratio is degraded by only 0.17 db. For sixteen quantizing levels (4 bit quantization), the degradation is only 0.05 db. For the case of high input signal-to-noise ratios similar results using relatively coarse quantization can be expected, although the exact values will change herewhat since the probability density of the composite input signal is no longer exactly Gaussian.

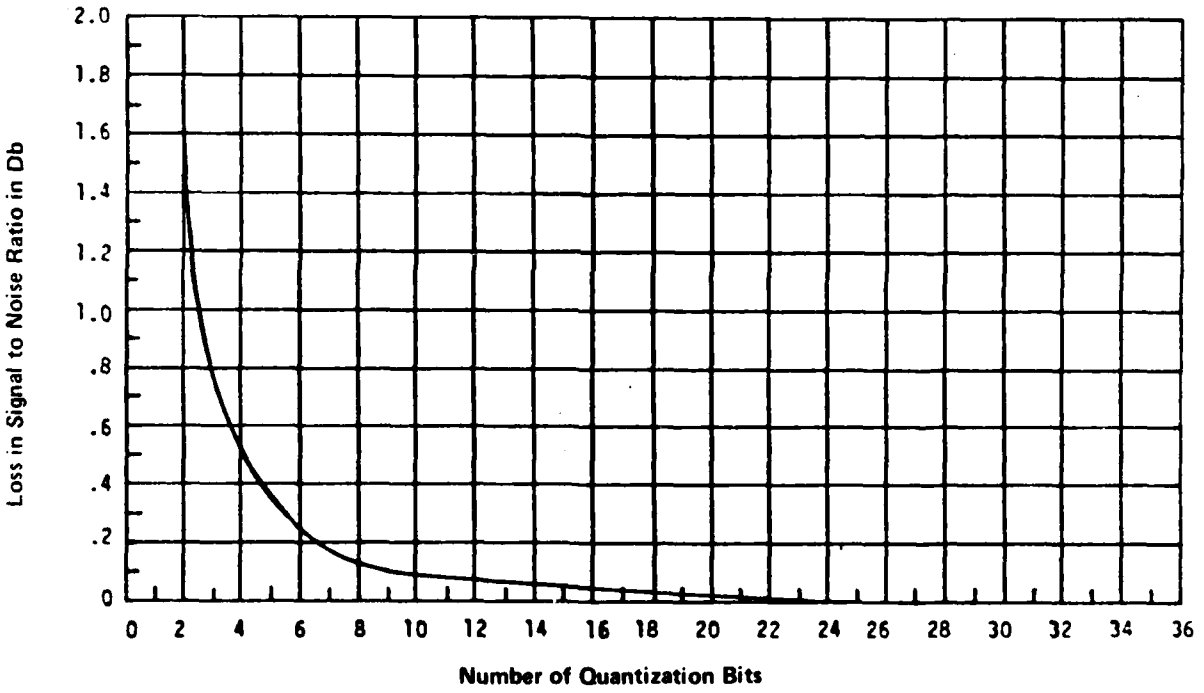


Figure B3 - Loss in Signal-to-Noise Ratio in Quantizing a Signal With a Small Input Signal to Noise Ratio

APPENDIX C

NOISE IN QUADRATURE SAMPLING

In quadrature sampling, the two channels are sampled 90° apart. The samples obtained for a time wave input will have the values

$$f_I\,(nt) = A \sin\theta$$

$$f_Q\,(nT) = A \cos\theta$$

for the real and quadrature channels respectively. Since quadrature sampling is done at one frequency, frequencies above and below the center frequency will not have the required 90° phase shift. To exactly produce the quadrature component all frequencies within the band of interest must be shifted 90°. This effect is given by the Hilbert transform,

$$\hat{x}(t) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau}\, d\tau$$

Directly performing the convolution above is undesirable because of the complexity of its implementation. However, for a narrow band signal,

$$\hat{x}(t) \approx x\left(t - \frac{\pi}{2\omega_c}\right)$$

That is, a delay of $\pi/2\omega_c$ is near 90° phase shift for frequencies close to the center frequency ω_c. Thus by taking samples with delay $\pi/2\omega_c$, then processing them in conjunction with the original samples, an effective 90° phase shift is approximated in the Q channel.

The degradation in signal-to-noise ratio from quadrature sampling is negligible. Inaccurate phase shift in the Q channel produces phase shifted correlation of the non-90° components, or a reduction in correlation at the correlation peak time. If we let the IF frequency to be at 70 Mhz with a bandwidth of ± 5 Mhz and an approximate 90° phase shift of $T = \pi/2\omega_c$, then, the phase distortion at the band edges is

$$\phi_B = \left(\frac{f_H}{f_c} - 1\right) 90°$$

$$= \left(\frac{75}{70} - 1\right) 90°$$

$$= 6.43°$$

The loss in detection due to improper phase shift at the band edge is

$$L = 10 \log_{10} [\cos^2 \phi_B]\ \text{db}$$

$$= 0.0548$$

The same result can be obtained by mixing the incoming signal with cos $\omega_c t$ and sin $\omega_c t$ to obtain the two channels. However, the result will bear the same noise as the delayed sampling approach since the mixing of the sin $\omega_c t$ and the cos $\omega_c t$ function with the incoming signal will produce a real and quad signal which is 90° out of phase only at the carrier frequency.

APPENDIX D

QUADRATURE SAMPLING OF BANDPASS SIGNAL

A bandpass signal can be represented by the following equation,

$$x(t) = x_c(t) \cos \omega_c t - x_s(t) \sin \omega_c t \qquad \text{(D-1)}$$

where $x_c(t)$ and $x_s(t)$ are the lowpass in phase and quadrature modulation components respectively, and $\omega_c = 2\pi f_c$ is the reference angular frequency which is called the carrier frequency. If x(t) is confined to the frequency band $(f_c - W/2, f_c + W/2)$, then $x_c(t)$ and $x_s(t)$ will be confined to the interval $(-W/2, W/2)$. Therefore, $x_c(t)$ is completely determined by a sequence of samples 1/W apart in time. Similarly, $x_s(t)$ is completely determined by sequence of samples 1/W apart in time. The result is that a bandpass function of bandpass bandwidth W is completely determined by a set of 2W samples per second.

In order that a sampled version of x(t) be obtained from the inphase and quadrature modulation components it is necessary to extract $x_c(t)$ and $x_s(t)$ from x(t). This is done by inphase and quadrature demodulation, sometimes called "quadrature" demodulation or "coherent" demodulation. This is shown in figure D1 together with the sampling to get the two sampled sequences, one for $x_c(t)$ and the other for $x_s(t)$. The process is called "quadrature sampling." It should be pointed out that the sampling rate for either modulation component is 1/W, although the total number of samples per second is 2W in order to represent the bandpass waveform.

More insight may be obtained by considering the pre-envelope of the bandpass waveform. Let $\underline{x}(t)$ be the pre-envelope and let its Fourier transform $\underline{X}(f)$ be confined to the frequency interval $(f_c - (1/2)\,W, fc + (1/2)\,W)$. $\underline{X}(f)$ can be considered to be the product of a periodically repeated version of X(f) and the rectangle function in frequency, as follows;

$$X(f) = \text{rep}_W\, \underline{X}(f)\, \text{rect}\,\frac{f-fc}{W} \qquad \text{(D-2)}$$

Therefore, the pre-envelope $\underline{x}(t)$ will be given by the correlation of the inverse transform $\underline{X}(f)$ and of rect [(f-fc)/W] as follows;

$$\underline{x}(t) = \sum_{n=-\infty}^{\infty} \underline{X}(n/W)\, \text{sinc}\,(\omega t - n)\, e^{j2\pi f_c\left(t - \frac{n}{\omega}\right)} \qquad \text{(D-3)}$$

Therefore, $x(t) = \text{Re}\,[\underline{x}(t)]$ is completely determined by the complex-valued samples $\underline{x}(n/W)$ taken 1/W apart. That is, if the sampled values are obtained properly, 2W real values per second specify its signal completely even if it is a bandpass function rather than a lowpass function.

Equation D-3 can be rewritten as follows:

$$\underline{x}(t) = \sum_{n} \underline{x}\left(\frac{n}{W}\right) e^{-j2\pi f_c n/W}\, \text{sinc}\,(2\omega t - n)\, e^{-j2\pi f_c t} \qquad \text{(D-3)}$$

It can be seen from appendix A that

$$\tilde{x}\left(\frac{n}{W}\right) = \underline{x}\,(n/W)\, e^{-j2\pi f_c\, n/W}$$

where $\tilde{x}(t)$ is the complex envelope of $x(t)$.

Then equation D-4 becomes,

$$\underline{x}(t) = \sum_n \tilde{x}\left(\frac{n}{W}\right) \text{sinc}\,(\omega t - n)\, e^{j2\pi f_c t}$$

Taking the real part yields the following,

$$x(t) = \sum_n \text{sinc}\,(\omega t - n)\,\left[x_c\left(\frac{n}{W}\right) \cos 2\pi f_c t - x_s\left(\frac{n}{W}\right) \sin 2\pi f_c t\right]$$

where

$$\tilde{x}(t) = x_c(t) + jx_s(t)$$

The reconstruction of $x(t)$ is shown in figure D2.

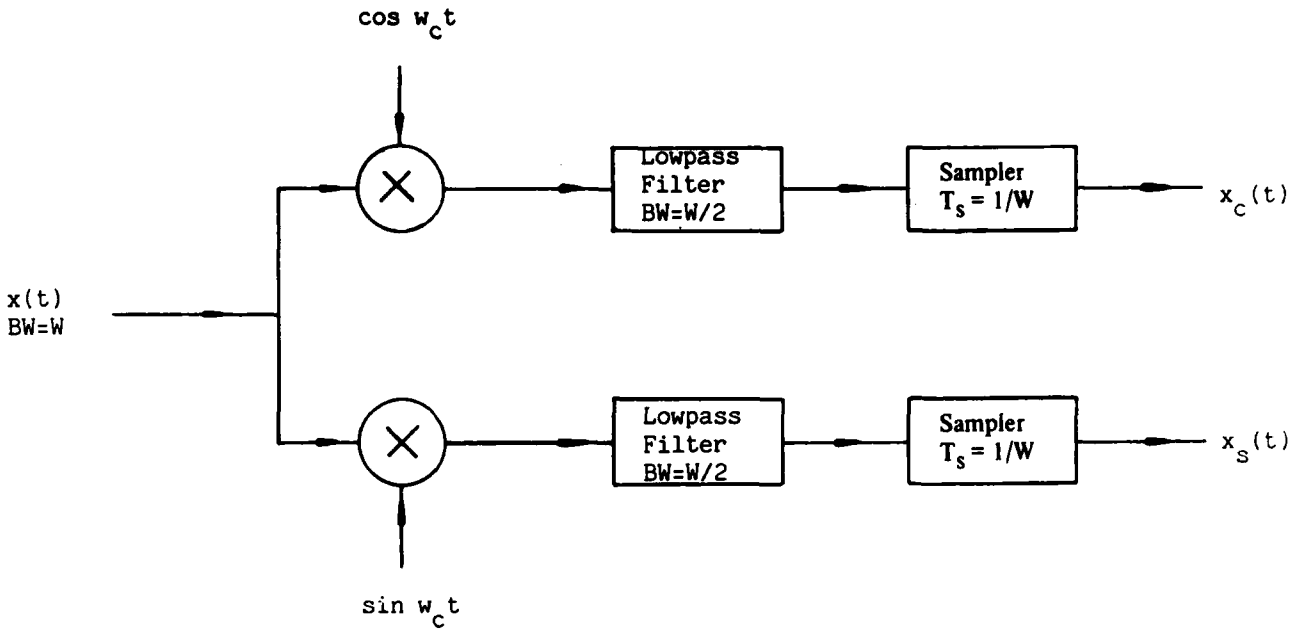


Figure D1 — Quadrature Sampling of a Bandpass Function

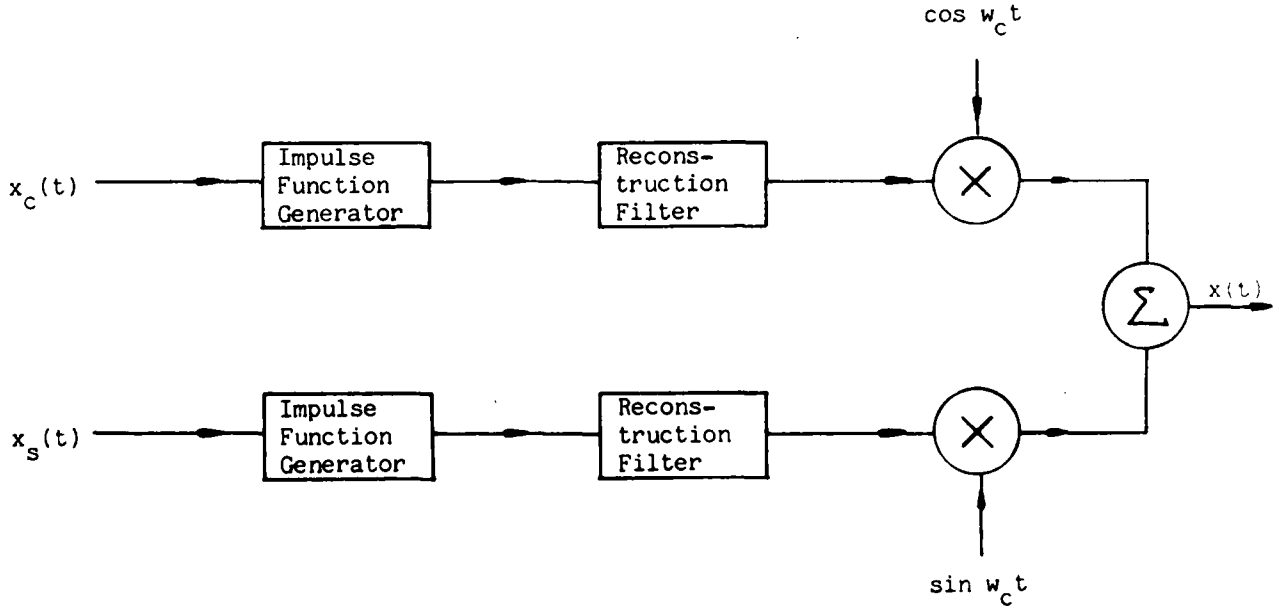


Figure D2 — Reconstruction of a Bandpass Signal

Figure A2 - Spectra of a Bandpass Signal