Chapter 4

Transversal Filters

4.1 INTRODUCTION

We are all familiar with the simple low-pass, high-pass, and bandpass filters formed out of resistances, capacitances, and inductances. Such filters are convenient for eliminating signals at unwanted frequencies, and for choosing signals in a desired frequency range, such as those used for television or radio.

As radar and communication techniques have developed, far more sophisticated applications of filtering techniques have become of interest. These can now be realized with *transversal filters*, which are basically tapped delay lines with each tap connected to a common input or output line. If we want to detect a signal that is weaker or lower in amplitude than an interfering noise signal, it is a great advantage to know the nature of the signal we want to find beforehand. For instance, a narrowband filter selects a particular narrowband signal and increases its amplitude relative to noise while eliminating interfering signals, thus improving the signal-to-noise ratio. This filter is a familiar example of one that operates in the frequency domain. If, instead, the signal is a digital code, a time-domain filter matched to this code, known as a *correlation filter*, will improve the signal-to-noise ratio. In general, such filters enable us to insert a long code into a matched transversal filter and obtain a narrow output pulse. Since the output signal contains the same amount of energy as the input signal, the peak amplitude of the pulse will be much higher than that of the input signal. Noise is treated differently because the filter is not matched to it; thus the signal-to-noise ratio is improved.

Another type of transversal filter removes distortion. If a television signal, for instance, suffers from ghost echoes because of reflections from buildings, it is

possible to construct an *inverse filter* to reverse the distortion process and remove the unwanted echo. In a linear system, if the filter removes an unwanted echo from a known signal, such as the synchronizing pulse of a television signal, it will also remove unwanted echoes from an unknown television signal that has been distorted in the same way.

There are many more interesting examples. We will devote much attention to *frequency-modulated chirps* (FM chirps), which are signals of approximately constant amplitude whose frequency varies linearly with time over a limited frequency range. We can construct a matched filter for such signals to produce a narrow output pulse in the same way as we described for digital codes. Thus a filter can be used to emphasize the wanted chirp signal relative to noise, thereby improving the signal-to-noise ratio.

FM chirp filters have far more general applications. We shall show, for instance, that when a modulated chirp signal is injected into such a filter, its output is an instantaneous Fourier transform of the modulation. Thus such filters are very powerful tools for carrying out real-time Fourier transforms. This process in turn leads to a wide range of devices that can carry out Fourier transforms for various types of filtering operations, such as eliminating unwanted frequency components while retaining desirable ones.

One reason FM chirps are widely used in radar systems is because it is easy to produce a chirp signal of relatively long duration with the same amount of energy as a short, intense pulse. Another reason is that Doppler shift changes the frequency of the chirp. This change in frequency, which can be measured instantaneously with a chirp filter, indicates the target's velocity [1].

Sophisticated pseudorandom digital codes are similarly used in spread-spectrum communication systems. The very long code of a matched filter can improve the signal-to-noise ratio by many orders of magnitude, and because so many long codes are possible, intercepting an unknown code becomes extremely difficult.

It is impractical, of course, to design such sophisticated filters using only lumped circuits, for a sophisticated code consisting of many chips could require almost as many lumped elements for its construction. Such filters require a simpler construction technique, preferably one employing the photolithographic methods now common in integrated circuits. Fortunately, filters suitable for both analog and digital signal processing have become available in the last few years. The original *transversal filters* of this type were first demonstrated in the form of *surface acoustic wave devices* (SAW devices), which typically operate in the frequency range 10 MHz to 1 GHz [2, 3].

More recently, the same principles have been applied to the *charge-coupled device* (CCD) [4–6], the *single transfer device* (STD) [7], a type of switched capacitor filter, and the *bucket brigade device* (BBD) [8]. Other types of transversal filters using acousto-optic interactions have also been demonstrated. Research is under way on transversal filters that employ tapped fiber-optic delay lines [9–11] and tapped superconducting lines [12]; these systems should lead to devices with bandwidths of several gigahertz.

The *switched capacitor filter* is a related device [7]. In its original form it is not a transversal filter, and so is outside the scope of this book; however, the STD,

a development of the switched capacitor filter, can be regarded as a type of transversal filter and will be described together with the CCD, BBD, and SAW devices.

For completeness, we shall give here a short description of the operation of the conventional form of the switched capacitor filter, which works on the principle that a capacitor switched rapidly on and off, as shown in Fig. 4.1.1, can simulate a resistor. Suppose that the capacitor in Fig. 4.1.1(a) is rapidly switched to the right. The charge flowing from V_2 is $Q = C(V_2 - V_1)$. If the switch is switched back and forth at a clock rate f_c, the current flowing is $I = Qf_c$. Hence the switched capacitor circuit behaves like a series resistor of value $R = 1/Cf_c$. Provided that the clock frequency is much higher than the signal frequency, we can use a circuit with two MOSFETs (metal-oxide semiconductor field-effect transistors), shown in Fig. 4.1.1(b), to synthesize the resistor. We can easily construct such a circuit by using a silicon integrated-circuit technology, together with operational amplifiers which are used as fundamental components for analog low-pass, high-pass, and bandpass filters at frequencies usually below 50 kHz.

In this section we briefly review some of the operating principles of transversal filters. Later we describe the various types of devices and their applications in more detail. We concentrate our attention on transversal filter types that either are now or will ultimately be suitable for radar or communication applications at frequencies above 100 kHz. We will also cover examples of lower-frequency signal processing applications that illustrate the state of the art as it stands at the time of this writing. We start with SAW devices for our illustrations because of their basic simplicity and flexibility, and because many applications of transversal filters were first demonstrated with SAW devices. Later we show the connection between these devices and the wide range of analog devices based on silicon technology that are now beginning to be widely employed. We will also briefly discuss some concepts for tapped fiber-optic and superconducting delay lines, on which research is now being conducted [9–12].

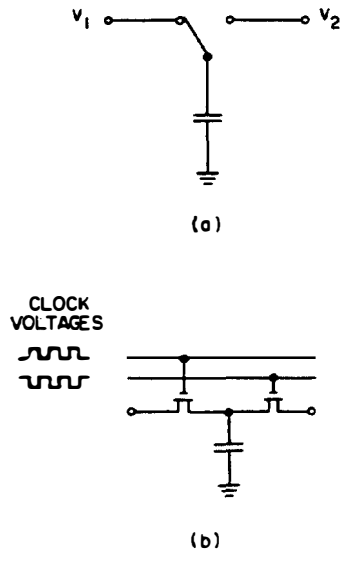


Figure 4.1.1 (a) Capacitor circuit switched to simulate a resistor. (b) MOS implementation of this circuit driven by a square-wave clock.

Surface acoustic waves provide a means for storage, delay, and complex parallel processing of long-duration wideband signals. It is possible to operate them at frequencies up to 500 MHz quite easily, and, with some care, up to 2 GHz, with bandwidths of as much as 30 to 40% of the center frequency. The corresponding data rates are as high as several hundred megabits per second. Certain types of these devices will store signals for as much as a second, and it is easy to construct tapped delay lines for surface acoustic waves. By arranging the spacing and the strength or weight of the taps correctly, it is possible to design broadband and narrowband filters, digital filters, and analog filters that recognize or generate particular signals. With further refinements, it is also possible to make programmable filters whose characteristics can be altered at will.

Analogous devices that can be constructed on silicon, such as CCDs, BBDs, and STDs, are at present more suitable for lower-frequency applications; with the development of gallium arsenide CCDs, however, the operating frequency is moving into the UHF range (100 MHz to 1 GHz). Silicon devices can carry out most of the signal processing functions provided by SAW devices, but in frequency ranges typically below 5 MHz, with correspondingly longer time delays. A major advantage of this technology is that the rate at which signals can be read into and out of these devices can be varied, so it is possible to expand or contract a given signal entering the device by reading it out at a different rate from that at which it was read in.

Interactions between two types of input signals are sometimes employed. For instance, because any semiconductor device can be made sensitive to light, it is possible to construct a light-sensitive CCD where a signal corresponding to the light intensity along the length of the device can be read out as one line of an optical image. By employing interactions of surface acoustic waves with semiconductors, it is possible to carry out Fourier transforms and other signal processing functions on the optical image.

Nonlinear interactions between two surface acoustic waves in SAW devices can lead to real-time processing functions in which the convolution or correlation of the two surface acoustic waves is obtained. Thus one signal can be used as a reference, instead of using a transversal filter with fixed properties. In this manner a transversal filter whose properties can be varied at will can be constructed merely by changing a reference code. A very similar device can also be constructed by using two tapped CCD delay lines connected to each other. One delay line controls the outputs from the taps of the other line; thus signals passed into one line act as a reference for signals passing through the other line [4].

Alternatively, we can utilize the interaction between a light beam and an acoustic wave to deflect the light beam, since the acoustic wave forms a type of optical diffraction grating. Such techniques have made it possible to devise devices that can carry out Fourier transforms or can correlate acoustic signals, either with each other or with a signal used to modulate a light beam. The basic principle of this device can also be used for deflecting a light beam to form a multiple-address system or an optical beam deflection system suitable for projection television systems.

The common feature of all these devices is the *transversal filter*, a delay line on which an acoustic or electrical signal propagates and then is accessed electrically, optically, or acoustically, at different points along it. These points form delay-line "taps." If we use a light beam or acoustic beam to tap the system, we can regard the taps as almost continuous; if we use electrical taps instead, we are essentially sampling the wave propagating along its surface. In either case, we can choose the amplitude and the phase weighting of the taps to perform a wide variety of signal processing functions.

4.2 LINEAR PASSIVE SURFACE WAVE DEVICES

4.2.1 Interdigital Transducer

The SAW filter has the advantage of being constructed by depositing thin films of metal on a piezoelectric substrate; thus the technology required is identical to that used for integrated circuits. This makes it possible to design filters with desirable characteristics that need none of the adjustments required by inductive or capacitive filters; these filters exhibit excellent characteristics over very long lifetimes and are also relatively simple and economical to manufacture.

These filters (whose basic structure is described here and in Sec. 4.2.2) employ a substrate of a piezoelectric material, such as quartz or lithium niobate, on which an interdigital transducer is deposited, as illustrated in Fig. 4.2.1. This transducer excites a surface acoustic wave whose fields fall off exponentially into the interior of the substrate, with a penetration depth on the order of one wavelength. At 40 MHz, with a wave velocity of approximately 3 km/s, the penetration depth of the waves is of the order of 75 μm. The interdigital transducer has neighboring fingers connected to the two bus bars (the input or output lines). Normally, the finger pair spacing (center-to-center spacing between pairs of fingers) is approximately one wavelength; for example, a 40-MHz television intermediate frequency (IF) filter has a finger pair spacing of 75 μm. Typical finger widths are approximately one-fourth of this value, or in this case 20 μm, so the devices can easily be manufactured by standard integrated-circuit technology.

As we shall see, devices of this kind can be tailored to yield a given amplitude and phase response over the frequency range of interest. They can also be designed to yield a given time-domain response, or even to respond to special analog or digital codes. Such devices can be made for operation anywhere in the range from 10 MHz to 2 GHz.

Digitally coded devices. It is very easy to forecast the time-domain response of a SAW transducer from simple geometrical considerations, and once the time-domain response of a filter is known, it is again easy to determine its frequency response by Fourier transform techniques. Thus we can construct transversal filters with a given frequency response quite simply, once we know the equivalent time-domain response.

We first consider physically the simple example of a digitally coded filter,

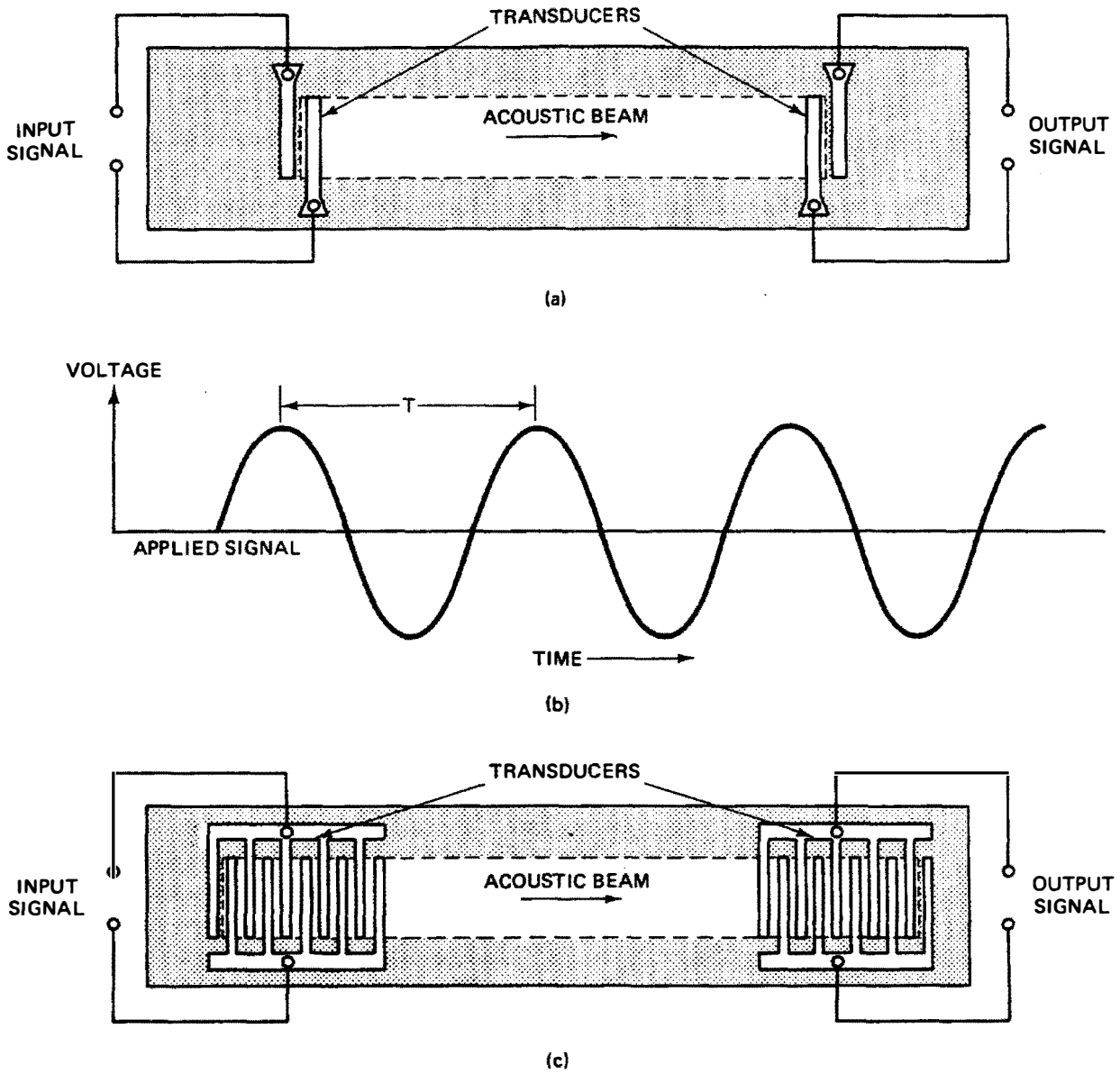


Figure 4.2.1 Excitation and detection of Rayleigh waves can be accomplished by using a simple transducer consisting of two metal electrodes deposited on a piezoelectric crystal (a). When a sinusoidal electric signal (b), which repeats itself in time T, is applied to the input electrodes, the alternating electric field sets up alternating vibrations in the piezoelectric material that give rise to Rayleigh waves. When the waves reach the output electrodes, they generate an alternating voltage between the two metal fingers. If the input electrodes have a uniform interdigital pattern (c), the separate waves excited by each pair of fingers will reinforce one another if the time required for the Rayleigh wave to travel between electrode pairs corresponds to the frequency of the electrical signal. If the output electrodes have the same spacing as the input electrodes, they will be "tuned" to receive the passing acoustic waves. (After Kino and Shaw [13].)

used both to produce given digital codes and as a matched receiver for a particular digital code. We then demonstrate mathematically that the geometry of the transducer is directly related to the time domain response of the filter.

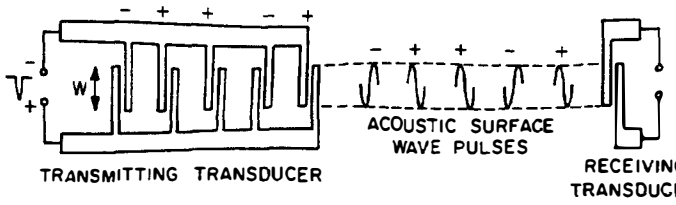


Figure 4.2.2 Digitally coded SAW filter.

Consider the interdigital transducer illustrated in Fig. 4.2.2. Metal fingers are laid down on a piezoelectric substrate and are connected to one of two input bus bars. A simple two-finger pair receiving transducer is laid down on the same substrate. When a voltage is applied across the two input bus bars, charges of opposite sign are induced on the two fingers. These charges, in turn, excite surface acoustic waves in the piezoelectric material because of the electric fields associated with them. We expect the amplitude of the surface acoustic wave signal excited by a metal finger to be proportional to the charge on the finger, or to the normal component of electric field at the surface of the finger.

In the configuration shown, the first pair of fingers on the right-hand side of the transmitting transducer may be regarded as positively polarized, and each finger pair may be regarded as a capacitor. Thus when a voltage is applied to the transducer, the charge on the finger will be proportional to the applied voltage. The charge on the first finger excites a positive acoustic pulse, followed by a negative pulse excited by the charge of opposite sign on the second finger. We note that the second pair of fingers is connected to the bus bars of opposite sign from the first pair. The first RF acoustic pulse travels down the delay line in the form of a single-cycle tone burst; it is followed by a second similar pulse of opposite sign, excited by the second pair of fingers, whose polarity is reversed from the first pair. In turn, each pair of fingers excites an RF pulse, so that a coded series of pulses, determined by the geometry and connections of the fingers, is excited on the SAW delay line. This digitally coded series of RF pulses may be detected at the receiving transducer, and the transversal filter yields a particular coded output, $+ - + + -$, determined by its geometrical configuration. This is known as a *biphase digital code*. Each element of the code, an RF pulse or tone burst, is called a *chip*.

The same device can be used as a coded receiver. In this case the charge excited on a finger of the output transducer is proportional to the amplitude of the acoustic signal passing underneath it, and the voltage output from the transducer is proportional to this charge. When an electrical signal $- + + - +$, corresponding to the time-reversed version of this digital code, is inserted into the left-hand transducer, as illustrated in Fig. 4.2.2, the first element of the code emitted from the left-hand pair of fingers will be a negative RF electrical pulse injected into a transducer of negative sign; thus it will produce a positive RF acoustic pulse. It reaches the receiving transducer at the same time the second element of the code is received from the second finger pair. As the input code matches the transducer code, all elements of the code reach the corresponding parts of the receiving

transducer simultaneously and yield signals of the same sign. These signals add to give a pulse five times the amplitude of an individual RF pulse from a single-finger pair.

If a different signal, having a different sequence of positive and negative RF pulses, is introduced into the delay line, there will be no instant at which it will match the polarities of all the tapping transducers. In other words, there is very little correlation between the signal and the filter response; therefore, the peak output signal will be reduced. This is a simple example of the ability of transversal filters to perform pattern recognition and to select a signal with a given code from all other signals. A so-called orthogonal set of codes is defined as one for which the output is of unit amplitude when the codes are not matched, and of N units for matched codes N chips long.

More generally, it will be seen that the minimum output pulse width τ is determined by the bandwidth B of the system, and is approximately $\tau = 1/B$. If T is the length of the coded pulse train, τ will be the length of one chip, or of the pulse of minimum length (in our example, this is the pulse corresponding to the signal from one finger pair or one chip). The ratio of the length of the input pulse to that of the output pulse is then T/τ, or approximately TB. This is the product of the bandwidth and the time delay of the filter, the *time–bandwidth product*, or, equivalently, a parameter known as the *pulse compression ratio*. The magnitude of the output pulse is increased over that of the input pulse by the same ratio (see Prob. 5); this is referred to as the *processing gain*. We note that for a digital code, the pulse compression ratio is N times the number of chips in the code, and $N \approx TB$.

A useful way of looking at this idea of pulse compression is to consider a digitally coded input signal N chips long, each of value ± 1, to a matched filter in the presence of a random code (noise). The output from the matched filter will only be one chip of time τ long. Thus its amplitude will be increased by a factor N and the peak output power by a factor of N^2. A very long random noise signal—an uncorrelated digital code, for instance—will not be correlated with the matched filter response, and the amplitudes of the individual chips will not necessarily add. On the other hand, the noise power output will still be N times as large as the power output from a single chip of a purely random code or noise signal, because it is being received by an N-chip-long filter and the powers are added at the individual elements of the filter. Thus the average noise power output from the random code will be increased by a factor N over that from a single chip, which will increase the peak signal-to-noise power ratio at the output by a factor N^2/N, or by the compression ratio $N = T/\tau$. More rigorous proofs of this property of matched filters are given in Secs. 4.4.2 and 4.4.3.

Experimental input and output signals from a tapped delay line of this type are shown in Fig. 4.2.3. This delay line has 127 taps, spanning a time delay of 25.4 μs, and the pulse compression ratio is approximately 127. In this system we produce a digital code by sending a short pulse into a coded delay line. The receiver delay line must have a time-reversed response to that of the input delay line. Therefore, it is usually called the *conjugate filter*.

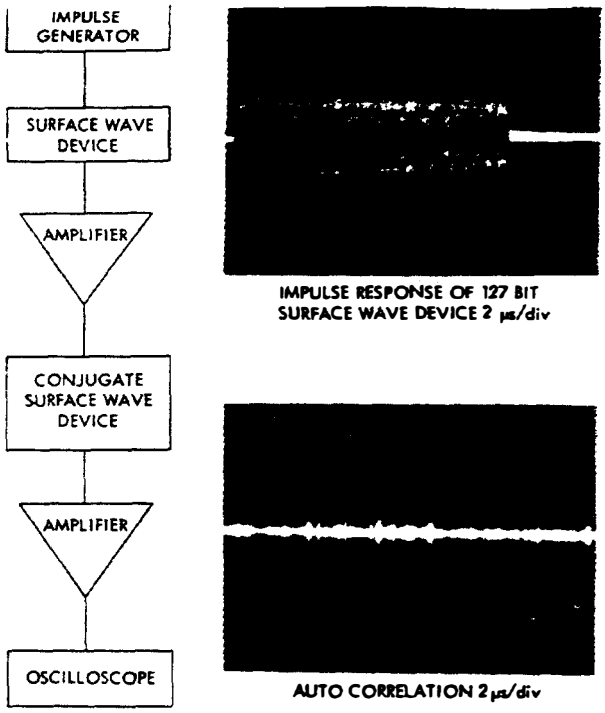


Figure 4.2.3 Impulse response and autocorrelation response obtained with a 127-chip PN-coded SAW device, using the system shown.

Mathematical model of the transducer. We now consider the response of an interdigital transducer using a simple mathematical formulation that is often called the *impulse model* [15]. As we shall see, this model applies in an almost unchanged form to CCD analog delay line filters. We shall use the symbols t' and z' to apply to the source, and t and z to apply to the observation point.

Suppose that at time t', a voltage $V(t')$ is applied to an electroded region between z' and $z' + dz'$ on the surface of the piezoelectric substrate, as shown in Fig. 4.2.4. Because of the capacity of this element to other electrodes or ground, the input voltage will induce a charge $\sigma(t', z')\,dz'$ proportional to the voltage $V(t')$ applied to this element. We might also expect the charge $\sigma(t', z')$ to be proportional to the width $w(z')$ in the x direction of the electroded region. The acoustic wave signal at the plane z at time t is defined as having an amplitude $a(t, z)$, where the power of the acoustic wave is $P = |a|^2$. This signal reaches the plane z from the element between z' and $z' + dz'$ after a delay time $(z - z')/V_R$, where V_R is the Rayleigh wave velocity. In a linear medium, the acoustic amplitude $a(t, z)$ must be proportional to the excitation charge $\sigma(t', z')$. The total acoustic signal at the plane z, $a(t, z)$, must therefore be of the form

$$\begin{aligned} a(t, z) &= \alpha \int_{-\infty}^{z} \sigma\left(t - \frac{z - z'}{V_R}, z'\right) dz' \\ &= \alpha \int_{-\infty}^{z} \sigma(t', z')\, dz' \end{aligned} \tag{4.2.1}$$

where α is a coupling constant independent of the value of $\sigma(t', z')$, $t' = t - (z - z')/V_R$, and $\sigma(t', z')$ is the charge density induced at the plane z' at a time t'.

If we take $\sigma(t'\ z') = \sigma(z')\ \sigma(t' - t_0)$, which corresponds to a very short input voltage pulse at $t' = t_0$, we find that

$$a(t, z) = \alpha \int_{-\infty}^{\infty} \sigma(z')\delta\left(t - t_0 - \frac{z - z'}{V_R}\right) dz' \tag{4.2.2}$$

and

$$a(t, z) = \alpha\sigma[z - V_R(t - t_0)] \tag{4.2.3}$$

where we have assumed that $\sigma(z') = 0$ for $z' > z$, and $\delta(t)$ is the Dirac delta function. Thus the impulse response $a(t, z)$ of the transducer is proportional to the charge distribution $\sigma(-V_R t + \text{constant})$ along it, where $z' = z - V_R(t - t_0)$. Note that as t is increased, the value of z' decreases; in fact, the impulse response $a(t, z)$ depends directly on $\sigma(-z')$.

This property is very useful because it enables us to design a transducer for a given frequency response merely by spacing or shaping its fingers to give a certain impulse response. We now have only a problem of simple geometrical design.

We shall assume, for simplicity, that the charge distribution is uniform over each finger, although it is fairly easy to modify this assumption to account for a charge distribution closer to reality, like the electrostatic charge distribution of a periodic array of strips. We shall define $\sigma(z')$ to be finite only where there are metal fingers.

More generally, if the finger width or position is altered, and the input voltage is $V(t')$, we can write

$$\sigma(t', z') = \mu V(t')w(z') \tag{4.2.4}$$

where μ is a constant.

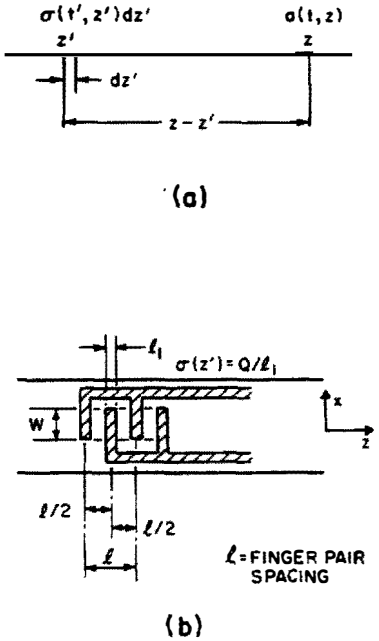


Figure 4.2.4 Notation used (a) to estimate the response of a transducer, and (b) for the finger pair spacing and width.

We shall call $w(z')$ the *finger weighting*. Typically, $w(z')$ is proportional to the finger width w, but changes sign depending on how the finger is connected to the external source.

It follows from Eqs. (4.2.1) and (4.2.4) that

$$a(t, z) = \kappa \int_{-\infty}^{z} V\left(t - \frac{- - -'}{\qquad}\right) w(z')\, dz' \tag{4.2.5}$$

where $\kappa = \alpha\mu$ is a constant. The respon ıort pulse at time $t' = 0$, which may be represented as $V(t') = \delta(t')$, is

$$h(t) = \kappa \int_{-\infty}^{z} \delta\left(t - \frac{-}{V_R}\right) w(z')\, dz' \tag{4.2.6}$$

or

$$h(t) = \kappa w(z - V_R t) \tag{4.2.7}$$

If we put $t' = t - (z - z')/V_R$ in Eqs. (4.2.5) and (4.2.7), it follows that

$$\begin{aligned} a(t, z) &= \kappa V_R \int V(t') w[z - V_R(t - t')]\, dt' \\ &= V_R \int V(t') h(t - t')\, dt' \end{aligned} \tag{4.2.8}$$

We may compare Eq. (4.2.5) or (4.2.8) with the usual forms for correlation and convolution:

$$\begin{aligned} \text{Correlation:} \quad s(t) &= \int g(t + \tau) f(\tau)\, d\tau = g(t) * f(t) \\ \text{Convolution:} \quad s(t) &= \int s(t - \tau) f(\tau)\, d\tau = g(t) * f(t) \end{aligned} \tag{4.2.9}$$

It will be seen that the signal induced in the acoustic delay line is either the *correlation* of the transducer weighting in the $+z$ direction with the input signal, or the *convolution* of the transducer weighting in the $-z$ direction with the input signal. This property makes the SAW device a powerful tool for signal processing. To understand it in more physical terms, the reader is referred to the example of digitally encoded transducers discussed earlier in this section.

The frequency response of the transducer is the Fourier transform of its time-domain response. Thus we can write

$$A(\omega, z) = \mathcal{F}[a(t, z)] = \int_{-\infty}^{\infty} a(t, z) e^{-2j\pi ft}\, dt \tag{4.2.10}$$

where the symbol $\mathcal{F}$ denotes the Fourier transform, and $f = \omega/2\pi$ is the frequency.

The frequency response $A(\omega, z)$ of the excited acoustic signal will be the product of the frequency response of the input signal $V(\omega)$ and the frequency response of the transducer $H(\omega)$:

$$A(\omega, z) = V(\omega) H(\omega) \tag{4.2.11}$$

It will be seen from Eq. (4.2.7) that $H(\omega)$ is given by the relation

$$H(\omega) = \kappa \int_{-\infty}^{\infty} e^{-j\omega t} w(z - V_R t)\, dt \tag{4.2.12}$$

Putting $t = (z - z')/V_R$, it follows that

$$H(\omega, z) = \gamma e^{-j\omega z/VR} \int_{-\infty}^{\infty} e^{j\omega z'/V_R} w(z')\, dz' \tag{4.2.13}$$

where $\gamma = -\kappa/V_R$. We have now obtained the important relation that *the frequency response of the transducer is the Fourier transform of the finger weighting*.† For a more direct derivation of this relation, see Sec. 2.5.

4.2.2 Bandpass Filter

Uniform transducer. We shall consider the example of a standard uniform transducer (Fig. 4.2.1) with constant finger length and uniform finger pair spacing, and derive how its response varies with frequency. Suppose that the fingers have a pair spacing of l and a finger width l_1, and that the charge on every other finger is Q and $-Q$, respectively. We shall assume that l_1 is very small (i.e., $kl_1 << 1$), where $k = \omega/V_R = 2\pi/\lambda$ and λ is the wavelength of the acoustic wave. Thus each finger is taken to be very narrow in comparison to a wavelength. We shall also assume, for simplicity, that σ is uniform over a finger, so that $\sigma l_1 = \pm Q$. With the correct normalization, since we are only interested in the relative amplitude of the response as a function of frequency, we can put $w(z') = \pm 1/l_1$. It follows from Eq. (4.2.13) that the contribution to the response $H(\omega)$ by a pair of fingers spaced $l/2$ apart, with the center of the pair at $z' = nl$, and the centers of the fingers at $z' = nl + l/4$ and $z' = nl - l/4$, respectively, is

$$H(\omega) = \frac{\gamma}{l_1}\left(-\int_{nl - l/4 - l_1/2}^{nl - l/4 + l_1/2} e^{jkz'}\, dz' + \int_{nl + l/4 - l_1/2}^{nl + l/4 + l_1/2} e^{jkz'}\, dz' \right) \tag{4.2.14}$$

or

$$H(\omega) = \gamma(e^{jkl/4} - e^{-jkl/4}) \frac{\sin (kl_1/2)}{kl_1/2} e^{jknl} \tag{4.2.15}$$

where we have omitted the exp $(-j\omega z/V_R)$ term. This is equivalent to defining the response at the plane $z = 0$. When $kl_1 << 1$, it follows that

$$H(\omega) = 2j\gamma \sin \frac{kl}{4} e^{jknl} \tag{4.2.16}$$

†This is actually the inverse Fourier transform of $w(z)$. However, it is the Fourier transform of $w(-z)$, or the finger weighting in the direction of decreasing z.

The contribution from the N-finger pair of the transducer is therefore the sum of the contributions from the individual finger pairs, or

$$\begin{aligned} H(\omega) &= 2j\gamma \sin \frac{kl}{4} \sum_{0}^{N-1} e^{jknl} \\ &= 2j\gamma \sin \left(\frac{kl}{4}\right) \frac{\sin (kNl/2)}{\sin (kl/2)} e^{jk(N-1)l/2} \\ &= j\gamma \frac{\sin (kNl/2)}{\cos (kl/4)} e^{jk(N-1)l/2} \end{aligned} \tag{4.2.17}$$

The parameter $H(\omega)$ is the Fourier transform of the charge distribution on the fingers.

The center frequency of the transducer $\omega = \omega_0$, $k = k_0$ is where $kl = 2\pi$ and the wavelength is $\lambda = l$. Using l'Hospital's rule on Eq. (4.2.17), it follows that $|H(\omega_0)| \rightarrow 2N\gamma$. The response is zero where $kl = 2\pi[1 \pm (1/N)]$; thus the bandwidth is narrowed as N increases. The bandwidth between the zeros in the response is given by the relation

$$\frac{\Delta\omega(\text{zero})}{\omega_0} = \frac{2}{N} \tag{4.2.18}$$

It is convenient to normalize Eq. (4.2.17) in terms of kNl, so that in the normalized form all transducers have the same response and have a maximum response when the normalized parameter is zero. We write

$$x = \frac{N\pi(k - k_0)}{k_0} = \frac{N\pi(\omega - \omega_0)}{\omega_0} \tag{4.2.19}$$

In this case, if $|\omega - \omega_0| \ll \omega_0$, we can see that when $x/N \ll 1$, $\cos kl/4 \rightarrow -x/2N$ and

$$|H(x)| \approx 2\gamma N \left|\frac{\sin x}{x}\right| \tag{4.2.20}$$

Thus all transducers have the same normalized response. It follows that the total bandwidth to the 4-dB points ($x = \pm\pi/2$) is

$$\frac{\Delta\omega(4 \text{ dB})}{\omega_0} = \frac{1}{N} \tag{4.2.21}$$

The 3-dB points occur where $\Delta\omega(3 \text{ dB})/\omega_0 = 0.89/N$. This behavior is illustrated in Fig. 4.2.5.

Apodized transducer. It is apparent that if we taper the length of the fingers and their spacing, we can tailor the response to any given characteristic. Such variation of the finger length is called *apodization*. For example, with a uniform finger spacing, but the finger lengths varying as $|\sin \mu z'/\mu z'|$ and the finger polarity depending on the sign of $\sin \mu z'$, the total charge on the finger will vary

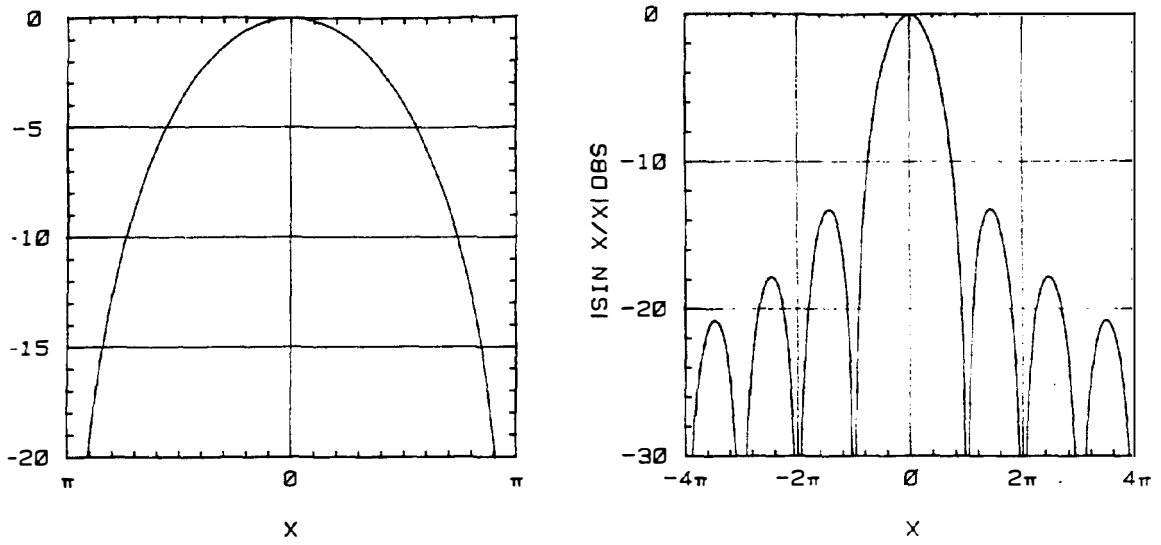


Figure 4.2.5 Plots of $|(\sin x)/x|$ in decibels.

as sin $\mu z'/\mu z'$. The transducer output will be the Fourier transform of the finger excitation, and the result will be a square-topped frequency response. The difficulty, of course, is that the transducer length L must theoretically be infinite. Furthermore, the fingers at each end of the transducer must be progressively shorter in length, and acoustic wave diffraction effects then become dominant.

An illustration of how filters with such a configuration are constructed is given in Fig. 4.2.6(a) and (b). The simple apodized transducer shown in Fig. 4.2.6(a) can yield a distorted response, for the velocity of a surface acoustic wave passing under a metal finger will be slightly different from the velocity in an unmetalized region of the piezoelectric delay line; this causes the acoustic beam to have phase distortion across its width. One way to eliminate this distortion is to place dummy fingers in the gap that are connected to the bus of the same polarity as the neighboring fingers. Then all parts of the acoustic beam suffer the same phase delay,

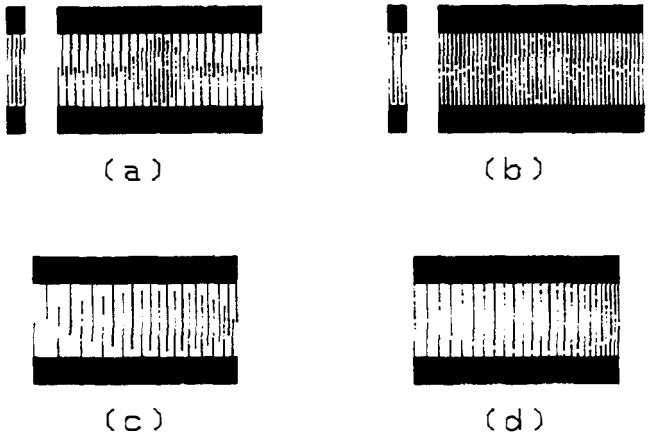


Figure 4.2.6 (a) Nondispersive bandpass filter configuration: one transducer is unapodized and one transducer is apodized with a sin x/x variation; (b) similar filter corrected with "dummy" electrodes to produce straight-crested surface waves (i.e., waves with no phase distortion across their width due to the presence of fingers); (c) electrode configuration of apodized dispersive transducer with uncorrected comb; (d) electrode configuration of apodized dispersive transducer with corrected comb.

and we obtain a beam with straight wavefronts, a *straight-crested wave.* A dispersive filter with variable finger spacing is shown in Fig. 4.2.6(c) and (d). This approach, taking into account such modifications as the charge distribution over a finger, diffraction, and weak reflections from the fingers, is the basis of design for interdigital transducer filters.

4.2.3 FM Chirp Analog Filter

Another important example of signal processing within an SAW delay line is one that uses an analog signal of the type illustrated in Fig. 4.2.7. The signal shown is an FM chirp signal whose amplitude is constant, but whose instantaneous frequency varies linearly with time. The finger spacing of the transducer array is varied along its length to match the frequency variation across the chirp (i.e., to correlate with it). The left-hand end of the array responds to the higher frequencies and the right-hand end to the lower frequencies. At the instant shown, the chirp signal, which is traveling to the right, registers exactly with the array, much as we have observed for the digital filter. An intense output pulse is obtained at the right-hand terminals. This is a dispersive filter in which the low-frequency end of the signal is delayed more than the high-frequency end, allowing the trailing edge of the long input pulse to catch up with the leading edge, thus collapsing the pulse.

Pulse compression techniques of this type are of great importance in a variety of systems. Perhaps the best known examples are radar systems using pulse compression in which the signal transmitted from the radar is an FM chirp; after returning from a target, it is passed through a pulse compression filter which compresses it into a short pulse [1]. In this way, it is possible to use a long pulse containing large amounts of energy for long-distance ranging, and compress it in a receiver into a short intense pulse for accurate timing and range resolution.

In secure radar and communications systems, an FM chirp represents one way of coding a signal. The listener needs to know the code (i.e., the chirp rate) and needs to have a chirp compression filter which operates at this rate in order

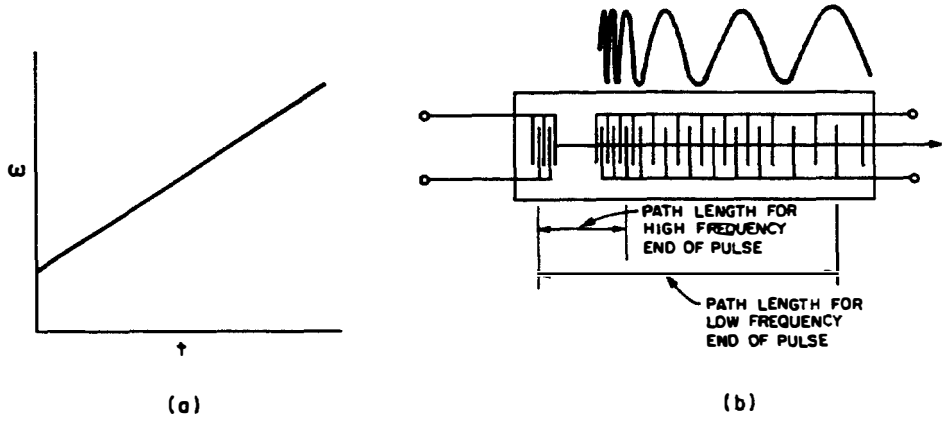


Figure 4.2.7 FM chirp pulse compression filter: (a) chirp frequency variation as a function of time; (b) schematic representation of the SAW filter.

to receive the signal. Indeed, the signal can be below the thermal or background noise level and, when received by a compressive receiver containing a filter matched to its chirp rate, can still be extracted from the background noise level because of the improvement in signal-to-noise ratio resulting from pulse compression. On the other hand, an ordinary receiver would not be able to detect the presence of the signal.

Surface acoustic waves fit naturally into this picture in applications requiring high chirp rates (i.e., a high rate of change of frequency versus time across the chirp), together with a large total frequency excursion (bandwidth). Compression ratios in the range 1000 to 10,000 can be reached with SAW systems [2]. SAW pulse compression filters have been built with bandwidths exceeding 500 MHz and with time delays (chirp lengths) of the order of a microsecond. The range resolution of a radar system is the inverse of its bandwidth, and this bandwidth (500 MHz) corresponds to a target range resolution capability of 1 ft (0.31 m). Far larger time delays have been achieved (more than 100 μs) at smaller bandwidths.

Fourier transform chirp filters are also applicable to nonscanning real-time spectrum analyzers. In fact, they are capable of calculating the complete complex Fourier transform of an arbitrary incoming analog signal. Consider the chirp signal illustrated in Fig. 4.2.8, whose frequency is $\omega = \omega_0 + \mu t$, and thus varies linearly with time. Suppose now that the chirp is modulated by a sinusoidal signal $\sin \Omega t$. This will introduce two sidebands of frequencies $\omega = \omega_0 + \mu t \pm \Omega$, in addition to the carrier of frequency $\omega = \omega_0 + \mu t$. Thus the system behaves as if there are three identical chirps delayed from each other by times Ω/μ. When these chirps are inserted into the correct matched dispersive delay line, three output pulses will be obtained. The differences in their time delays will indicate their frequency differences. More generally, an arbitrarily modulated chirp, when inserted into a chirp filter matched to the carrier, gives an output that is the Fourier transform of the modulated signal.

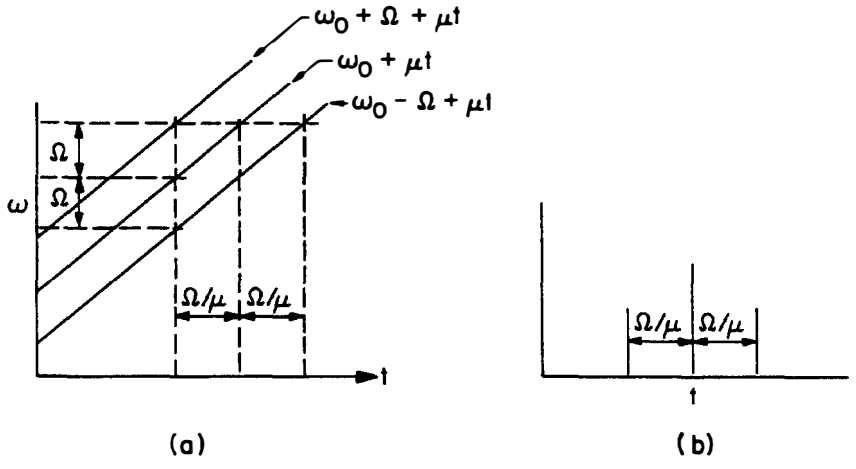


Figure 4.2.8 (a) Chirp modulated by a sinusoidal signal; (b) output from the matched filter.

4.2.4 Resonators

The conducting electrodes used to form interdigital arrays cause small reflections of surface waves which can lead to spurious low-level signals. In designing long filters, these reflections must be limited to specified levels. On the other hand, it is possible to use the reflection phenomenon constructively, by forming an SAW resonator from an array of reflectors [16, 17]. These are useful for electronic circuits, where it is advantageous to have very small, inexpensive, accurate, and pretuned resonators that can be mass-produced by photolithographic methods. Arrays of isolated parallel conducting strips deposited on the surface, as shown in Fig. 4.2.9, can behave as efficient reflectors for surface waves if a large number of strips (typically hundreds) are used, and if they are spaced so that the reflected waves from individual strips reinforce each other. As shown in Fig. 4.2.9(b), this occurs when the strips are $\lambda/2$ apart, for the waves reflected from two neighboring strips will trace distances differing by λ, and will therefore be in phase. Thus all the waves reflected from the strips will be in phase when the strips are spaced $\lambda/2$ apart.

In a surface wave resonator, the surface waves are trapped between the two reflectors of Fig. 4.2.9(a), making multiple transits between them and creating a standing wave, like electromagnetic waves in a cavity resonator or optical waves in a Fabry–Perot interferometer. Interdigital transducers are used to couple this resonant standing wave to an external electrical circuit. Grooves etched into the substrate can also be used effectively to form the reflective gratings.

Values of resonant Q ($Q = f_0/\Delta f$, where f_0 is the resonant frequency and Δf is the 3-dB bandwidth) as high as 20,000 have been achieved, with such resonators operating at frequencies in the range from 50 MHz to several gigahertz. The maximum Q is limited by the finite reflectivity of the arrays, for the propagation loss on quartz substrates (at least in the lower-frequency part of the range) is small enough to allow approximately one more order-of-magnitude increase in Q. As we might expect, these resonators can also be used as circuit elements to build ladder networks and other types of classical filter networks.

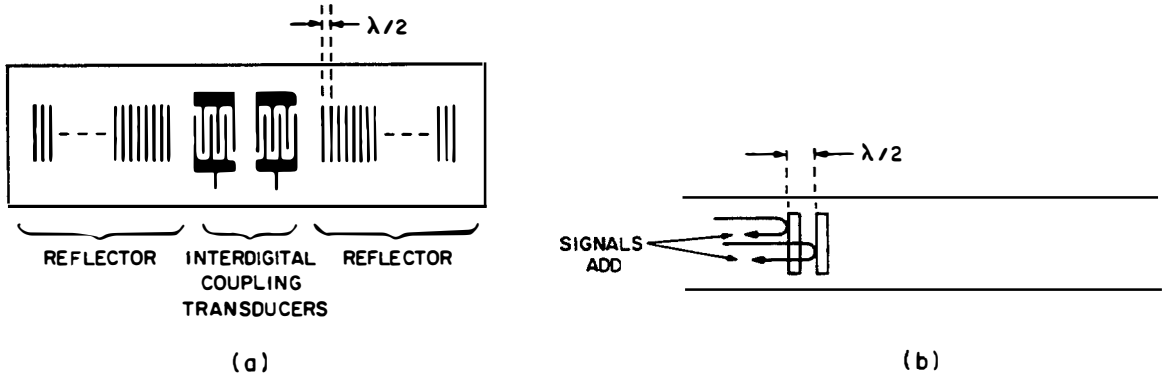


Figure 4.2.9 (a) SAW resonator; (b) waves reflected from neighboring fingers.

4.2.5 RAC Filters

The *reflective array compressor* (RAC) is closely related in its principle of operation to the surface wave resonator [18–20]. This device is an alternative form of transversal filter in which arrays of reflecting parallel grooves or metal strips on the delay-line surface perform the function of tapping, which is normally performed by interdigital electrode arrays. Just as with the resonator, we take advantage of the reflections resulting from the grooves or strips, instead of regarding them as a difficulty to be eliminated. The key idea is that a groove acts as a tap for the surface acoustic wave, because the portion of the surface acoustic wave reflected from the groove can be collected elsewhere by an interdigital transducer. The basic system is shown in Fig. 4.2.10: a surface acoustic wave is excited by transducer A and reflected by the two sets of grooves back to transducer D.

We can understand the operation of such devices by considering two extreme surface acoustic wave paths. A signal from the interdigital transducer A will be partially reflected into a surface acoustic wave traveling from B to C, where it will again be partially reflected and travel to interdigital transducer D. If rays reflected from the two sets of neighboring grooves in the two reflective arrays have a total difference in path length of λ, the acoustic wavelength, the effect will be strong. In this case, the rays traveling from B to C are in phase. The total time delay for this signal will be proportional to the path length $ABCD$. Thus if the spacing between the grooves is tapered, high-frequency signals will be reflected strongly from the front region of the grating array. However, a signal of lower frequency will follow the path $AEFD$ and experience a longer time delay. Thus this device has the same type of dispersion characteristic, with time delay as a function of frequency, as the FM chirp interdigital structure of Fig. 4.2.7, and can be used to perform the same functions. The folded paths in Fig. 4.2.10, however, give twice the time delay for a given substrate length, and chirp pulse compression filters with time delays exceeding 100 μs can be achieved.

Reflective arrays are less defect sensitive than interdigital arrays. They have an additional advantage for very high frequency operation, because the grooves in reflective arrays are generally spaced by the order of a half wavelength, compared to the quarter-wavelength spacing, which is more characteristic of split-finger, nonreflecting, interdigital arrays (see Prob. 3), the dimensional requirements are less stringent; furthermore, internal reflections no longer present a problem, so very long arrays may be made. This system has a major advantage over one using a very long interdigital transducer, because the many small reflections inherent in such transducers tend to build up and make it impossible to obtain the characteristics

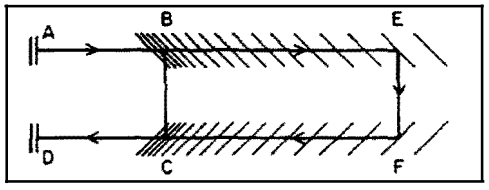


Figure 4.2.10 Reflective array compressor.

predicted by theory. RAC delay lines have been made with compression ratios of as much as 10,000 to 1 [18–20]; these results are probably a factor of 10 better than those obtained by any interdigital array system.

4.2.6 Stabilized SAW Oscillators

Another device of great importance is the SAW oscillator, which consists of a standard amplifier whose output is fed back to its input through a surface wave delay line, as illustrated in Fig. 4.2.11. This forms an oscillator operating in the UHF range (100 to 1000 MHz), whose frequency and frequency stability are determined by the properties of the SAW delay line crystal. This oscillator is simpler and cheaper to make than alternative approaches producing highly stable signals in this frequency range, such as relatively low frequency quartz crystal–controlled oscillators followed by multiplier chains.

The key point is that the delay line can be thousands of wavelengths long between the input and output transducers because of the very low propagation velocity of acoustic waves. As is well known, the frequency of an oscillator with an external feedback loop adjusts itself so that the phase shift around the loop is $2N\pi$, where N is an integral number. The larger the value of N, the better the short-term stability of the oscillation frequency. Both the basic frequency selectivity associated with the length of the delay-line path and the frequency-filtering characteristics of interdigital transducers can be brought into play. The stability of the delay-line crystal gives high stability for any of a number of different longitudinal modes (different values of N) with different center frequencies, and the interdigital transducers are used to select one oscillation mode from the entire set possible.

Methods of compensating for those phase shifts within the oscillator that result from voltage and temperature variations have been demonstrated. These methods make it possible to come very close to the ultimate frequency stability of the delay line itself. Techniques have also been devised to tune the voltage of the oscillator frequency over a small range by changing the phase in the external loop, for use in tracking or frequency modulation applications. Frequency synthesizers, which can be programmed to operate at any of a number of equally spaced frequencies with high short-term stability, have been demonstrated as well [22, 23].

At the time of writing, stable oscillators operating in the frequency range 100 MHz to 4 GHz have been demonstrated. Another class of SAW oscillators employs the SAW resonator; this form of oscillator, with an internal feedback loop, can be as useful as one with an external feedback loop.

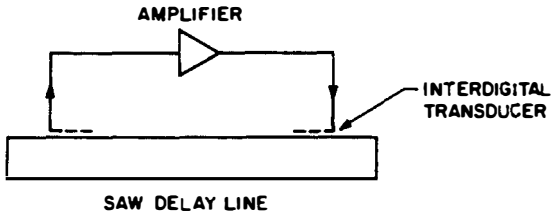


Figure 4.2.11 SAW oscillator with an external feedback loop. (From Crabb et al. [21].)

PROBLEM SET 4.2

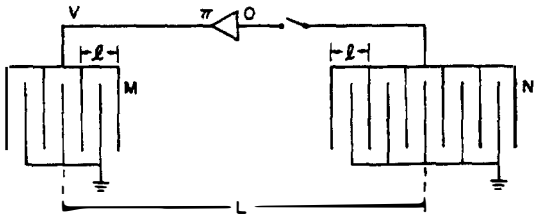
1. **(a)** Work out the response of an N-finger-pair interdigital transducer with fingers of finite width l_1 (kl_1 finite) and finger pair spacing l. Assume that the charge distribution on the fingers is uniform. The result should be a slightly more general version of Eq. (4.2.17).
   (b) Show that if $l_1 = l/4$, the response at the third harmonic $\omega = 3\omega_0$ is zero.

   Note: In practice the charge density on the fingers is not uniform, so the third harmonic response is finite.

2. An acoustic resonator uses N equally spaced strips as reflectors. Let each strip be of width d and their spacing be l. Suppose that we take the reflected wave from the nth strip to be of amplitude b and the incident wave, of amplitude a. Suppose also that each strip is of negligible width d ($kd \ll 1$), and that the reflection coefficient of each strip is $\Gamma = b/a$. Assuming, for simplicity, that $a_{n+1} = a_n \exp(-jkl)$, where $k = \omega/V_R$ (i.e., the forward wave is unperturbed, with a similar phase delay for the reflected waves). Find the total reflection coefficient from N strips. At what freqeuncy is this $N\Gamma$? Your result should be similar in form to Eq. (4.2.17). By using the same arguments as given after Eq. (4.2.17), find the 4-dB bandwidth of the reflected wave.

3. **(a)** Following the analysis of Prob. 2, modify the theory to take account of fingers of finite width d. Assume that k is unchanged, both under the fingers and in the region in between. Assume that the reflection coefficient of the wave from the back edge of the finger is Γ and from the front edge, $-\Gamma$ (this assumption can be justified from transmission-line analogs, where the region under the strip is taken to have an impedance different from that between the strips).
   (b) At what frequency is the total reflected wave amplitude identically zero?
   (c) Note that an interdigital transducer is normally operated with a finger spacing of approximately $\lambda/2$, where the reflection from the fingers is maximum. Suggest a configuration in which each finger is split down its middle with a gap between its two halves (both connected to the same bus bar) and which eliminates the reflection problem.

4. An SAW oscillator is made with the configuration shown below. Two transducers with M and N finger pairs, respectively, each with a finger pair spacing l, are used. The centers of the two transducers are a distance L apart, as shown in the figure. The amplifier should be taken to be an inverter that gives a π phase shift.



   (a) If the M-finger transducer is excited with a voltage $V \exp(j\omega t)$, find an expression for the frequency response of the output signal from the N-finger transducer when the switch to the amplifier is open. Assume that the signal induced on a finger whose center is at the plane z is proportional to $a(z, t)$, and that the wave velocity on the substrate is V_R.

   Hint: By reciprocity, the response of a receiving transducer is the same as a transmitting transducer.

(b) Suppose that $M = N$. Show how to choose the length L so that the phase delay from the input to the output of the delay line at the center frequency is $(2R + 1)\pi$, where R is an integer. This causes the device to oscillate when the switch is closed. If the response of the delay line is known at the center frequency, work out an expression for the response when the phase delay changes to $(2R - 1)\pi$.

(c) With $M = N$, suggest a choice for M that will make the loss very large at all other possible oscillation frequencies so as to avoid oscillations at these spurious frequencies.

5. A 7-bit Barker code has the form $+ + + - - + -$. A matched SAW receiving filter is made up for this code. Show that the correlation output of this filter to the code consists of a single peak of amplitude 7 and subsidiary pulses of amplitudes -1 or 0. How many subsidiary pulses are there?

6. A radar system uses an FM chirp signal of length T and frequency $\omega = \omega_0 + \mu t$. It is used to detect a target traveling with a velocity V away from the radar. The velocity of electromagnetic waves in air is c, and $c >> V$. Find the delay time to the target which is at a distance $R = R_0 + Vt$. Then find the delay time of the return echo from the transmitter back to the receiver.

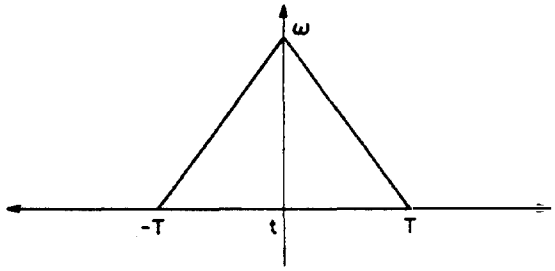
(a) If the signal emitted by the radar is $f(t)$, find the form of the return echo signal. Assume that $V << c$ and the bandwidth of the signal is much smaller than its center frequency. Use the formula for a Fourier transform to show that there is a simple linear Doppler shift of frequency proportional to V/c (keep only terms linear in V/c).

Hint: a signal $f(t)$ after travelling a distance $2R$ has the form $f(t - 2R/c)$. Consider the effect of R varying with time.

(b) Now suppose that a chirp signal of length T is inserted into a matched filter of length T. Estimate how the output falls off with target velocity. Show that for $T = 2$ ms, $f_0 = 10$ GHz, a bandwidth $B = 2$ MHz, and a velocity V of 100 m/s, there is essentially no change in the output peak pulse amplitude. Thus we cannot differentiate between a change in range and a change in velocity of the target.

Hint: The output is proportional to the number of active taps in the filter (i.e., the taps that respond to their corresponding frequencies). If there is a large frequency shift in the chirp, some taps in the filter are not excited.

(c) Consider a V-FM chirp waveform of the type shown in the figure below.



If this signal of length $2T$ is inserted into a matched filter of length $2T$, it will give a simple pulse output. Show that by using this waveform, it should be relatively easy to distinguish a target with a small velocity $V << c$, as well as its range (the time width of a peak of a correlated FM chirp signal is approximately $\tau = 1/B$, corresponding to a frequency change in the input signal of B, where $B = \mu T/2\pi$. Take $T = 1$ ms, $V = 200$ ms, $B = 500$ kHz, and $f_0 = 10$ GHz. This example is a simple

illustration of the use of the ambiguity function of radar processing. The problem is to obtain good range definition as well as good velocity or Doppler resolution.

Note: The numbers given here are more appropriate for a CCD filter than for an SAW filter.

4.3 CHARGE-TRANSFER DEVICES

4.3.1 Introduction

Here we describe the principles of three types of *charge-transfer devices* (CTDs), which can be employed to process analog signals in much the same manner as the SAW devices discussed in Secs. 4.1 and 4.2. These devices employ the principle that a free electron or charge can be stored in a *metal-oxide semiconductor* (MOS) capacitor or in a PN diode capacitor. In the *charge-coupled device* (CCD) [4, 5] and the *bucket brigade device* (BBD) [8], the charge is transferred along a chain of capacitors; in the *single transfer device* (STD) [24], which is basically a switched capacitor system [7], the charge is read into a capacitor and then read out from the same capacitor at a later time.

In all these devices, the charge read into the system can be made to be linearly proportional to an input signal voltage, and the output signal can similarly be made proportional to the charge arriving at the output register. Because only a finite number of registers exists in such systems, all these devices must basically be sampled data signal processing devices (i.e., they store several samples during one RF cycle of the analog signal). Sampled data signal processing devices have been developed for use with digital signal processing. Here, however, because the charge can be made proportional to the input and output voltages, these devices can be used as analog sampled data systems. Thus CCDs, BBDs, and STDs offer the possibility of performing many sampled data filtering functions in the analog domain. While these devices combine some of the best features of digital and analog techniques, they also encounter some of the same difficulties. Like digital filters, CTDs are controlled by a master oscillator, but the need for analog-to-digital conversion is eliminated and all functions are performed in the analog domain. Furthermore, analog devices are usually much smaller and consume less power than those required for the equivalent digital signal processing systems. For instance, a digital device must use eight separate stages for an 8-bit system, while the equivalent sampled analog system requires only one stage. One advantage of analog devices is that they can be used as image sensors for television line scans, because light can generate charge within the registers or in neighboring registers. On the other hand, digital devices can be made with greater accuracy, and far more flexibility and programmability, than analog devices.

The analog signal processing functions that can be implemented with these devices are very similar to those made available by SAW devices. The nature of silicon technology gives us much more flexibility, albeit at a somewhat lower frequency range. A disadvantage, however, is the need to provide power, for these

are "active" rather than "passive" devices. A further disadvantage is that the "clock" signal used to control the transfer of charge from one register to another, or from input to output, can give rise to interference with the signals passing through the device. This is not a problem in digital devices, although the clock signals used in them can interfere with analog signals.

There are five broad applications of interest:

1. Analog delay, where the input waveform is sampled at a sampling frequency f_s and delayed by a time T_d.
2. Multiplexing, where a number of inputs are loaded in parallel into a register and clocked out serially. The complementary function of demultiplexing is also important, as is a combined multiplexer and demultiplexer, as used in the STD.
3. Transversal filtering, where the CTD device is nondestructively tapped at a number of points and the outputs are weighted and summed together.
4. Recursive filtering, where the CTD is tapped as in a transversal filter and the weighted outputs are fed back to the CTD input.
5. Correlation or convolution, in which two analog waveforms are multiplied and integrated together to form an output signal that is the correlation or convolution of these two signals.

Because these analog devices can easily be operated at lower frequencies than SAW devices, they are in a class by themselves for use at frequencies below 10 MHz. Furthermore, because the time delay through these devices can be varied at will, they can perform functions that would be impossible with surface acoustic waves. For many possible applications, there is an overlap between the two types of delay devices, so that either could be used. Relative cost of technological development and production determines which one shall be used; so do trade-offs in detailed performance characteristics important to the designer.

Some examples of simple delay-line applications are:

1. Line delays for the PAL and SECAM TV systems. In these systems the signals from neighboring lines are compared to assure perfect color reproduction. Fused quartz bulk wave acoustic wave delay lines have been employed for this purpose because of their low cost, but CCD and BBD delay lines are now being used in the same application.
2. Ghost suppression in television. Signals result from multipath interference (i.e., from signals reflected from buildings which arrive with a time delay T). The signal passing through this delay line is reversed in sign, and its amplitude is changed appropriately and added to the original signal. Thus it will cancel out the ghost, leaving a still weaker signal at a time $2T$. The signal produced by this process will normally be weak enough not to be noticed. Thus the use of a variable time-delay line with appropriate circuitry and possibly more complex algorithms with microprocessor control can, in principle, lead to a simple ghost eliminator. SAW delay lines have also been used for this purpose.

3. Time-delay beam-forming in sonar and acoustic imaging applications. By combining the appropriately delayed return signals from an array of acoustic transducers, a high-resolution beam can be formed in any direction [25]. Sonar and radar applications have been described by Jack et al. [26].
4. Audio delay lines for artificial reverberation and for changing the speed of taped speech without changing its frequency.
5. Devices for time stretching or compression. If a signal at one frequency is clocked into the delay line and then clocked out at a different rate, an input signal can be changed in frequency and total time length so as to reproduce it faithfully at either a higher or lower frequency. This allows us to inject high-frequency signals or fast transients into low-frequency devices, such as a computer. Alternatively, it enables us to take very low frequency signals and display them rapidly.
6. Time axis equalization in a videotape playback unit. If we monitor a constant-frequency tone recorded on tape, we can correct variation (resulting from stretching of the tape or of motor speed) by passing the video through a delay line whose clock frequency is controlled to keep a pilot tone at a constant frequency.
7. Analog memories for electronic counter measures. There are also closely related delay-line functions, such as rejection of *clutter* (i.e., interference from stationary ground targets) in moving target indicator (MTI) radar. The required operations can be performed with recursive filtering or transversal filters and will be discussed more fully when we deal with applications.

4.3.2 Bucket Brigade Devices

A schematic of the BBD configuration is shown in Fig. 4.3.1. In the integrated-circuit version illustrated in Fig. 4.3.1(b), it consists of a chain of *field-effect transistors* (FETs). Charge is transferred through this chain of FETs, which are connected source to drain–drain to source, by applying a clock signal to the gates. The system is arranged so that when an FET is cut off by a voltage on its gate electrode, the charge is stored at the drain in the so-called overlap capacity between the drain and the gate. As illustrated in Fig. 4.3.1(a), the input signal is applied at the first source and taken out at a last drain. In the configuration shown in Fig. 4.3.1(b), a layer of silicon dioxide is laid down over the silicon and extends over the $n+$ islands and the p-type channel regions that separate them. The only contacts that need to be made are to the first and last $n+$ islands; generally, these islands act as both source and drain.

We may understand the operation of the device by referring again to Fig. 4.3.1(b). We suppose that the clock signals V_1 and V_2 are square waves of opposite phase. Thus V_1 is positive when V_2 is negative, and vice versa. We take $|V_1| = |V_2|$. The gates themselves are supplied with a dc potential V_0 in addition to the clock signal.

We suppose that the input and output are both held positive. If V_1 is sufficiently positive, an n-type channel (inversion layer) under the odd numbered

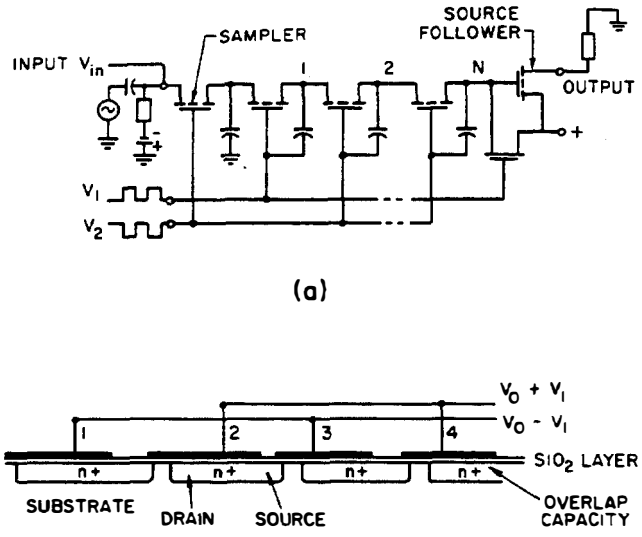


Figure 4.3.1 (a) Circuit configuration of an MOS delay line; (b) Sectional view of the integrated circuit.

gates will be induced (i.e., the odd channels will be conducting). At the same time, the *n*-type islands to the right of the odd-numbered gates will be driven positive because of the overlap capacity C, while the even-numbered gates are driven negative. Thus their overlap capacity with the *n* islands on the right will drive the islands more negative than those on the right-hand side of the odd-numbered gates. The more negative *n*-type islands therefore act as sources, and the more positive islands, as drains.

Referring now to Fig. 4.3.1(a), it will be seen that when the first FET is switched on, current can flow from the drain; thus the charge stored in its source capacitor moves from the source to the drain capacitor. When the next FET is turned on, the first is turned off, and the charge flows through it from its source (the drain of the first FET) to the capacitor at its drain.

We call a device of this kind a *bucket brigade device* (BBD) because the individual capacitors can be regarded as buckets that store the charge and the device transfers charge from bucket to bucket.

Suppose that V_T is the threshold voltage between the gate and the source for a conducting channel to form. Charge is injected into the conducting channel when the input is pulsed more negative than $V_0 + V_1 - V_T$ during the time that V_1 is positive. Thus the clock pulses sample the input waveform and the charge transfers along the system. The quantity of charge induced at the input FET is proportional to the potential $V_1 - V_T - V_{in}$. At the output the charge is detected as a momentary voltage drop across the output load. Provided that there are at least two output current pulses per RF input cycle, Nyquist's sampling theorem indicates that the input signal can be reproduced. This implies the use of a clock frequency at least twice that of the maximum input frequency.

Performance limitation of the BBD. We now consider the performance limitations of the BBD. We consider the BBD to be operated at a clock frequency

f_c. The maximum time for each transfer is therefore

$$\tau = \frac{1}{2f_c} \tag{4.3.1}$$

There will always be a finite amount of charge left behind after a time τ, and the charge left behind at each register will be cumulative. Ultimately, the signal passing through the registers becomes unacceptably degraded, and a limitation on either the number of registers or the clock frequency will result. It is apparent that the higher the clock frequency, the worse the degradation of the signal.

To provide an understanding of some of the basic concepts of CTDs, let us carry out a mathematical treatment of the performance of this device. We first consider the situation when no charge is transferred along the system. If there is no charge on a drain just before the clock voltages switch from one polarity to the other, the potential on the new source V_s just after switching is

$$V_s = V_0 - V_1 \tag{4.3.2}$$

However, assuming current saturation, the current that flows in an FET channel (the inversion layer) can be shown to be of the form

$$\begin{aligned} I &= \beta(V_G - V_s - V_T)^2 \\ &= \beta(V_0 + V_1 - V_T - V_s)^2 \end{aligned} \tag{4.3.3}$$

where $V_G = V_0 + V_1$ is the gate voltage, $\beta = \mu k_1 Z C_{Ox}/L$, μ is the *mobility* of the charge carriers, k_1 is a constant ~ 0.5, Z is the width of the channel (inversion layer), C_{Ox} is the capacity per unit area of the oxide, and L is the length of the channel. It will be apparent that the current is zero if

$$V_s = V_0 + V_1 - V_T \tag{4.3.4}$$

It therefore follows from Eqs. (4.3.2) and (4.3.4) that there is no charge on the source initially, and no charge flows from source to drain if the clock voltage is such that

$$V_1 = \frac{V_T}{2} \tag{4.3.5}$$

We now consider what occurs when charge is inserted at the input of this FET, and we keep $V_1 = V_T/2$. The charge on a source whose voltage is V_s is

$$\begin{aligned} Q &= -C(V_0 - V_1 - V_S) \\ &= -C(V_0 + V_1 - V_T - V_s) \end{aligned} \tag{4.3.6}$$

where C is the capacity between the gate and the source. It follows that the current flowing from source to drain is

$$I = \frac{-dQ}{dt} \tag{4.3.7}$$

Equations (4.3.3), (4.3.6), and (4.3.7) yield the following relation for Q:

$$\frac{dQ}{dt} = \frac{-\beta Q^2}{C^2} \tag{4.3.8}$$

This expression can be integrated to obtain the following formula for the charge as a function of time:

$$Q = \frac{Q_0}{1 + \beta Q_0 t/C^2} \tag{4.3.9}$$

Thus at a time $\tau = 1/2f_c$, the charge remaining on the source is

$$Q(\tau) = Q_0\left(1 + \frac{\beta Q_0}{2f_c C^2}\right)^{-1} \tag{4.3.10}$$

where Q_0 is the initial charge on the source.

We assume that $(\beta Q_0/2f_c C^2) >> 1$ [i.e., $Q(\tau) << Q_0$]. Thus most of the charge is transferred from source to drain in time τ. Using a Taylor expansion in the parameter $2f_c C^2/\beta Q_0$, it follows that

$$\begin{aligned} Q(\tau) &= \frac{2f_c C^2}{\beta}\left(1 + \frac{2f_c C^2}{\beta Q_0}\right)^{-1} \\ &= \frac{2f_c C^2}{\beta} - 4\left(\frac{f_c C^2}{\beta}\right)^2 \frac{1}{Q_0} \end{aligned} \tag{4.3.11}$$

Remember that $Q(\tau)$ is the charge left behind on the source. The first term is a constant charge that depends on the frequency (i.e., the average charge left behind because a finite time is used for transfer of charge). The second term is dependent on the initial charge on the source Q_0.

Note that if we use a different clock voltage, $V_1 > V_T/2$, the charge transferred along the system will still be given by Eq. (4.3.9). The only difference will be that the source will assume an average dc level corresponding to a nontransferable charge Q_F, for which the current in the channel is zero, and V_s is given by Eq. (4.3.4).

We consider the results of Eq. (4.3.11). When we take representative values of $\beta = 10^{-5} A/V^2$, $Q_0 = 1$ pC, $C = 0.5$ pF, and $f_c = 1$ MHz, the remaining charge $Q(\tau)$ will be approximately 0.05 pC; this is mainly due to the first term. Thus 5% of the charge remains. On the other hand, with $Q_0 = 2.5$ pC and $Q(\tau) = 0.05$ pC, as before, only 2% of the charge remains because there is an approximate linear relation between $Q(\tau)$ and f_c. Thus after 20 transitions, 36% of the charge would be transmitted in the first case and 67% in the second, and the BBD would not be useful.

The basic problem is that the source charge cannot easily be reduced to zero; this would take a very long time. Instead, it is better to work with what is called a "fat zero." Suppose that we do not require the charge to be reduced to zero, and consider instead what occurs if we work with a finite minimum level of charge.

If the injected charge is Q_1, the corresponding remaining charge will be

$$Q_1(\tau) = \frac{2f_cC^2}{\beta} \quad 4\left(\frac{f_cC^2}{\beta}\right)^2 \frac{1}{Q_1} \tag{4.3.12}$$

If the injected charge is changed t

$$Q_2(\tau) = \frac{2f_cC^2}{\beta} - 4\left(\frac{f_cC^2}{\beta}\right)^2 \frac{1}{Q_2} \tag{4.3.13}$$

If we regard the charge Q_1 as the minimum charge, the dc level, or the fat zero, we now need only consider $\Delta Q = Q_2(\tau) - Q_1(\tau)$ as the remaining difference charge. Thus our criteria for efficiency will now be based on a parameter ε, defined by the relation

$$\varepsilon = \frac{Q_2(\tau) - Q_1(\tau)}{Q_2 - Q_1} \tag{4.3.14}$$

We wish to have ε as small as possible. We see that

$$Q_2(\tau) - Q_1(\tau) = 4\left(\frac{f_cC^2}{\beta}\right)\left(\frac{1}{Q_1} - \frac{1}{Q_2}\right) \tag{4.3.15}$$

Hence

$$\begin{aligned} \varepsilon &= 4\left(\frac{f_cC^2}{\beta}\right)^2 \frac{1}{\overline{Q_1Q_2}}\left(\frac{1}{Q_1} - \frac{1}{Q_2}\right) \\ &= 4\left(\frac{f_cC^2}{\beta}\right)^2 \frac{1}{\overline{Q_1Q_2}} \end{aligned} \tag{4.3.16}$$

Originally, a similar criterion was

$$\varepsilon_0 = \frac{2f_cC^2}{Q_0\beta} \tag{4.3.17}$$

Now, with the use of a fat zero, $\varepsilon \sim \varepsilon_0^2$. For example, with $Q_2 = 2.5$ pC, $Q_1 = 1$ pC, and the same parameters as before, $\varepsilon = 0.094\%$, where $\varepsilon_0 = 5\%$ and 2%, for $Q_0 = 1$ pC and 2.5 pC, respectively. Thus the difference charge passed through the register is 0.99906 of the difference charge entering it, and the signal is transmitted faithfully. After N transfers, the signal charge transferred through the system will be

$$(\Delta Q)_N = \Delta Q(1 - \varepsilon)^N = \Delta Q e^{-N\varepsilon} \tag{4.3.18}$$

Thus ΔQ has dropped to 98% of its original value after 20 transfers. Note, however, that ε now varies as f_c^2. Thus if we increase the frequency to 2 MHz, $\varepsilon = 0.37\%$, we can only allow five transfers for a 2% degradation. We can make further improvements by using shaped clock waveforms and far shorter gates.

In real BBD devices, the configuration is somewhat more complicated, as shown in Fig. 4.3.2. In the device shown in Fig. 4.3.2(a), the input signal usually

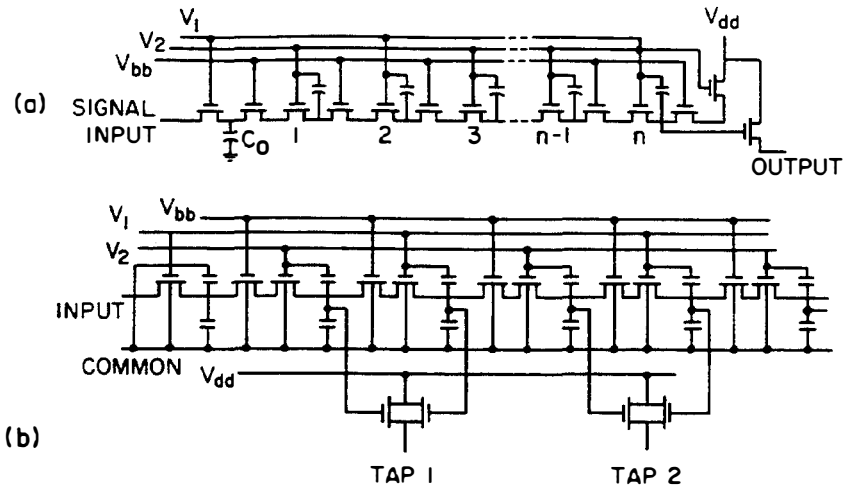


Figure 4.3.2 BBD structure used by Reticon: (a) input circuit; (b) method of tapping. (From Reticon Corporation [24].)

consists of a band-limited analog signal which varies around a dc bias level. This automatically introduces a fat zero and gives better incremental transfer efficiency. A further change in the configuration in Fig. 4.3.2(a) can be made by using tetrode rather than triode type FETs (i.e., intermediate FETs with fixed voltages on the gates). This isolates the drain capacitor from the source capacitor, eliminating the Miller effect (i.e., the feedback between output and input).

As we shall see, the CCD may be simpler to make in principle but a great deal more difficult in practice. In certain forms, however, it does lend itself to higher-speed operation. BBD devices have an advantage over CCDs because they are easily employed for making externally tapped devices. The configuration used by Reticon in their tapped device is shown in Fig. 4.3.2(b). The output is taken from a pair of balanced source followers at the tap, giving an output that is the difference between the charges on two neighboring registers (i.e., between the source and the drain). Thus the fat zero is removed from the output on these taps. This means it is possible to construct a tapped delay line with individual taps that have adjustable weights. Typical tapped delay lines have 32 or more taps.

A similar system can be constructed in the BBD configuration to give an input, instead of an output, at each tap, so that the output can be taken out of the delay line with parallel inputs. This configuration was used by Reticon and is shown in Fig. 4.3.3. In this device, two parallel BBD lines are used to provide isolation between adjacent even and odd inputs. In both delay lines, each signal sample is isolated on either side by a reference sample. The two parallel bucket brigade delay lines are driven from two-phase complementary clocks. The delay line receives the input signal charges in parallel, from corresponding even or odd input storage capacitors that have previously stored samples of the input signal. These new input signal charges are transferred into the bucket brigade capacitor storage sites under alternate V_1 and V_2 gates. The even-numbered input taps transfer signal samples to nodes under alternate even-numbered V_2 gates of the

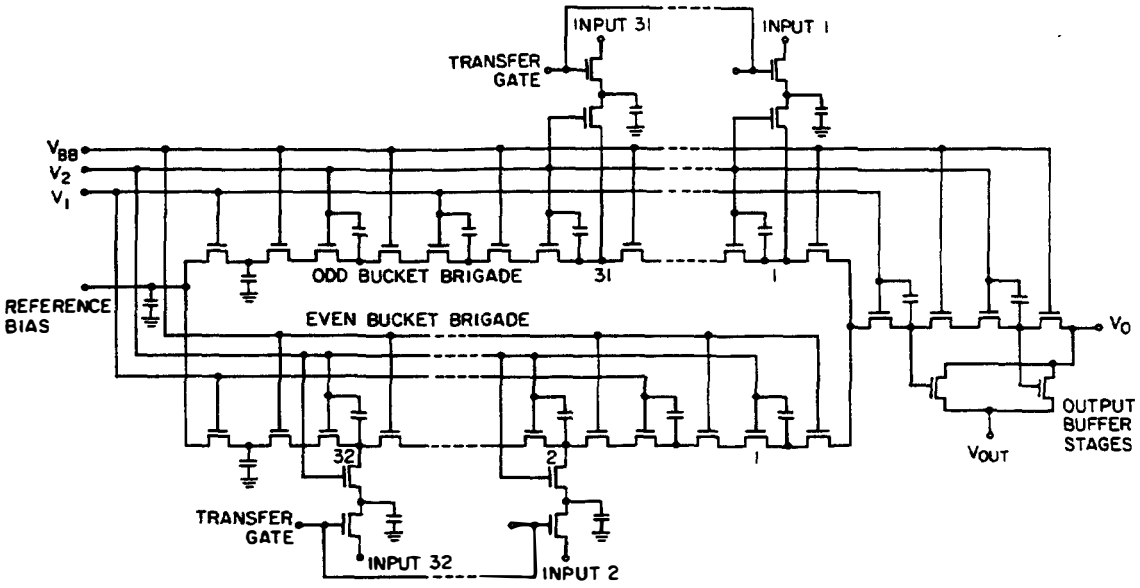


Figure 4.3.3 Equivalent circuit of Reticon tapped BBD. (From Reticon Corporation [27].)

even bucket brigade, while the odd-numbered taps transfer this sample to nodes of the odd-numbered gates of the odd bucket brigade.

The sampling to odd and even stages permits the outputs of the two bucket brigades to be multiplexed (i.e., switched alternately into a single channel) and yields a serial stream of pulses at the output. The combined signal channel output appears at two successive storage sites. It has the advantage of giving a balanced output, and multiplexing makes it possible to use fewer stages in a series for a given number of registers. This implies that the clock frequency can be reduced by a factor of 2, so that the requirements of high-frequency operation are reduced by such multiplexing schemes. Thus the multiplexing arrangement is useful not only for tapped systems, but for any system where high-frequency operation is a problem.

4.3.3 Charge-Coupled Devices

The CCD is even simpler in concept than the BBD. The essence of its operation is to store minority carriers in a spatially defined depletion region (defined by the clock potentials) at the surface of a semiconductor, and to move this charge about by moving the potential well in which the charges are stored. The charge can be injected at one end of the device, moved along it, and detected at the output. Alternatively, charge can be injected electrically through taps at points along the device, or by incident light, which generates charge. This makes it possible to construct a camera that can detect one line of an image and then read out the information sequentially from the output of the storage device [4, 5, 28].

To understand the storage mechanism, let us first consider a single *metal insulator semiconductor* (MIS) structure on an *p*-type semiconductor. We could, for instance, consider one element of a system like that shown in Fig. 4.3.4, in

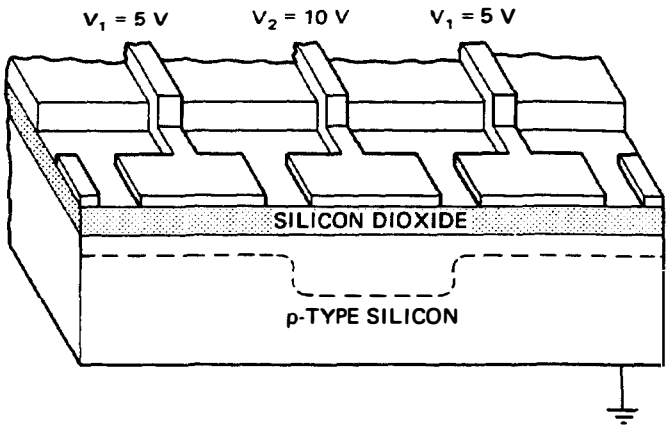


Figure 4.3.4 Cutaway of a CCD in the storage condition. The dashed line represents both the edge of the depletion region and the potential distribution. Voltages are typical. (From Sequin and Tompsett [5].)

which the center electrode is kept at a potential more positive than that of the two surrounding electrodes, thus forming a potential well. Normally, in thermal equilibrium, an inversion layer or surface charge of electrons is generated and forms just under the silicon dioxide. Such a layer may take seconds to form. Under nonequilibrium conditions, when there is no charge present in this inversion layer, the surface is depleted of minority charge, a potential well that is available to accept electrons. If we now suppose that charge (in this case, electrons) is introduced into the depletion region, the presence of electrons will cause the surface to assume a more negative potential and the depletion layer will reduce in width. At a time $\tau = \infty$, after thermal equilibrium has been reached by generating electrons from surface taps, the interference potential reaches its equilibrium value. Any further introduction of electrons, which will make the interface potential still more negative, will eventually lead to injection of electrons into the bulk; in other words, we cannot store more charges than would possibly be formed in the inversion layer in thermal equilibrium.

We now consider the three-phase charge-coupled device shown in Figs. 4.3.4, 4.3.5, and 4.3.6. In this type of device, every third electrode is connected to a common conductor. Each of these common conductors is a separate clock line and forms a three-phase clocking system. Initially, a voltage V_2 is applied to electrodes 1, 4, 7, and so on, and a voltage V_1, $|V_2| > |V_1|$ is applied to the other electrodes. The semiconductor is held at zero potential.

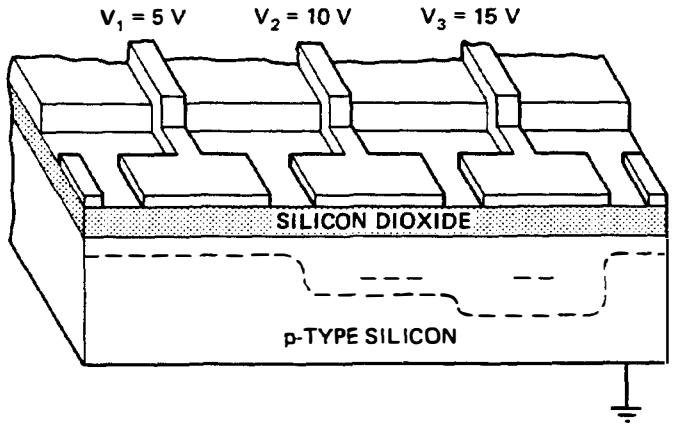


Figure 4.3.5 Cutaway of a CCD in the transfer condition. (From Sequin and Tompsett [5].)

It is assumed that $|V_1| > |V_T|$, where V_T is the threshold voltage for inversion under thermal equilibrium conditions. Under conditions shown in Figs. 4.3.4 and 4.3.6(a), and those that we have described, charge can be stored in the potential well so that electrons which have been introduced into this region will be stored at the oxide interface. We now suppose that a voltage V_3 is applied to electrodes 2, 5, 8, and so on. As shown in Figs. 4.3.5 and 4.3.6(b), $V_3 > V_2$. The charge will now transfer from electrode 1 to the potential minimum under electrode 2, and so on. The voltages are changed to those shown in Fig. 4.3.6(c). Charge is stored under electrodes 2, 5, 8, and so on, and the process is repeated once more to transfer charge along the system.

In this discussion, it is of course assumed that charge is transferred along the system in times much shorter than the storage time or generation time τ_g, where $\tau_g = Q/I_g$ is the time for generation of the charge Q by the thermally generated current I_g in the depletion layer and the surface states. In carefully constructed devices with high-quality bulk material and a small number of surface states, such times can be of the order of several seconds, for the main source of generated current is due to surface states.

So far, we have discussed devices with a three-phase clock. The disadvantages of such a system are apparent. For one thing, the minimum clock frequency must be three times the maximum frequency of the signal passed through the device, which puts more stringent criteria on the storage time in the device and the highest frequencies used. When the devices were first developed, another difficulty resulted from bringing three clock lines to the electrodes: one of the clock lines had to cross over another, which meant that a more complicated multiple-layer technology was needed. However, this technology is now routinely accomplished.

The problem with a two-phase system is that when the system is symmetrical, charge can pass equally well in either direction. Therefore, if a two-phase system is used, an asymmetry must be built into it. One method of doing this is illustrated in Fig. 4.3.7. Here, by depositing the electrodes on top of layers of silicon dioxide

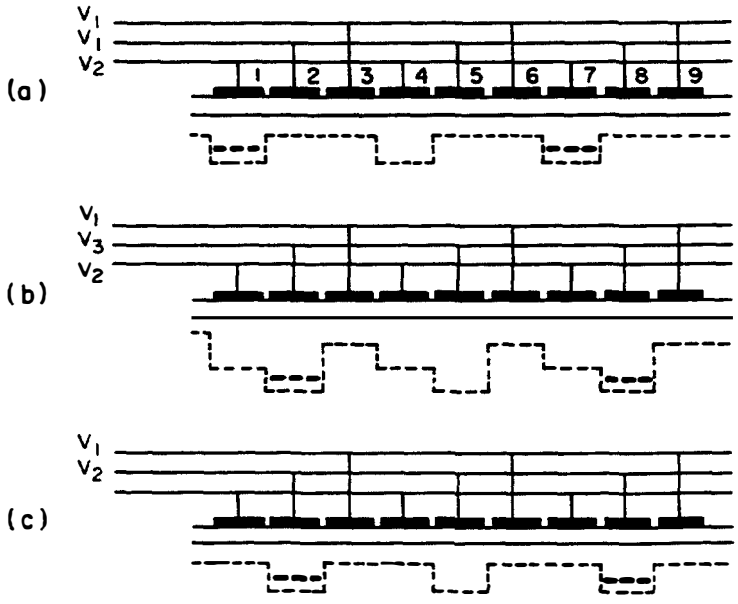


Figure 4.3.6 Three-phase CCD; illustration of its operation.

with appropriate windows, an asymmetry is built into the electrodes themselves, and a two-phase system moves carriers to the right. Other techniques involve ion implantation at one end of the electrode, so as to produce a built-in surface charge and to provide a bias that is built in at one end of the electrode. All in all, the two-phase system still tends to be as complex to manufacture as three-phase systems.

The signal carriers basically move between electrodes by diffusion, somewhat aided by their own space-charge fields and the fields between the electrodes, as discussed in Appendix J. If there were a large gap between the electrodes, the transfer would be increased. Thus it is often convenient to overlap the electrodes, as shown in Fig. 4.3.7, and so introduce transverse field components. Such a system is the natural one for a two-phase clock.

The charge transfer rate is determined by the mobility of the carriers; the higher the mobility, the better. For this reason it is preferable to use *n*-type carriers. Sometimes we can remove electrons from the surface region by trapping them in a potential maximum at a buried *p*-type layer formed by ion implantation; this is called a *buried channel*. This makes it far easier to provide a large transverse component of field for high charge-transfer rates. A further development is to use a material such as gallium arsenide, where the mobility of the electrons is of an order of magnitude greater than for silicon. This means that a buried channel must be used, together with so-called Schottky gates, to eliminate problems with surface states in gallium arsenide. However, this method has resulted in CCDs capable of handling signal frequencies of several hundred megahertz [29].

In all cases, the transfer efficiency is never perfect, so the fat zero concept applies equally well to CCD and BBD operation. A short discussion of the mathematical theory of operation of these devices is given in Appendix I. The detailed theory tends to be more complicated than that of the BBD and is outside the scope of this book.

A tapped CCD delay line can be implemented by taking advantage of the fact that when charge passes under an electrode, it induces an equal and opposite charge in the electrode. Thus by reducing the current flowing into the clock lines, and integrating it, the charge on a possible electrode can be determined. A technique is also needed for implementing a weighted transversal filter in which the weights can be positive or negative.

The technique employed is known as the *split-electrode method.* In one phase, the electrodes are split into two sections of different length, as shown in Fig. 4.3.8, with a schematic of the circuit used shown in Fig. 4.3.9. Two clock lines are employed instead of the V_3 clock line. These are supplied through the capacitors C_1^+ and C_1^-. Transistors T_1 and T_2 are used to discharge the clock lines after the clock is turned off. We shall call these the V_3^- and V_3^+ clock lines. These clock lines are individually connected to each side of a differential current integrator, so that if the strips on each side are of equal length, the output from the

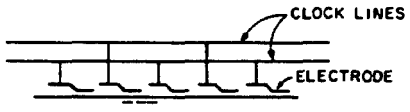


Figure 4.3.7 Two-phase CCD structure.

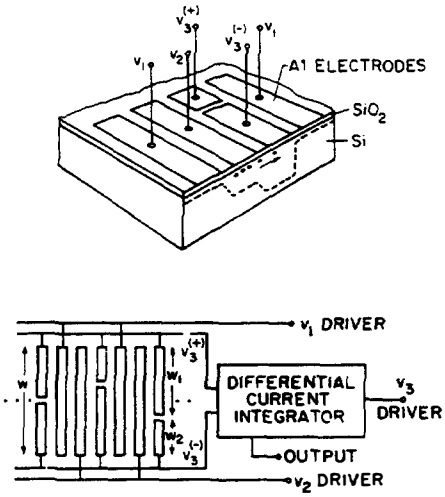


Figure 4.3.8 Split-electrode technique for implementing weighting coefficients of a transversal filter. (From Buss et al.[4].)

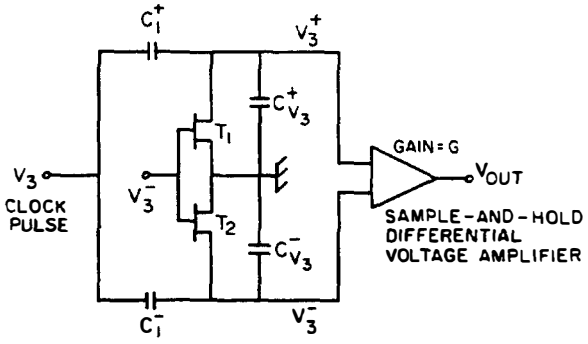


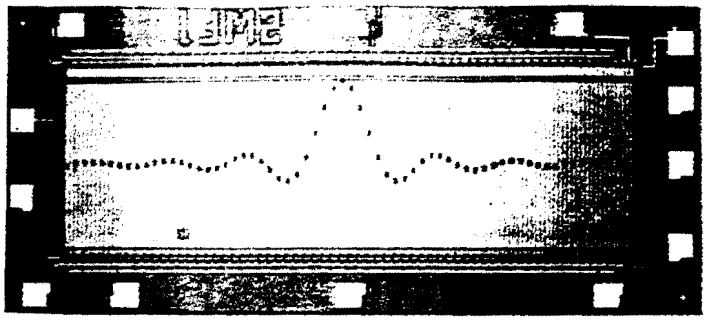
Figure 4.3.9 Differential current integrator (DCI) output amplifier. (From Buss et al. [4].)

differential current integrator will be zero when the charge passes underneath a particular strip. Now suppose that the total length of a strip is w, and the individual parts of a strip are of lengths w_1 and w_2, such that $w_1 + w_2 = w$. The total output from the integrator would then be proportional to $w_1 - w_2$, and both positive and negative weightings could be easily obtained.

By using a sample-and-hold circuit in the output (essentially, a capacitor connected through the FET switches T_1 and T_2 to the output circuit only when a signal is present), we can differentiate against higher-frequency clock noise, and the device will act faithfully as a transversal filter in exactly the manner already described for the SAW delay line. The charge induced on the individual electrodes is proportional to the weighting and is obtained at a time corresponding to the time delay through the delay line. The system may thus be used to obtain the convolution of the input signal with the weighting of the CCD. The design problem is *exactly* like that of SAW devices, except that the device is now operated at baseband. It follows that the device lends itself well to the design of a wide range of digital and analog transversal filters.

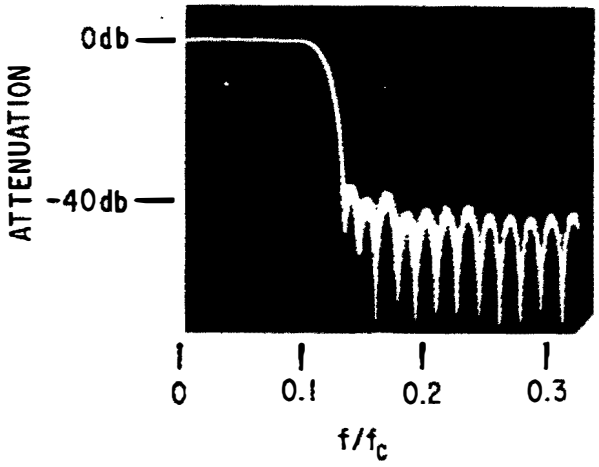
This method differs in one respect from the SAW delay line—it is not normally convenient to vary the distance between electrodes. Because of this, these devices are usually operated with a fixed spacing between electrodes, and the signals are sampled at these fixed points. We shall discuss the difference presented by this design problem later, when we discuss the chirp z transform implementation in both CCD and SAW delay lines.

It is apparent, however, that it is possible to construct low-frequency CCD and BBD equivalents of SAW delay lines for use as digital filters and bandpass filters, as described above. Figure 4.3.10 shows an example of a bandpass filter using the $(\sin x)/x$ response already described for SAW delay lines. Figure 4.3.10(a) shows the pattern cut into the electrodes, while Fig. 4.3.10(b) shows the time



(a)

(b)



(c)

Figure 4.3.10 (a) Photomicrograph of a 63-stage CCD low-pass filter; (b) measured impulse response of the filter; (c) measured frequency response of the filter. (After Baertsch et al., as noted in Buss et al. [4, 30, 31].)

response of the clock delay line. Note that the frequency response of the filter in Fig. 4.3.10(c) is of excellent quality.

4.3.4 Effect of Imperfect Charge-Transfer Efficiency

When charge is transferred through a CCD or BBD delay line, a fraction ε of the charge is left behind after each clock pulse. Consequently, we might expect that after transfer through N stages, the output amplitude will be diminished from that

of an ideal delay line. In addition, because charge from earlier transfers is still present in the registers, weak interfering signals will be present and will cause phase distortion of the input signal. The fractional losses of commercial CCDs and BBDs in their operating frequency ranges are usually of the order of $\varepsilon \sim 10^{-4}$, and at worst, $\varepsilon \sim 10^{-3}$. Although such fractional losses appear to be small, with a large number of registers with transfers of as many as $N = 1000$, even a small loss per stage can be vitally important. Examples of typical transfer efficiencies of Reticon commercial devices as a function of frequency are shown in Fig. 4.3.11.

In this section we derive an expression for the response of a charge-transfer device, assuming that ε is invariant with the frequency of the signal being transferred, although it will normally be a function of the transfer time (i.e., the clock period $T_c = 1/f_c$, where f_c is the clock frequency). We consider the properties of a device N registers long.

We consider the amplitude of a signal of the form exp $(j\omega t)$. We call the time at which charge is injected into the input of the CTD, t'. After passing through N perfect registers, the signal at time $t = t' + NT_c$ is still exactly exp $(j\omega t')$, or exp $[j\omega(t - NT_c)]$. Thus the original signal input is reproduced at the output, and is of the form

$$H_N(\omega) = e^{-j\omega NTc} \tag{4.3.19}$$

We note that with sample time T_c apart, there is aliasing; that is, the response at frequency ω is the same as it is at frequency $\omega + 2m\pi/T_c$, where m is an integer. It is convenient to use a notation suitable for such sampled systems; as we shall see, this will simplify the algebra considerably. Adopting the z transform notation, we write [4, 32, 33]

$$z = e^{j\omega T_c} \tag{4.3.20}$$

Here we have essentially expressed the frequency ω in terms of z. Thus for a perfect system, in the z-transform notation,

$$H_N(z) = z^{-N} \tag{4.3.21}$$

In this notation, the ideal response after a delay T_c (one register) is $1/z$, and through N registers, z^{-N}.

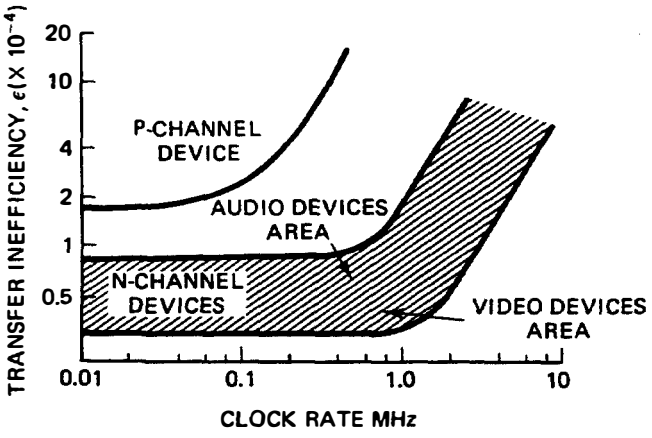


Figure 4.3.11 Charge-transfer inefficiency versus clock frequency. (After Reticon Corporation [24].)

Consider now the transfer function of the nonideal charge-transfer device. Suppose that the voltage at the kth register at a sampling time nT_c is $v_k(nT_c)$. This voltage consists of a term due to the incomplete transfer of charge from the previous register of the form $(1 - \varepsilon)v_{k-1}[(n - 1)T_c]$, and the charge remaining in the register from the previous transfer is $\varepsilon v_k[(n - 1)T_c]$. Thus we may now write

$$v_k(nT_c) = (1 - \varepsilon)v_{k-1}[(n - 1)T_c] + \varepsilon v_k[(n - 1)T_c] \tag{4.3.22}$$

Assuming that all quantities vary with time as $\exp(j\omega t)$, we could write that at a time $t = nT_c$, a signal component of frequency ω at the kth register would be of the form $V_k(\omega) \exp(j\omega nT_c)$. It follows that Eq. (4.3.22) can be written in the form

$$V_k(\omega)e^{j\omega nT_c} = (1 - \varepsilon)V_{k-1}(\omega)e^{j\omega(n-1)T_c} + \varepsilon V_k(\omega)e^{j\omega(n-1)T_c} \tag{4.3.23}$$

or

$$z^n V_k(z) = (1 - \varepsilon)V_{k-1}(z)z^{n-1} + \varepsilon V_k(z)z^{n-1} \tag{4.3.24}$$

Equation (4.3.24) reduces to the expression

$$zV_k(z) = (1 - \varepsilon)V_{k-1}(z) + \varepsilon V_k(z) \tag{4.3.25}$$

It follows from Eq. (4.3.25) that

$$V_k(z) = \left(\frac{1 - \varepsilon}{1 - \varepsilon/z}\right)\frac{1}{z} V_{k-1}(z) \tag{4.3.26}$$

Thus the response of the line after the signal has passed through N registers is

$$H_N(z) = \left(\frac{1 - \varepsilon}{1 - \varepsilon/z}\right)^N z^{-N} \tag{4.3.27}$$

We can simplify Eq. (4.3.27) by using the relation $\ln (1 + \alpha)^N \approx N\alpha$ when $\alpha \ll 1$, so that $(1 + \alpha)^N \approx \exp(N\alpha)$. Assuming that $\varepsilon \ll 1$, Eq. (4.3.27) can be written in the form

$$H_N(z) = e^{-N\varepsilon(1-1/z)}z^{-N} \tag{4.3.28}$$

or

$$H_N(\omega) = e^{-N\varepsilon(1-e^{-j\omega T_c})}e^{-j\omega NT_c} \tag{4.3.29}$$

In turn, we can write $H_N(\omega)$ in terms of its amplitude and phase variations:

$$H_N(\omega) = A_N(\omega)e^{-j\phi_N(\omega)} \tag{4.3.30}$$

where

$$A_N(\omega) = e^{-N\varepsilon(1-\cos\omega T_c)} \tag{4.3.31}$$

and

$$\phi_N(\omega) = N(\varepsilon \sin \omega T_c + \omega T_c) \tag{4.3.32}$$

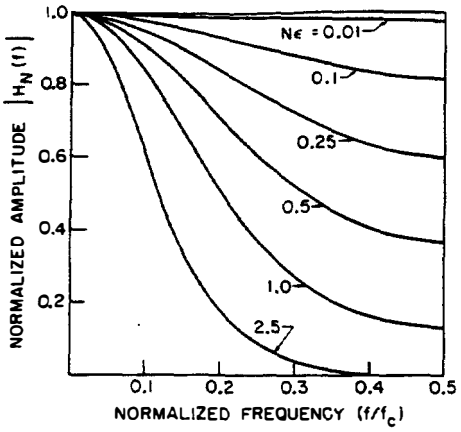


Figure 4.3.12 Magnitude of the frequency response of a CCD delay line. Eq. (4.3.29) is plotted for various values of $N\varepsilon$. This curve assumes that impulse sampling at the output must, in general, be multiplied by the frequency characteristic of the output amplifier stage. (From Buss et al. [4].)

It is convenient to write $\omega T_c = 2\pi f/f_c$ and then to plot Eq. (4.3.30) in the form shown in Fig. 4.3.12. In order to satisfy the sampling theory, the clock frequency must be at least $f_c = 2f$. But these results imply that the amplitude loss is maximum at this frequency. Thus it is usually wiser to operate a charge-transfer device with a clock frequency considerably larger than the maximum signal frequency.

It is also of interest to determine the time response or impulse response of the system after N transfers. We do this by calculating the time response from the inverse Fourier transform of the frequency response, and write

$$h_N(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{j\omega t}H_N(\omega)\,d\omega \tag{4.3.33}$$

Although it is difficult to evaluate Eq. (4.3.33) exactly, it is convenient to expand it in a series form as an expansion in powers of $1/z$. Thus, using Eq. (4.3.27), we write

$$\begin{aligned} H_N(z) &= e^{-N\varepsilon(1-1/z)z-N} \\ &= z^{-N}e^{-N\varepsilon}\left[1 + \frac{N\varepsilon}{z} + \frac{1}{2}\left(\frac{N\varepsilon}{z}\right)^2 + \cdots + \frac{1}{n!}\left(\frac{N\varepsilon}{z}\right)^n + \cdots\right] \end{aligned} \tag{4.3.34}$$

By writing $z^n = \exp(j\omega nT_c)$ and substituting Eq. (4.3.34) in Eq. (4.3.33), it follows that the output can be written in the form

$$h_N(t) = e^{-N\varepsilon}\{\delta(t - NT_c) + N\varepsilon\delta[t - (N+1)\,T_c]\cdots\} \tag{4.3.35}$$

where $\delta(t - nT_c)$ is a δ function, the inverse transform of $\exp(j\omega nT_c)$. Thus we might expect a series of output pulses reduced in amplitude from the expected pulse at the time $t = NT_c$. More exactly, we could have used the formula of Eq. (4.3.27) and expanded it to obtain an nth term in the form $(1 - \varepsilon)^N\varepsilon^n z^{-(N+n)}$ $(N + n)!/n!N!$. The corresponding expansion for the time response is similar to that of Eq. (4.3.35), and is plotted in normalized form in Fig. 4.3.13 for an input function consisting of a short pulse. As we might expect, when $N\varepsilon$ is small, the initial response will be very close to that of a single pulse, but followed by a weak

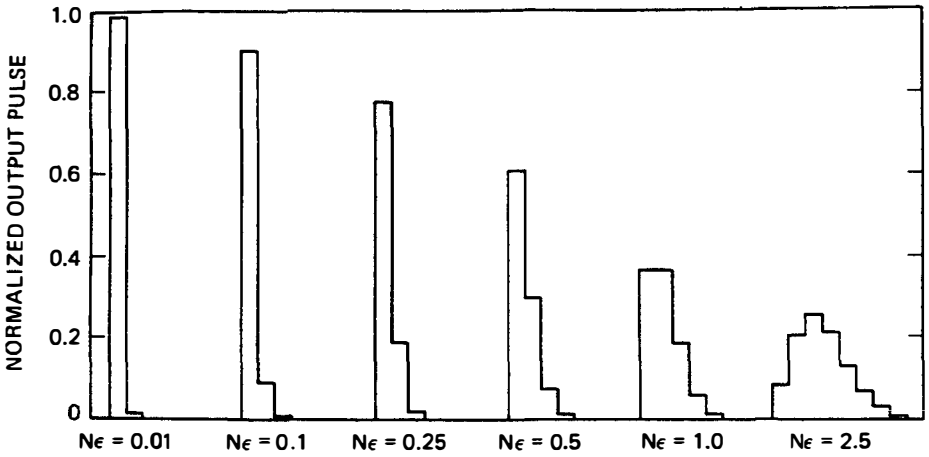


Figure 4.3.13 Impulse response of an N-stage CTD delay line having loss per stage ε. The output given by using Eq. (4.3.27) to obtain a more exact form of Eq. (4.3.35) is plotted for various values of $N\varepsilon$. (After Buss et al. [4].)

second pulse reduced in amplitude by $N\varepsilon$. As $N\varepsilon$ is increased so that $N\varepsilon \rightarrow 1$, the initial pulse may be weaker than the later pulses coming out of the device, and the signal is radically distorted (see Prob. 2).

4.3.5 Single Transfer Devices (Switched Capacitors)

One way to eliminate the difficulty caused by loss of charge in CTD devices when the charge passes through a very large number of registers is to employ M delay lines, each with N registers, and to inject the signal in turn into each delay line by multiplexing. Thus the system can be made to behave like a CTD with NM registers, with the corresponding loss through only the N registers.

An extreme example of this uses only one register, injects a signal into this register, and stores it, taking it out when needed. Such devices are known as *single transfer devices* (STDs). STDs essentially integrate sets of multiplexed sample-and-holds. Each successive sample is stored in a separate, discrete memory cell, where it stays until read out, as illustrated in the simplified diagram of Fig. 4.3.14. It is now only necessary to use digital shift registers to switch from one sample-and-hold to the next, as illustrated in Fig. 4.3.15. In the device illustrated, which is manufactured by Reticon, the input signal is read into a series of sample-and-hold capacitors, by means of FET switches controlled from the read-in shift register. The signal can then be read out in turn from the capacitors by means of the read-out shift register. This technique allows us to construct a simple delay line in which the clock rate of the output signal may be different from the input signal. Thus the device lends itself to time-base modification or correction of analog signals.

Devices of this kind are useful in the audio–video range and provide simple and inexpensive analog memory delays. Because of the single transfer, they do not have the same problem of transfer inefficiency as the CTDs, but they do have performance limitations associated with clocking noise and *fixed pattern noise*; that

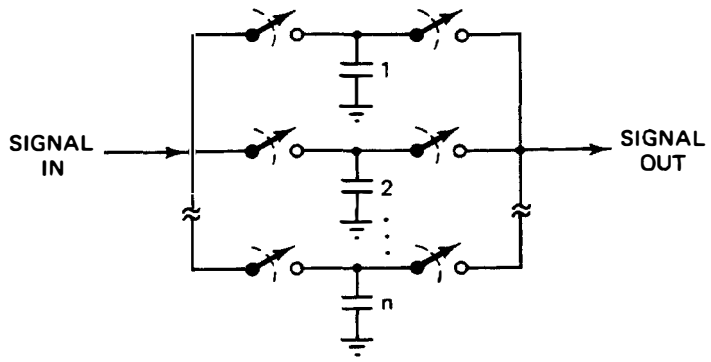


Figure 4.3.14 Sample-and-hold model of an STD.

is, there is a fixed pattern in the variation of the output signal when the memory cells or switches are not uniform, which implies some limitation on the dynamic range of the device.

The problem of clock noise can be radically decreased by using balanced circuits for readout. As an example, consider the self-scanning photodiode array system, developed by Reticon and illustrated in Fig. 4.3.16. In this device, the sample-and-hold diode is photosensitive, so that when it is illumunated with light, charge is generated within it; thus the potential across the diode is altered by the illumination during the given integration time. The charge read into a diode can then be read out by switching the output line to the diode.

In a manner similar to the CCDs and BBDs already discussed, both odd and even shift registers can be employed to lower the clock frequency. A further addition is to use a set of dummy photodiodes which are not exposed to light. Each photosensitive photodiode has a corresponding dummy diode, and the outputs from the two diodes are read out at the same time into opposite sides of a balanced circuit. The output obtained from the balanced circuit is therefore the difference in the charge on the two diodes and is proportional to the illumination. However, as the same clock pulses are used for both diodes, the clock output is eliminated, or at least differentiated against.

Note that a similar balanced system can be used with tapped BBD or CCD delay lines. Now, however, the charge is read from the diode into a tap on a CCD or BBD delay line and the output is read out serially through the delay line, rather

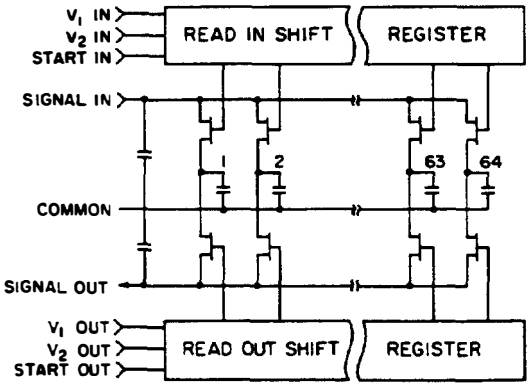


Figure 4.3.15 Simple equivalent circuit of the Reticon SAM-64. (From Reticon Corporation [24].)

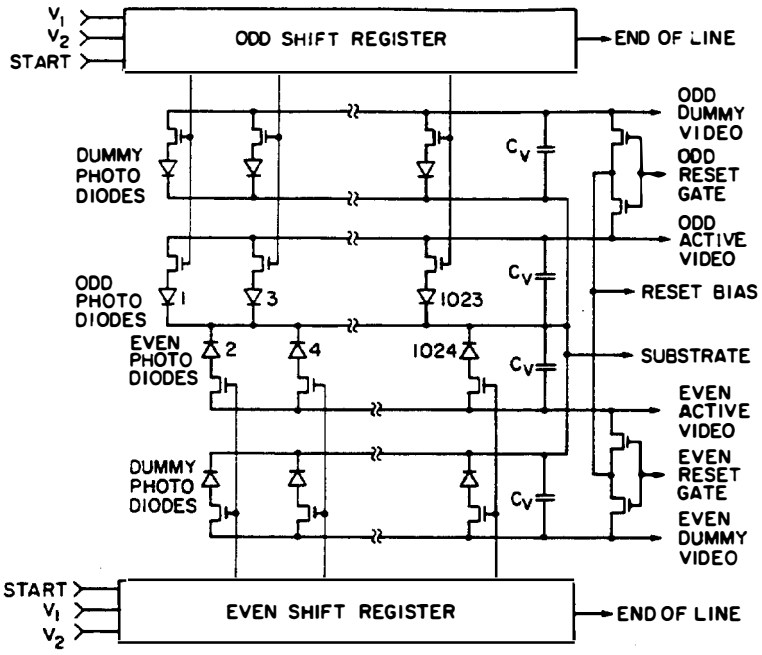


Figure 4.3.16 Equivalent circuit of Reticon self-scanning photodiode system. (From Reticon Corporation [28].)

than in the manner shown for the STD device. Such a system lends itself better to a two-dimensional configuration, although combinations of both systems are sometimes used in two-dimensional imaging devices.

PROBLEM SET 4.3

1. A unit step function $u(t) = 0, t < 0; u(t) = 1, t \geq 0$ is injected into an imperfect charge transfer delay line. Assuming ε is small use Eq.(4.3.34) to find the form of the output signal from the Nth register. Carry out the problems in two parts
   (a) With a short pulse input find the output, as in the text. Plot the output for a short pulse and $N\varepsilon = 2.5$ for the three cycles before and after the largest pulse. Note that the largest pulse is not the first output pulse.
   (b) Plot the amplitudes for the step function input for $t = (N - 1)T_c$ to $t = (N + 8)T_c$ with $N\varepsilon = 0.01, 2.5$. You can do this problem by convolution.
2. Derive the result of Fig. 4.3.13 for $N\varepsilon = 2.5$ by using an expansion of Eq. (4.3.27) without the approximation of Eq. (4.3.28) [see the discussion after Eq. (4.3.35)].
3. Compare the use of one CTD, N registers long, operating at a clock frequency f_c and signal frequency f, with the use of two multiplexed CTDs, $N/2$ registers long, operating at a clock frequency of $f_c/2$. The signal of frequency f is switched alternately from one CTD to the other on every half-cycle of the signal, and the output signals are added. Determine the amplitude and phase response of the second system, and compare its performance with that of the first system when $N\varepsilon = 0.1$. Plot your results as in Fig. 4.3.12.
4. Consider the use of a sample-and-hold circuit to eliminate clock signals and aliasing in the output waveform. Suppose that a signal $f(t)$ is sampled at times $t - nT_c$ and injected in a sample-and-hold circuit. Between the times $t = nT_c$ and $t = (n + 1)T_c$, the sample-

and-hold circuit maintains the output at a constant level $f(nT_c)$. So we can write the sampled signal in the form

$$f_s(t) = \sum_n f(nT_c)\Pi\left(\frac{t - nT_c - T_c/2}{T_c}\right)$$

where

$$\Pi(x) = \begin{cases} 1 & 0 < |x| < 0.5 \\ 0 & |x| > 0.5 \end{cases}$$

When the input signal is $f(t) = \exp(j\omega t)$, find the amplitudes of the frequency components $\omega + 2\pi m/T_c$ as a function of ωT_c. How much are the $m = 1$ and $m = 2$ components reduced from that of the $m = 0$ component when $fT_c = 0.5$, 0.25, and 0.1, where $f = \omega/2\pi$?

One way to carry out the analysis is to regard the output as of the form

$$f(t) = e^{j\omega t} \sum_m A_m e^{2j\pi m t/T_c}$$

and to find the amplitude of the mth harmonic component. Another way to solve the problem is to carry out the Fourier transform directly.

5. The outputs from an array of sonar receivers used to detect submarines are inserted into a tapped BBD delay line, as shown. Suppose that the signal frequency is f and the clock frequency, f_c. For simplicity, regard the taps as connected to neighboring registers. Assume that the parallel rays coming from a distant source travel at an angle θ to the normal to the plane of the array. Let there be N elements in the array spaced a distance L apart, and let the wave velocity in the medium be c.
   (a) Find how the output varies as a function of the angle θ and frequency f_c when there are N transducers.
   (b) Show how to choose f_c ($f_c \neq 0$) to be such that the signal from this electronically steerable array is maximum at $\theta = 0$. What are the possible values of f_c for $L = 1$ m, $c = 1.5$ km/s, and $f = 300$ Hz, for $\theta = 0$ and $\theta = 10°$?

4.4 MATCHED FILTERS

4.4.1 Introduction

SAW devices and charge-transfer devices can be designed as filters for a particular passband, as already described. We have seen that they can also be designed (by changing the spacing between the taps in addition to the polarity of the taps) as *transversal filters*, which can respond to or generate a particular digital code. The merit of such devices is that they are particularly simple in form, and so can be made by the standard photolithographic integrated-circuit technology.

Another broad application of SAW devices is to recognize analog codes rather than digital codes. A good example of such an analog code is the so-called linear FM chirp, a signal of constant amplitude with a frequency that varies linearly with time. The reason we are interested in the use of FM chirps and other analog codes is best illustrated by the example of radar, for which this kind of coding was initially developed. As we shall see later, there are also other broad applications of FM

chirps, the prime example being the *chirp z transform*, a method of obtaining a fast Fourier transform.

The initial form of radar consists of a transmitter emitting a short pulse, which is then reflected from an object and detected at the receiver. The range of the object is determined by measuring the transit time of the pulse to the object and back. For good range resolution, we must obviously use as short a pulse as possible; for the best signal-to-noise ratio, we must use as high a peak power as possible. Thus the necessity for a transmitter with a very large peak power can limit the system.

The development of radar techniques was stimulated by an important monograph by Woodward [34], which proved that it is not necessary to work with high peak powers. Instead, we can use complex waveforms with much less peak power and longer time duration. A linear filter that maximizes the return signal-to-noise power ratio is called a *matched filter*. For noise with a uniform frequency spectrum, the matched filter has a time response that is the time reverse of the input signal. The output signal is a narrow pulse that is the autocorrelation of the input signal. The maximum peak power output from such a matched filter, and hence the sensitivity and signal-to-noise ratio of the radar, is determined by the total energy in the input signal [1, 34–36].

The range accuracy of a radar is inversely proportional to the bandwidth, while the Doppler shift velocity accuracy is inversely proportional to the time length of the signal. In both cases, the accuracy improves with signal-to-noise ratio. Therefore, another important advantage of a matched filter is that it yields the optimum range and velocity accuracy; this accuracy depends on the total energy in the input signal.

The choice of input signal waveform and an appropriately matched filter is an extremely important parameter to be chosen by the radar designer. On one hand, a large bandwidth provides good range resolution; on the other, a large signal length provides good frequency resolution. As we shall see in Sec. 4.4.5, a Gaussian pulse is therefore the worst possible choice, for it makes the product of these two resolutions minimum.

The key to obtaining optimum performance is to use a signal with as large a time–bandwidth product as possible. An important advantage of a system using a matched filter is that it has a built-in preference for its own signals rather than for other, competing signals. Thus the use of a matched filter gives optimum performance in the presence of coherent interference as well as random noise.

Other filter functions are also possible. For instance, if the signal-to-noise ratio is high, the criterion may be to obtain as narrow a pulse as possible, and hence the best range definition. The ideal filter for this purpose is known as an *inverse filter* or *equalizer*. Its basic purpose is to compensate for any distortions in the signal from the ideal delta function (δ function) response [36].

As the ideal inverse filter requires infinite bandwidth, a practical system must be designed on the basis of different criteria. One possible criterion is to design a filter that gives an output signal with the best mean-square fit to a desired signal (this may be a δ-function pulse, i.e., an infinitesimally narrow pulse) over a given time range or frequency range. Such a filter is known as a *Wiener filter* [36].

In this section we derive the theoretical properties of the matched filter. We then give examples of how such properties are realized in practice with analog tranversal filters. In Sec. 4.8 we follow the same procedures for inverse and Wiener filters.

4.4.2 Matched Filter Theory

The maximum peak signal-to-noise ratio for a given input signal in the presence of random noise is obtained by the use of a matched filter, that is, one for which the time response of the filter is the time-reversed version of the input signal. Here we demonstrate this general property mathematically.

We consider a signal $X(\omega)$ in the frequency domain passed into a filter with a response $H(\omega)$. The output from the device will be

$$Y(\omega) = X(\omega)H(\omega) \tag{4.4.1}$$

We may define the following Fourier transform pairs for an input signal $x(t)$ using a notation similar to Bracewell's [32]:

$$\begin{aligned} X(\omega) &= \mathcal{F}[x(t)] = \int_{-\infty}^{\infty} x(t)e^{-j\omega t}\,dt \\ &= \int_{-\infty}^{\infty} x(t)e^{-2j\pi ft}\,dt \\ x(t) &= \mathcal{F}^{-1}[X(\omega)] = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)e^{j\omega t}\,d\omega \\ &= \int_{-\infty}^{\infty} X(f)e^{2j\pi ft}\,df \end{aligned} \tag{4.4.2}$$

with similar relations for the output signal $y(t)$. We shall define the transforms in terms of the radian frequency $\omega = 2\pi f$, and write, for example,

$$\begin{aligned} Y(\omega) &= \mathcal{F}[y(t)] = \int_{-\infty}^{\infty} y(t)e^{-j\omega t}\,dt \\ y(t) &= \mathcal{F}^{-1}[Y(\omega)] = \frac{1}{2\pi}\int_{-\infty}^{\infty} Y(\omega)e^{j\omega t}\,d\omega \end{aligned} \tag{4.4.3}$$

For a causal system, with a signal $x(t)$ and filter response $h(t)$ starting at a time $t = 0$, the response of the system in the time domain, equivalent to Eq. (4.4.1), is given by the convolution integral

$$\begin{aligned} y(t) = x(t) * h(t) &= \int_0^t x(\tau)h(t-\tau)\,dt \\ &= \int_0^t x(t-\tau)h(\tau)\,d\tau \end{aligned} \tag{4.4.4}$$

where $h(t)$ is the response of the filter to an impulse $\delta(t)$. The upper limit of the

integrals is t so that $x(t - \tau) = 0$ and $h(t - \tau)$ if $t - \tau < 0$. We define the signal-to-noise ratio of the output from the filter as

$$\frac{S}{N} = \frac{\text{peak instantaneous output signal power}}{\text{output noise power}}$$

Our objective will be to determine the optimum form of $h(t)$ or $H(\omega)$, which yields the maximum value of S/N.

Let us suppose that the signal output is maximum at the time $t = T_d$, where T_d is the delay time through the filter. At this time

$$y(T_d) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)H(\omega)e^{j\omega T_d}\,d\omega \tag{4.4.5}$$

It follows from Eq. (4.4.4) that we could equally well write

$$\begin{aligned} y(T_d) &= \int_0^{T_d} x(\tau)h(T_d - \tau)\,d\tau \\ &= \int_0^{T_d} x(\tau)h(T_d - \tau)\,d\tau \end{aligned} \tag{4.4.6}$$

The real instantaneous output power into a 1-Ω load is

$$S = y^2(t) \tag{4.4.7}$$

Suppose that the amplitude of the noise input to the filter between the frequencies f and $f + df$ is $X_N(f)\,df$. Then the noise output from the filter at frequency f is

$$Y_N(f) = X_N(f)H(f) \tag{4.4.8}$$

The total normalized mean-square noise power σ^2 at the filter output is the expectation value of the total noise output power, or

$$\sigma^2 = \left\langle \int_{-\infty}^{\infty} |Y_N(f)|^2\,df \right\rangle = \left\langle \int_{-\infty}^{\infty} |X_N(f)H(f)|^2\,df \right\rangle \tag{4.4.9}$$

where $X_N(f)$ is defined for both positive and negative frequencies. Since the noise is assumed to be uncorrelated, or random, we can write

$$\begin{aligned} \sigma^2 &= \frac{1}{2}\int_{-\infty}^{\infty} N_0(f)|H(f)|^2\,df = \int_0^{\infty} N_0(f)|H(f)|^2\,df \\ \sigma^2 &= \frac{1}{4\pi}\int_{-\infty}^{\infty} N_0(\omega)|H(\omega)|^2\,d\omega \end{aligned} \tag{4.4.10}$$

where $N_0(f) = N_0(-f)$ is the noise power spectral density in W/Hz, defined for positive frequencies. It follows that

$$\tfrac{1}{2}N_0(f) = \langle |X_N(f)|^2 \rangle \tag{4.4.11}$$

We observe that $N_0(f)$ has the dimensions of energy, the product of power and time. As an example for Gaussian noise from a blackbody at temperature T, $N_0(f) = kT$, where k is Boltzmann's constant. It is usually assumed that filters

have equal responses at positive and negative frequencies. It also follows from Parseval's theorem that when $N_0(f) = N_0$ is constant,

$$\sigma^2 = \frac{N_0}{2} \int_0^\infty h^2(t)\, dt \tag{4.4.12}$$

Our aim is to maximize $y^2(T_d)/\sigma^2$. From now on we shall assume that $N_0(f) = N_0 =$ constant (see Prob. 1 for a more general relation), and write

$$\frac{S}{N} = \frac{\left| \int_0^{T_d} x(T_d - \tau) h(\tau)\, d\tau \right|^2}{\frac{1}{2} N_0 \int_0^\infty h^2(t)\, dt} \tag{4.4.13}$$

We have taken the upper limit of the integral in the numerator as T_d in order to make $x(T_d - \tau) = 0$ if $T_d - \tau < 0$. It follows from Schwarz's inequality† that if $x(t)$ and $h(t)$ are real, then

$$\frac{S}{N} \leqslant \frac{\int_0^{T_d} x^2(T_d - t)\, dt \int_0^{T_d} h^2(t)\, dt}{\frac{1}{2} N_0 \int_0^\infty h^2(t)\, dt} \tag{4.4.14}$$

The maximum value of S/N is given when the right-hand side of Eq. (4.4.14) is equal to the right-hand side of Eq. (4.4.13). Therefore, the maximum value of S/N is obtained at a time T_d if

$$\begin{aligned} h(t) &= \alpha x(T_d - t) \qquad t < T_d \\ &= 0 \qquad\qquad\quad t > T_d \end{aligned} \tag{4.4.15}$$

where α is a constant. Eqs. (4.4.13) and (4.4.14) become identical and the upper limit of the integral in the denominator is T_d.

The filter given by Eq. (4.4.15) is known as a *matched filter*. Its response is a time-delayed version of the time-reversed signal, and the output is the autocorrelation function of the signal to which it is matched [Eq. (4.4.19)]. Equations (4.4.13) and (4.4.14) show that the matched filter gives a maximum signal-to-noise ratio of

$$\left(\frac{S}{N}\right)_{\max} = \frac{\int_0^{T_d} x^2(T_d - t)\, dt)}{N_0/2} = \frac{\int_0^{T_d} x^2(t)\, dt}{N_0/2} \tag{4.4.16}$$

†One form of Schwarz's inequality is

$$\left| \int f(x) g(x)\, dx \right|^2 \leqslant \int |f(x)|^2\, dx \int |g(x)|^2\, dx$$

The left- and right-hand sides of this equation are equal only if $|g(x)| = \alpha |f(x)|$, where α is a constant. [If $f(x)$ and $g(x)$ are real, the modulus symbols may be omitted.]

We conclude that as the numerator in Eq. (4.4.16) is the total energy in the signal, the maximum signal-to-noise ratio of a signal, after passing through a matched filter, is determined only by its *total energy* E, not by the detailed structure of the signal. Thus we can write

$$\left(\frac{S}{N}\right)_{\max} = \frac{2E}{N_0} \tag{4.4.17}$$

where for a signal of length T, it follows from Rayleigh's theorem that the energy in the waveform is

$$E = \int_0^T x^2(t)\, dt = \int_{-\infty}^{\infty} |X^2(f)|\, df \tag{4.4.18}$$

In practice, as the numerator of Eq. (4.4.16) must have the maximum possible value, the time T_d should be chosen to equal the length of the pulse T. This also implies that the delay time through the filter must be the length of the pulse for optimum signal-to-noise ratio. An infinitely long signal would ideally need an infinitely long delay line as a matched filter. In all cases, if the length of the signal is of a finite length T, the delay line must have a delay time $T_d \geq T$, in order to obtain the convolution of the signal within the time response of the delay line. When the tranversal filter is longer than necessary for a matched filter, it picks up more noise than the minimum possible value, and its response is decreased. We have already seen examples of devices using such principles in Sec. 4.2, where we discussed FM chirp filters and time-delay filters.

The response of a matched filter, with a delay time T to a pulse of length T, is the autocorrelation function of $x(t)$. We can write that the output from a matched filter is

$$y(t) = \int_{t_0}^{t} x(\tau)x(T - t + \tau)\, d\tau \tag{4.4.19}$$

The upper limit of the integral is determined by the condition that $h(t - \tau) = 0$ for $\tau > t$. The lower limit of the integral is determined by one of two conditions: When $t < T$, then $x(\tau) = 0$ if $\tau < 0$; thus $t_0 = 0$ for $t < T$. Alternatively, when $T < t < 2T$, then $x(T - t + \tau) = 0$ if $\tau < t - T$; thus $t_0 = t - T$ for $T < t < 2T$. Finally, if $t > 2T$, then $x(T - t + \tau) = 0$ for $\tau < t$; thus the output is zero.

Example: Matched Filter for a Square Pulse

We can put this concept on a more physical basis by considering a square pulse of length T, injected into a tapped delay line with a delay time $T_d = T$, as illustrated in Fig. 4.4.1. We assume that there is a large number of closely spaced taps which sample a small portion of the signal passing through the line. The output from these taps is then summed. The line is a matched filter because its response to a δ function pulse is a square pulse of length T. When the pulse is injected into the delay line, the output builds linearly with time, as the signal from an increasing number of taps is summed when the pulse reaches them. This increase in output signal continues until a time $t = T$. Suppose that z is the distance along the delay line, and the pulse moves into the delay line with a velocity V. Then the output from the line is from

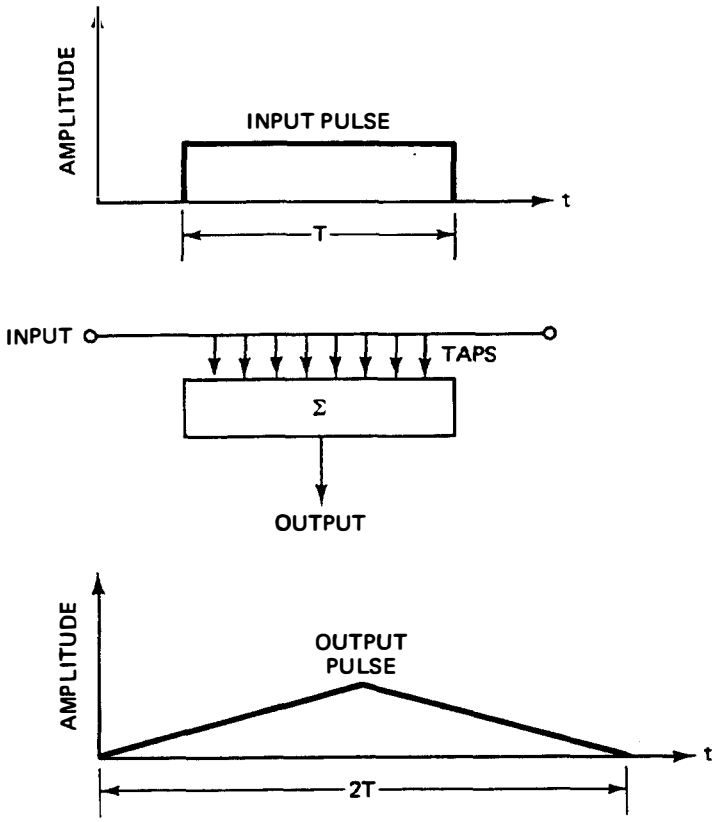


Figure 4.4.1 Tapped delay line of delay length T excited by a square pulse of length T.

a point $z = 0$ to $z = Vt$, so the limits of the integral in Eq. (4.4.19) are at 0 and t. Beyond this point, the taps at the front end of the line are no longer excited by the pulse of length T, and the output decreases linearly with time. Correspondingly, the lower limit of the integral corresponds in time to the point $z = V(t - T)$, where T is the pulse length, or in Eq. (4.4.19), $t_0 = t - T$. Finally, after a time $2T$, the pulse has passed through the line and there is no output.

We note that if the tapped delay line has a longer delay T_d than T, the output will increase to its maximum value, stay constant for a time $T_d - T$, and then decrease linearly with time to zero at a time $T_d + T$. As more taps are present, the noise output will be larger. Furthermore, the unused taps will tend to place a load across the output and decrease it.

If, on the other hand, the length of the pulse is $T > T_d$, the delay time through the line, the output signal cannot increase to as large a value as it would if $T_d = T$. Thus, once more, the line is not the ideal matched filter.

Matched filter in the frequency domain. We could just as well have carried out this derivation in the frequency domain, using Eqs. (4.4.5) and (4.4.9). It then follows that

$$\frac{S}{N} = \frac{1/(2\pi)\left|\int_{-\infty}^{\infty} X(\omega)H(\omega)e^{j\omega T_d}\,d\omega\right|^2}{(N_0/2)\int_{-\infty}^{\infty}|H(\omega)|^2\,d\omega} \tag{4.4.20}$$

By using Schwarz's inequality, we find that

$$\left|\int_{-\infty}^{\infty} X(\omega)H(\omega)e^{j\omega T_d}\,d\omega\right|^2 \leq \int_{-\infty}^{\infty} XX^*\,d\omega \int_{-\infty}^{\infty} HH^*\,d\omega \qquad (4.4.21)$$

where $X^*(\omega)$ is the complex conjugate of $X(\omega)$. Thus S/N is maximum when

$$H(\omega) = \alpha X^*(\omega)e^{-j\omega T_d} \qquad (4.4.22)$$

where α is a constant. The matched filter has a frequency response that is the complex conjugate of the signal, and an additional phase delay associated with its length ωT_d. Note that a matched filter has an output in the frequency domain $Y(\omega) = \alpha XX^* \exp(-j\omega T_d)$; the maximum signal-to-noise ratio obtained with such a filter is

$$\left(\frac{S}{N}\right)_{\max} = \frac{2\int_{-\infty}^{\infty} X(f)X^*(f)\,df}{N_0} = \frac{2E}{N_0} \qquad (4.4.23)$$

Thus the signal-to-noise ratio at the output of a matched filter is determined by the total energy in the input signal.

4.4.3 Pulse Compression

Now let us consider how the signal-to-noise ratio is improved by using a matched filter. We have shown, by a heuristic argument in Sec. 4.2, that if the input signal consists of a code of equal amplitude, positive or negative tone bursts, or an analog signal (such as an FM chirp of uniform peak amplitude), the output pulse will be compressed and will have a higher peak value than any part of the input signal. The improvement in signal-to-noise ratio is essentially determined by the time–bandwidth product of the input signal.

We may use the formulas for a matched filter to prove this concept directly. Suppose that the peak value of $x(t)$ is x_M. If the detector is band-limited, the input signal-to-noise ratio can be determined by taking $H(f) = 1$ over a bandwidth B. From Eq. (4.4.10), with $S(\max) = x_M^2$, we get the result

$$\left(\frac{S}{N}\right)_{\text{in}} = \frac{x_M^2}{N_0 B} \qquad (4.4.24)$$

It follows, however, from Eq. (4.4.16), that at the output of a matched correlation filter, the signal-to-noise ratio is

$$\left(\frac{S}{N}\right)_{\text{out}} = \frac{\int_0^T x^2(t)\,dt}{N_0/2} \qquad (4.4.25)$$

The *correlation gain* G, or the improvement in signal-to-noise ratio due to the

filter, is therefore

$$G = \frac{(S/N)_{\text{out}}}{(S/N)_{\text{in}}} = \frac{2B\int_0^T x^2(t)\,dt}{x_M^2} \tag{4.4.26}$$

Thus it follows that the improvement in signal-to-noise ratio, obtained by using a matched filter with a code of uniformly weighted tone bursts $\langle x^2(t)\rangle = x_M^2/2$, is

$$G = BT \tag{4.4.27}$$

Therefore, the ideal improvement in signal-to-noise ratio gained by using a matched filter is equal to the time–bandwidth product of the signal. We note that if the amplitude weighting of the signal over the time T is not uniform or the frequency response of the filter is not uniform over the bandwidth of the system, the improvement in maximum signal-to-noise ratio will be less than BT, the time–bandwidth product of the signal.

4.4.4 Matched Filters for Complex Signals

Before proceeding further, it is worthwhile considering how a correlation filter processes a complex signal of the form

$$u(t) = a(t)e^{j\phi(t)} \tag{4.4.28}$$

where $a(t)$ is the amplitude of the signal and ϕ is its phase. We normally take the real signal as

$$\begin{aligned} x(t) &= \text{Re}\,[u(t)] \\ &= a(t)\cos\,[\phi(t)] \end{aligned} \tag{4.4.29}$$

Following the derivation of Eq. (4.4.16), we consider the power in a complex signal $u(t)$, across a 1-Ω load, to be $S = u(t)u^*(t)/2$. Hence the signal-to-noise ratio of a signal $u(t)$, when it is inserted into a filter $h(t)$, is

$$\left(\frac{S}{N}\right)_{\max} = \frac{\frac{1}{2}\int_{-\infty}^{\infty} u(T_d - \tau)h(\tau)\int_{-\infty}^{\infty} u^*(T_d - \tau)h^*(\tau)\,d\tau}{(N_0/2)\int_{-\infty}^{\infty} h(t)h^*(t)\,dt} \tag{4.4.30}$$

From Schwarz's inequality, the signal-to-noise ratio is maximum (i.e., the system is a matched filter) when a complex matched filter is used with the property

$$h(t) = \alpha u^*(T_d - t) \tag{4.4.31}$$

or

$$h_R(t) = \alpha x(T_d - t) \tag{4.4.32}$$

as before. In Sec. 4.5 we see how a real filter processes a complex signal, and how a true complex filter can be constructed by using two filters.

When the signal is of finite length T, it follows that the maximum signal-to-noise ratio is

$$\left(\frac{S}{N}\right)_{\max} = \frac{\frac{1}{2}\int_0^T u(t)u^*(t)\,dt}{N_0/2} = \frac{\frac{1}{2}\int_0^T a^2(t)\,dt}{N_0/2} \tag{4.4.33}$$

If we substitute the form of $u(t)$ given in Eq. (4.4.28) in the numerator of Eq. (4.4.33), at a time $t = T$, a matched filter exactly compensates for phase errors in the input signal, and the real output signal is just

$$y(t' = 0) = \frac{\alpha}{2}\int_0^T a^2(t)\,dt \tag{4.4.34}$$

Since $a^2(t)$ is always positive, all components of the signal add together to give the maximum output. The energy in the waveform is

$$y(t - T) = \tfrac{1}{2}\int_0^T a^2(t)\,dt \tag{4.4.35}$$

The maximum signal-to-noise ratio is therefore

$$\left(\frac{S}{N}\right)_{\max} = \frac{2E}{N_0} \tag{4.4.36}$$

Thus the signal-to-noise ratio at the output of a matched filter is determined by the total energy in the signal.

4.4.5 Range and Velocity Accuracy of a Radar System

We are not necessarily interested in measuring the amplitude of a radar or communication signal. For example, a radar is used to determine the time delay of the return echo from a target as precisely as possible. Radar systems are also used to determine the Doppler frequency shift of a return echo; the Doppler frequency shift Δf is

$$\Delta f = \frac{2Vf_0}{c} \tag{4.4.37}$$

where V is the velocity of the target, c is the velocity of light, and f_0 is the carrier frequency of the transmitted signal (see Prob. 4.2.6).

The signal returning from the target is corrupted by noise; this decreases the accuracy of both the range and frequency resolutions. We shall show that the resolution depends directly on the signal-to-noise ratio, and hence on the energy in the signal, and that the optimum time resolution is obtained with a matched filter.

Simple physical treatment. This result can be illustrated by a simple physical argument. Consider a rectangular pulse of amplitude A and rise time t_r, as illustrated in Fig. 4.4.2. Suppose that there is a random noise signal $n(t)$ present, as illustrated by the dashed line.

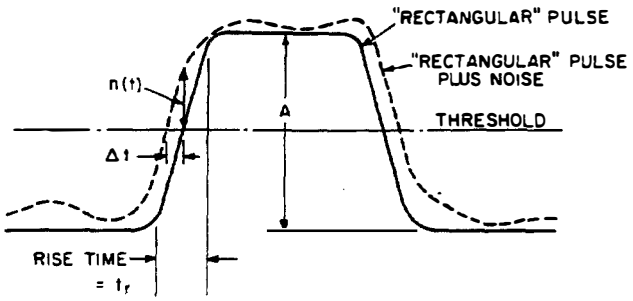


Figure 4.4.2 Measurement of time delay using the leading (or trailing) edge of the pulse. Solid curve, echo pulse uncorrputed by noise; dashed curve, effect of noise.

The slope of the leading edge of the pulse is approximately A/t_r. Thus at the point of measurement of the leading edge of the pulse, as shown in the figure, the error in rise time is

$$\Delta t = \frac{t_r n(t)}{A} \tag{4.4.38}$$

The mean-square error in Δt after many pulses is therefore

$$\langle (\Delta t)^2 \rangle = \frac{t_r^2 \langle n^2(t) \rangle}{A^2} \tag{4.4.39}$$

The maximum signal power into a 1-Ω load is $S = A^2$, and the average noise power is $N = \langle n^2(t) \rangle$. Hence we can write

$$\begin{aligned} [\langle (\Delta t)^2 \rangle]^{1/2} &= [\langle N^2 \rangle]^{1/2} \frac{t_r}{A} \\ &= \frac{t_r}{(S/N)^{1/2}} \end{aligned} \tag{4.4.40}$$

If the pulse is the output from a matched filter, Eq. (4.4.36) yields the result

$$[\langle (\Delta t)^2 \rangle]^{1/2} = \left(\frac{N_0}{2E}\right)^{1/2} t_r \tag{4.4.41}$$

If the rise time is limited by the bandwidth B of the system, the output is of the form $A(\sin \pi Bt)/(\pi Bt)$.† The maximum slope is approximately $1.3\ A/B$. The result is

$$[\langle (\Delta t)^2 \rangle]^{1/2} \approx \frac{1.3(N_0/2E)^{1/2}}{B} \tag{4.4.42}$$

Thus the error in measured range is least when the signal-to-noise ratio is as large as possible, and so is determined directly by the energy in the pulse. We therefore expect the optimum range resolution to be obtained with a matched filter.

More rigorous treatment. A more rigorous argument for the error in the measured time can be established. Let the total signal and noise-free signals

†We are concerned with radar signals, which are modulated carriers. Therefore, we have defined B as the total bandwidth of the radar signal and B/2 as the bandwidth of the video signal.

returning from the target be $x(t)$ and $x_0(t)$, respectively. Then

$$x(t) = x_0(t) + n(t) \tag{4.4.43}$$

where $n(t)$ is the noise signal. The output signal from a radar with a matched filter response $h(t) = x_0(-t)$, and maximum output at $t = 0$, is

$$y(t) = \int_{-\infty}^{\infty} x_0(\tau)x_0(\tau - t)\,d\tau + \int_{-\infty}^{\infty} n(\tau)x_0(\tau - t)\,d\tau \tag{4.4.44}$$

where the noise-free output signal, $y_0(t)$, is defined by the relation

$$y_0(t) = \int_{-\infty}^{\infty} x_0(\tau)(x_0(\tau - t)\,d\tau \tag{4.4.45}$$

Let the maximum value of $y_0(t)$ be $y_0(0)$. For short times Δt from the maximum, we use a Taylor expansion of $y_0(t)$ to obtain the result

$$y_0(\Delta t) = y_0(0) + \frac{(\Delta t)^2}{2} y_0''(0) + \cdots \tag{4.4.46}$$

where $y_0'(0) = 0$ when $y_0(t) = y_0(0)$, its maximum value. We will need $y_0^2(\Delta t)$ to second order in Δt. It follows from Eq. (4.4.46) that

$$y_0^2(\Delta t) = y_0^2(0) + (\Delta t)^2 y_0''(0)y_0(0) + \cdots \tag{4.4.47}$$

From Eq. (4.4.45),

$$y_0(0) = \int x_0^2(\tau)\,d\tau = E \tag{4.4.48}$$

and

$$y_0''(t) = \int x_0''(\tau - t)x_0(\tau)\,d\tau \tag{4.4.49}$$

Thus

$$y_0''(0) = \int_{-\infty}^{\infty} x_0''(\tau)x_0(\tau)\,d\tau = -4\pi^2 \int_{-\infty}^{\infty} f^2 X_0(f)X_0^*(f)\,df \tag{4.4.50}$$

It is convenient to define a quantity β, related to the bandwidth B as follows:

$$\beta^2 = \frac{4\pi^2 \int_{-\infty}^{\infty} f^2 X_0(f)X_0^*(f)\,df}{\int_{-\infty}^{\infty} X_0(f)X_0^*(f)\,df} \tag{4.4.51}$$

or

$$\beta^2 = \frac{4\pi^2}{E} \int_{-\infty}^{\infty} f^2 X_0(f)X_0^*(f)\,df \tag{4.4.52}$$

where E is the total energy in the pulse.

This definition of bandwidth was first introduced by Gabor and then used by Woodward in his treatment of range accuracy by inverse probability [34, 35]. As

an example, a Gaussian pulse of the form $\exp[-(1.18t/\tau)^2]$ has its 3-dB points τ apart, a 3-dB bandwidth of $B = 0.44/\tau$, and a value of $\beta = 1.18/\tau$. We note that β^2 is the normalized second moment of the spectrum $|X(f)|^2$ and is expressed in rad/s. From its definition, for a fair comparison with B, we should use β/π. For a Gaussian pulse, $\beta/\pi = 0.38/\tau$.

It follows from Eqs. (4.4.50) and (4.4.52) that

$$y_0''(0) = -\beta^2 E \tag{4.4.53}$$

and

$$y_0^2(\Delta t) - y_0^2(0) = -\beta^2 E^2 (\Delta t)^2 \tag{4.4.54}$$

We wish to determine, as accurately as possible, the time when $y_0(t)$ is maximum. We therefore estimate the mean square error $\langle \Delta t^2 \rangle$. We define this time from the condition that the mean square error $\langle \varepsilon^2 \rangle$, in estimating $y_0(0)$, is equal to the noise, or

$$\langle \varepsilon^2 \rangle = \langle y_0^2(0) - y_0^2(\Delta t) \rangle = \beta^2 E^2 \langle (\Delta t)^2 \rangle \tag{4.4.55}$$

We can write from Eqs. (4.4.43), (4.4.44), and (4.4.55) that as the noise is uncorrelated with the signal,

$$\langle \varepsilon^2 \rangle = \left\langle \left| \int_{-\infty}^{\infty} n(\tau) x_0(\tau) d\tau \right|^2 \right\rangle = \beta^2 E^2 \langle (\Delta t)^2 \rangle \tag{4.4.56}$$

We use the following relation, obtained from Parseval's theorem [32] [Eqs. (4.4.11) and (4.4.48)]:

$$\begin{aligned} \left\langle \left| \int_{-\infty}^{\infty} n(\tau) x_0(\tau)\, d\tau \right|^2 \right\rangle &= \left\langle \left| \int_{-\infty}^{\infty} N^2(f) X_0(f) \right|^2 df \right\rangle \\ &= \tfrac{1}{2} N_0 \int_{-\infty}^{\infty} |X_0^2(f)|\, df = \frac{N_0 E}{2} \end{aligned} \tag{4.4.57}$$

where N_0 is the noise power per Hz in the positive frequency range, and $N(f)df$, which is the same as $X_N(f)df$ in Eqs. (4.4.8) and (4.4.11), is the noise amplitude between frequencies f and $f + df$. It follows from Eqs. (4.4.56) and (4.4.57) that

$$($$ $\frac{N_0}{}$ (4.4.58)

or

$$[(($$ $\frac{1}{\beta(2E/N_0)^{1/2}}$ (4.4.59)

This result is therefore the more rigorously derived equivalent of Eq. (4.4.41). Note that the use of this more rigorous formula tends to lead to a slightly better time error than the approximate formula of Eq. (4.4.41).

We could have carried out a more complete derivation by taking the response of the radar receiver to be $h(t)$, and finding the optimum value of $h(t)$ for $\langle \Delta t^2 \rangle^{1/2}$

to be minimum. With the proper choice of normalization for the frequency response, it can be shown that the optimum filter is indeed a matched filter.

Accuracy of frequency measurement. The accuracy with which we can determine the Doppler shift of an echo depends on the signal duration T, rather than its bandwidth. By analogy to his normalized bandwidth function, Gabor has suggested a normalized signal duration defined by the relation [34, 35]

$$\alpha^2 = \frac{4\pi^2 \int_{-\infty}^{\infty} t^2 x_0^2(t)\, dt}{\int_{-\infty}^{\infty} x_0^2(t)\, dt} \tag{4.4.60}$$

We note that $(\alpha/2\pi)^2$ is the normalized second moment of $x_0^2(t)$, and that the effective signal duration time is α/π. It can be shown by methods very similar to those we have used already (see Prob. 5) that the minimum rms error in the measurement of frequency is

$$[\langle(\Delta f)^2\rangle]^{1/2} = \frac{1}{\alpha(2E/N_0)^{1/2}} \tag{4.4.61}$$

Thus the longer the time duration of the measuring signal, the better the accuracy in determining the Doppler shift in radar frequency; the larger the bandwidth, the better the range resolution.

Uncertainty relation. It is possible to obtain an uncertainty relation for radar systems [see the derivation following Eq. (4.4.65)], which states that the product of the effective bandwidth β and effective time duration α of a signal is such that

$$\beta\alpha > \pi \tag{4.4.62}$$

This implies, from Eqs. (4.4.58) and (4.4.61), that as

$$[\langle(\Delta t)^2\rangle]^{1/2}\,[\langle(\Delta f^2)\rangle]^{1/2} = \frac{1}{\alpha\beta(2E/N_0)} \tag{4.4.63}$$

then

$$[\langle(\Delta t)^2\rangle]^{1/2}\,[\langle(\Delta f^2)\rangle]^{1/2} < \frac{N_0}{2E\pi} \tag{4.4.64}$$

We conclude that the time delay and frequency of a radar signal can be measured simultaneously to any desired accuracy by designing the system to yield a sufficiently large ratio of energy in the pulse-to-noise power per unit bandwidth. To measure both of these quantities accurately, the signal must have a long duration and large bandwidth. This conclusion, of course, is almost the opposite of the uncertainty principle in quantum mechanics, since in the latter case the $\beta\alpha$ product is fixed by the energy of one quantum rather than an energy quantity chosen by the observer.

Proof of the inequality. We can derive the inequality we have used, fairly simply, by noting that

$$\beta^2 = \frac{-\int_{-\infty}^{\infty} x_0''(t)x_0(t)\,dt}{\int_{-\infty}^{\infty}} \tag{4.4.65}$$

Integrating by parts, and assuming $x_0(t) \to 0$ as $t \to \pm\infty$, we find that

$$\beta^2 = \frac{\int_{-\infty}^{\infty} [x_0'(t)]^2\,dt}{\int_{-\infty}^{\infty}} \tag{4.4.66}$$

and

$$\beta\alpha = \frac{\left\{2\pi \int_{-\infty}^{\infty} [x_0'(t)]^2\,dt \int_{-\infty}^{\infty} t^2 x_0^2(t)\,dt\right\}^{1/2}}{\int_{-\infty}^{\infty} x_0^2(t)\,dt} \tag{4.4.67}$$

From Schwarz's inequality, it follows that

$$\int_{-\infty}^{\infty} x_0'(t)^2\,dt \int_{-\infty}^{\infty} t^2 x_0^2(t)\,dt \geq \left[\int_{-\infty}^{\infty} t x_0'(t)x_0(t)\,dt\right]^2 \tag{4.4.68}$$

Integrating by parts, and assuming that $x_0(t) \to 0$ as $t \to \pm\infty$, we can write

$$\begin{aligned} \int_{-\infty}^{\infty} t x_0'(t)\,x_0(t)\,dt &= \tfrac{1}{2}\int t\frac{d}{dt}[x_0^2(t)]\,dt \\ &= -\tfrac{1}{2}\int x_0^2(t)\,dt \end{aligned} \tag{4.4.69}$$

Thus it follows from Eqs. (4.4.66)-(4.4.69) that

$$\beta\alpha \geq \pi \tag{4.4.70}$$

The equality sign in Eq. (4.4.68) holds only when $-Atx(t) = x'(t)$, where A is a constant. This relation yields a solution in the form of a Gaussian pulse, $\exp(-At^2/2)$. Therefore, the Gaussian pulse gives poorer simultaneous measurements of range and frequency than any other signal.

Narrowband waveforms. It is usually more convenient to measure the maximum amplitude or zero crossing of a waveform by working with the detected video signal rather than the carrier itself. This procedure avoids aliasing problems that result from the repetitive nature of the carrier. In this case, we may use the results we have already obtained with only a slight modification, provided that the signal has a "narrowband" waveform centered about a frequency ω_0. It is con-

venient to write

$$u(t) = a(t)e^{j\omega_0 t} \tag{4.4.71}$$

or

$$u(t) = f(t)e^{j[\omega_0 t + \theta(t)]}$$

where the real part of $u(t)$ is $x(t)$, and we define $f(t)$ as a real quantity. For a *narrowband waveform*, $f(t)$ and $\theta(t)$ are defined as amplitude and phase functions that vary relatively slowly compared to $\omega_0 t$. We can write the real part of $u(t)$ in the form

$$\begin{aligned} x(t) &= f(t)\cos[\omega_0 t + \theta(t)] \\ &= \tfrac{1}{2}[u(t) + u^*(t)] \end{aligned} \tag{4.4.72}$$

The employment of the modulation function $a(t)$ in Eq. (4.4.71) is often no more convenient analytically than employing $u(t)$, unless $x(t)$ or $u(t)$ is a narrowband waveform. A narrowband waveform can be defined as one for which

$$\begin{aligned} U_+(\omega) &= 0 \qquad \omega < 0 \\ U_-(\omega) &= 0 \qquad \omega > 0 \end{aligned} \tag{4.4.73}$$

where

$$U_+(\omega) = \tfrac{1}{2}\int_{-\infty}^{\infty} u(t)e^{-j\omega t}\,dt \tag{4.4.74}$$

it follows from Eq. (4.4.71) that

$$U_+(\omega) = \tfrac{1}{2}\int_{-\infty}^{\infty} f(t)e^{-j(\omega-\omega_0)t}e^{j\theta(t)}\,dt \tag{4.4.75}$$

Similarly, we define $U_-(\omega)$ as follows:

$$U_-(\omega) = \tfrac{1}{2}\int_{-\infty}^{\infty} u^*(t)e^{-j\omega t}\,dt \tag{4.4.76}$$

or

$$U_-(\omega) = \tfrac{1}{2}\int_{-\infty}^{\infty} f(t)e^{-j(\omega+\omega_0)t}\,dt \tag{4.4.77}$$

where $u^*(t)$ is the complex conjugate of $u(t)$. Thus a narrowband waveform, as illustrated in Fig. 4.4.3(a), is one for which there is no spillover of $U_+(\omega)$ into the negative frequency region, due to a modulation signal $a(t)$. This can be seen from Eq. (4.4.75) if we assume that the only contributions to $U_+(\omega)$ are from the region near $\omega = \omega_0$, where the phase of the integrand varies only slowly. The narrowband assumption will be satisfied for most radar signals, as illustrated in Fig. 4.4.3(a), but not always for sonar or acoustic imaging devices that are operated with extremely short baseband pulses; this condition is illustrated in Fig. 4.4.3(b).

A linear detector, or envelope detector, can be regarded as obtaining $a(t)$ directly, that is, because of the narrowband nature of $u(t)$ or, more correctly, $x(t)$,

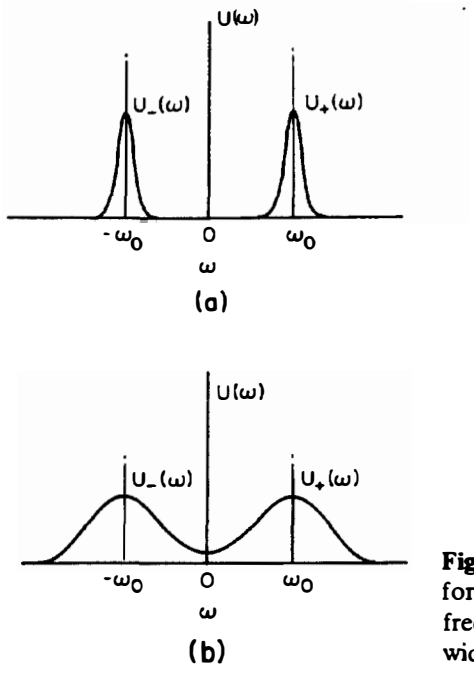


Figure 4.4.3 (a) Narrowband waveform with no "spillover" from positive frequencies to negative frequencies; (b) wideband waveform.

we can carry out sufficient filtering. After passing $u(t)$ through a matched filter, we obtain an output

$$\begin{aligned} y(t) &= e^{j\omega_0 t} \int_{-\infty}^{\infty} a(\tau)a(\tau - t)\, d\tau \\ &= \frac{1}{2\pi} e^{j\omega_0 t} \int_{-\infty}^{\infty} |A(\Omega)|^2 e^{j\Omega t}\, d\Omega \end{aligned} \tag{4.4.78}$$

where $A(\Omega)$ is the Fourier transform of $a(t)$ and Ω is the radian modulation frequency. Thus we can write

$$\begin{aligned} A(\Omega) &= \int_{-\infty}^{\infty} a(t)e^{-j\Omega t}\, dt \\ &= \int_{-\infty}^{\infty} f(t)e^{j\theta(t)}e^{-j\Omega t}\, dt \end{aligned} \tag{4.4.79}$$

The stored energy in the real signal is

$$\begin{aligned} E &= \int_{-\infty}^{\infty} x^2(t)\, dt \\ &= \tfrac{1}{4} \int_{-\infty}^{\infty} \{u^2(t) + [u^*(t)]^2 + 2u(t)u^*(t)\}\, dt \end{aligned} \tag{4.4.80}$$

It follows from Parseval's theorem [32] that

$$E = \tfrac{1}{2} \int_{-\infty}^{\infty} [U(f)U(-f) + U(f)U^*(f)]\, df \tag{4.4.81}$$

For a narrowband signal, $U(f) = 0$ when $U(-f)$ is finite, and vice versa. Hence

$$\begin{aligned} E &= \tfrac{1}{2}\int_{-\infty}^{\infty} U(f)U^*(f)\,df \\ &= \tfrac{1}{2}\int_{-\infty}^{\infty} A(f)A^*(f)\,df \end{aligned} \tag{4.4.82}$$

We define a bandwidth parameter β_M for the modulated signal as follows:

$$\beta_M^2 = \frac{4\pi^2 \int_{-\infty}^{\infty} (f - f_0)^2 U(f)U^*(f)\,df}{\int_{-\infty}^{\infty}} \tag{4.4.83}$$

When the waveform is narrowband, there is no spillover from the negative frequency terms near $f = -f_0$. Thus we can write

$$\beta_M^2 = \frac{4\pi^2 \int_{-\infty}^{\infty} (f - f_0)^2 |A(f)|^2\,df}{\int_{-\infty}^{\infty}} \tag{4.4.84}$$

4.4.6 Ambiguity Function

In Sec. 4.4.5 we showed that the use of a matched filter provides optimum range and Doppler shift resolution. The choice of waveform to optimize both these parameters is a principal task of the radar designer. The designer's aim is to choose a waveform that will leave the least amount of ambiguity after the matched filtering process. To do this, it is convenient to define a function, the *ambiguity function* $|\chi(\Delta t, \Delta f)|^2$, where

$$\chi(\Delta t, \Delta f) = \int_{-\infty}^{\infty} u(s)u^*(s - \Delta t)e^{2j\pi s\Delta f}\,ds \tag{4.4.85}$$

Here we have used s as a floating variable instead of τ, which we shall reserve for other purposes. The ambiguity function is the basis of modern radar technology in the systematic search for the optimum waveform. As we shall show, it is proportional to the peak power output from the matched filter when the input signal time changes by Δt, and its frequency changes by Δf from the optimum values for peak power output from the matched filter.

We generalize Eq. (4.4.45) for a complex signal and write that the output at a time Δt due to an input signal $u(t)$ is

$$\chi(\Delta t, 0) = \int_{-\infty}^{\infty} u(s)u^*(s - \Delta t)\,ds \tag{4.4.86}$$

The input signal can be written in the form

$$u(t) = a(t)e^{j\omega_0 t} \tag{4.4.87}$$

If the carrier frequency is changed to a frequency ω so that

$$\omega = \omega_0 + 2\pi \Delta f \tag{4.4.88}$$

then when the input is changed to a radian carrier frequency ω, the output from a matched filter designed for the original signal of radian carrier frequency ω_0 is the parameter $\chi(\Delta t, \Delta f)$ defined in Eq. (4.4.85).

In accordance with the conventional notation for the ambiguity function, we put $\phi = -\Delta f$ and $\Delta t = -\tau$, and write

$$\chi(\tau, \phi) = \int_{-\infty}^{\infty} u(s)u^*(s + \tau)e^{-2j\pi s\phi}\, ds \tag{4.4.89}$$

By using Parseval's theorem [32], it follows that

$$\chi(\tau,\phi) = \int_{-\infty}^{\infty} U^*(f)U(f + \phi)e^{-2j\pi f\tau}\, df \tag{4.4.90}$$

From Eqs. (4.4.89) and (4.4.90) it is easy to show that

$$|\chi(\tau, \phi)|^2 = |\chi(-\tau, -\phi)|^2 \tag{4.4.91}$$

Thus the ambiguity function is symmetric around the origin, as might be expected, because the output power obtains its peak value when $\tau = 0$ and $\phi = 0$. We can also show that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |\chi(\tau, \phi)|^2\, d\tau\, d\phi = |\chi(0, 0)|^2 \tag{4.4.92}$$

This property is usually referred to as the radar uncertainty principle, which is closely related to the uncertainty principles already derived in Sec. 4.4.5; it shows that the total potential ambiguity is the same for all signals that possess the same energy. The radar designer's aim is to distribute the ambiguity in the optimum way for a particular system.

Some examples of the distribution of ambiguity, plotted as functions of τ and ϕ, are given in Fig. 4.4.4. The first example is a short monotone pulse, shown in Fig. 4.4.4(a), for which a contour of constant amplitude is plotted. This takes the form of an ellipse with major and minor axes of lengths $\Delta\tau = T/2$ and $\Delta\phi = 1/2T$, respectively. Multiple peaks occur when a periodic pulse train is used, as shown in Fig. 4.4.4(c). These correspond to aliasing, or sidelobes, which must be considered in the design of the optimum radar system. Similar sets of ambiguity functions for FM pulses are shown in Fig. 4.4.4(b) and (d). The uniform pulse train shown in Fig. 4.4.4(c) degenerates to a distribution of points when the pulse train takes the form

$$u(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT) \tag{4.4.93}$$

The corresponding ambiguity function is known as a "bed of nails." It is illustrated in Fig. 4.4.5.

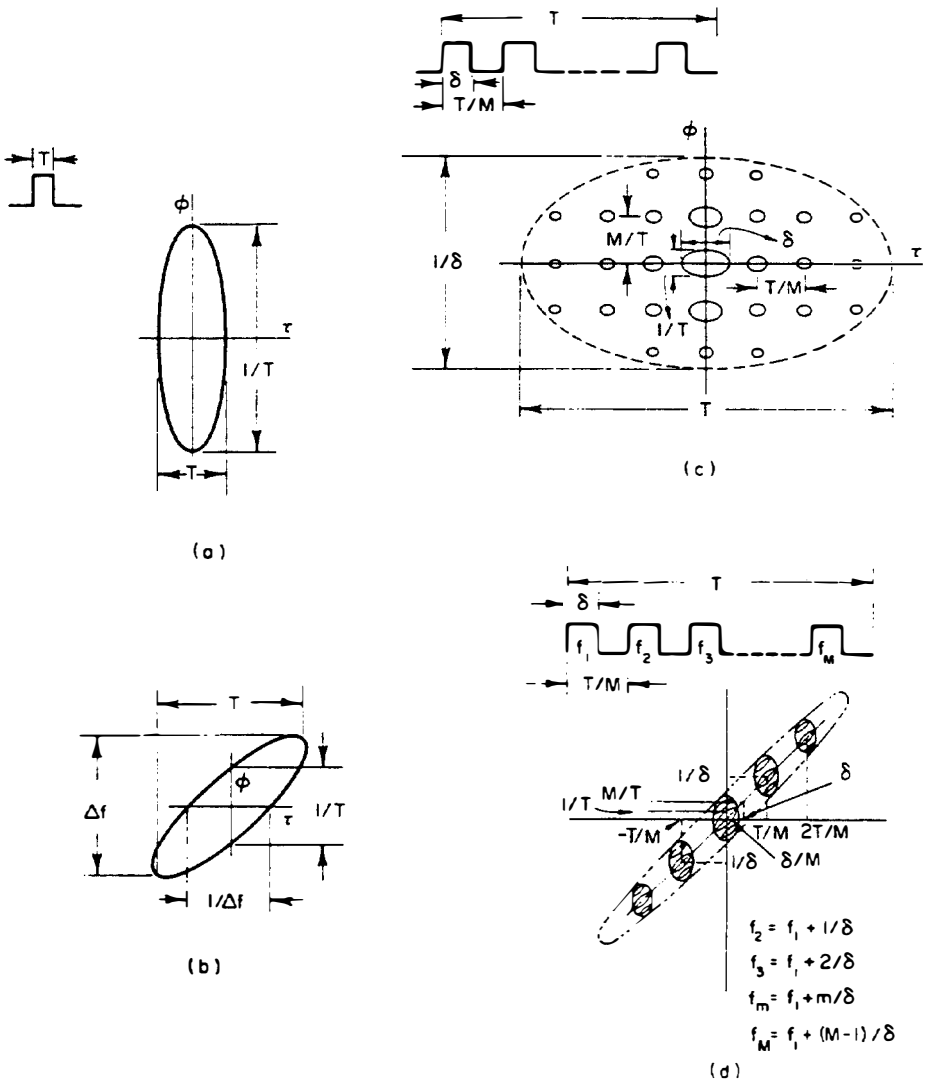


Figure 4.4.4 Examples of ambiguity function distributions, with contours at a constant amplitude level relative to $|\chi(0, 0)|$: (a) short monotone pulse; (b) linear FM pulse; (c) uniform pulse train; (d) pulse-to-pulse stepped FM pulse train. (After Bernfeld as noted in Cook and Bernfeld [1, 37].)

We can generalize the analysis for the range and velocity accuracy of a radar system given in Sec. 4.4.5, and write

$$\frac{|\chi(\tau, \phi)|^2}{|\chi(0, 0)|^2} = 1 + \frac{1}{|\chi(0,0)|^2}\left[\tau^2\left(\frac{\partial^2\chi}{\partial\tau^2}\right)_{\tau=\phi=0} + 2\tau\phi\left(\frac{\partial^2\chi}{\partial\tau\partial\phi}\right)_{\tau=\phi=0} + \phi^2\left(\frac{\partial^2\chi}{\partial\phi^2}\right)_{\tau=\phi=0}\right] \tag{4.4.94}$$

Following the same kind of analysis as in Sec. 4.4.5, it can be shown that

$$\frac{|\chi(\tau, \phi)|^2}{|\chi(0, 0)|^2} = 1 - \beta^2\tau^2 + 2\alpha\beta\rho\tau\phi + \alpha^2\phi^2 \tag{4.4.95}$$

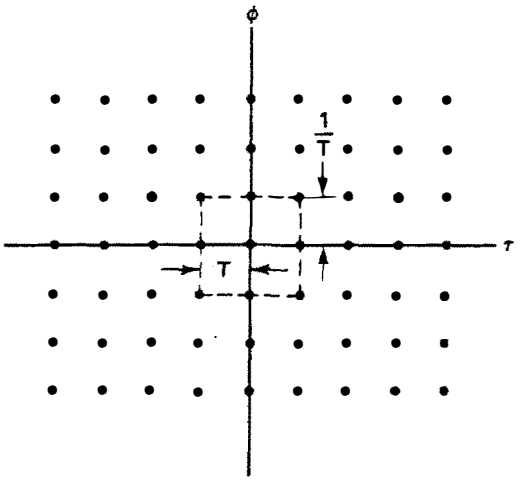


Figure 4.4.5 "Bed of nails" ambiguity function as an infinite train of impulses.

where

$$2E\alpha^2 = 4\pi^2 \int_{-\infty}^{\infty} t^2 |u(t)|^2 \, dt \tag{4.4.96}$$

$$2E\beta^2 = 4\pi^2 \int_{-\infty}^{\infty} f^2 |U(f)|^2 \, df \tag{4.4.97}$$

and

$$2E\rho = -\frac{2\pi}{\alpha\beta} \operatorname{Im}\left\{ \int_{-\infty}^{\infty} tu(t)[u^*(t)]' \, dt \right\} \tag{4.4.98}$$

The curve formed by the intersection of $|\chi(\tau, \phi)|^2$ with a level plane close to the maximum of $|\chi(0, 0)|^2$ is

$$\beta^2\tau^2 + 2\alpha\beta\rho\tau\phi + \alpha^2\phi^2 = \gamma^2 \tag{4.4.99}$$

This curve is an ellipse. By putting $\gamma^2 = N_0/2E$, Eq. (4.4.99) describes the *uncertainty ellipse*, corresponding to a generalized form of the mean-square error in range and frequency described in Sec. 4.4.5. A plot of this uncertainty ellipse is given in Fig. 4.4.6. It will be seen that, in general, the maximum errors in time and frequency are

$$\Delta\tau_{\max} = \pm \frac{1}{2\beta\sqrt{E/N_0}} \frac{1}{\sqrt{1 - \rho^2}} \tag{4.4.100}$$

and

$$\Delta\phi_{\max} = \pm \frac{1}{2\alpha\sqrt{2E/N_0}} \frac{1}{\sqrt{1}} \tag{4.4.101}$$

These results correspond to generalizations of our earlier results for time and frequency error.

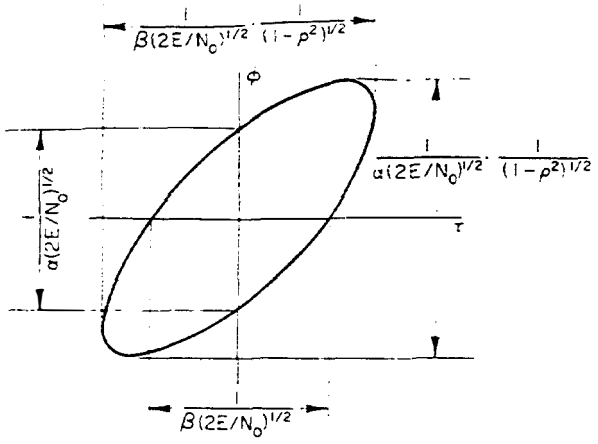


Figure 4.4.6 Uncertainty ellipse. (After Cook and Bernfeld [1].)

We have discussed general properties of the ambiguity function. The idealized ambiguity function would have the form of a spike at $\tau = 0$, $\phi = 0$, with zero amplitude elsewhere. Obviously, this is not possible.

We have discussed many times the linear FM pulse waveform, whose properties are illustrated in Fig. 4.4.7 together with a plot of $|\chi(\tau, \phi)|$. In this case it is very difficult to distinguish between changes in τ and ϕ. On the other hand, the V-FM pulse waveform consisting of a down-chirp followed by an up-chirp, as illustrated in Fig. 4.4.8, makes it relatively easy to distinguish between time and

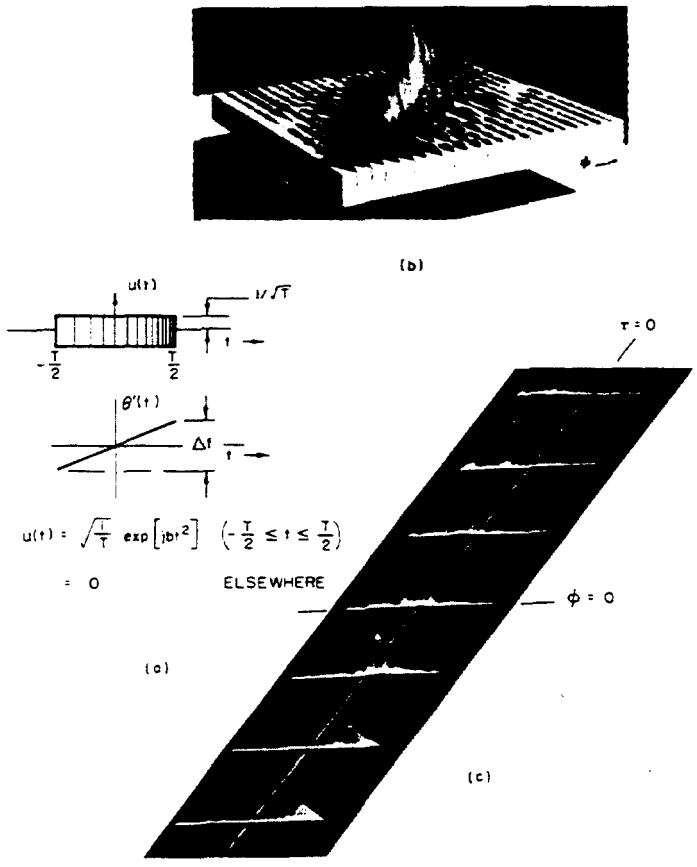


Figure 4.4.7 Linear FM pulse waveform and response function properties $T\Delta f = 10$: (a) waveform; (b) calculated surface; (c) experimental composite surface. (After Cook and Bernfeld [1].)

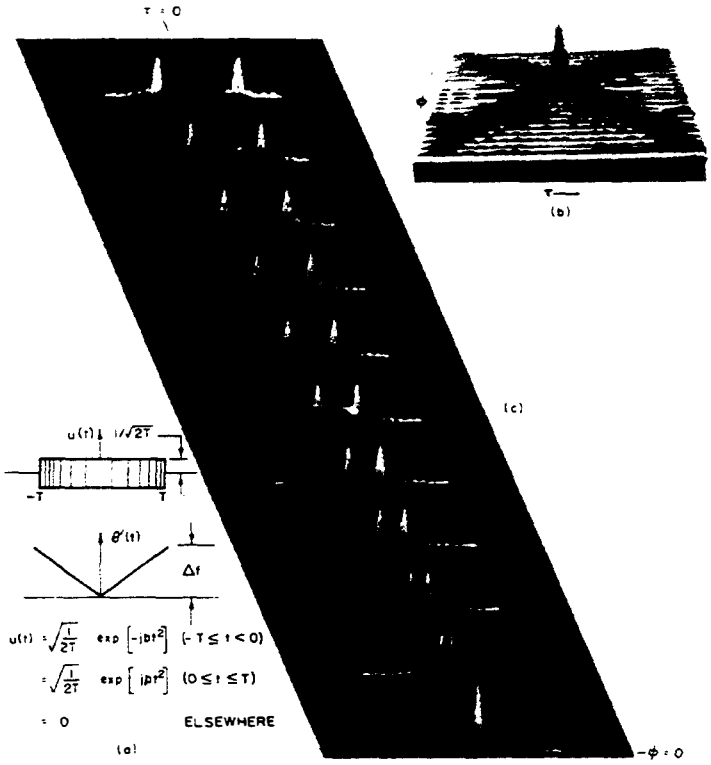


Figure 4.4.8 V-FM pulse waveform and response function properties $2T\Delta f = 200$: (a) waveform; (b) calculated composite response surface (note that pedestal is not included); (c) experimental composite response surface. (After Cook and Bernfeld [1].)

frequency changes. Time differences correspond to the difference in the τ direction between two peaks, while frequency differences correspond to movements along the ϕ axis. Parabolic FM pulse waveforms and other kinds of waveforms can provide still better improvements in the form of the ambiguity function (see Prob. 9).

A possible set of idealized waveforms are ones that are rotationally invariant in the ambiguity plane. The *Hermite polynomials* can be used for such a set. For the nth-order Hermite polynomial $H_n(x)$, the required signal has the form

$$u_n(t) = \frac{2^{1/4}}{\sqrt{n!}} e^{-\pi t^2} H_n(2\sqrt{\pi}t) \tag{4.4.102}$$

where $H_0(x) = 1$; $H_1(x) = x$; $H_2(x) = x^2 - 1$; $H_3(x) = x^3 - 3x$; . . . . A plot of the composite response surface for the Hermite waveform, with $n = 10$, is shown in Fig. 4.4.9.

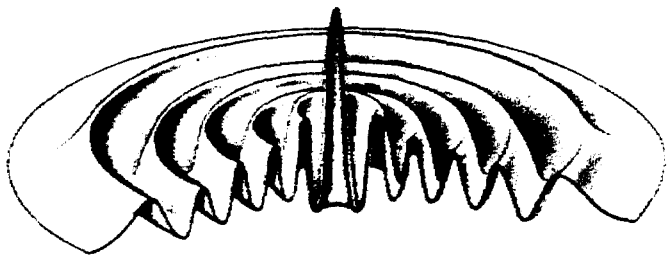


Figure 4.4.9 Composite response surface $\chi(\tau, \Phi)$ for Hermite waveform, $n = 10$. (After Klauder, as noted in Cook and Bernfeld [1, 38].)

Obviously, there is no ideal ambiguity function. The ambiguity function is a measure of how the radar system distinguishes between similarly coded waveforms at the receiver that differ in time and frequency. The reader is referred to the standard references for more information on ambiguity functions [1].

PROBLEM SET 4.4

1. Determine the optimum matched filter and maximum signal-to-noise ratio when the noise power spectral density $N_0(f)$ in W/Hz is nonuniform over the frequency range.

   Hint: You will find it convenient to define a new function $M(f) = H(f)[N_0(f)]^{1/2}$.

2. The example in Sec. 4.4.2 of a matched filter for a square pulse of length T and frequency response $\sin(\omega T/2)/(\omega T/2)$ implied the use of a filter with an identical frequency response $\sin(\omega T/2)/(\omega T/2)$. Find an expression to determine how much an ideal low-pass filter, with a response

$$H(\omega) = \begin{cases} 1 & 0 < |\omega| < \omega_c \\ 0 & |\omega| > \omega_c \end{cases}$$

   will reduce the signal-to-noise ratio from that of the ideal matched filter, when used with the same input pulse of length T. Plot the reduction factor R as a function of $\omega_c T$. From your numerical calculations, show that the signal-to-noise ratio is reduced by 0.84 dB, when $\omega_c T = 4.4$, where R is maximum. You can carry out a computation or use tables for

$$\int_0^a \text{——}$$

   You will also need the result

$$\int_{-\infty}^{\infty} \left(\frac{\sin x}{x}\right)^2 dx = \pi$$

3. Repeat Prob. 2 for a Gaussian filter with a response characteristic $\exp(-\omega/\omega_c)^2$. Find $\omega_c T$ for the maximum signal-to-noise ratio.

4. Consider the value of the root-mean-square (RMS) time error $[\langle \Delta t^2 \rangle]^{1/2}$ for an input signal $x(t) = x_0(t) + n(t)$, received by a radar with a response $h(t)$. Define a generalized bandwidth parameter β' as follows:

$$(\beta')^2 = \frac{4\pi^2 \int_{-\infty}^{\infty} f^2 X_0(f) H(f)\, df}{\int_{-}^{\infty}}$$

   Show that with the effective bandv the optimum response for minimum time error [i.e., a minim (near a time $t = 0$] will be minimum when

$$H(f) = KX_0^*(f)$$

   where K is a constant.

5. Consider a narrowband radar signal of the form $u(t) = a(t)e^{j2\pi f_0 t}$. Write down the form of the matched filter for this waveform. Find $y_0(0)$, the maximum output at $t = 0$ when there is no noise present. Now suppose that we change the carrier frequency of the input signal to this matched filter, from f_0 to $f_0 + \Delta f$. Using the methods leading to Eq. (4.4.59), write down an expansion for $|Y_0(0, \Delta f)|^2$ (i.e., the change in output power when the input frequency is changed by Δf). Use this result to show that in the presence of noise, the mean-square error in determining f by measuring the point where the output amplitude is maximum is

$$\langle (\Delta f)^2 \rangle^{1/2} = \frac{1}{\alpha \sqrt{2E/N_0}}$$

where

$$\alpha^2 = \frac{4\pi^2}{2E} \int_0^T t^2 u(t) u^*(t)\, dt$$

and

$$E = \tfrac{1}{2} \int_0^T u(t) u^*(t)\, dt = \tfrac{1}{2} \int_0^T a^2(t)\, dt$$

6. Prove Eq. (4.4.92).
7. Prove Eq. (4.4.99).
8. **(a)** Show that for a Gaussian envelope $u(t) = (2k^2/\pi)^{1/4} \exp(-k^2 t^2)$, the function $\chi(\tau, \phi)$ has the form

$$\chi(\tau, \phi) = \exp\left[-\tfrac{1}{2} \left(k^2 \tau^2 + \frac{\pi^2 \phi^2}{k^2} \right) \right]$$

(b) What are α, β, and ρ for this signal when τ and ϕ are small?

9. **(a)** Consider a radar waveform in the form of a parabolic FM chirp of the form

$$u(t) = \sqrt{\frac{1}{T}} \cos(ct^3) = \sqrt{\frac{1}{T}} \operatorname{Re}\left(e^{jct^3} \right)$$

By using the method of stationary phase, show that

$$|\chi(\tau, \phi)| = \frac{1}{T} \sqrt{\frac{\pi}{3c\tau}} \cos\left(\frac{\pi^2 \phi^3}{3c\tau} - c\tau^3 \right)$$

Note that the result is similar to that for the V-FM waveform.

(b) Find an expression for the contours of maximum $|\chi(\tau, \phi)|$, that is, the peaks of $|\chi(\tau, \phi)|$.

4.5 FM CHIRP FILTERS AND CHIRP TRANSFORM PROCESSORS

4.5.1 Introduction

In Sec. 4.2.3 we discussed the basic concept of an FM chirp filter. As we saw, if a dispersive delay line (i.e., delay line whose time delay varies with frequency) is used as a matched filter for a linear FM chirp of length T, an output pulse of length

τ_p will be obtained. The pulse width τ_p is determined by the bandwidth B of the delay line, and $\tau_p \approx 1/B$. The signal-to-noise ratio is therefore improved by a factor of $T/\tau_p \approx BT$, the compression ratio.

Furthermore, as we have seen, the system can be used to obtain real-time Fourier transforms of modulated chirp signals. In this section we first consider a mathematical treatment of these concepts; we then describe various SAW and CCD realizations for chirp filters, and their applications to filtering complex signals.

4.5.2 Mathematical Treatment of FM Chirp Filter

We now examine in more detail what occurs when a linear FM chirp signal is inserted into a matched filter, as well as the problems associated with sidelobes and the techniques used to decrease them. We consider more general filters derived from the basic matched filter, which have time delays not necessarily equal to the length of the input pulse. We also consider how real physical filters can be used for complex signals.

A linear FM chirp signal, with a radian frequency that varies with time as $\omega = \omega_0 + \mu t$, has a phase

$$\phi(t) = \int \omega \, dt = \omega_0 t + \frac{\mu t^2}{2} \tag{4.5.1}$$

where the slope of the chip μ has the dimensions of rad/s^2. We can therefore write a modulated chirp signal in the form

$$u(t) = f(t)e^{j(\omega_0 t + \mu t^2/2)} \tag{4.5.2}$$

where $f(t)$ is, in general, the complex amplitude modulation of the FM chirp. If we insert this signal in a matched filter, it follows from Eq. (4.4.31) that the matched filter characteristic is of the form

$$h(t) = \alpha u^*(T_0 - t) = f^*(T_0 - t)e^{j(\omega_0 t - \mu t^2/2)} \tag{4.5.3}$$

where α is a constant of the filter and T_0 is the time at which the output signal will reach its peak value. To simplify the algebra, we shall assume from now on that $T_0 = 0$, and that the FM chirp can start at a time earlier than $t = 0$, as dictated by the value of $f(t)$. We will generalize the form of the filter from that of a simple matched filter and use it as a matched filter only for the carrier exp $[j(\omega_0 t + \mu t^2/2)]$. Therefore, we write

$$h(t) = \alpha w_0(t)e^{j(\omega_0 t - \mu t^2/2)} \tag{4.5.4}$$

The parameter $\omega_0(t)$ is called the *weighting* of the filter, while $f(t)$ is called the *modulation of the chirp*. In the simplest case, $f(t) = 1$ and $w_0(t) = 1$ for $-T/2 < t < T/2$ with $f(t) = 0$, and $w_0(t) = 0$ for $|t| > T/2$. This is an exact matched filter.

It is convenient to use one of the convolution formulae for the output signal

$$y(t) = \int_{-\infty}^{\infty} u(\tau)h(t - \tau) \, d\tau \tag{4.5.5}$$

or

$$y(t) = \int_{-\infty}^{\infty} u(t - \tau)h(\tau)\, d\tau \tag{4.5.6}$$

When the filter is of finite length T and the chirp length is much longer than T, we take $f(t) = 1$ for $-\infty < t < \infty$, with $w_0(t) = 1$ for $-T/2 < t < T/2$, and $w_0(t) = 0$ for $|t| > T/2$. In this case, it is convenient to use Eq. (4.5.6). If the output is maximum at $t = 0$, and $h(t) = \alpha \exp[j(\omega_0 t - \mu t^2/2)]$, then

$$y(t) = \alpha \int_{-T/2}^{T/2} e^{j(\omega_0(t-\tau) + \mu(t-\tau)^2/2]} e^{j(\omega_0\tau - \mu\tau^2/2)}\, d\tau \tag{4.5.7}$$

It follows that

$$y(t) = \alpha e^{j(\omega_0 t + \mu t^2/2)} \int_{-T/2}^{T/2} e^{-j\mu t\tau}\, d\tau \tag{4.5.8}$$

This expression can be integrated to give the result

$$y(t) = \alpha T \frac{\sin(\mu Tt/2)}{\mu Tt/2} e^{j(\omega_0 t + \mu t^2/2)} \tag{4.5.9}$$

Width of main lobe. The output is now sharply peaked around the time $t = 0$, with a maximum amplitude proportional to the time length of the delay line or filter T, as we might expect. Furthermore, the amplitude of the signal drops 4 dB at the points where $\mu Tt = \pi$. Thus we can define the 4-dB width of the pulse in seconds as

$$\tau_p(4\text{ dB}) = \frac{2\pi}{\mu T} = \frac{1}{B} \tag{4.5.10}$$

and the 3-dB width as

$$\tau_p(3\text{ dB}) = \frac{0.89}{B} \tag{4.5.11}$$

where $B = \mu T/2\pi$ is the bandwidth of the chirp in hertz and μT is its bandwidth in rad/s. Thus the pulse compression ratio (to the 3-dB points) is $\tau_p/T = 0.89/BT$, close to what we would have expected from general considerations.

Sidelobes. We observe that the first sidelobe is at $t = \pm 3/2B$, with an amplitude $2/3\pi$ smaller than the main lobe. Thus the first sidelobe has a level only 13 dB down from the main lobe. In a radar system (or any other signal processing system) this can cause serious problems: If a later weak signal (from a weak reflector) enters the filter and has an amplitude more than 13 dB below that of the main signal, it may be masked by a sidelobe from the main signal, or else a sidelobe might be mistaken for an echo from a weaker target. For this reason it is usually necessary to design radar systems with a modified filter or chirp signal, which is used to decrease the sidelobe level at the expense of slightly increasing

the width of the main lobe and slightly worsening the optimum signal-to-noise ratio. The procedures used are similar to those employed in designing antennas or interdigital transducers to obtain low sidelobe levels.

Physics of FM chirp filter. To put this derivation on a more physical basis, we can consider the output obtained directly from an SAW chirp filter of the type illustrated in Fig. 4.2.7. Suppose that the input to the filter were of the form given in Eq. (4.5.2). After the signal has traveled along the delay line a distance z from the input toward the matching dispersive transducer, as shown in Fig. 4.2.7, it will have an amplitude

$$u\left(t - \frac{z}{V}\right) = f\left(t - \frac{z}{V}\right) e^{j\omega_0(t - z/V)} e^{j\mu(t - z/V)^2/2} \tag{4.5.12}$$

where V is the velocity of the wave along the delay line, and we have included a general modulation function $f(t)$. Suppose that the length of the finger at the plane z is $w(z)$, and that the weighting $w(z)$ may be positive or negative. Using the apodizing technique illustrated in Fig. 4.2.6, with closely spaced fingers of infinitesimal width, we would expect the current induced on a finger to be proportional to $w(z)$. Thus the total output is of the form

$$y(t) = \gamma \int_{-L/2}^{L/2} w(z) u\left(t - \frac{z}{V}\right) dz \tag{4.5.13}$$

where L is the spatial length of the delay line, taken to start at $z = -L/2$, and γ is a constant.

We choose $w(z)$ to have the physically realizable form

$$w(z) = w_0(z) \cos\left(\frac{\omega_0 z}{V} - \frac{\mu z^2}{2V^2}\right) \tag{4.5.14}$$

so that $w(z)$ is positive, negative, or zero. We can split the expression for $w(z)$ into two parts

$$w(z) = \frac{w_0(z)}{2}\left[e^{j(\omega_0 z/V - \mu z^2/2V^2)} + e^{-j(\omega_0 z/V - \mu z^2/2V^2)}\right] \tag{4.5.15}$$

and we now take $w_0(z) = w_0 =$ constant for $|z| < L/2$. Then the contribution of the dominant cumulative term from Eqs. (4.5.14) and (4.5.15) [i.e., the first term in Eq. (4.5.15)] to the integral in Eq. (4.5.13) is

$$y(t) \approx \frac{\gamma w_0}{2} \int_{-L/2}^{L/2} f\left(t - \frac{z}{V}\right) e^{j[\omega_0(t - z/V) + \mu(t - z/V)^2/2]} e^{j(\omega_0 z/V - \mu z^2/2V^2)}\, dz \tag{4.5.16}$$

where, in analogy with our previous example, we have taken the chirp to be very long, and the delay line to have its input at $-L/2$ and to be of length L. Setting $L = VT$ and $z = V\tau$, Eq. (4.5.16) reduces to the form

$$y(t) = \frac{\gamma V w_0}{2} e^{j(\omega_0 t + \mu t^2/2)} \int_{-T/2}^{T/2} f(t - \tau) e^{-j\mu t\tau}\, d\tau \tag{4.5.17}$$

If $f(t) = 1$, Eq. (4.5.17) will be identical in form to Eq. (4.5.8). In this case, the maximum amplitude of the output is at $t = 0$ and of value $y(0) = \lambda V w_0 T/2$. But note that if we had chosen the delay line to have a time delay much longer than the length of the chirp, our results would have been different (see Prob. 1).

In using a physical delay line, we have chosen the elements to have a cosine weighting. Therefore, in deriving the filter response, we had to neglect the contribution of the second term in Eq. (4.5.15). It can be shown that in a SAW device, the output signal due to this effect is negligible (see Prob. 2). A device that operates at baseband, however, like a CCD filter, can give rise to spurious signals associated with the interaction of an unwanted cosine or sine component of the chirp signal with the delay line. In Sec. 4.5.5 we describe how a pair of sine and cosine weighted delay lines can be used to synthesize a complex filter to eliminate this problem.

We have described here a system with fingers that are infinitesimally spaced (i.e., that continuously sample the signal). A real SAW or CCD delay line would employ sampling elements a finite length apart. One possibility would be to choose the elements to be equally spaced and amplitude-weighted in the manner prescribed by Eq. (4.5.14), with the elements at positions $z_n = z = nl$, where l is the spacing of the elements. Alternatively, as shown in Fig. 4.2.7, the elements would be placed at positions $z = z_n$ corresponding to the maxima of the cosine function, such that

$$\frac{\omega_0 z_n}{V} - \frac{\mu z_n^2}{2V^2} = n\pi \tag{4.5.18}$$

In both cases, there are usually at least two elements per RF wavelength to satisfy the Nyquist sampling criterion.

4.5.3 Filter Weighting for Sidelobe Reduction

Before we proceed, let us consider how we can weight the filter to give a response with low sidelobe levels [1, 39]. Otherwise, if the sidelobe levels are high, as shown in Eq. (4.5.9), two chirp signals arriving at different times from two different targets will both give $\sin(\mu tT/2)/(\mu tT/2)$ responses after passing through the matched filter. If the sidelobe level of this response is too high, the main lobe of the weak signal may be obscured by the sidelobe from the strong signal. The basic reason for this difficulty comes from the fact that the FM chirp signal and/or the matched filter may be finite in length.

It is obviously desirable to smooth out the response of the matched filter; therefore, we consider weighting the filter by the weighting function $w_0(t)$ given in Eq. (4.5.4). Similarly, if we were using a physical transversal filter, we would have chosen an equivalent filter of the form given by Eq. (4.5.14). But in this case, it follows from Eq. (4.5.6) that a similar derivation to that of Eq. (4.5.17) yields the output from the matched filter in the form

$$y(t) = \alpha e^{j(\omega_0 t + \mu t^2/2)} \int_{-\infty}^{\infty} e^{-j\mu t\tau} w_0(\tau)\, d\tau \tag{4.5.19}$$

Here we have used infinite limits for the integral by assuming that $w_0(\tau)$ has a finite length, and we have taken $\tau = z/V$ in the physical filter as before. We call the weighting of a finite-length filter *apodization*.

Note that the output signal is the *Fourier transform* of the weighting of the filter. To obtain optimum weighting, we want to choose $w_0(\tau)$ such that $S(\mu t)$, defined by the integral

$$S(\mu t) = \int_{-T/2}^{T/2} e^{-j\mu t\tau} w_0(\tau)\, d\tau \tag{4.5.20}$$

with a main lobe centered at $t = 0$, has a minimum width and minimum sidelobe levels.

As a simple example, suppose that we take $w_0(\tau)$ to be a Gaussian function with the filter infinitely long, such that

$$w_0(\tau) = e^{-8(\tau/T)^2} \tag{4.5.21}$$

The effective length of this filter between the $1/e^2$ amplitude points is T, and the bandwidth between these points is $B = \mu T/2\pi$. The transformed output then takes the form

$$U(\mu t) = \frac{T\sqrt{\pi}}{2\sqrt{2}} e^{j(\omega t + \mu t^2/2)} e^{-(\mu T)^2/32} \tag{4.5.22}$$

The output is a Gaussian pulse with no sidelobes, and with an effective pulse length $\tau_p(3\text{ dB})$ between 3-dB points of $\tau_p(3\text{ dB}) = 6.66/\mu T$, or $\tau_p(3\text{ dB}) = 1.06/\text{B}(1/e^2)$. Thus the use of an infinitely long Gaussian waveform, of approximately the same bandwidth or time length between $1/e^2$ points as that of the rectangularly weighted chirp, increases the width of the main lobe by a factor of 1.06/0.89, or 1.19.

In practice, of course, it is not possible to use a *Gaussian* weighting, for this would imply a filter of infinite length. We might approximate the filter by cutting it off where the amplitude drops to $1/e^2$; to do so, however, would increase the sidelobe level considerably, and this design method is still inefficient because it requires a much longer filter length than necessary. So techniques identical to those used for antenna array design and digital filter design have been developed. The aim is to design a finite-length weighted filter with the lowest possible sidelobe level and the narrowest possible main lobe. Such a filter normally has a fairly close approximation to a Gaussian weighting.

We conclude that the basic aim of apodization is to use extremely smooth functions for the windowing. The difficulty, of course, is that the filter must be finite in length. Some examples of apodization functions are given in Table 4.5.1. Plots of these functions, and of the main lobes and sidelobes arising from them, are given in Fig. 4.5.1; Table 4.5.2 summarizes the results obtained with them.

As an example, observe that *Hanning* weighting with $w_0(\tau) = \cos^2\pi\tau/T$ is smooth and fairly close in form to a Gaussian function. It yields a maximum sidelobe level of -32 dB, with an increase in compressed 3-dB pulse width of 1.62 times that for *rectangular* weighting. Its advantage compared to rectangular weighting is that the far-out sidelobe amplitudes fall off as $1/t^3$. On the other hand, *Hamming*

Table 4.5.1 SOME EXAMPLES OF APODIZATION FUNCTIONS

Dirichlet (rectangular)	$w_0(\tau) = 1$	
Bartlett	$w_0(\tau) = 1 + 2\tau/T$	$-T/2 < \tau < 0$
(triangular)	$w_0(\tau) = 1 - 2\tau/T$	$0 < \tau < T/2$
Hanning	$w_0(\tau) = \cos^2(\pi\tau/T)$	$-T/2 < \tau < T/2$
	$= 0.5[1 + \cos(2\pi\tau/T)]$	
Hamming	$w_0(\tau) = 0.08 + 0.92\cos^2(\pi\tau/T)$	$-T/2 < \tau < T/2$
	$= 0.54 + 0.46\cos(2\pi\tau/T)$	
Blackman	$w_0(\tau) = 0.42 + 0.5\cos(2\pi\tau/T) + 0.08\cos(4\pi\tau/T)$	$-T/2 < \tau < T/2$
Finite		
	$w_0(\tau) = \exp[-12.5(\tau/T)^2]$	τ
	$w_0(\tau) = \exp[-12.5(\tau/T)^2]$	

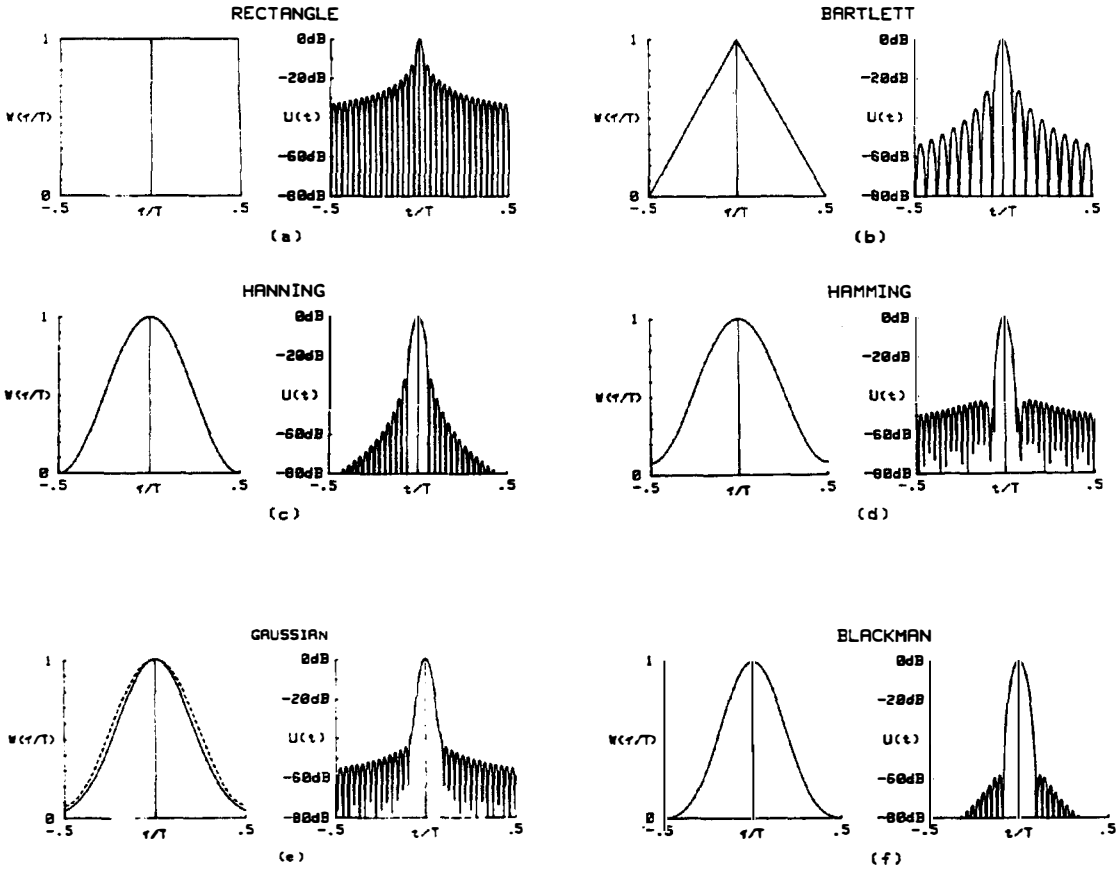


Figure 4.5.1 Some examples of windows and their Fourier transforms. All examples are for a time–bandwidth product of $BT = 50$. (a) Rectangular window and log magnitude of transform; (b) Bartlett or triangular window and log magnitude of transform; (c) Hanning window and log magnitude of transform; (d) Hamming window and log magnitude of Fourier transform; (e) finite Gaussian window $\exp[-12.5(\tau/T)^2]$ (solid line) and log magnitude of its transform compared to the Hamming window (dashed line); (f) Blackman window and log magnitude of transform.

Table 4.5.2 PROPERTIES OF WEIGHTING FUNCTIONS

Weighting function	Peak sidelobe level (dB)	$B\tau_p$(3 dB)[a]	$B\tau_p$(6 dB)[a]	R (dB) [Eq. (4.5.26)]	Far sidelobe falloff rate
Rectangular	−13	0.89	1.21	0	$1/t$
Bartlett (triangle)	−27	1.28	1.78	−1.23	$1/t^2$
Hanning $\cos^\alpha(\pi\tau/T)$					
$\alpha = 1$	−23	1.20	1.65	−0.90	$1/t^2$
$\alpha = 2$	−32	1.44	2.00	−1.76	$1/t^3$
$\alpha = 3$	−39	1.66	2.32	−2.20	$1/t^4$
$\alpha = 4$	−47	1.86	2.59	−2.88	$1/t^5$
Hamming	−43	1.3	1.81	−1.34	$1/t$
Finite Gaussian $\exp[-12.5(\tau/T)^2]$	−42	1.33	1.86	−1.43	$1/t$
Infinite Gaussian $\exp[-12.5(\tau/T)^2]$	—	1.32	1.87	—	$\exp\left[\frac{-(\pi Bt)^2}{12.5}\right]$
Blackman	−58	1.68	2.35	−2.38	$1/t^3$

[a]The bandwidth in this table is $B = \mu T/2\pi$.

weighting, a commonly used weighting function, yields a somewhat narrower pulse with lower sidelobe levels, with the disadvantage of a lower rate of amplitude falloff for the far-out sidelobes. We shall see later that improvement of this latter characteristic can be useful for designing high-quality bandpass filters with very low spurious levels outside the passband. Note that Hamming weighting is very close in form and gives similar results to the cutoff Gaussian $\exp[-12.5(\tau/T)^2]$.

Weighting functions other than the ones given here are sometimes used to obtain an optimum response. The reader is referred to the literature for discussion of *Dolph–Tchebysheff* weighting, to which Hamming, Hanning, and *Blackmann* weightings form approximations [1, 39]. A very flexible form of weighting is *Kaiser* weighting, defined by the relation

$$w(\tau) = \frac{I_0[(\omega_a T/2)(1 - 4\tau^2/T^2)^{1/2}]}{I_0(\omega_a T/2)} \tag{4.5.23}$$

where I_0 is a modified Bessel function of zero order, and ω_a is a parameter that can be adjusted to trade the main lobe width against the sidelobe amplitude. Kuo and Kaiser have shown that these windows tend to yield the largest energy in the main lobe for a given peak sidelobe amplitude [40].

Effect of apodization on signal-to-noise ratio. Now that the filter is no longer perfectly matched, we might expect nonuniform weighting to decrease its

signal-to-noise ratio. The signal-to-noise ratio is

$$\frac{S}{N} = \frac{\left| \int_0^\infty u(T - t)h(t)\, dt) \right|^2}{\frac{1}{2}N_0 \int_0^\infty |h^2(t)|\, dt} \tag{4.5.24}$$

With $t = 0$, we take

$$h(t) = \alpha w_0(t)u^*(-t) \tag{4.5.25}$$

and assume that the input signal has a rectangular envelope. The reduction in signal-to-noise ratio, due to weighting, will be

$$R = \frac{(S/N)_{\text{weighted}}}{(S/N)_{\text{rectangular}}} = \frac{\langle w_0 \rangle^2}{\langle w_0^2 \rangle} \tag{4.5.26}$$

In most cases, as we can see from Table 4.5.2, the penalty paid in signal-to-noise ratio with the use of common apodizations is small.

4.5.4 Fourier Transform Operations with FM Chirp Filters

In Sec. 4.5.2 we showed that the output from a weighted chirp filter is the Fourier transform of the weighting. We have also shown heuristically in Sec. 4.2.3 that if an FM chirp signal is modulated by a signal $f(t)$ injected into a matched filter, we might expect to obtain an output that is the Fourier transform of $f(t)$. We shall now consider this operation mathematically and give some examples of its use.

We shall take the input signal to be a modulated "down-chirp," that is, a chirp whose frequency is decreasing with time

$$u(t) = f(t)e^{j(\omega_0 t - \mu t^2/2)} \tag{4.5.27}$$

We assume that the matched filter is unapodized [$w(\tau) = 1$] and very long. Thus we use Eq. (4.5.5) [i.e., $y(t) = \int u(\tau)h(t - \tau)\, d\tau$] to show that the output from the chirp filter, with a modulated input chirp of length T, is

$$\begin{aligned} Y(\mu t) &= \alpha \int_{-T/2}^{T/2} f(\tau)e^{j(\omega_0 \tau - \mu\tau^2/2)}e^{j(\omega_0(t-\tau) + \mu(t-\tau)^2/2]}\, d\tau \\ &= \alpha e^{j(\omega_0 t + \mu t^2/2)} \int_{-T/2}^{T/2} f(\tau)e^{-j\mu t\tau}\, d\tau \end{aligned} \tag{4.5.28}$$

We put

$$F(\mu t) = \alpha \int_{-T/2}^{T/2} f(\tau)e^{-j\mu t\tau}\, d\tau \tag{4.5.29}$$

Then in this case, with the chirp length much shorter than the filter length, the output from the chirp filter is just the Fourier transform $F(\mu t)$ of the input signal,

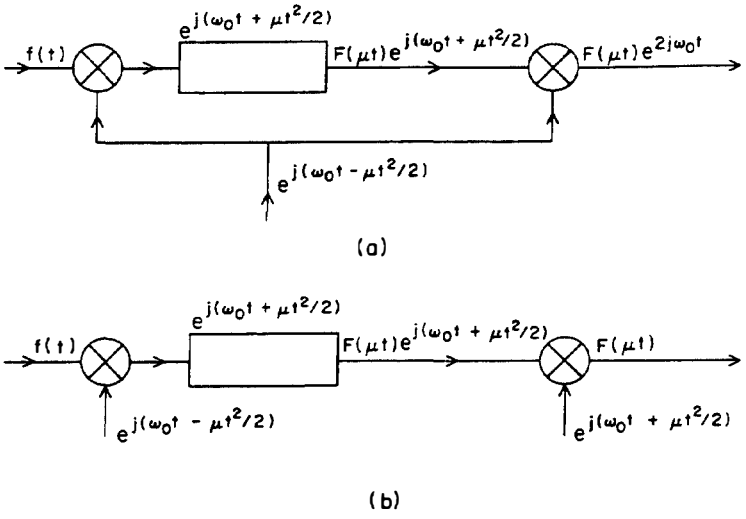


Figure 4.5.2 Two alternative schemes for obtaining a complex Fourier transform: (a) frequency transform obtained as the modulation of a carrier of frequency $2\omega_0$. (b) Fourier transform obtained as a baseband signal.

multiplied by an "up-chirp." The output signal is now a function of time, but μt represents the radian frequency in the Fourier transforms. If we require only $|F(\mu t)|$, it is sufficient to determine $|Y(\mu t)|$. However, if we need both the phase and amplitude of $F(\mu t)$, we can post-multiply the output signal $Y(\mu t)$ by a further down-chirp, $\exp[j(\omega_0 t - \mu t^2/2)]$, and obtain $F(\mu t)\exp(2j\omega_0 t)$. In this case, $F(\mu t)$ is the modulation of a carrier of frequency $2\omega_0$. Alternatively, as a mixer multiplies two signals together, signals of the forms $\cos(\omega_1 t + \phi_1)$ and $\cos(\omega_2 t + \phi_2)$ yields outputs at their sum and difference frequencies and phases, that is, with form $\cos[(\omega_1 + \omega_2)t + \phi_1 \pm \phi_2]$. We could therefore multiply by an up-chirp and obtain the output $F(\mu t)$ at baseband.

We summarize the required system in Fig. 4.5.2. In the first case, the signal is multiplied by a chirp $\exp[j(\omega_0 t - \mu t^2/2)]$, inserted into the matched filter for this chirp, and then post-multiplied by the same chirp. Observe that if the input is at baseband ($\omega_0 = 0$), the output obtained will be the direct Fourier transform of $f(t)$. Alternatively, with ω_0 finite, the modulation of the carrier is $F(\mu t)$, provided that the mixing chirp is delayed by a time T. Otherwise, we obtain a different modulated carrier as the output Fourier transform.

To determine the resolution of this Fourier transform device, let us consider what occurs when we insert a signal $f(t) = \exp(j\Omega t)$ into the Fourier transform system. The output we obtain is

$$Y(\mu t) = \alpha e^{j(\omega_0 t + \mu t^2/2)} \int_{-T/2}^{T/2} e^{j(\Omega - \mu t)\tau}\, d\tau \tag{4.5.30}$$

This formula may be integrated to give the result

$$|Y(\mu t)| = \alpha T \left| \frac{\sin(\Omega - \mu t)T/2}{(\Omega - \mu t)T/2} \right| \tag{4.5.31}$$

As we predicted from the simple physical model discussed in Sec. 4.2.3, the output will be maximum at a time $t = \Omega/\mu$. Thus the time at which the maximum output occurs depends on the input frequency.

An important critical parameter is the number of separate frequency components that can be resolved. The problem is closely analogous to the definition of resolution in an optical or acoustic image, discussed in Chapter 3. The simplest way to define resolution is to take it as the spacing between the 3-dB points of the response [i.e., $\tau_p(3\text{ dB}) = 0.89/B$], where $B = \mu T/2\pi$ is the bandwidth. In this case, from Eq. (4.5.31), the frequency variation $\Delta f = \Delta\Omega/2\pi$ between 3-dB points is

$$\Delta f(3\text{ dB}) = \frac{0.89}{T} \tag{4.5.32}$$

Thus the number of resolvable frequencies in the spectrum is $N = 1.12BT$. However, this definition of resolution is not entirely adequate. Consider, for instance, what occurs when the two equal amplitude signals of frequencies $f_1 = \Omega_1/2\pi$ and $f_2 = \Omega_2/2\pi$ are present. Let $\Delta f = f_2 - f_1$. Then the magnitude of the Fourier transformed signal is

$$|Y(t)| = \alpha T \left| \left[\frac{\sin \pi(Bt - \Delta fT/2)}{\pi(Bt - \Delta fT/2)} \right] + e^{j\phi} \left[\frac{\sin \pi(Bt + \Delta fT/2)}{\pi(Bt + \Delta fT/2)} \right] \right| \tag{4.5.33}$$

where $t = 0$ is now defined to correspond to the frequency $f_0 = (f_1 + f_2)/2$ and ϕ is the phase difference between the two signals. This function is plotted for $\phi = 0$, $\pi/2$, and π in Fig. 4.5.3, with $\Delta f = 0.89/T$. As we might expect, peak signals corresponding to the two frequencies are easily resolved when $\phi = \pi$; for $\phi = \pi/2$, they are barely resolved; and when $\phi = 0$, there is actually a peak in amplitude at the midway frequency (corresponding to $t = 0$) of value 2. To ensure good resolution of equal-amplitude in-phase signals, we must increase Δf. Calculation shows that in this case, a value of $\Delta fT = 1.33$ is just adequate for resolution of the two frequencies.

A simple way to estimate the optimum value of Δf, required to resolve two in-phase equal-amplitude components, is to have each component be half its maximum amplitude point at the midway (6-dB point). These values are tabulated in Table 4.5.2 for various apodizing functions. They tend to be slightly optimistic,

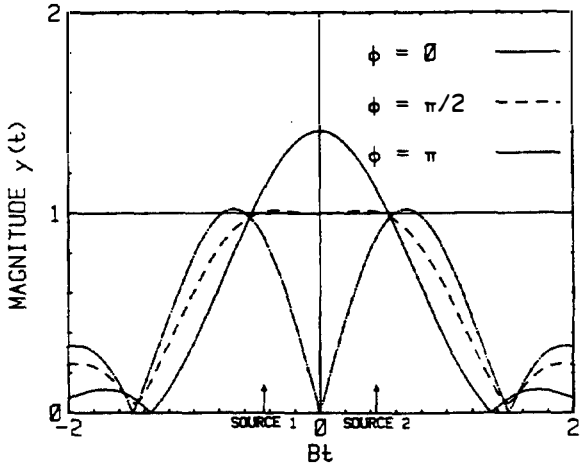


Figure 4.5.3 Plot of the magnitude of the Fourier transform of two equal-amplitude signals separated by a frequency $\Delta f = 0.89/T$, with phase differences $\phi = 0$, $\phi = \pi/2$, and $\phi = \pi$, respectively.

because the observable maximum amplitude of each component tends to be slightly lowered by the out-of-phase term contributed by the other component.

It follows from this treatment that the number of resolvable points in the spectrum is given by the approximate relation

$$N \approx BT \tag{4.5.34}$$

This is an important general result. The number of resolution points in any transversal filter system basically depends on the time–bandwidth product of the system. The exact figure will depend on the relative amplitudes and phases of the frequency components of interest, on the shape of the main lobe, and on the amplitudes of the sidelobes.

We must approach this result with caution when referring to the limitations of a real system, because we are concerned with the time length and bandwidth of both the chirp and the filter. Up to now, we have assumed that the filter is much longer than the chirp. The frequencies present in a single sideband chirp waveform, partially modulated by a frequency Ω, as in Eq. (4.5.30), are $\omega_0 - \mu t$ and $\omega_0 + \Omega - \mu t$. Thus the total bandwidth of the filter is required to be $f + B$, where $f = \Omega/2\pi$ is the signal bandwidth or bandwidth of the modulation and $B = \mu T/2\pi$ is the bandwidth of the chirp. This implies that if f can vary from 0 to B, the entire Fourier transform can only be produced if the filter has a bandwidth $B_d = 2B$ and a time length of $T_d = 2T$. In this case the maximum frequency excursion of the modulation can be B. Therefore, the limitation on the number of resolvable frequencies becomes, in practice,

$$N \approx \frac{B_d T_d}{4} \tag{4.5.35}$$

Unless we make other modifications to the system, such as with the sliding transform, where the input chirp has twice the bandwidth and time length of the filter, as discussed in Sec. 4.5.5, the system must have a much larger bandwidth than that required by the signal alone. As we will see, such stratagems destroy the phase information in the transform, but are useful when only the magnitude of the transform is required.

We first consider applications of this Fourier transform technique to situations where we require only the magnitude or real part of the Fourier transform $|Y(t')|$.

In practice, the Fourier transform of a complex signal requires that the signal be divided into its real and imaginary parts. Real [$\cos(\omega_0 t + \mu t^2/2)$] and imaginary [$\sin(\omega_0 t + \mu t^2/2)$] filters must be used, each having a response for each real and imaginary input signal component. This means that we need four filters in all. To put it another way, extra filters are required to reproduce both the amplitude and phase of the Fourier transform. We shall describe such processes in Secs. 4.5.5 and 4.5.6. If the signal is real and its Fourier transform is also real (a symmetric input waveform), only one real transform is required. Thus all the necessary operations can be obtained by using a premultiplied chirp of the form $\cos(\omega_0 t - \mu t^2/2)$ and a filter of the form $\cos(\omega_0 t + \mu t^2/2)$.

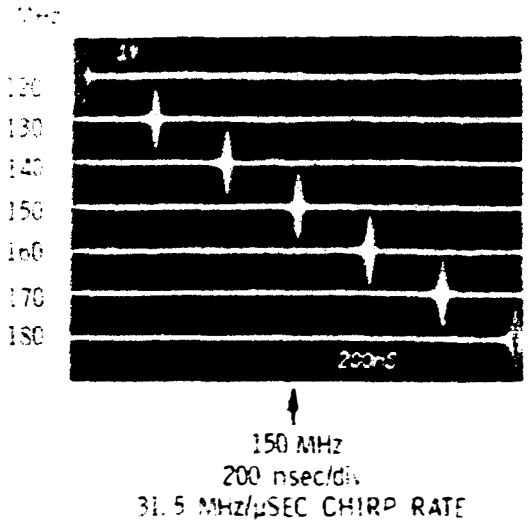


Figure 4.5.4 Prototype chirp transform results for seven successive CW input signals, stepped from 120 to 180 MHz in 10-MHz steps. (After Hays et al. [41].)

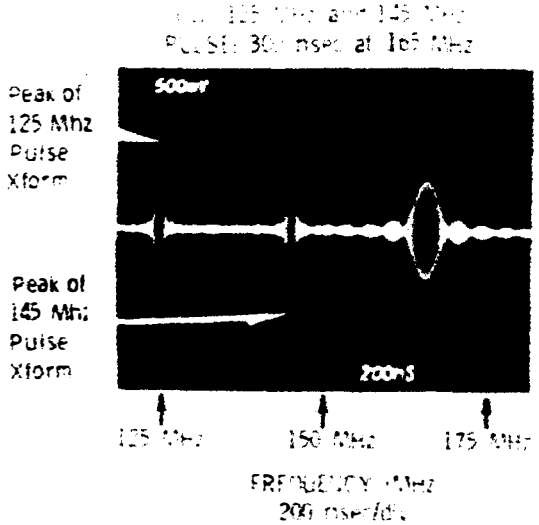


Figure 4.5.5 Chirp transform of three simultaneous input signals, including CW and a tone burst. (After Hays et al. [41].)

Compressive receiver. Perhaps the simplest application of a real transform is to the so-called *compressive receiver*. This device instantaneously detects and determines the frequencies of unknown signals. There are obvious applications to radar detection systems, where only a few RF pulses might be available for this purpose. In this case, an input signal is mixed with a chirp; the chirp can be generated by injecting a pulse into a dispersive delay line. The modulated chirp signal is then passed through the matched filter and into a detector. Thus we obtain an output that is the spectrum of the input modulating signal. This output will occur at a time that is independent of the time of the input signal, provided that the input occurs during the time the chirp is entering the matched filter. Thus the device is capable of interrogating several signals at once and displaying the frequencies of the separate input signals. Results obtained with a system of this nature are shown in Figs. 4.5.4 and 4.5.5.

Variable bandpass–bandstop filters and bandshape filters. A second application of this chirp transform technique is to use the device for variable bandpass–bandstop filtering, as illustrated in Figs. 4.5.6 and 4.5.7. In this case, because a particular frequency in the waveform corresponds to an output signal at a certain time, it is possible to gate the output signal to eliminate a particular frequency. Then by taking the inverse Fourier transform, the original signal may be obtained with the interfering signal filtered out.

More generally still, if we multiply the transformed signal by a pulse that is the Fourier transform of the filter function required, we can design a general filter function system. An illustration of such a filter realization for differentiation of the input signal [multiplying by ω (or by t in the transform domain) corresponds to differentiation] is shown in Fig. 4.5.8.

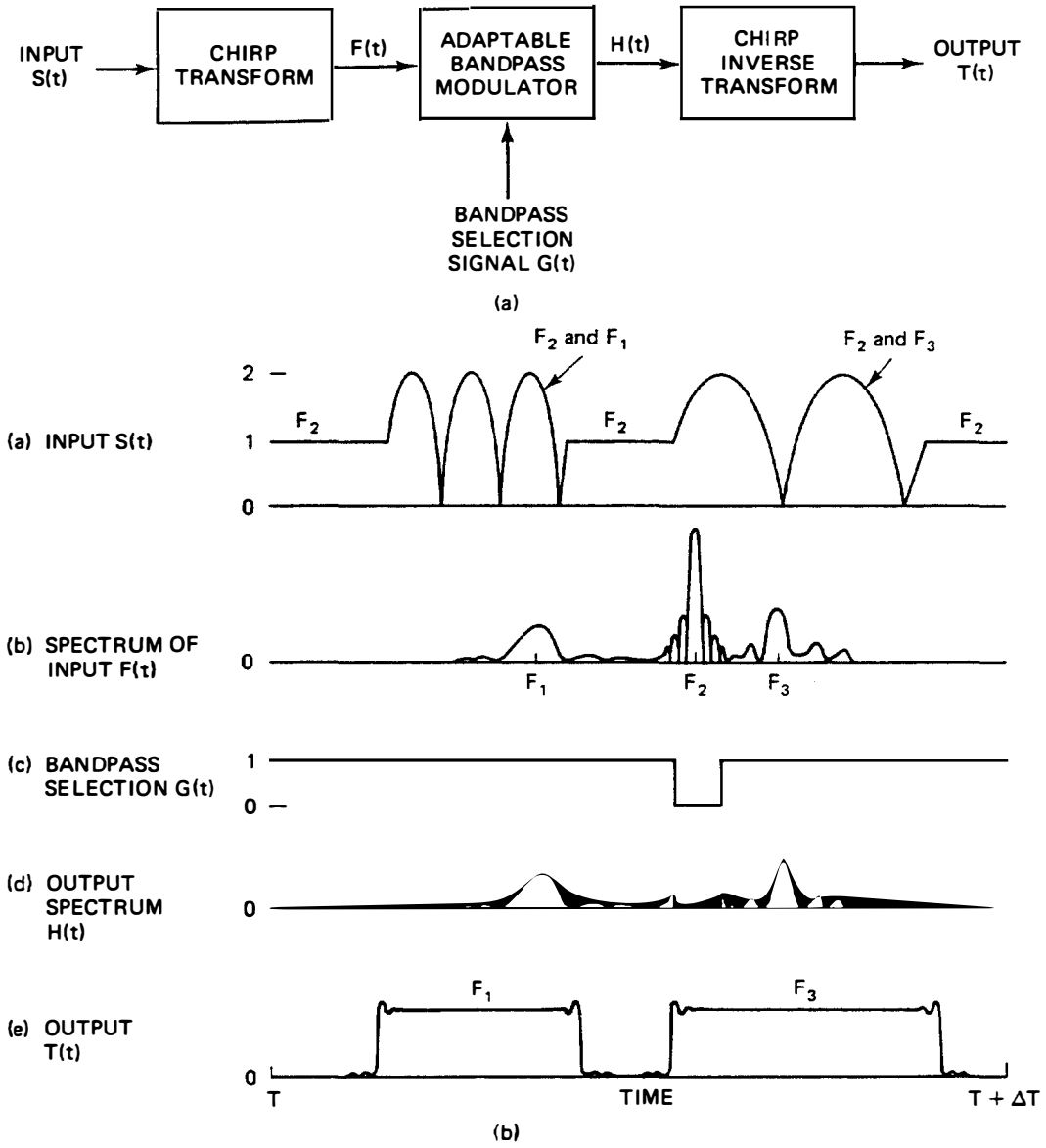


Figure 4.5.6 (a) Bandpass–bandstop filter; (b) chirp transform use for variable bandpass/bandstop filtering. (After Hays et al. [41].)

Variable-time-delay filters. The Fourier transform filter can give a variable time delay. Translation by T_0 of a waveform $f(t)$ results in the multiplication of its Fourier transform by a phase shift term exp $(j\omega T_0)$ or, in our case, exp $(j\mu\tau T_0)$. To put it more simply, if we translate the frequency of the transform by $\omega_T = \mu T_0$, the time of the output signal will change by T_0. Thus all that is needed to vary the delay is to multiply the Fourier transform by a signal of frequency ω_T, and then to find the inverse transform. Results of this kind are illustarted in Fig. 4.5.9.

Correlation or convolution of two signals. A final application is the correlation or convolution of signals. Suppose the two signals are $f(t)$ and $g(t)$.

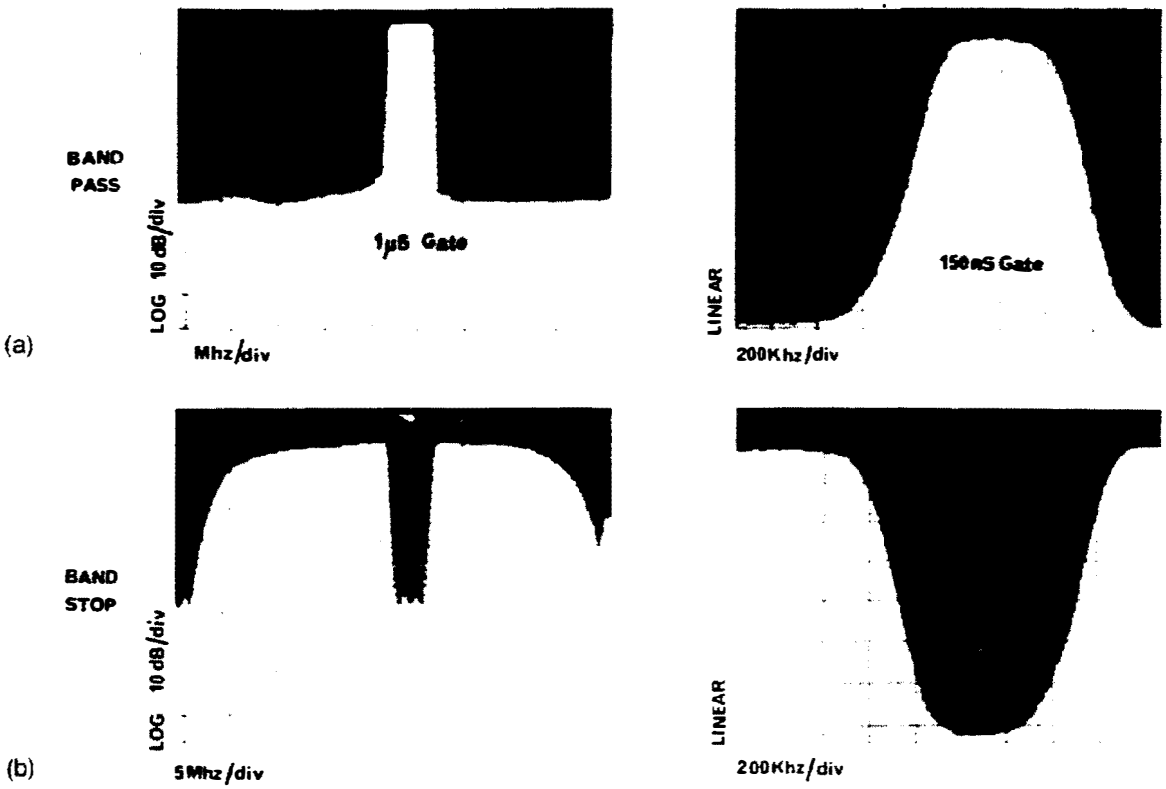


Figure 4.5.7 (a) Wide-gate and narrow-gate passband and stopband results for a time-domain filter transmission as a function of input frequency, with gate fixed (experiment). (After Maines et al. [42].) (b) Lowpass and bandpass filtering of the chirp transform of two different-frequency (225 kHz and 600 kHz) time-overlapped signals. The outputs are still modulated by a chirp factor. (After Atzeni et al. [43].)

Their convolutions are

$$h(t) = \int f(t - \tau)g(\tau)\, d\tau \tag{4.5.36}$$

The Fourier transform of this result is

$$H(\omega) = F(\omega)G(\omega) \tag{4.5.37}$$

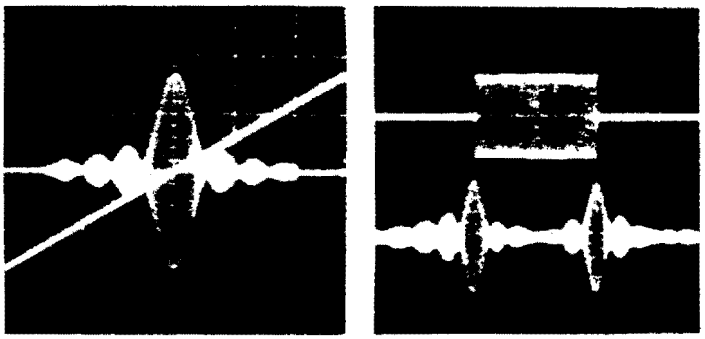
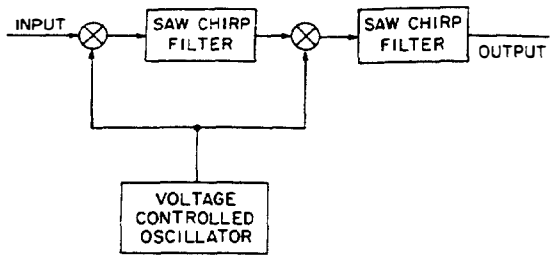


Figure 4.5.8 Linear ramp synthesizing the differentiator transfer function, superimposed to the CT of a rectangular pulse (left) and derivative of the pulse (right). Chirp modulation is not removed. (After Atzeni et al. [43].)



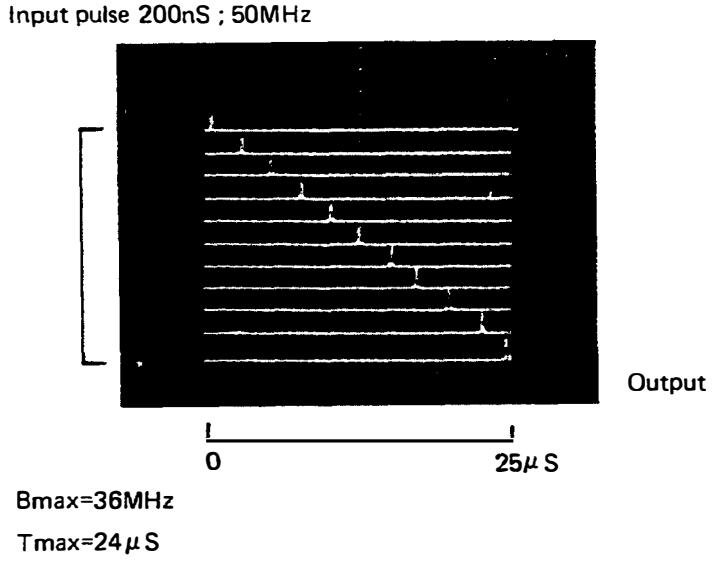


Figure 4.5.9 (a) Schematic of SAW variable delay line using chirp filters. (b) SAW variable-delay-line operation. The input pulse is 200 ns at 50 MHz. The remaining traces show the delayed long 50-MHz outputs. (After Maines and Paige [44].)

Thus if we take the Fourier transforms of the two signals, multiply them together, and then find the inverse Fourier transform of the result, the output will be the convolution of the two signals. Similarly, if we multiply $F(\omega)$ by $G^*(\omega)$, the complex conjugate of $G(\omega)$, we will obtain the correlation of the two signals after finding the inverse transform of the product.

To carry out convolution with chirp transform devices, we simply carry out the Fourier transforms of $f(t)$ and $g(t)$, multiply the Fourier transforms $F(\mu t)$ and $G(\mu t)$ together, and find the inverse Fourier transform of the resultant signal.

Alternatively, as shown in Fig. 4.5.10, if correlation is required, the Fourier transform $G(\mu t)$ can be mixed with a signal $\exp[j(\omega_1 t + \mu t^2/2)]$ + c.c., where c.c. is shorthand for the complex conjugate. The output signal from the mixer has a form

$$S(\mu t) = G(\mu t)e^{j[(\omega_0+\omega_1)t+\mu t^2]} + G^*(\mu t)e^{j(\omega_1-\omega_0)t} + \text{other terms} \qquad (4.5.38)$$

If this signal is passed through a bandpass filter to keep only the latter term, we can obtain the necessary conjugate signal. Results obtained with such a pair of Hamming-weighted chirps and a 127-chirp biphase code are shown in Fig. 4.5.11.

We can also employ the technique to measure the difference in arrival time of two arbitrary signals from the same source. Suppose that we find the autocorrelation of the signals $f(t)$ and $g(t - T)$, from two antennas, for example. The

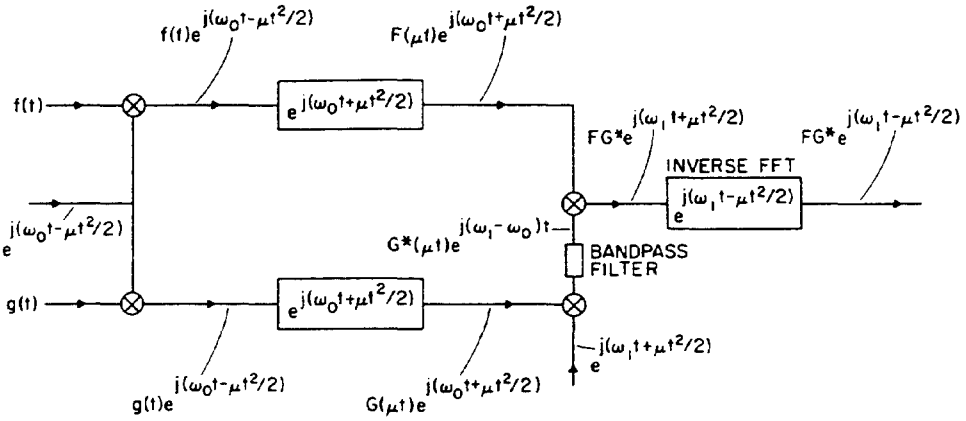


Figure 4.5.10 Circuit required for cross-correlations.

time delay T from a noisy source such as a radio star will vary with the position of the source. When the source is located on a line midway between the antennae, $T = 0$. The correlation will be $\int f(\tau)f(\tau + t - T)\,d\tau$. The output correlation peak occurs at a time $t = T$. Thus we can measure the time delay and the inclination of the source. Figure 4.5.12 shows a result of this kind.

Spectral whitening and nonlinear processing. A third application of these techniques is to eliminate an interfering CW signal, as illustrated in Fig. 4.5.13. A CW signal will give rise to a large peak in the Fourier transform. Therefore, by clipping the output of the Fourier transform processor to limit the amplitude of this peak and then taking the inverse transform of the resultant signal, we can, to a large extent, eliminate the interfering signal. In practice, this can give a reduction of as much as 40 dB in an interfering signal.

4.5.5 Implementation of Chirp *z* Transforms with SAW Devices

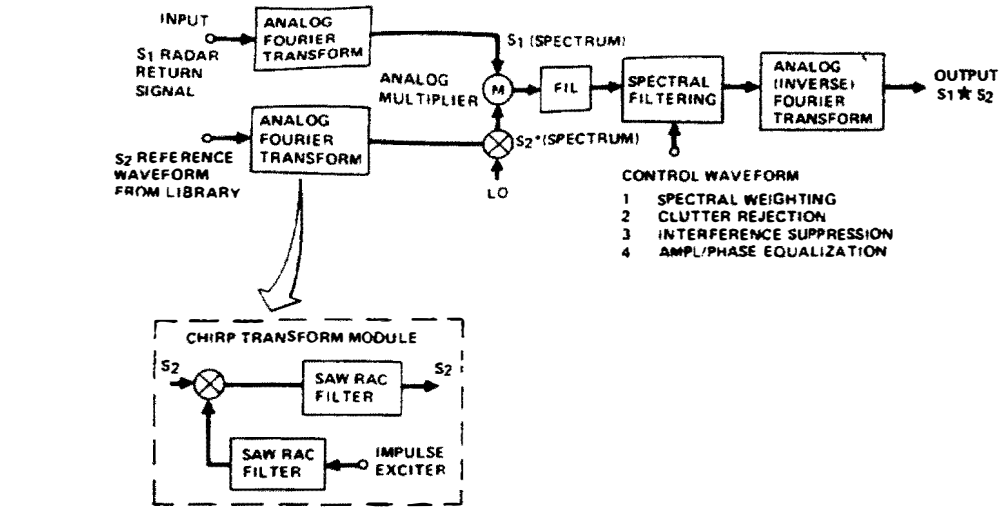
Here we describe the implementation of complex Fourier transform processes. We call these transforms *chirp z transforms* because of their close relation to digitally implemented sampled data transforms. The relation to z transform theory will be discussed in Sec. 4.5.6. Before doing this, however, let us discuss why it is necessary to employ separate transform processors to process the real and imaginary parts of an input waveform, and the advantages of such processors, even for processing real input waveforms.

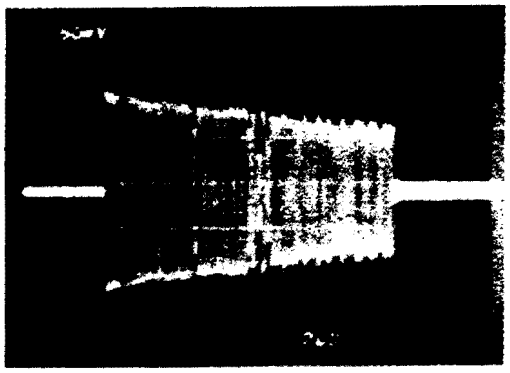
A complex function $f(t)$ can be represented by two real components $f_R(t)$ and $f_I(t)$, with

$$f(t) = f_R(t) + jf_I(t) = a(t)e^{j\phi(t)} \tag{4.5.39}$$

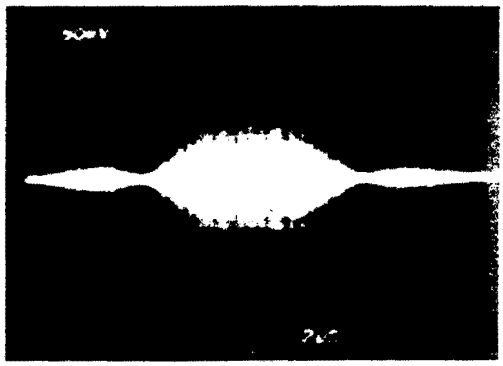
where

$$\begin{aligned} f_R(t) &= a(t)\cos[\phi(t)] \\ f_I(t) &= a(t)\sin[\phi(t)] \end{aligned} \tag{4.5.40}$$

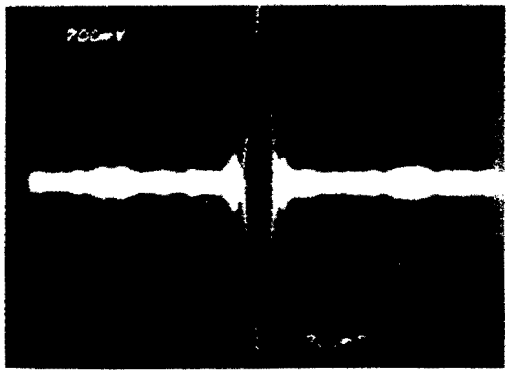




127 CHIP BIPHASE CODED WAVEFORM (0.1 μSEC CHIPS)



CHIRP TRANSFORM OF BIPHASE CODE HORIZ SCALE → 2.0 MHz/μSEC



CORRELATION PULSE SIDELOBE LEVEL = -23 dB

Figure 4.5.11 Autocorrelation of two signals: (a) block diagram of the system used. (b) correlation of a 127-chip biphase-coded waveform. (After Gerard et al. [45].)

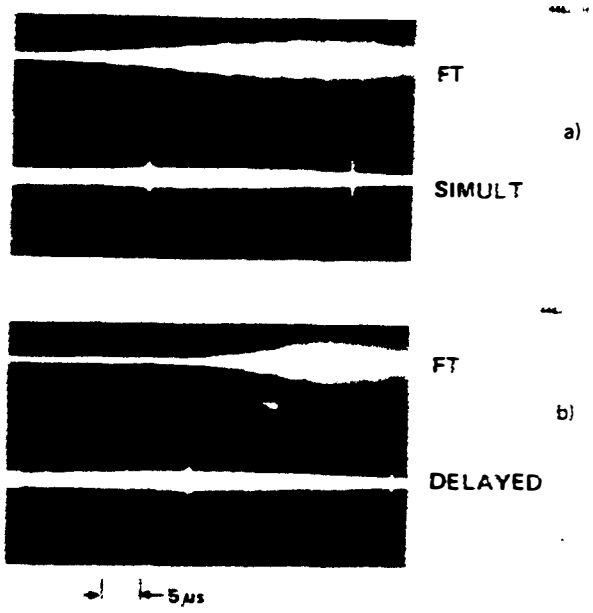


Figure 4.5.12 Relative-time-of-arrival determination. (After Nudd and Otto [46].)

Both the real and imaginary parts of $f(t)$ are needed if we wish to obtain the amplitude and phase of its Fourier transform.

Conversely, if we wish to take the inverse Fourier transform of a function and reconstruct the original signal, we must know both the amplitude and phase of the function.

Consider, as an example, a rectangular pulse with a time duration T. This has a real Fourier transform $\sin(\omega T/2)/(\omega T/2)$. However, if the pulse were passed through a delay line with a time delay T_D, its Fourier transform would be $\exp(-j\omega T_D)\sin(\omega T/2)/(\omega T/2)$, with

$$F_R = \cos \omega T_D \sin \frac{(\omega T/2)}{(\omega T/2)}$$
$$F_I = -\sin \omega T_D \sin \frac{(\omega T/2)}{(\omega T/2)} \tag{4.5.41}$$

It is apparent that in order to reconstruct the delayed pulse using an inverse Fourier transform, we would have to know both F_R and F_I, although in this case $f(t)$ is real. In general, we must be able to deal with complex signals or the amplitude

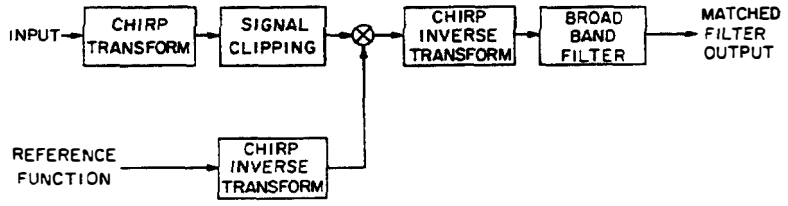


Figure 4.5.13 Chirp transform programmable matched filter, with suppression of narrowband interference by transform clipping to approximate spectral pre-whitening.

and phase of real signals, because the output of a real-time Fourier transform device itself is one or more time-varying signals.

Now let us consider how to process an input signal $f(t)$ modulated by a down-chirp $\exp[j(\omega t - \mu t^2/2)]$, which we write in the form

$$u(t) = f(t)e^{j(\omega t - \mu t^2/2)} \tag{4.5.42}$$

where, from Eqs. (4.5.39) and (4.5.42), the real physical signal is

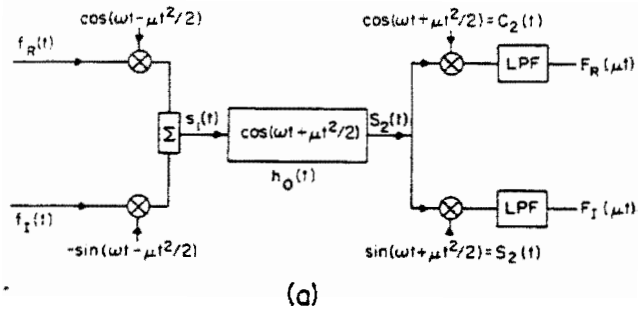
$$s_1(t) = f_R(t)\cos\left(\omega t - \frac{\mu t^2}{2}\right) - f_I(t)\sin\left(\omega t - \frac{\mu t^2}{2}\right) \tag{4.5.43}$$

Thus we can obtain the physical signal required by multiplying $f_R(t)$ and $f_I(t)$ by a cosine and sine chirp, respectively, and adding the resultant signals, as shown in Fig. 4.5.14.

The matched filter for the complex chirp is of the form $\exp[j(\omega t + \mu t^2/2)]$. Here we use a physical matched filter of the form $\cos(\omega t + \mu t^2/2)$. The output from the system, when a signal $u(t)$ is inserted into it, can be written in the form

$$\begin{aligned} S_2(t) &= \int f(\tau)e^{j(\omega\tau - \mu\tau^2/2)}\cos\left[\left[\omega(t-\tau) + \frac{\mu(t-\tau)^2}{2}\right]\right] d\tau \\ &= \tfrac{1}{2}\int f(\tau)\left[e^{j(\omega t + \mu t^2/2)}e^{-j\mu t\tau} + e^{-j(\omega t + \mu t^2/2)}e^{j\mu(t-\tau)\tau}e^{2j\omega\tau}\right] d\tau \end{aligned} \tag{4.5.44}$$

The second term in the integrand contains a rapidly varying term that varies as $\exp(-j\mu\tau^2)$. This term will yield very little contribution to the integrand and may



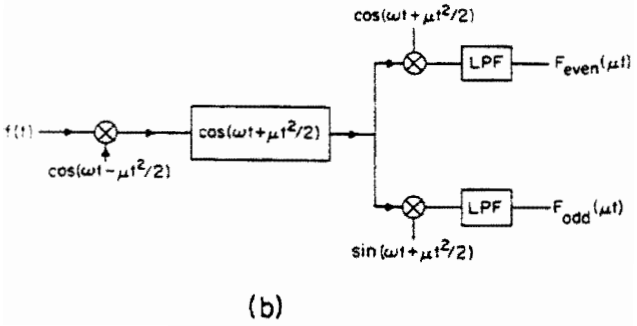


Figure 4.5.14 (a) Fourier transform processor for a complex function. The real input is represented by $f_R(t)$, the imaginary input is represented by $f_I(t)$, and F_R and F_I are the real and imaginary parts of the Fourier transform. (b) Fourier transform processor for odd and even real functions.

be neglected. Therefore, we write

$$S_2(t) \approx \tfrac{1}{2} e^{j(\omega t + \mu t^2/2)} \int f(\tau) e^{-j\mu t\tau}\, d\tau \tag{4.5.45}$$

Thus the output from this signal processing operation is the Fourier transform of $f(\tau)$ multiplied by an up-chirp.

We are interested in the real and imaginary parts of this Fourier transform, which are

$$\begin{aligned} F_R(\mu t) &= \text{Re} \int f(\tau) e^{-j\mu t\tau}\, d\tau \\ &= \int f_R(\tau) \cos \mu t\tau\, d\tau + \int f_I(\tau) \sin \mu t\tau\, d\tau \\ F_I(\mu t) &= \text{Im} \int f(\tau) e^{-j\mu t\tau}\, d\tau \\ &= \int f_I(\tau) \cos \mu t\tau\, d\tau - \int f_R(\tau) \sin \mu t\tau\, d\tau \end{aligned} \tag{4.5.46}$$

respectively, where

$$\text{Re}[S_2(t)] = F_R(\mu t) \cos\left(\omega t + \frac{\mu t^2}{2}\right) - F_I(t) \sin\left(\omega t + \frac{\mu t^2}{2}\right) \tag{4.5.47}$$

If $f(t)$ is real, we need only multiply the original signal by a cosine chirp, as shown in Fig. 4.5.14(a), to obtain the result

$$\begin{aligned} F_R(\mu t) &= \int f_R(\tau) \cos \mu t\tau\, d\tau \\ F_I(\mu t) &= \int f_R(\tau) \sin \mu t\tau\, d\tau \end{aligned} \tag{4.5.48}$$

In this case, $F_R(\mu t) = F_{\text{even}}$ is the Fourier transform of the even part of $f_R(t)$, and $F_I(\mu t)$ is the Fourier transform of the odd part of $f_R(t)$.

Suppose now that we multiply the output $S_2(t)$ of the filter by a chirp signal of the form $\cos(\omega t + \mu t^2/2 + \phi)$, and pass the resultant output signal of the mixer through a low-pass filter (i.e., look at the output at baseband). The term of interest, then, is

$$\text{Re}\,(Y_{\text{out}}) = \cos\phi\, F_R(\mu t) + \sin\phi\, F_I(\mu t) \tag{4.5.49}$$

By choosing $\phi = 0$ or $\pi/2$ [i.e., by multiplying by $\cos(\omega t + \mu t^2/2)$ or $\sin(\omega t + \mu t^2/2)$], we can separate the real and imaginary parts of $F(\mu t)$, as shown in Fig. 4.5.14(a).

Thus the output from the top filter in Fig. 4.5.14(a) is the real part of the complex Fourier transform $F_R(\mu t)$, and the output from the bottom filter is the imaginary part $F_I(\mu t)$ of the complex Fourier transform of $f(t)$. When $f(t)$ is real, $F_R(\mu t)$ is the transform of the even part of $f(t)$ and $F_I(\mu t)$ is the transform of the odd part of $f(t)$, as shown in Fig. 4.5.14(b). When $f(t)$ is complex, it follows from

Eq. (4.5.46) that $F_R(\mu t)$ can be stated in terms of Fourier transforms of $f_R(t)$ and $f_I(t)$.

An example of how this type of SAW processor is operated has been given by Jack and Collins. The following quote (slightly modified) is from their paper "Fast Fourier Transform Processor Based on the SAW Chirp Transform Algorithm" (slight modifications have been made to correspond with our earlier examples) [47].

Experimental SAW FFT Processor Realization. The SAW FFT processor shown in schematic form in Figs. 4.5.14(a) and 4.5.15 was configured to demonstrate computation of the Fourier transform in accordance with the theory of the previous section. The processor employs three commercially available SAW chirp filters (MESL types WB041 B-2 and WB045B-2). Here the input chirp C_1 is centered at 22 MHz and sweeps over 4 MHz bandwidth in 25 μs (TB = 100, dispersive slope $\mu = 2\pi \times 160$ kHz/μs). The chirp filter C_2 and the post-multiplier chirp C_3 are centered at 32 MHz and sweep over 8 MHz in 50 μs (TB = 400, $\mu = 2\pi \times 160$ kHz/μs). In view of the offset center frequencies of C_1, C_2 a 54-MHz reference carrier must be inserted at the input to permit acceptance of baseband input signals. However, since the center frequencies of C_2, C_3 are identical, the output—after the required low-pass filtering discussed previously—is automatically at baseband. The timing electronics associated with the SAW FFT processor are arranged to operate in synchronism with the inserted 54-MHz reference carrier using divider circuits. Pre- and post-multiplier chirps in phase quadrature

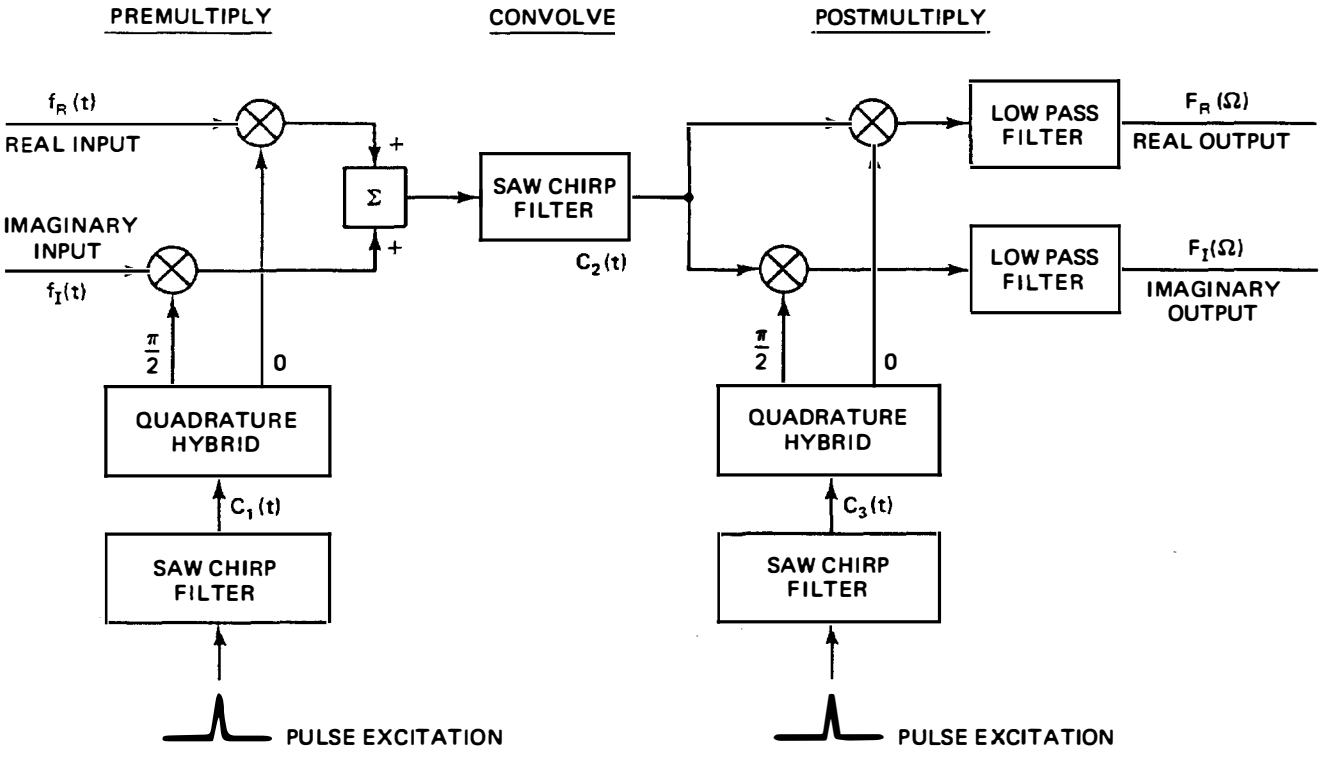


Figure 4.5.15 SAW chirp transform arrangement for baseband operation.

were derived from C_1 and C_3, respectively, by means of a commercial quadrature hybrid circuit (Merrimac type QHS-3-30). The input waveforms used in this experiment were purely real, analog and continuous. These waveforms were derived from a function generator (Tektronix type FG504), which permits synchronous operation (i.e., stationary relative to the premultiplier chirp) with variable input signal phase.

The results shown in Fig. 4.5.16(a) relate to a *real* and *even* (cosine) input, with period 2.08 μs (480 kHz). Choice of such an input period (a basic vector) means that an integer number of cycles of the input waveform, fit within the time window defined by the premultiplier chirp. The center and bottom traces represent the real, $F_R(\Omega)$, and imaginary, $F_I(\Omega)$, components, respectively, of the Fourier transform as measured at the outputs of the SAW FFT processor. $F_R(\Omega)$ and $F_I(\Omega)$ represent frequency-domain functions over both positive and negative regions of the frequency domain relative to zero frequency (dc). The horizontal scale for these traces can be calculated as 160 kHz/μs × 1 μs/div = 160 kHz/div. The center and bottom traces share a common oscilloscope delaying time base. The delay has been adjusted such that the center of the display corresponds to zero frequency (dc). As stated earlier, the upper trace in Fig. 4.5.16(a) represents the input time-domain signal. Here in view of the propagation delay through the SAW processor, it is not possible to show input and outputs simultaneously. How-

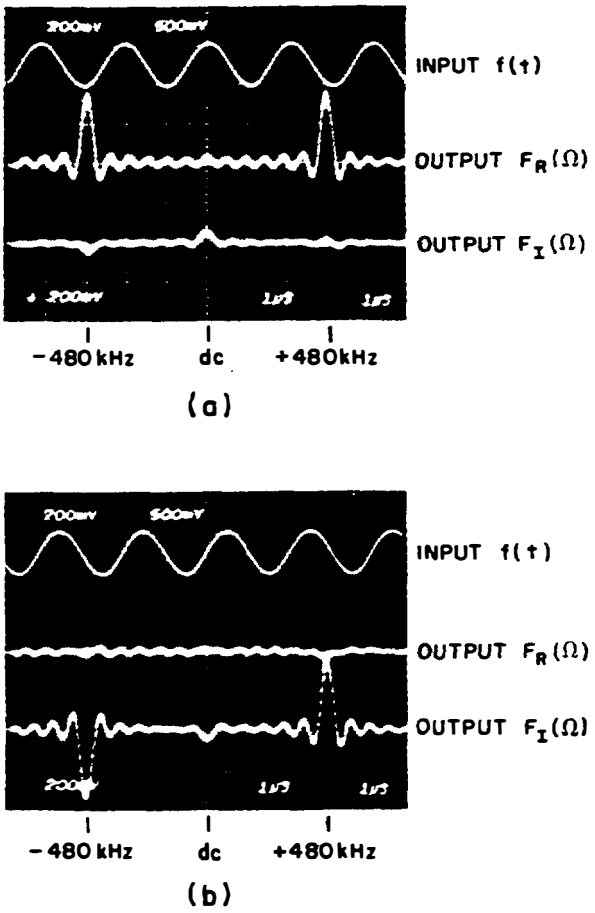


Figure 4.5.16 SAW FFT performance with (a) 480-kHz cosine input, and (b) 480-kHz sine input. (After Jack and Collins [47].)

ever, the upper trace uses a separate oscilloscope delaying time base where the (arbitrary but fixed) delay has been set such that the time-domain signal, $f(t)$, in Fig. 4.5.16(a) is *even* relative to the center of the trace.

The upper trace of Fig. 4.5.16(b) shows a *real* and *odd* (sine) input, with period 2.08 μs (480 kHz). It is important to note that the photographs of Fig. 4.5.16 were taken with the two time-base delays unchanged. The photographs confirm the presence of cosine and sine inputs, respectively, these being produced by variation of the start phase of the function generator output.

The output Fourier transform of the (truncated) cosine input (480 kHz), $f(t)$, shown in Fig. 4.5.16(a) displays *real* and *even* responses as can be predicted theoretically. These responses peak at ± 3 μs $\times$ 160 kHz/μs $= \pm 480$ kHz and consist of $(\sin x)/x$ responses with nulls at spacings of 250 ns $\times$ 160 kHz/μs. This structure can be directly related to signal truncation by the input window of 25-μs duration. In comparison, the output Fourier transform of the (truncated) sine input (480 kHz) as shown in Fig. 4.5.16(b), displays *imaginary* and *odd* responses at ± 3 μs $\times$ 160 kHz/μs $= \pm 480$ kHz. The unwanted signals visible in the respective "zero" output components of Fig. 4.5.16 have been investigated, and these appear to be due predominantly to the practical difficulty of ensuring the existence of a purely even or purely odd input.

A very similar set of results obtained with a larger bandwidth system have been given by Jack and Paige [48], who demonstrated the Fourier transform of 7.5-MHz signals.

4.5.6 Chirp *z* Transform with a CCD

A baseband system cannot be used in the same manner as an SAW chirp z-transform system. In the latter case, we could use low-pass filters and a high carrier frequency to eliminate unwanted components of the output. However, a baseband system must employ separate chirp filters to obtain the complex Fourier transform. Consider, for example, the implementation of the same process in a CCD that operates at baseband. The system may now be regarded as a sampled system where we can write the discrete Fourier transform in the form

$$\begin{aligned} F_k &= \sum_{n=0}^{n=N-1} f_n e^{-2jn\pi k/N} \\ &= \sum_{n=0}^{n=N-1} f_n z^{2n} \end{aligned} \tag{4.5.50}$$

where $z = \exp[-(j\pi k/N)]$.

Using the substitution

$$2nk = n^2 + k^2 - (n - k)^2 \tag{4.5.51}$$

we find that

$$F_k = e^{-j\pi k^2/N} \sum_{n=0}^{n=N-1} f_n e^{-j\pi n^2/N} e^{j\pi(k-n)^2/N}$$
$$= z^k \sum_{0}^{n=N-1} f_n z^{n^2/k} z^{-(k-n)^2/k} \tag{4.5.52}$$

This operation of the chirp z transform is identical to that described earlier for a CW system in Eqs. (4.5.7) and (4.5.16). In this case, the input signal is multiplied by a sampled baseband chirp exp $(-j\mu t^2/2)$, where for the nth sample, $\mu t_n^2 = 2\pi n^2/N$. The resultant signal is then inserted into a chirp filter with the opposite slope and finally the output is multiplied by a chirp. As with SAW devices, final multiplication by a chirp can be eliminated only if the power density spectrum is required. To process this signal, we require four filters, as illustrated in the block diagram in Fig. 4.5.17. In this case, the input data would be premultiplied by a chirp of a frequency varying from $-f_1/2$ to $f_1/2$, with a duration T.

The expression in Eq. (4.5.52) shows that the filter, which has an exp $[j\pi(k - n)^2/N]$ response, must have $2N$ stages, that is, the maximum value of $k + n$, corresponding to a delay length $2T$ and a frequency variation from $-f_1$ to f_1. This ensures that all frequency components of the input, multiplied by the chirp, can be processed by the filter, and that the chirp is present in the filter at all times while the processing is being carried out. This requirement, of course, is identical to that of the SAW device, the only difference being that the chirp frequency varies from $f_0 - f_1/2$ to $f_0 + f_1/2$, where f_0 is the carrier frequency. As the chirp length is shorter than that of the filter, we must weight the chirp rather than the filter.

Provided that these processes are carried out, we can obtain, with a post-multiplier, the correct component of the transform (i.e., real or imaginary, even or odd). In principle, it is not always necessary to use four filters, because if the input signal were purely real, we could use only two filters. The advantage of a

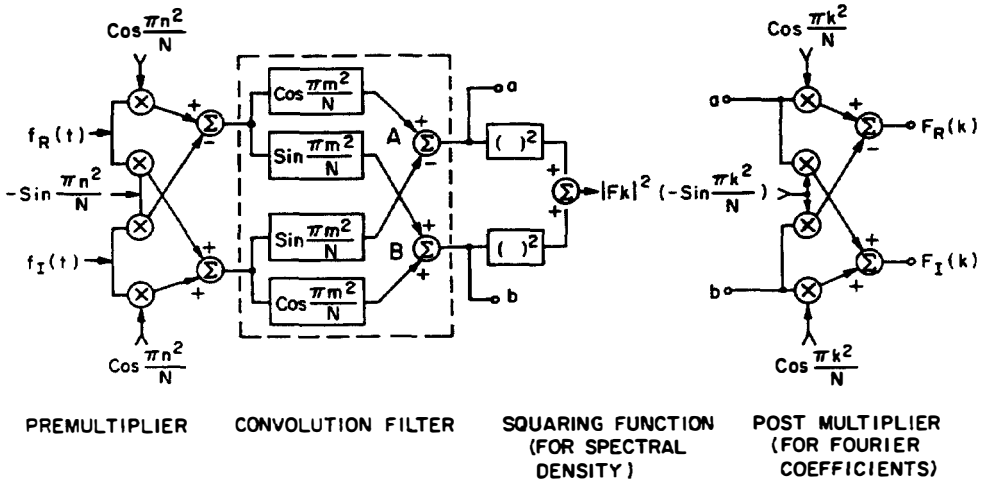


Figure 4.5.17 Implementation of FFT chirp transform or for obtaining spectral-density.

four-filter system, however, is that it can eliminate the sidelobes that arise from unwanted signal components, which correspond to the second term in the second equation of Eq. (4.5.44), and which we normally ignore.

Another way to look at this result might be to consider what happens when the input is a single-frequency exp ($j\Omega t$), so that $f_R(t) = \cos \Omega t$ and $f_I(t) = \sin \Omega t$. When these signals are premultiplied by the chirps cos ($\mu t^2/2$) and $-\sin(\mu t^2/2)$, respectively, and the resultants subtracted, we obtain a single-sideband chirp cos ($\Omega t - \mu t^2/2$) at the input to the upper set of filters in Fig. 4.5.17. Similarly, the lower set of filters has an input sin ($\Omega t - \mu t^2/2$). Therefore, the output from the summer A in the diagram is

$$\begin{aligned} Y_U(t) &= \int \left\{ \cos\left(\Omega\tau - \frac{\mu\tau^2}{2}\right) \cos\left[\Omega(t-\tau) + \frac{\mu(t-\tau)^2}{2}\right] \right. \\ &\quad \left. - \sin\left(\Omega T - \frac{\mu\tau^2}{2}\right) \sin\left[\Omega(t-\tau) + \frac{\mu(t-\tau)^2}{2}\right] \right\} d\tau \qquad (4.5.53) \\ &= \int \cos\left(\Omega t + \frac{\mu t^2}{2} - \mu t\tau\right) d\tau \end{aligned}$$

We note that there are no terms in the argument of the cosine function that vary as τ^2. Thus the equivalent of the second term of the integrand of the second equation of Eq. (4.5.44) has been eliminated in the output of the filter. A similar argument applies to the further processing through the post-multiplier section. The use of four multipliers yields the one required sideband without further filtering.

Furthermore, as real input components are composed of two complex components, one of positive frequency and one of negative frequency, the use of real filters alone, even for a real input signal, can give rise to aliasing. Each real input component, such as the signal cos Ωt, is composed of complex components exp ($j\Omega t$) and exp ($-j\Omega t$), whose imaginary parts cancel. If these signals are sampled over a range 0 to $f_s/2$, where f_s is the sampling frequency, negative frequency components occur in the band f_s to $f_s/2$ as an alias, thus restricting the useful input bands to components lying below the Nyquist limit $f_s/2$. However, a complex four-channel system will eliminate these alias terms and make it possible to work over a band to f_s without aliasing, for $2f_s$ samples are effectively employed per second. Another way to look at this is as a consequence of the fact that we are using a single-sideband chirp modulator, and are therefore eliminating the aliasing frequencies.

The CCD system normally operates with split-gate sampled weighting of the type already described [4]. For instance, the weighting amplitude would be cos $\pi k^2/N$ for the cosine chirp, as illustrated in Fig. 4.5.18.

Sliding transform. When only the power spectrum is required, it is often more convenient to employ longer chirps relative to the length of the filter. If the chirp length is longer than that of the filter, only a portion of the chirp waveform

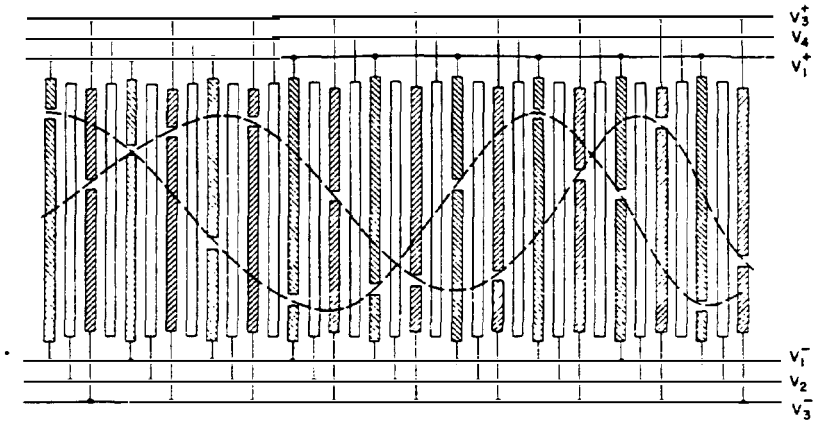


Figure 4.5.18 Layout of the CCD gates showing the split-gate weighting. (From Reticon Coroporation [49].)

is being sampled at any one time. The output from the filter then takes the form

$$F_k = \sum_{n=k}^{k+N-1} f_n e^{-2j\pi nk/N} = e^{-2j\pi k^2/N} \sum_{n=0}^{N-1} f_{n+k} e^{-2j\pi nk/N} \tag{4.5.54}$$

This is essentially equivalent to the continuous form of Eq. (4.5.17) for a matched filter response when the chirp is longer than the filter. In this case, the input chirp is modulated by $f(t)$, a down-chirp is used, and increasing n in this case is equivalent to decreasing z. The difference between this transform and the usual DFT is that the input data are shifted by one sample each time the spectral component is calculated. For a general waveform, this procedure destroys the phase information. However, because the power spectrum does not contain any phase information, this is unimportant. Now we can use a filter N registers long, with a bandwidth f_1, to process signals with the same bandwidth. The system is thus more efficient than one employing a chirp only half the length of the total filter. Furthermore, as we saw in our treatment of weighting (Sec. 4.5.3), its major advantage is that in order to decrease the sidelobe level, the output filter can be weighted instead of the chirp.

4.5.7 Chirp Transforms with Superconductive Delay Lines

Research has been and is being carried out on other technologies for transversal filters. Chirp transforms were demonstrated many years ago using dispersive wire or tape delay lines [1]. (Applications of fiber-optic delay lines are described in Sec. 4.6.) Here we discuss early research on superconductive delay lines, and demonstrate their use for carrying out chirp transforms [12].

So far, the application of superconductive delay lines to carry out chirp transforms has extended the operating frequency range of transversal filters from the UHF up into the range 2 to 20 GHz. The device concepts are very similar to

those of SAW delay lines, but the implementation must, of course, be changed to suit the particular attributes of superconductive delay lines.

The basic idea is to employ an electromagnetic delay line, instead of an SAW delay line, as a transversal filter. Since the velocity of an electromagnetic wave is approximately 10^5 times larger than that of an acoustic wave, the obtainable delays of an electromagnetic delay line will be far less than for an acoustic device. Thus if we want to obtain large delays of the order of 100 ns or more, we require an electromagnetic delay line at least 30 m long. But such delays lines are extremely bulky and their attenuation becomes very large as their diameter is reduced. To circumvent these difficulties, we may use superconductive materials, such as niobium, laid down on a high-quality, low-loss insulating substrate, such as sapphire or silicon, which will be an insulator at temperatures of the order of 4°K. By reducing the width of a strip of superconductive material to dimensions of the order of 25 μm, a relatively small line can be laid down in the form of a spiral, as illustrated in Fig. 4.5.19(a), or a meander line, as illustrated in Fig. 4.5.19(b). The attenuation as a function of distance along the line l is $\exp(-\alpha l)$, where

$$\alpha = \alpha_d + \alpha_c + \alpha_r \tag{4.5.55}$$

The parameters α_d, α_c, and α_r are the contributions of dielectric loss, conductor loss, and radiation losses, respectively. The dielectric loss for sapphire is very small and can be written in the form

$$\alpha_d = 27.3\sqrt{\varepsilon}\,(\tan\delta)/\lambda_0 \qquad \text{(dB/unit length)} \tag{4.5.56}$$

Reported values for the loss tangent $\tan\delta$ of crystalline sapphire range from 10^{-5} to 10^{-8}. The surface resistance of niobium at 4.2°K equals 2.6×10^{-5} Ω/square

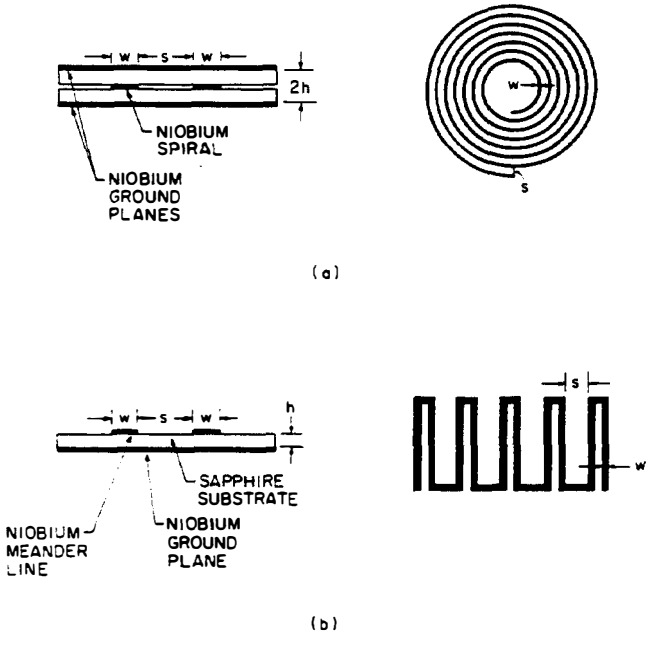


Figure 4.5.19 Superconductive delay lines: (a) spiral microstrip; (b) meander line.

at 10 GHz. The conductor loss α_c for strip line is

$$\alpha_c = \frac{8.68(R_{s1} + R_{s2})}{2wz_0} \quad \text{(dB/unit length)} \tag{4.5.57}$$

where R_{s1} and R_{s2} are the surface resistances of the strip and the neighboring ground planes, respectively. The loss per wavelength for a microstrip design with a strip width w and sapphire thickness h, with $w/h = 0.85$ and $h = 25$ μm, is plotted as a function of frequency, as shown in Fig. 4.5.20. As we might expect, this loss figure is considerably lower than for copper, and in fact the attenuation per wavelength is much lower than for an SAW delay line.

A simple delay line can be made out of a single spiral. To construct a tapped delay line, two interleaved spirals are laid down on the same substrate to make a double spiral dispersive delay line, as shown in Figs. 4.5.21 and 4.5.22. In Fig. 4.5.21, the input signal on electrode 1 to the first spiral is coupled out to excite a wave propagating in the opposite direction on the second. The coupling is between the wider regions, where the two spirals are close to each other.

The action of this coupler can be described by coupled wave theory (see Prob. 6). The coupling of two neighboring TEM modes propagating in the same direction

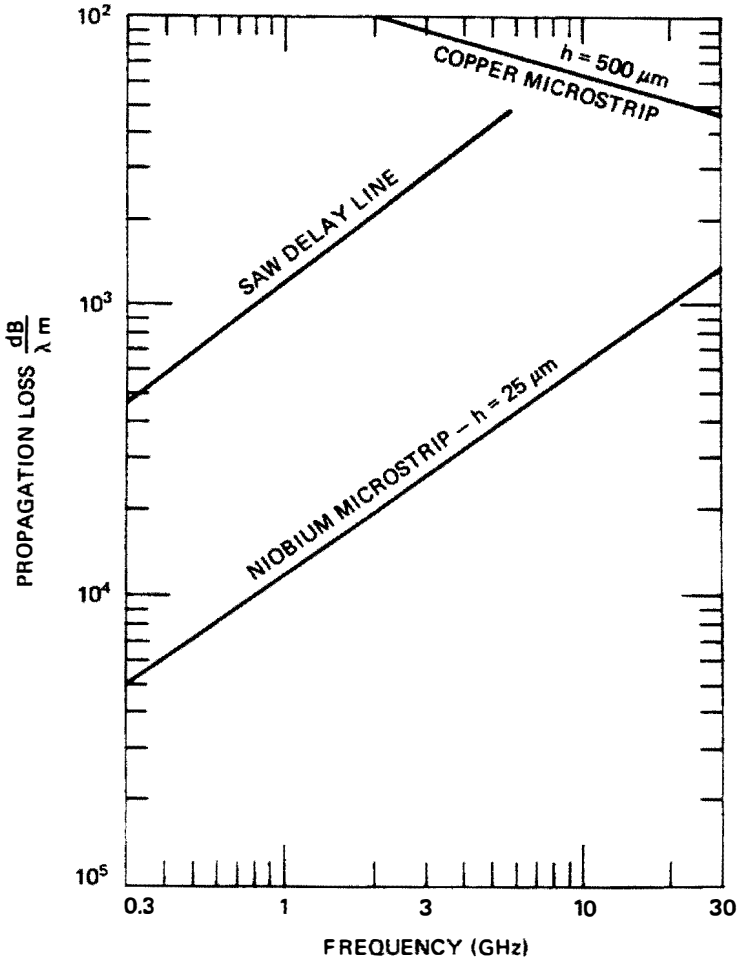


Figure 4.5.20 Propagation loss per wavelength for an SAW delay line and a superconductive delay line. (After Reible [12].)

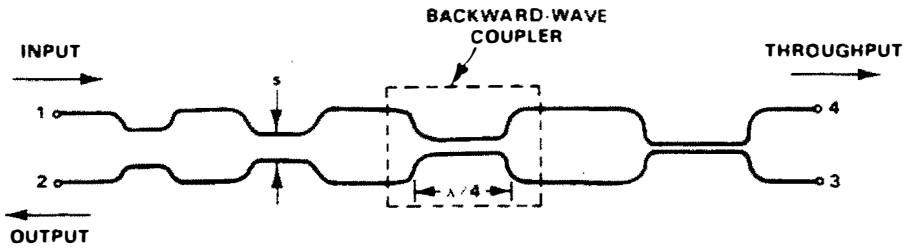


Figure 4.5.21 Proximity-tapped delay line. (After Reible [12].)

tends to be small because the inductive and capacitive coupling terms cancel each other out. The inductive and capacitive coupling from one line to another, of two waves propagating in opposite directions, adds. But the total excitation would also be very small if the coupling were over a long length. If the coupling is regarded as uniform over a length L, it can be shown that optimum coupling is obtained from one line to another when $L = \lambda/4$, where λ is the wavelength of the waves in both spirals. In early work, Reible made a delay line of this kind that was highly dispersive. It gave 27 ns of dispersion (i.e., of delay change) over a 2-GHz bandwidth centered at 4 GHz. The device had 101 proximity taps, with the potential for a time–bandwidth product of 50. The strip-line structure consisted of a 0.2-μm-thick niobium film sandwiched between two 5-cm-diameter, 125-μm-thick sapphire wafers. The pattern had two parallel lines, with a total length of 25 m, wound in a spiral pattern. Figure 4.5.23 shows the results for one of these devices used as an expander, when a 200-mV dc step with a 25-ps rise time is applied to its input. The resulting 27-ns chirp was amplified and time-gated to produce the pulse shown in the lower figure of Fig. 4.5.23(a). This was supplied to the input of the compressor, the up-chirp device. The resulting compressed pulse is displayed in the upper trace of Fig. 4.5.23(a) and a version expanded in time is shown in Fig. 4.5.23(b). The width of the main lobe of the compressed pulse is 2.5 ns, exactly what would be expected from an input with an 800-MHz bandwidth. The first three sidelobes levels were in good agreement with what would be expected from the theoretical sinc response.

This demonstration of a superconductive delay line is of great interest because it can obtain such large bandwidths. Its disadvantage is the low temperature required for its operation.

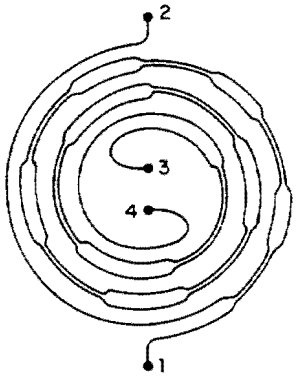


Figure 4.5.22 Double-spiral dispersive delay line. (After Reible [12].)

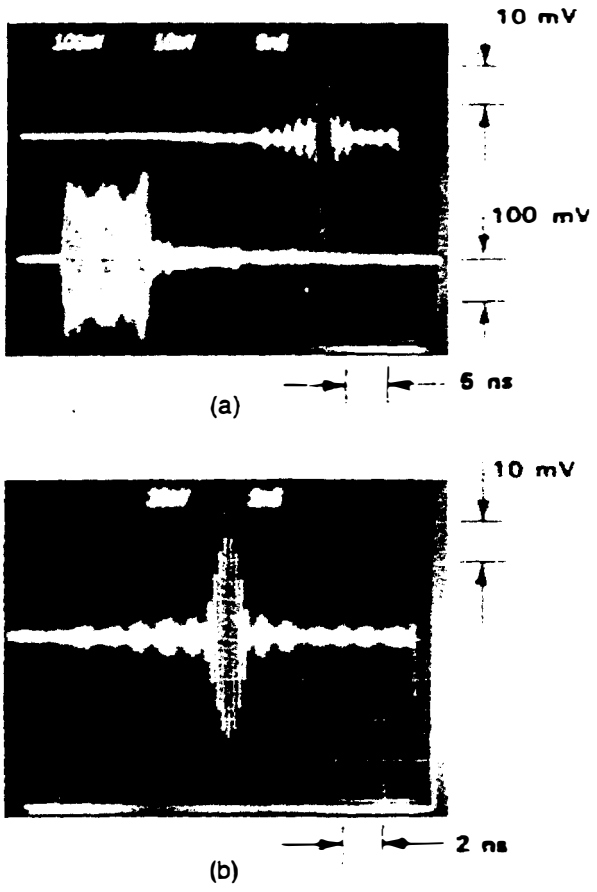


Figure 4.5.23 Outputs of superconductive dispersive filters: (a) top trace shows compressed pulse output, while bottom trace shows expanded waveform; (b) compressed pulse with expanded scale. (After Reible [12].)

PROBLEM SET 4.5

1. Consider Eq. (4.5.12) for a signal entering a physical filter. Following the derivation of Eq. (4.5.17), consider the FM chirp signal to be of finite length T, rather than infinitely long, as has been implicitly assumed in this derivation, and the length L of the delay line to be finite, with $L = VT$. Point z on the delay line can be excited by a signal only if it reaches that point at a time t. Thus if the signal has either not reached the plane z or already passed by it at a time t, there is no contribution to the output from the plane z. This implies that for a finite-length chirp signal, the limits on Eq. (4.5.13) must be reconsidered.
   (a) Determine the form of the output when the chirp is of uniform amplitude between times $-T/2 < t < T/2$ and zero, for $t < -T/2$ and $t > T/2$. Find the form of the output signal for $t < -T$, $-T < t < 0$, $0 < t < T$, and $t > T$.
   (b) Show that if $\mu T^2 >> 1$, the main and first few subsidiary lobes (peaks and valleys) of the output signal are not much affected by changing the length of the chirp, as compared to keeping the length of the delay line L constant and using an infinitely long chirp.
2. When an FM chirp signal is inserted into a matched dispersive delay line whose characteristics are given by Eqs. (4.5.14) and (4.5.15), the output consists of a correlation peak at $t = 0$ with some sidelobes. In the text, we neglected the noncumulative second

term in Eq. (4.5.15). However, this gives rise to sidelobes or a background signal level that may be of some importance. Consider the neglected term, and find its amplitude as a function of time when the delay line is of length L and extends from $-L/2$ to $L/2$, and the chirp length can be regarded as being much longer than L/V. It is of particular interest to find the amplitude of this background signal at times near $t = 0$. To carry out the problem, it will be convenient to make use of the method of stationary phase given in Appendix G. This is equivalent to completing the square in the argument of the exponential, as in Eqs. (4.5.51) and (4.5.52).

(a) Show that the position z, which makes the main contribution to the integral, moves along the delay line at a velocity $V/2$. Show also that at time $t = 0$, this point is outside the delay line if the bandwidth B is such that $B < \omega_0/2\pi$, so that the contribution of this term is negligible in a SAW chirp transform device.

(b) Consider what occurs for a baseband chirp in a CCD filter ($\omega_0 = 0$). Find the amplitude of the unwanted signal relative to the amplitude of the main lobe. (Regard the dispersive CCD filter as a simple continuous filter of time delay T or length L.)

3. Consider a sampled system with N elements spaced a time t_0 apart. Some examples are a CCD chirp transform system or an amplitude-weighted chirp transform system of the type described in Sec. 4.5.3. By taking the weighting of a chirp filter $w(t)$ in Eq. (4.5.4) to be of the form

$$w(t) = \sum_{0}^{N-1} \delta(t - nt_0)$$

find the output of this system with the appropriate input chirp. Show that aliasing can occur, that is, that the output waveform will repeat itself at times $t = M/B$ apart, where B is the bandwidth and M is an integer. Explain physically, in terms of a spatial harmonic expansion representing the response of a quasiuniform region of tapped Fourier delay line, why this effect occurs.

4. One way to look at the sidelobe design problem is to cancel out the sidelobes. Consider Hamming weighting in a general form such that the weighting function is

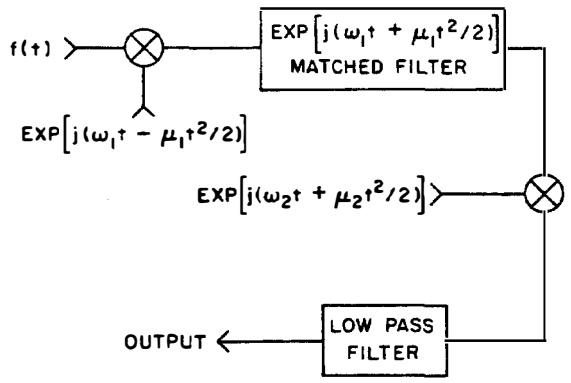
$$w_0(\tau) = (1 - A) + A \cos \frac{2\pi\tau}{T}$$

Show that this weighting function gives rise to a Fourier transform, that is, to a time response with three main lobes arising from the first term and the cosine term, respectively. Choose the amplitude of the main lobes resulting from the cosine term so that it will cancel out the first sidelobe at $t = 3\pi/\mu T$ that resulted from the contribution of the constant term. With this choice, what is the ratio of the sidelobe to main lobe response at $t = 5\pi/\mu T$? Compare this to the unweighted system with $A = 0$ and to the result at $t = 3\pi/\mu T$ and $t = 5\pi/\mu T$ for the following Hamming weighting:

$$w_0(\tau) = 0.54 + 0.46 \cos \frac{2\pi\tau}{T}$$

5. Consider that the scales of a Fourier transform and an inverse Fourier transform need not necessarily be the same. On this basis, carrying out the two operations in turn can give rise to a change in the scaling, or time length, of an input signal. With the circuit shown in the figure, note that the input signal $f(t)$ is multiplied by a chirp exp $[j(\omega_1 t - \mu_1 t^2/2)]$, and inserted into the matched filter, which has a response exp $[j(\omega_1 t + \mu_1 t^2/2)]$. The output signal is then mixed with (multiplied by) a chirp of different slope exp

$[j(\omega_2 t + \mu_2 t^2/2)]$, and the resultant signal is passed through a low-pass filter to keep only the difference frequency term. This signal is inserted into the appropriate matched filter. Show that the output chirp is modulated by a scaled form of $f(t)$, namely $f[t(\mu_2 - \mu_1)/\mu_1]$.



Hint: You will need to write the result in terms of a double integral (i.e., the two Fourier transforms). Invert the order of integration and use the relation

$$\int_{-\infty}^{\infty} e^{j\omega(t-t_0)}\, d\omega = \delta(t - t_0)$$

where $\delta(t)$ is a Dirac delta function.

6. Consider two coupled transmission lines coupled over a length L. Suppose that a wave of amplitude $a_1(z)$, propagating in the forward direction on line 1, is coupled to a wave of amplitude $a_2(z)$, propagating in the backward direction on line 2. We can write the following coupled mode equation for the excitation of a_2:

$$\frac{da_2}{dz} - jk_0 a_2 = j\kappa a_1$$

Assume that the coupling is weak ($\kappa << k_0$) so that we can write

$$a_1(z) = A_1 e^{-jk_0 z}$$

Show that the maximum value of $a_2(0)$ is obtained when the length L of the coupling region is $\lambda/4$. Assume that $a_2(L) = 0$.

4.6 BANDPASS FILTERS

4.6.1 Introduction

The basic method of designing bandpass filters was described in Sec. 4.2. It involved using the Fourier transform technique to predict the time response of the required transducer. In this section we discuss the use of this technique in more detail. In addition, we discuss other techniques, such as phase weighting and the design of unidirectional transducers.

The techniques employed in the earlier chapter change little when applied to CCD delay lines, except that CCDs operate at baseband in a lower-frequency range than is typical of SAW devices. They do not, however, suffer from reflections of signals at the taps; thus the design problem for CCDs is somewhat simpler than for SAW devices.

4.6.2 Strip Coupler

Before describing the amplitude-weighted design, let us discuss the strip coupler [50, 51], a device used to eliminate reflections, excitation of bulk waves, and diffraction problems in SAW filters. Although the strip coupler is an important component in the design of SAW devices, it is also interesting for its own sake, because it is an excellent illustration of the coupled mode principle. We have already seen how this technique can yield information on the excitation of waves by a metal structure without requiring a detailed knowledge of the field distribution of the waves of interest. Here we show how this formalism can be applied to describe the coupling between two waves.

The basic strip-coupled device transfers a surface acoustic wave beam from track 1 to a neighboring track 2, as illustrated in Fig. 4.6.1. The device requires a series of closely spaced metal strips laid down across the two acoustic beam paths. Beam 1 excites a signal on these strips; the strips in turn reexcite the second beam on track 2. This device eliminates problems with unwanted volume wave modes that can give rise to interfering signals; the strip coupler does not couple strongly to these volume waves. It also lets us employ very short fingers for apodized transducers, as illustrated in Fig. 4.6.2(a). If these short fingers are placed very near the multistrip coupler, then diffraction loss, due to the beam expanding sideways, is no longer a problem, as can be seen from the experimental results shown in Fig. 4.6.2(b) and (c).

It is fairly simple to demonstrate the principle of these strip couplers and of other coupled-mode devices. Consider the expression for excitation of a surface acoustic wave at the plane z, propagating in the forward direction [Eq. (4.2.1)]. This is

$$a(t, z) = \alpha \int_{-\infty}^{z} \sigma\left[t - \frac{(z - z')}{V_R}, z'\right] dz' \tag{4.6.1}$$

Here the upper limit of the integral is taken to be z, to account for the fact that although the strip coupler extends beyond the point z, a forward wave is excited only by strips for which $z' < z$.

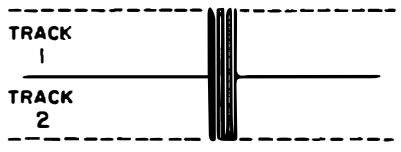


Figure 4.6.1 Transfer of an acoustic wave in a multistrip coupler from track 1 to track 2 for equal trace widths.

The Fourier transform of this expression is

$$A(\omega, z) = \alpha \int_{t=-\infty}^{\infty} e^{-j\omega t} \int_{z'=-\infty}^{z} \sigma\left[t - \frac{(z - z')}{V_R}, z'\right] dz'\, dt$$

$$= \alpha \int_{-\infty}^{z} e^{-j\omega(z-z')/V_R} \int_{t=-\infty}^{\infty} \sigma(t', z')\, e^{-j\omega t'}\, dt'\, dz' \qquad (4.6.2)$$

$$= \alpha \int_{-\infty}^{z} \sigma(\omega, z') e^{-jk(z-z')}\, dz'$$

where $k = \omega/V_R$ and $t' = t - (z - z')/V_R$. By differentiating this equation with respect to z,† it follows that

$$\frac{dA}{dz} = \alpha\sigma(z) - jk\alpha \int_{-\infty}^{\infty} \sigma(z') e^{-jk(z-z')}\, dz'$$
$$= \alpha\sigma(z) - jkA(z) \qquad (4.6.3)$$

or

$$\frac{dA}{dz} + jkA = \alpha\sigma(z) \qquad (4.6.4)$$

where, for convenience, we have dropped the ω in $\sigma(\omega, z)$, $A(\omega, z)$.

Now consider the beams traveling along two neighboring paths, as illustrated in Fig. 4.6.1, which we shall denote by subscripts 1 and 2, respectively. Either beam can be excited by charge on the strips. For simplicity, we suppose that the strips have infinitesimal spacing and are infinitesimally narrow. If the voltage at the surfaces of the beams on both paths are equal, and we take a to be proportional to these voltages, then $A_1 = A_2$, where the subscripts 1 and 2 correspond to the parameters of the waves on each beam path, respectively. In this case we expect the charge on the strips, $\sigma(z)$, to be zero, for no current flows along the strips or is supplied externally to them. More generally, we expect that if the beams are of equal width, the charge on the strips must be proportional to $A_1 - A_2$, as discussed in Prob. 2. Thus we can write a *coupled mode equation* to express the excitation of beam 1, in the form

$$\frac{dA_1}{dz} + jk_0A_1 = j\kappa(A_2 - A_1) \qquad (4.6.5)$$

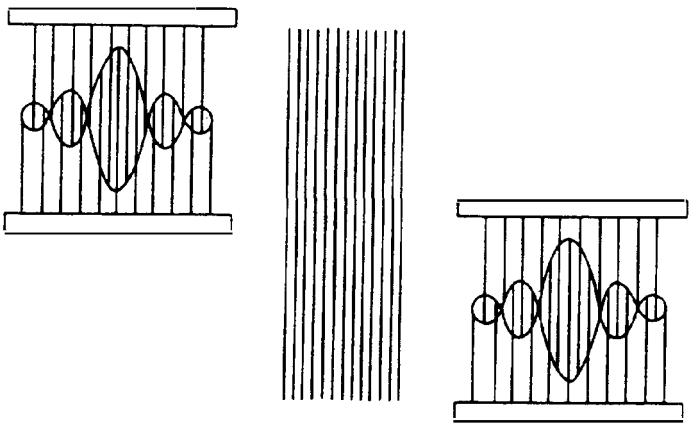
where κ is an arbitrarily chosen real coupling constant. The j is used on the right-hand side so that the eventual solution for the propagation constant of the waves will be real. For instance, if $A_2 = 0$, then A_1 must vary as exp $(-jkz)$, where k

†If

$$G(z) = \int_{-\infty}^{z} F(z, u)\, du$$

then it follows that

$$\frac{\partial G}{\partial z} = F(z, z) + \int_{-\infty}^{z} \frac{\partial F(z, u)}{\partial z}\, du$$



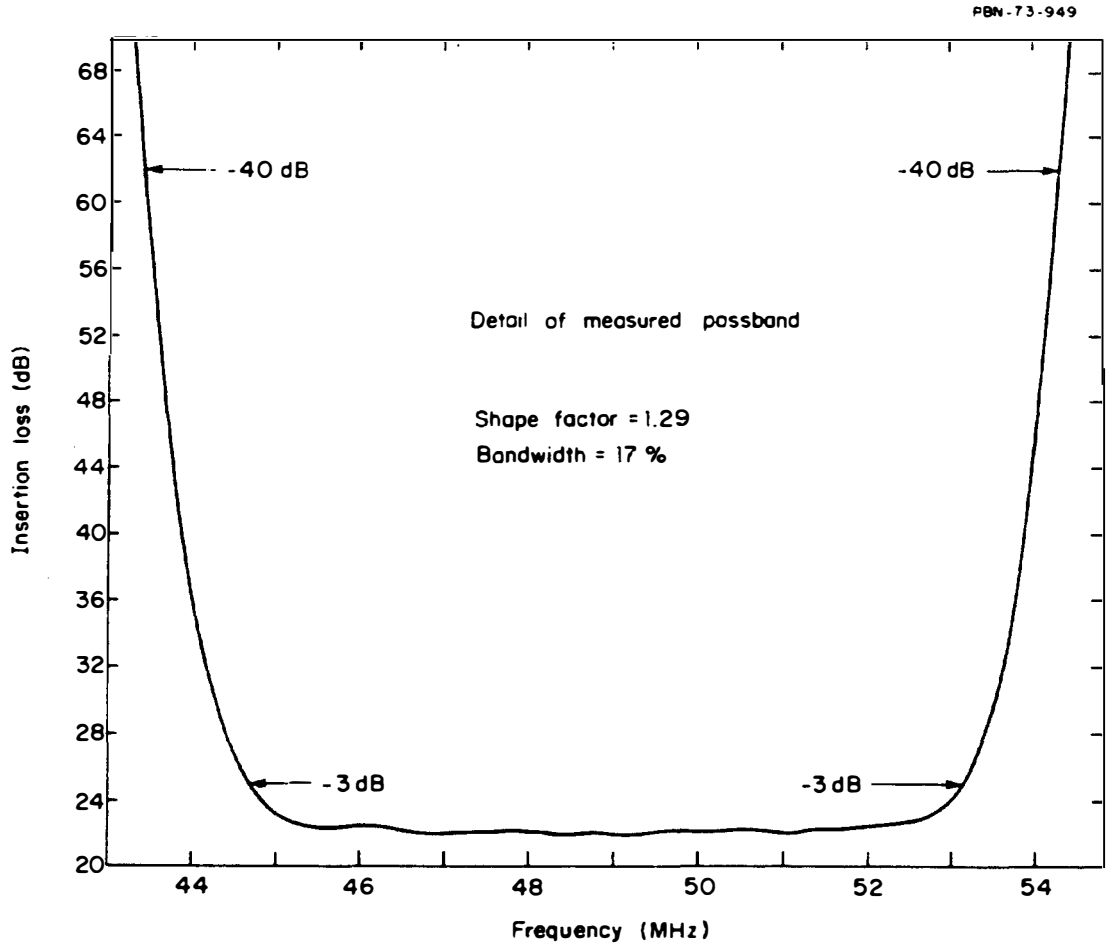


Figure 4.6.2 Filter incorporating a multistrip coupler, and the measured amplitude and phase responses for a filter with sharp skirts. The abrupt change in impedance at the band edges causes most of the phase error. (After Tancrell and Engan [52].)

is real. Additional justification for this choice is given in Prob. 2. By symmetry, we can also write

$$\frac{dA_2}{dz} + jk_0A_2 = j\kappa(A_1 - A_2) \tag{4.6.6}$$

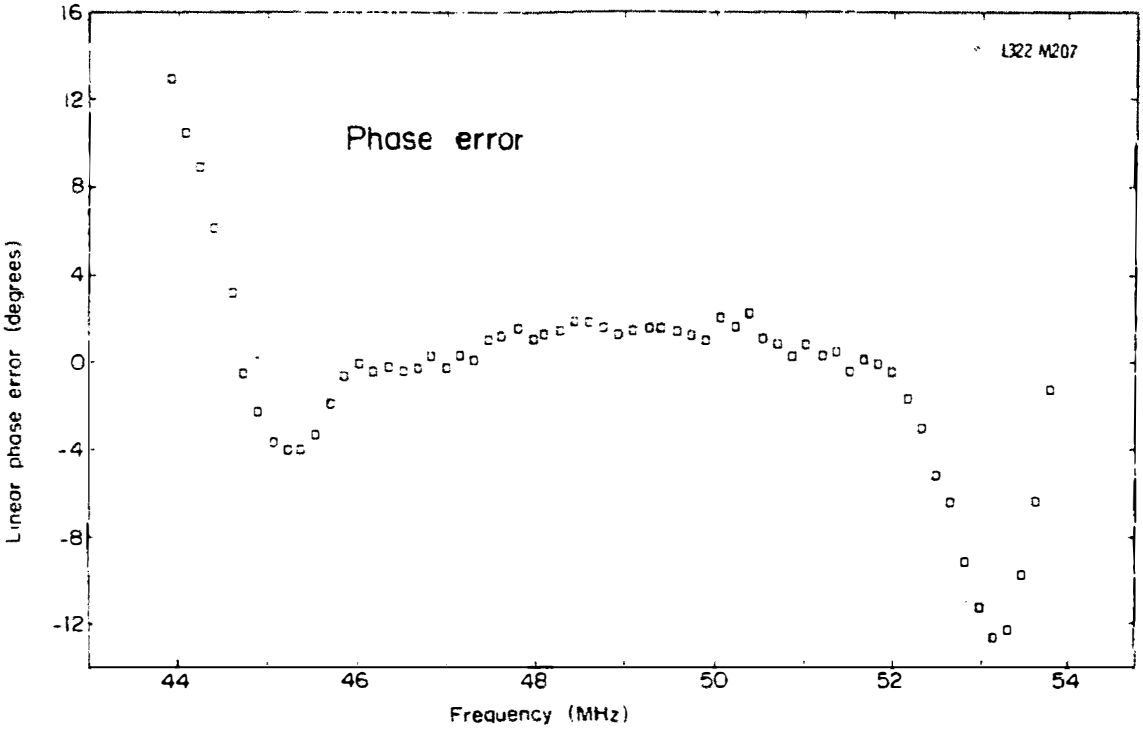


Figure 4.6.2 (continued)

Note that we have assumed that the propagation constants of the waves on both beam paths are equal and of a value k_0 when there are no strips present.

These equations can be solved by assuming a solution in the form $A_1 = A_{10} \exp(-jkz)$, $A_2 = A_{20} \exp(-jkz)$. This procedure leads to the results

$$A_{10}(k_0 - k + \kappa) = \kappa A_{20} \tag{4.6.7}$$

and

$$A_{20}(k_0 - k + \kappa) = \kappa A_{10} \tag{4.6.8}$$

Multiplying Eq. (4.6.7) by Eq. (4.6.8), we find that

$$(k_0 - k + \kappa)^2 = \kappa^2 \tag{4.6.9}$$

By taking the square root of Eq. (4.6.9), we find two solutions:

$$k = k_0 \tag{4.6.10}$$

and

$$k = k_0 + 2\kappa \tag{4.6.11}$$

In turn, this implies that for the two solutions, we can write

$$A_{10} = A_{20} \qquad (k = k_0) \tag{4.6.12}$$

and

$$B_{20} = -B_{10} \quad (k = k_0 + 2\kappa) \tag{4.6.13}$$

respectively, where we have replaced A_{10} by B_{10} and A_{20} by B_{20} for the second solution.

When the amplitudes of the two waves are equal and of the same sign, the propagation constant of the system is unperturbed, as we might expect. When the amplitudes of the two waves are equal and of opposite sign, the propagation constants of the waves are increased (i.e., the wave velocity is slowed down).

We may now solve for the situation when a signal is excited on track 1, and is initially of zero amplitude at $z = 0$ on track 2. The signal on track 1 generally excites both waves of the system; these waves are called the *normal modes* of the system. The normal-mode solutions, as we have seen, have propagation constants $k = k_0$ and $k = k_0 + 2K$, respectively. Thus the initial boundary condition is $A_2(0) = 0$. We can write the solution on track 2 in the form

$$A_2(z) = e^{-jk_0z}\left(A_{20} + B_{20}e^{-2j\kappa z}\right) \tag{4.6.14}$$

We need amplitude coefficients to solve for the total signals on tracks 1 and 2. We will call these A_{20} and B_{20}, respectively. To satisfy the boundary condition $A_2(0) = 0$, it follows that $A_{20} + B_{20} = 0$, or

$$A_2(z) = 2jA_{20}e^{-j(k_0+\kappa)z} \sin \kappa z \tag{4.6.15}$$

It also follows from Eq. (4.6.13) that

$$\begin{aligned} A_1(z) &= 2A_{20}e^{-j(k_0+\kappa)z} \cos \kappa z \\ &= A_1(0)e^{-j(k_0+\kappa)z} \cos \kappa z \end{aligned} \tag{4.6.16}$$

At the point where $z = 0$, $A_2 = 0$ and $A_1(0) = 2A_{20}$. Farther along the line, however, where $\kappa z = \pi/2$, $A_1 = 0$ and $|A_2| = 2A_{20} = A_1(0)$. Thus all the power is transferred from one beam path to the other in a transfer distance $z = L_T = \pi/2\kappa$.

Furthermore, if the strip coupler is made longer, such that $\kappa z = \pi$, all the power is transferred back again to track 1. Thus there is a periodic transfer of energy back and forth between the two beam paths.

We can obtain a measure of the coupling parameter κ by noting that in the situation where the signals on each track are of opposite sign, that is, the solution for which $A_1(z) = -A_2(z)$ or $A_{10} = -A_{20}$ [Eq. (4.6.13)], one beam tends to excite the strip so it is positive, while the other tends to excite the strip so that it is negative. Under these conditions, current flows along the strips but the voltage on them must, by symmetry, be zero; therefore, the strips behave as if they are shorted to each other. Thus if we know the relative perturbation in surface acoustic wave velocity $\Delta V/V$, resulting from short-circuiting the surface of the device with a thin metal film, we can write

$$\frac{2\kappa}{k_0} = \frac{-\Delta V}{V} \tag{4.6.17}$$

As this quantity $\Delta V/V$ is easily calculable and measurable for most piezoelectric substrates and it is possible to determine κ, and hence the length required to transfer energy from one beam to the other.

For a material such as lithium niobate, with $\Delta V/V = 0.023$, Eq. (4.6.17) indicates that the transfer length is $L_T \approx \lambda/(2\Delta V/V)$, which is approximately 22λ, where $\lambda = V/f$ is the acoustic wavelength. This corresponds to a transfer length of approximately 0.7 mm at 100 MHz. In practice, the transfer length is approximately a factor of 2 larger than this figure because the strips have a finite gap between them; consequently, the entire region occupied by the strip coupler is not employed in coupling. Again in practice, the maximum transfer efficiency is usually better than -0.5 dB, with the minimum signal left in track 1 corresponding to a level less than -30 dB below the input signal. The directivity of the signal transfer with the coupler is usually better than 30 dB.

The basic principles described here apply to a wide range of systems. For instance, two optical beams in a fiber-optic waveguide system can be coupled to each other by lapping the two optical waveguides down to their cores, and arranging them so that they have a close spacing over the coupling length required (see Sec. 4.6.6). Microwave waveguides can be coupled to each other in a similar manner, by arranging them so that they have a common wall, with either a slot or closely spaced holes cut in it. In addition, acoustic beams may be coupled to optical beams by arranging for a suitable interaction region. In all cases, we can describe the qualitative behavior of the coupling mechanism by the simple coupled mode equations we have derived. Estimating the coupling coefficient itself is normally more difficult.

4.6.3 Amplitude-Weighted Bandpass Filters

The basic steps in designing an amplitude-weighted baseband bandpass filter involve deciding on the frequency response required, taking its Fourier transform, and then arranging the apodization of the filter to give the time response needed. A problem arises because such filters should, in principle, be of infinite length. Suppose, for instance, that we require a filter with a flat frequency response from $\omega_0 - \Omega/2$ to $\omega_0 + \Omega/2$ or $\Pi(\omega - \omega_0)/\Omega$, where $\Pi(x)$ is the rectangle function defined as

$$\Pi(x) = \begin{cases} 1 & |x| < \frac{1}{2} \\ 0 & |x| > \frac{1}{2} \end{cases}$$

Thus the bandpass in the positive frequency domain extends from $\omega_0 - \Omega/2 < \omega < \omega_0 + \Omega/2$. In this case, the time response of the filter will be

$$h(t) = \frac{\Omega}{2\pi} \frac{\sin \Omega t/2}{\Omega t/2} e^{j\omega_0 t} \tag{4.6.18}$$

The time response required is therefore infinite in length.

Suppose that it is decided to cut off the time response of the filter at times $t = \pm T/2$, that is, to use a filter with a finite length T, as illustrated in Fig. 4.6.3(a).

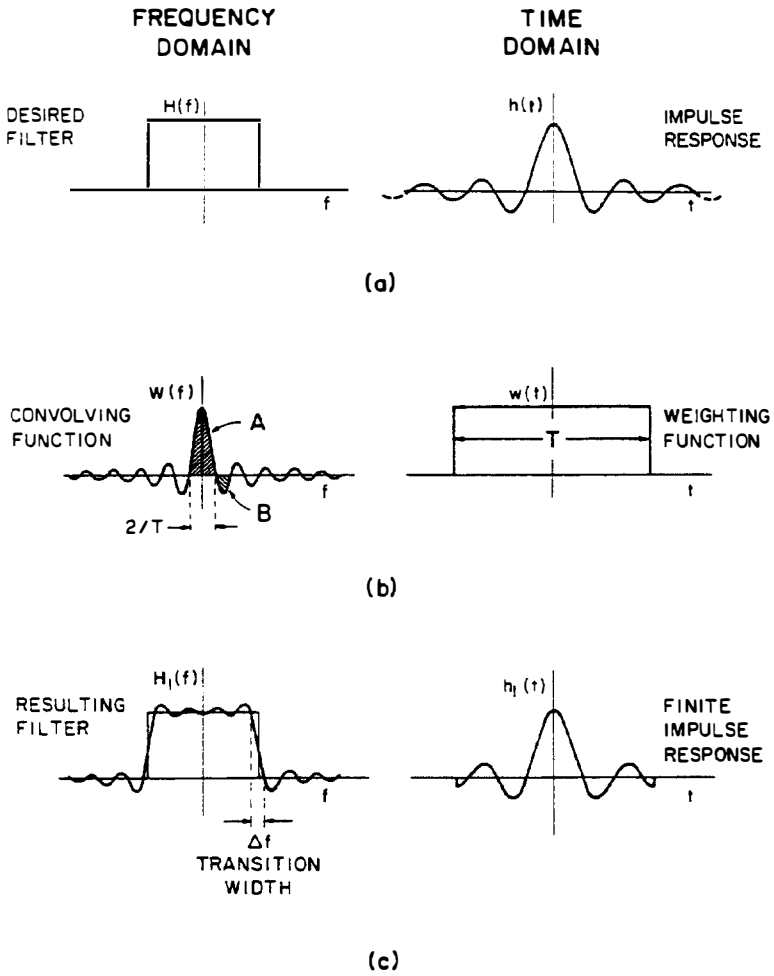


Figure 4.6.3 Convolution of the desired function $H(f)$ with the frequency spectrum of a finite impulse of length T: (a) ideal frequency and time responses; (b) weighting functions in the time and frequency domains for a finite-length filter; (c) behavior of a finite-length filter.

For an SAW device, this would imply using a filter of physical length $L = VT$, where V is the surface wave velocity. In a CCD, T would correspond to the total time delay in the device itself. This corresponds to a weighting function $w(t) = \Pi(t/T)$, whose Fourier transform is

$$W(\omega) = T\frac{\sin \omega T/2}{\omega T/2} \tag{4.6.19}$$

The output from the device in the frequency domain will therefore be the convolution of the frequency responses $W(\omega)$ and $H(\omega) = \pi(\omega - \omega_0)/\Omega$, that is, the frequency response of the device will be

$$H_1(\omega) = \int H(\omega - s)W(s)\,ds \tag{4.6.20}$$

We can illustrate this convolution process graphically by the diagram shown

in Fig. 4.6.3. The length of the weighting function $W(\omega)$, shown in Fig. 4.6.3(b), between its two zeros is $\Delta\omega = 4\pi/T$, or

$$\Delta f = \frac{2}{T} \tag{4.6.21}$$

Therefore, by limiting the length of the impulse response to a time T, we might expect the transition width of the frequency response to be approximately $2/T$, corresponding to the width between the zero of the main lobe of $W(\omega)$, as illustrated in Fig. 4.6.3. Furthermore, we can estimate the minimum attenuation outside the passband, from the area A of integration under the first sidelobe of $W(\omega)$, as compared to the area B of integration under the main lobe of $W(\omega)$, as indicated in Fig. 4.6.3(b). By taking the regions A and B to be triangular, we can show that the ratio of these two areas is approximately $(2/3\pi)/2 = 1/3\pi$, or 19.5 dB. The accurate result, given in Table 4.6.1, indicates -21 dB as the minimum stopband attenuation. On the same basis, the ripple at the edge of the passband would depend on the area under the first and second sidelobes, and would vary approximately from $1 - 1/3\pi$ to $1 - 1/3\pi + 1/5\pi$, or 0.6 dB.

If we could use a smoother weighting function, $W(\omega)$, the ripple within the passband would obviously be radically decreased. In addition, the attenuation outside the passband would be increased. Thus it is useful to employ the kinds of windowing functions described in Sec. 4.5.3. By doing this, we obtain the results given in Table 4.6.1. Typically, the stopband attenuation is improved at the expense of a wider transition width. The longer the filter length, the better the transition width of the filter.

An example of such a filter configuration is shown in Figs. 4.6.4 and 4.6.5. The results obtained with a low-frequency experimental filter terminated at the third lobe on either side of the main lobe (i.e., at $T = 8/B$) are shown in Fig. 4.6.5. In this case, with a Bartlett (triangular) weighting, the response outside the passband is improved considerably. Before weighting, the level outside the passband was -14 dB, worse than the -21 dB predicted by theory, while after weighting, the level outside the passband was -35 dB, better than the value of -25 dB predicted by theory. Since these early results were obtained, more carefully made devices have been constructed, and the results have usually been in very good agreement with theory.

TABLE 4.6.1 SIDELOBE LEVELS AND TRANSITION WIDTHS FOR PASSBAND FILTERS

Window	Transition width of main lobe	Minimum stopband attenuation (dB)
Rectangular	$2/T$	-21
Bartlett	$4/T$	-25
Hanning	$4/T$	-44
Hamming	$4/T$	-53
Blackman	$12/T$	-74

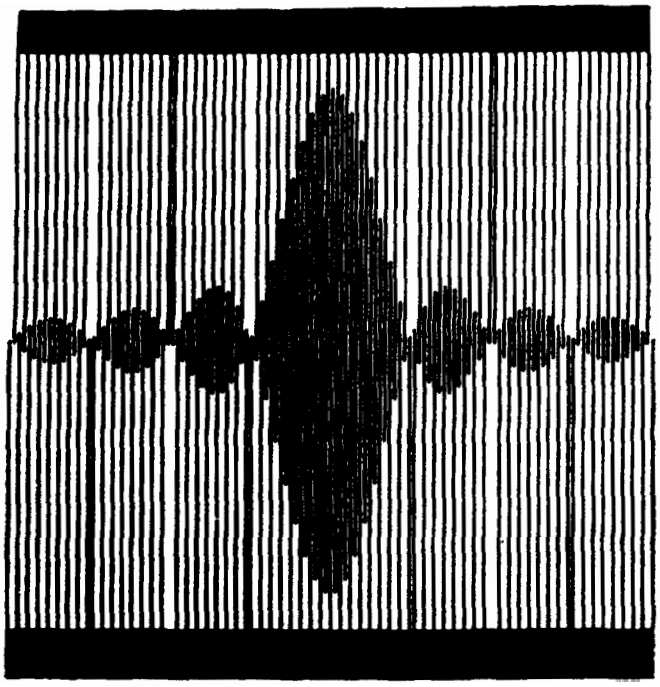


Figure 4.6.4 Interdigital electrode configuration in an SAW filter. (After Atzeni and Masotti [53].)

Other methods of weighting may also be used. For instance, the Dolph–Tchebysheff weighting technique allows us to maximize the attenuation outside the passband, and also to obtain the narrowest transition width possible for this attenuation. Because this particular technique is complicated to design and tends to lead to a weighting function with sharp maxima at the ends of the transducer, it is normally approximated by other methods, Hamming weighting being one of them. Thus the basic design approach is normally similar to that already described, with the added proviso that an attempt be made to optimize the transition width for a given attenuation outside the passband. Using these techniques, successful transducer designs for a wide range of bandwidths have been formulated with attenuations outside the passband as great as 60 dB.

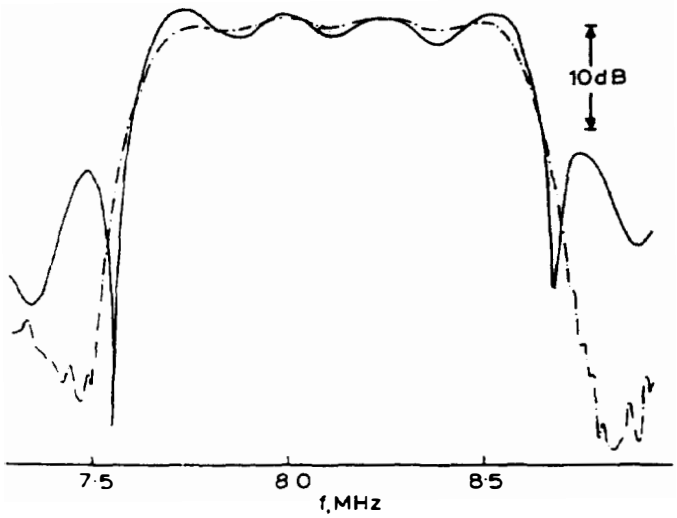


Figure 4.6.5 Amplitude against frequency response of the triangular weighted SAW filter (dashed line), compared with that of the unweighted filter (solid line). (After Atzeni and Masotti [53].)

Usually, the problem of obtaining sufficient attenuation outside the passband is associated with spurious effects. Simple designs, which might theoretically give rise to as much as a 70-dB attenuation outside the passband, often lead to attenuations of only 30 dB. This is due to unwanted excitation of volume waves, reflections from the fingers and the edges of the substrate, and undesirable diffraction phenomena.

The bulk wave excitation problem and the diffraction problem can, to a large extent, be eliminated by using strip couplers, because the strip coupler does not couple very strongly to bulk waves. Strip-coupled designs are thus often used for these purposes. Another approach, favored in Japan, is to use special cuts of lithium niobate and other materials for which the excitation of bulk waves is very weak. Successful intermediate-frequency (IF) filters for television receivers have been made with this method. It is normally fairly simple to eliminate reflections from the edges of the substrate. When a soft material, such as wax or a plastic resin, is deposited on the edges of the substrate, the surface waves are rapidly attenuated. Reflections from the fingers can usually be eliminated with the split-finger technique. Each finger is split into two parts, so that the periodic spacing of fingers is approximately $\lambda/4$, as illustrated in Fig. 4.6.6 (see Prob. 4.2.3). Neighboring fingers now give reflections π out of phase that cancel each other out. The excellent characteristic of a television IF filter, designed by Kodama using a combination of these techniques, is illustrated in Fig. 4.6.7.

The equivalent CCD devices do not suffer from most of these problems, but they can have problems with spurious effects resulting from charge-transfer efficiency. This tends to limit their design to a relatively low frequency, below 1 MHz, in addition to limiting the total length of the device. As the transition width depends on the total length of the device, this limits the transition times obtainable at a given frequency. As we have seen in Sec. 4.3 and Fig. 4.3.10, very successful low-frequency CCD filters have been made.

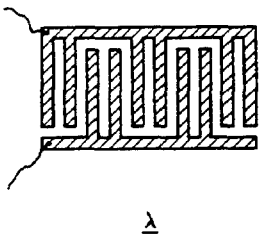


Figure 4.6.6 Transducer with split fingers to eliminate reflections.

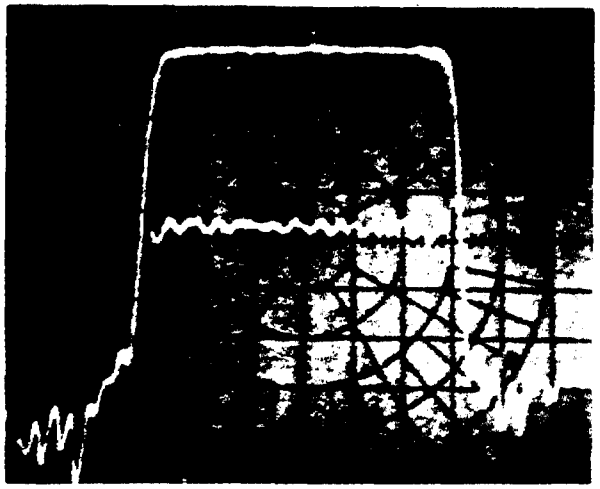


Figure 4.6.7 Measured vestigial sideband (VSB) filter amplitude and group-delay response of television IF filter designed by Kodama. (After Kodama [54].)

One problem with the standard interdigital transducer is that its symmetry excites equal-amplitude signals in both directions. Consequently, the transducer has a minimum loss of 3 dB; a device using two interdigital transducers has a minimum loss of 6 dB. A further consequence of this fact is that the interdigital transducer is a three-port device (one electrical port and two acoustic ports), and all three ports cannot be matched in a three-port network. For instance, suppose that a receiving transducer is perfectly matched at both acoustic ports (1 and 2), and an attempt is made to obtain maximum power into the electrical port (3). Then, by reciprocity, we know that if the input signal into port 1 is a_1 and the input power is a_1^2, the maximum power across the load at port 3 will be $a_1^2/2$ (a 3-dB loss). By reciprocity and symmetry, this signal across the load excites signals of value $a_1^2/4$ at ports 1 and 2. Thus a reflected signal is emitted from port 1 that is 6 dB lower in amplitude than the incident signal. The signal leaving port 3 is also 6 dB less than the input signal. Consequently, if both transducers of a delay line are matched as well as possible, there is a *triple transit echo*, that is, a signal reflected from the output transducer and then re-reflected from the input transducer, which is 12 dB down at the output electrical port 3 from the directly transmitted signal.

Reflections can be a serious problem when an interdigital transducer is matched for best efficiency. Thus the transducers often are not loaded for optimum efficiency. To improve the efficiency, a central input transducer is sometimes used to supply two output transducers, connected in parallel and placed on each side of the input transducer.

Another stratagem is to employ a three-phase transducer, illustrated in Fig. 4.6.8. In this transducer there are three fingers per wavelength, excited with phases 0, $2\pi/3$, and $4\pi/3$. This set of three fingers excites only a wave in the forward direction, since the wave excited by the three fingers in the backward direction has an amplitude $1 + \exp(-4j\pi/3) + \exp(-8j\pi/3) = 0$. Thus the device is a one-way transducer. The disadvantage is that a special electrical network must be used to excite this transducer; in addition, the geometry of the device becomes more difficult because it requires three separate connections. However, bandpass filters with excellent characteristics, high directivity, and only a 2-dB total loss from terminal to terminal have been made by this technique [55].

Finally, there are various permutations of the strip coupler configuration that can be used to give one-sided excitation. Such techniques, although useful, tend to introduce losses, so the improved performance desired is not always obtained (see Prob. 1) [50, 51].

4.6.4 Phase-Weighted Transducers

The treatment already carried out for chirp delay lines shows that such delay lines exhibit good frequency response over a finite bandwidth. However, because the input-output response of these delay lines is dispersive, their phase response or group delay may not be what is normally required in a signal processing system. A simple way to eliminate this difficulty is shown in Fig. 4.6.9(c). The dispersion

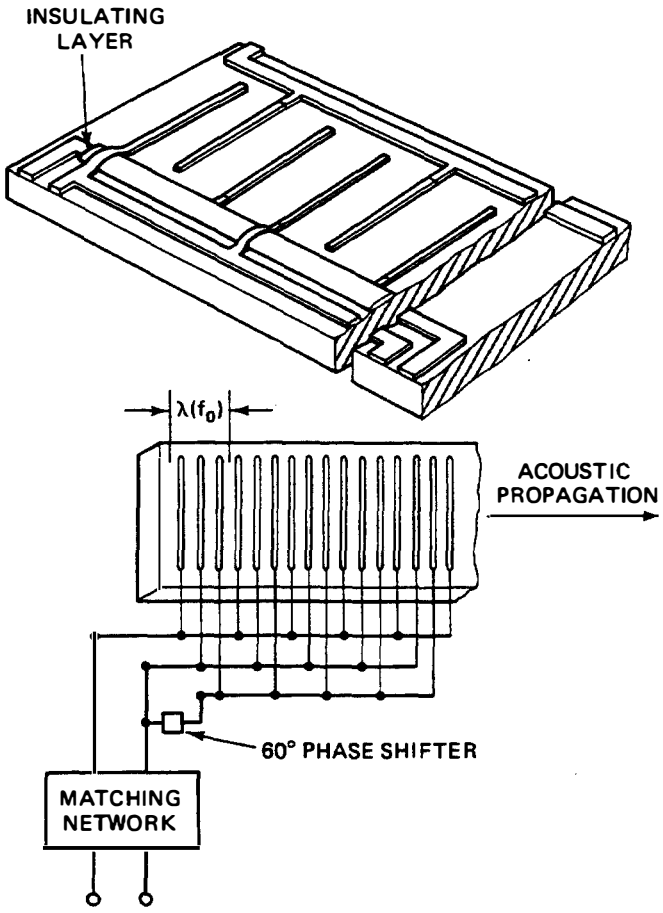


Figure 4.6.8 Geometry of a multilayer three-phase transducer. (After Hartmann et al. [55].)

is eliminated with two phase-weighted chirp transducers aligned in the same direction; one is used as a transmitting transducer and the other, as a receiving transducer. This implies that the delay time of any particular frequency component through the filter will be constant. The results obtained with such a filter are shown in Fig. 4.6.9(a) and (b), where the time response is of the form sin t/t, as we might expect. Thus this provides another simple way of designing filters (see Prob. 3).

4.6.5 Building-Block Design Technique

Many filters require not only a good response within the passband, but also extremely high attenuations near specific frequencies outside the passband; that is, they require the provision of "traps." A technique that has been utilized in the design of television IF filters, where deep traps are required for the neighboring channel and the sound channel, is illustrated in Fig. 4.6.10. Several transducers are placed in parallel, and the energy excited by them is transferred by a strip coupler to a single output transducer. Thus the individual transducers are designed for a given response and the responses are added.

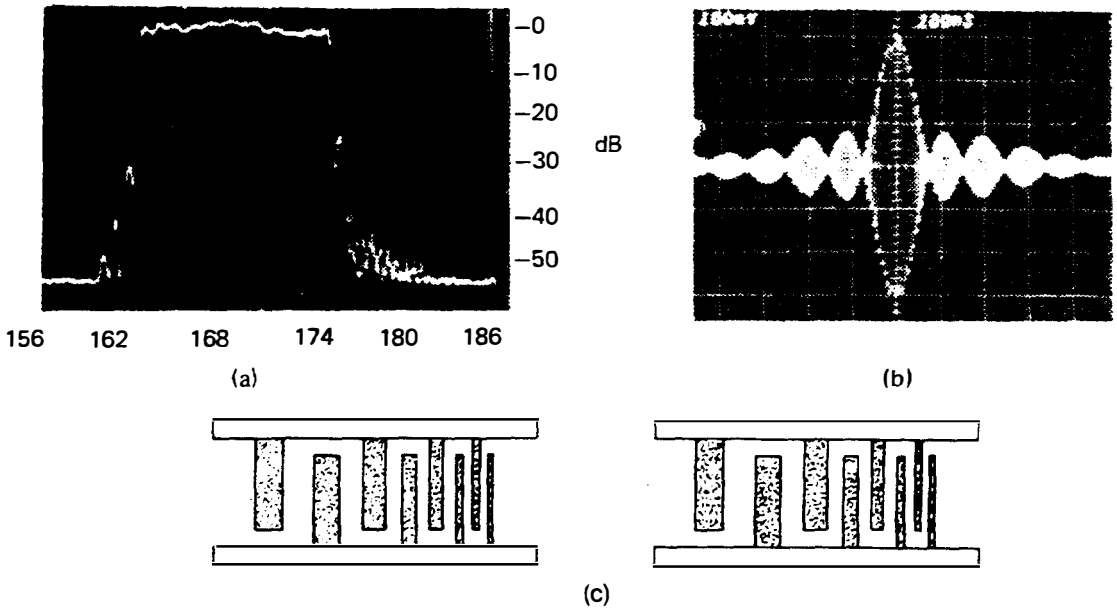


Figure 4.6.9 Low-shape-factor surface wave filter module: (a) 12 MHz at 168-MHz frequency response; (b) $(\sin x)/x$ impulse response indicates linear phase response and rectangular bandpass; (c) two phase-weighted transducers with opposite dispersion for canceling phase nonlinearity. (After Hartmann et al. [15].)

Suppose that we have two transducers, of center frequencies ω_1 and ω_2, with N_1 and N_2 fingers, respectively. Then the total response is

$$H(\omega) = W_1 \operatorname{sinc} \frac{N_1(\omega - \omega_1)}{\omega_1} + W_2 \operatorname{sinc} \frac{N_2(\omega - \omega_2)}{\omega_2} \tag{4.6.22}$$

where $\operatorname{sinc} x = (\sin \pi x)/\pi x$, and W_1 and W_2 are the weights of the two transducers; that is, W_1 and W_2 are proportional to the finger lengths of the two transducers, respectively.

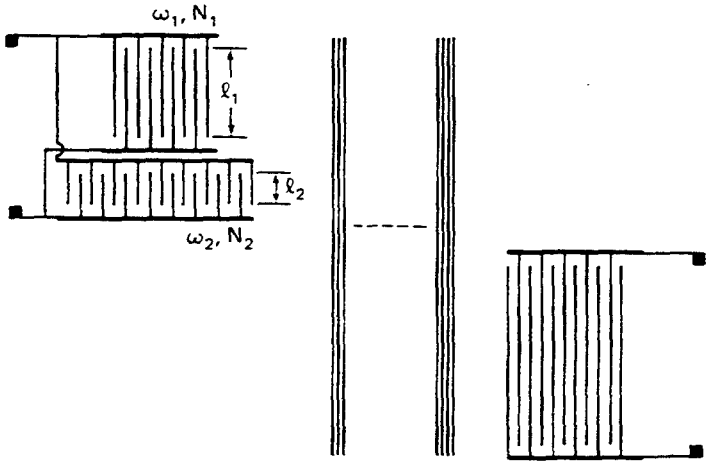


Figure 4.6.10 System for utilizing multiple transducers to affect frequency response. (After DeVries [56].)

Suppose, for instance, that we wished to compensate for the first upper sidelobe of the transducer with a center frequency ω_1. If we choose the second transducer to have a finger number N_2, such that the zero of the response of the second transducer is at the same point as the zero of the response of the first transducer, we must have $\omega_2 = \omega_1(1 + 3/2N_1)$, $N_2 = 2N_1 + 3$, and $W_2 = 2W_1/3\pi$ or $W_2 = 0.21W_1$. This will give a calculated response for the first upper sidelobe like that of Fig. 4.6.11, with a zero response or trap in its original position.

A similar technique can be used to broaden the frequency response of the main lobe. We may now choose $\omega_2 = \omega_1$, $N_2 = 2N_1$, and $W_2/W_1 < 0$. We will obtain a response of the type shown in Fig. 4.6.12.

Thus, in general, we design a system out of building blocks, so that the total response is of the form

$$H(\omega) = \sum_n W_n \text{sinc}\, N_n \frac{(\omega - \omega_n)}{\omega_n} \tag{4.6.23}$$

By taking the Fourier transform of this response, we note that each term consists of a finite length RF burst, as illustrated in Fig. 4.6.13. By summing these responses, we can make a simple filter corresponding to this total time response. A configuration of a TV IF filter designed by this technique is shown in Fig. 4.6.14, where the zeros of the response have been chosen as the trap frequencies. The results obtained are shown in Fig. 4.6.15.

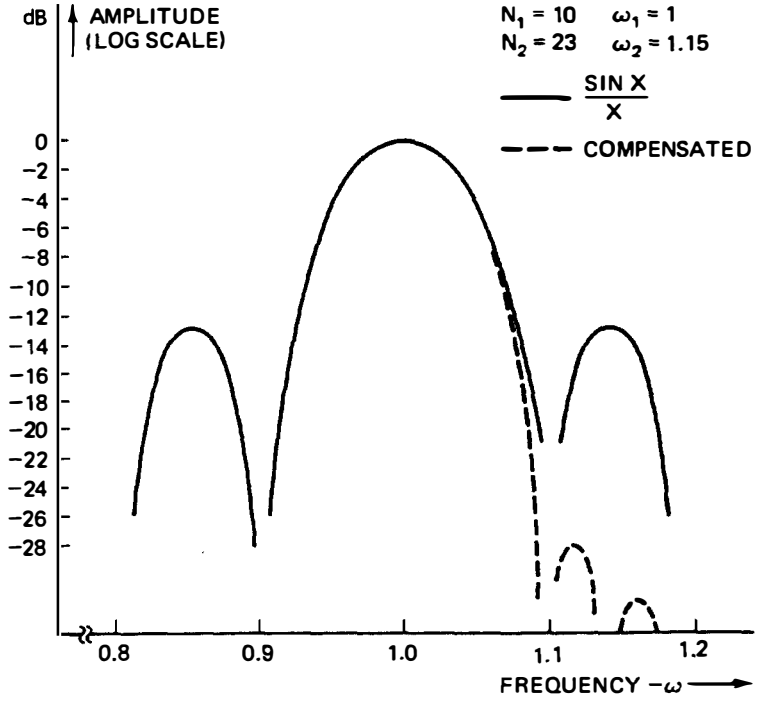


Figure 4.6.11 Calculated response curve for first upper sidelobe compensation. (After DeVries [56].)

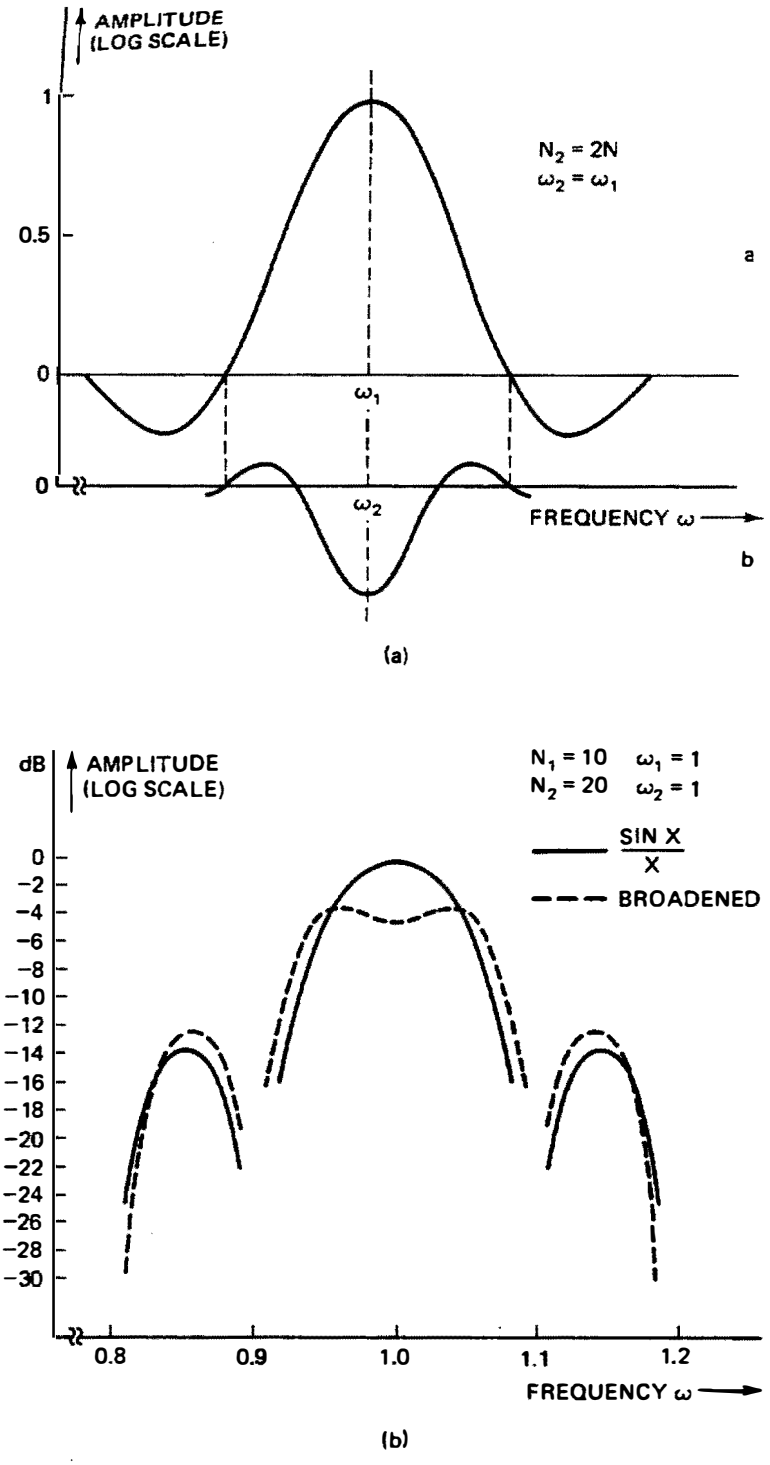


Figure 4.6.12 (a) System for broadening of frequency response; (b) calculated broadened frequency response. (After DeVries [56].)

Note that in this design both the finger spacing and the finger length vary. A multistrip transducer of a hundred strips was employed to eliminate bulk wave interactions. The output transducer did not use split fingers because the design was simplified by using fingers with a width of approximately $3\lambda/8$, which generally tends to eliminate spurious reflections.

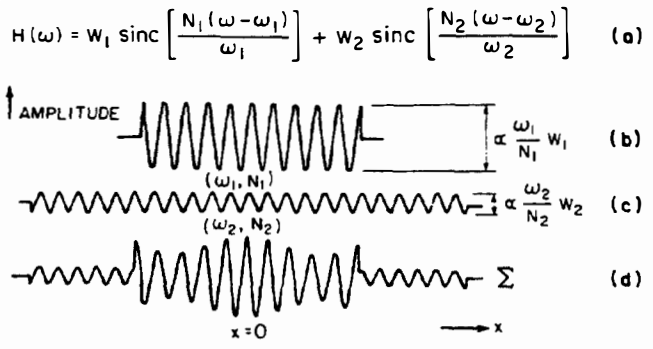


Figure 4.6.13 Method for synthesizing transducers. The desired frequency response is given by part (a). The respective Fourier transforms of the first and second terms are shown in parts (b) and (c). In part (d), the composite waveform from which the transducer configuration can be derived is shown. (From DeVries [56].)

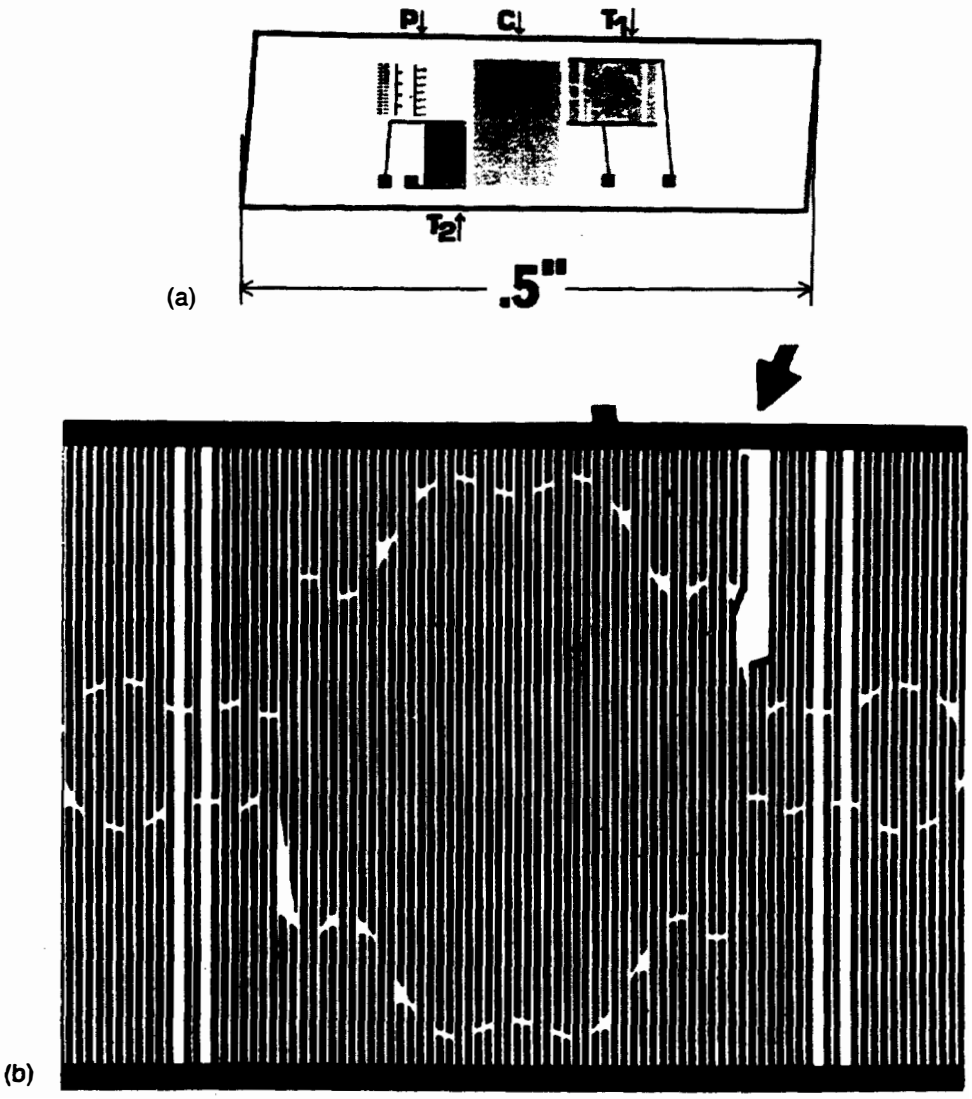


Figure 4.6.14 (a) Filter element showing transducer structure on lithium niofate substrate. The structure above the left transducer identifies the position of the element on the plate. (After DeVries and Adler [57].) (b) Configuration of weighted input transducer. An added transducer section that corrects for a finite trap depth is indicated by an arrow. (After DeVries et al. [58].)

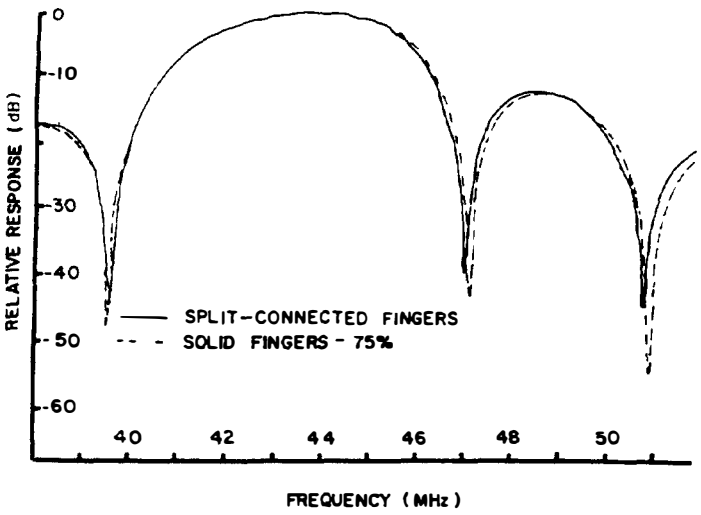


Figure 4.6.15 Comparison of the frequency response of a transducer with split connected fingers with a duty factor of 75%. (After DeVries et al. [58].)

4.6.6 Recursive Filters and Comb Filters

A. Introduction

Most transversal filters, because of their limited length, do not have as sharp a transition bandwidth as is sometimes desired. Better transition bandwidths can often be obtained in standard filters, because in z transform terms there are poles as well as zeros in the characteristic, which can give rise to very sharp changes in the transfer function. Such a pole in the characteristic can be generated in a transversal filter by using feedback between the output and the input. Zeros in the characteristic can be obtained using feedforward from input to output. Filters based on this principle are called *recursive filters*.

Suppose that the delay time in a simple delay line is T, as illustrated in Fig. 4.6.16; we feed the signal forward as shown, and add the input and output signals. We can write

$$y(t) = x(t) + \alpha_1 x(t - T) \tag{4.6.24}$$

If we choose $\alpha_1 = -1$, it follows that

$$y(t) = x(t) - x(t - T) \tag{4.6.25}$$

For an input signal that varies as $x(t) = \exp(j\omega t)$,

$$y(t) = e^{j\omega t}(1 - e^{-j\omega T}) \tag{4.6.26}$$

$$y(t) = 2je^{j\omega(t-T/2)} \sin \frac{\omega T}{2}$$

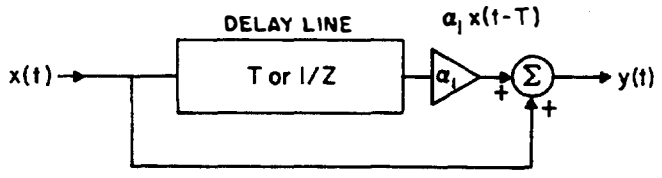


Figure 4.6.16 Recursive filter.

In z transform terms, we would write $z = e^{j\omega T}$, so that

$$H(z) = H(\omega) = (1 - e^{-j\omega T}) = 1 - \frac{1}{z} \tag{4.6.27}$$

This filter has a zero at $z = 1$, corresponding to notches in the output at $f_s = 1/T$ and all integral multiples of this frequency. Such a filter is called a *comb filter.*

This is the basic delay-line canceler element used for clutter rejection in moving target indicator (MTI) radar. The purpose is to eliminate *clutter*, that is, signals from stationary targets such as the ground or the sea. The frequency reflected from stationary targets is the transmitted frequency. A moving target, however, will give a Doppler shift of the return signal frequency and therefore can be observed while the clutter from the stationary targets is removed in the ideal system. As an example, a BBD delay line has been constructed for this purpose with 400 stages. This is clocked at a frequency $f_c = 1.6$ MHz, making the time delay through the device $T = 250$ μs. The effective sample rate for the data is $f_s = 1/T = 4$ kHz, and the delay-line canceler has notches at $f = 0$, 4 kHz, 8 kHz, and so on. Thus the radar is operated with a repetition frequency for the transmitted pulse of $f_s = 4$ kHz.

Note that signals arriving from different ranges arrive at different times. The 400-stage delay line described here has 400 possible "buckets," each spaced by 1/1.6 MHz (i.e., 0.67 μs), to store the information from the different ranges approximately 100 m apart. Thus we regard this device as having 400 "range bins." The effective sampling rate for each range bin is, in fact, 4 kHz (i.e., each range bin is filled every 250 μs). We can arrange to multiplex or switch the output $y(t)$ to 400 different storage units, where the information (time averaged over many samples, for example) can be processed.

In operation, the Nth pulse from the transmitter is received as a series of radar echoes spaced over a time of 250 μs. These echo signals pass into the delay line and fill individual buckets. As they leave the delay line, after being present in it for a time of 250 μs, they are added to the equivalent echo from the $(N + 1)$th pulse. Each echo can be switched or multiplexed from the delay line into a separate storage bin; echoes processed later can be added to the earlier ones to average the time of the signal.

We note that if we add M pulses of amplitude V, we obtain a total voltage output MV. However, as noise is incoherent, the total noise signal only adds as $\sqrt{M}$. To put it another way, the signal power into a 1-Ω load is

$$P_S = M^2V^2 \tag{4.6.28}$$

The noise power out is

$$P_N = M\langle N^2\rangle \tag{4.6.29}$$

where $\langle N^2\rangle$ is the expectation value of the noise power. Thus the signal-to-noise ratio is improved by a factor M after averaging M times.

The signal entering the BBD is, in fact, a short pulse. Thus at best it is a sampled version of the RF signal. Suppose, for simplicity, that the return echo

signal is an RF pulse or tone burst of length T_p and Doppler shifted carrier frequency ω, and that the repetition time is T. Then the nth RF pulse has the form $\cos \omega(t - nT)$. If the signal is synchronously detected by mixing it with the carrier signal $\cos [\omega_0(t - nT)]$, a baseband product signal $y_{out}(t)$ will be obtained at a frequency $\omega - \omega_0$, where

$$\begin{aligned} y_{out}(t) &= \int_{-T_p/2}^{T_p/2} \cos [\omega(\tau - nT)] \cos [\omega_0(\tau - nT)]\, d\tau \\ &\approx \frac{\cos (\omega - \omega_0)nT \sin [(\omega - \omega_0)T_p/2]}{\omega - \omega_0} \end{aligned} \tag{4.6.30}$$

Thus the output from the radar detector has an amplitude that varies as $\cos (\omega - \omega_0)nT$. We now observe that each pulse entering a bucket is a sampled version of the detected signal $\cos (\omega - \omega_0)t$. Thus by averaging over many repetition cycles, the recursive filter treats the signal as if it were a continuous wave of frequency $\omega - \omega_0$, and the cancellation of signals for which $\omega = \omega_0$ can be obtained from stationary targets.

B. Recursive Filter with Feedback

The basic problem with the simple delay-line canceler described above is, of course, that it lacks an extremely sharp cutoff, which means that the signals from slowly moving targets will also tend to be attenuated. A recursive filter with feedback as well as feedforward has a characteristic in z transform terms as follows:

$$H(z) = \frac{1 + \alpha_1/z}{1 + \beta_1/z} \tag{4.6.31}$$

In our first example, we took $\beta_1 = 0$, $\alpha_1 = -1$. Suppose that we had used the circuit shown in Fig. 4.6.17. It is fairly easy to show that this circuit has the characteristic given in Eq. (4.6.31), with a pole at

$$z = e^{j\omega T} = -\beta_1 \tag{4.6.32}$$

Therefore, for $\beta_1 \to 1$, it has sharply peaked resonances at multiples of $f_s = 1/T$. By varying β_1, we can change the position of these resonances and their Q. At the same time, the recursive filter has zeros at

$$z = e^{j\omega T} = \alpha_1 \tag{4.6.33}$$

Thus it is possible to obtain a sharply peaked resonance very near zero frequency.

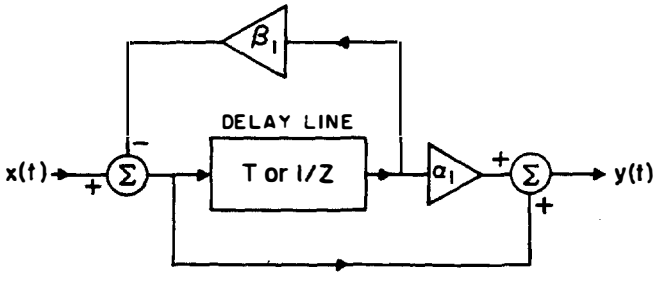


Figure 4.6.17 Single-pole recursive filter.

By cascading delay lines of this kind, it is possible to make far more sophisticated filters with well-understood characteristics using relatively short delay lines. We can write a general characteristic of such a recursive filter in the form

$$H(z) = \frac{\sum_{0}^{N} \alpha_n z^{-n}}{\sum_{0}^{N} \beta_n z^{-n}} \tag{4.6.34}$$

Consider the filter shown in Fig. 4.6.18, which consists of a first-order filter with $\alpha_1 = -1$, $\beta_1 = -K_1$, followed by a second-order filter with $\alpha_1 = -2$, $\alpha_2 = 1$, $\beta_1 = -K_2$, and $\beta_2 = K_3$. The operation of this filter can best be understood by setting the feedback equal to zero; then $K_1 = K_2 = K_3 = 0$. In this case

$$H(z) = (1 - z^{-1})^3 \tag{4.6.35}$$

It follows that the frequency characteristic is

$$|H(\omega)| = 8 \sin^3 \frac{\omega T}{2} \tag{4.6.36}$$

Thus this filter rejects zero frequency and all multiples of the pulse repetition frequency (the reciprocal of the transit time). However, it severely attenuates nonzero frequencies near zero that correspond to slow-moving targets, although it has a better characteristic than the first-order filter. More generally, when feedback is present, the filter has the characteristic

$$H(z) = \frac{(1 - z^{-1})^3}{(1 - K_1 z^{-1})(1 - K_2 z^{-1} + K_3 z^{-2})} \tag{4.6.37}$$

By selecting the amplifier gains $K_1 = 0.0881$, $K_2 = 1.2001$, and $K_3 = 0.7012$, we achieve the frequency response shown by the solid curve in Fig. 4.6.19. The incorporation of the feedback results in very little attenuation for $f > 0.1 f_s$ and in a uniform amplitude response over a wide range of frequency between the notches.

The MTI canceler of Fig. 4.6.17 has been implemented by using eight-stage CCDs for the delay line. In this configuration, eight range bins can be filtered by clocking the CCD eight times each pulse repetition frequency. Feedback amplifiers are implemented with resistor networks around operational amplifiers. When the CCD is clocked to simulate a 38.4-kHz pulse repetition frequency, the filter rejection at $f = 0$ and $f = 38.4$ kHz is -32 dB, relative to the maximum output. The great advantage of this CCD over the equivalent acoustic delay-line cancelers is that its time delay can be controlled at will: therefore, the timing for the CCD

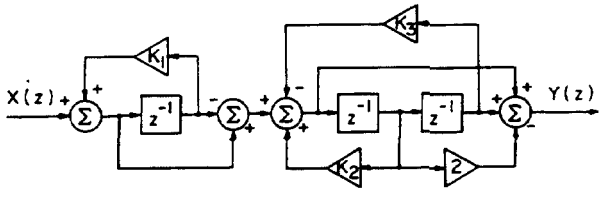


Figure 4.6.18 Block diagram of the three-delay canceler for MTI. It consists of a first-order recursive filter cascaded with a second-order recursive filter. (From Buss et al. [4].)

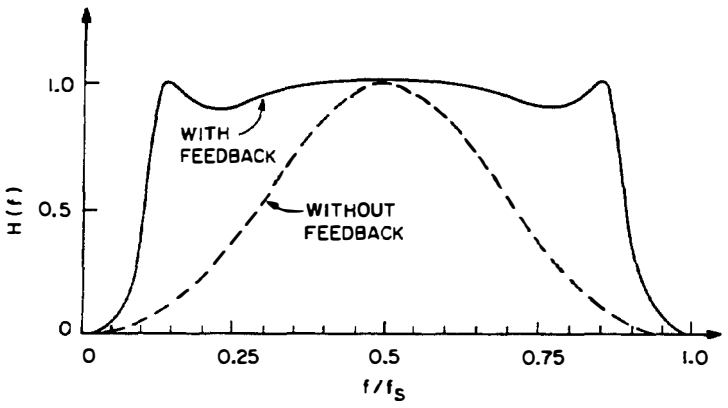


Figure 4.6.19 Calculated frequency response of the three-delay canceler of Fig. 4.6.18. The solid line shows the result for $K_1 = 0.0881$, $K_2 = 1.2001$, and $K_3 = 0.7012$. The dashed line shows the result with $K_1 = K_2 = K_3 = 0$. (From Buss et al. [4].)

can be synchronized with the timing of the transmitter to achieve a delay of exactly one pulse repetition frequency. On the other hand, the CCD has problems because of its imperfect charge-transfer efficiency, which does alter the filter characteristics. SAW delay lines have also been used in this application, with the clocks of the radar essentially controlled by the delay time through the delay line.

Example of an SAW recursive comb filter. Recursive filters may be constructed with SAW devices in the manner illustrated in Fig. 4.6.20. A relatively short input transducer is employed with two output transducers that have a center-to-center time separation T, and hence a separation between response maxima of $\Delta f = 1/T$. The passband will have the form of a "comb response," as illustrated in Fig. 4.6.21(b). Because of the characteristics of both the input and the separate output transducers, this comb response will be multiplied by their bandpass responses, as shown in Fig. 4.6.21(a).

One application of this comb response characteristic, which has been employed commercially, is for a channel indicator on a TV set. In a TV set, the input signal is mixed with the output of a varactor tuned oscillator to produce the IF frequency. When this oscillator frequency is swept through the TV channels in turn, the response of the comb filter passes through a series of peaks corresponding to the different channels, which are equally spaced in frequency. The detected output is connected to a counter which indicates the channel. An additional grating reflector trap is used to indicate the starting frequency for counting, as shown in Fig. 4.6.21(c) and (d). Four sets of filters each are required for the

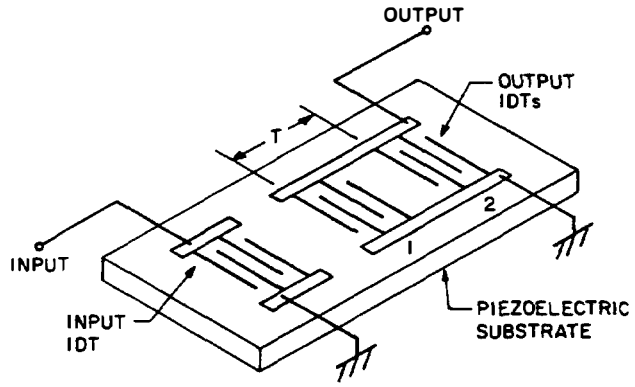


Figure 4.6.20 Fundamental structure of SAW comb filter. (From Kishimoto et al. [59].)

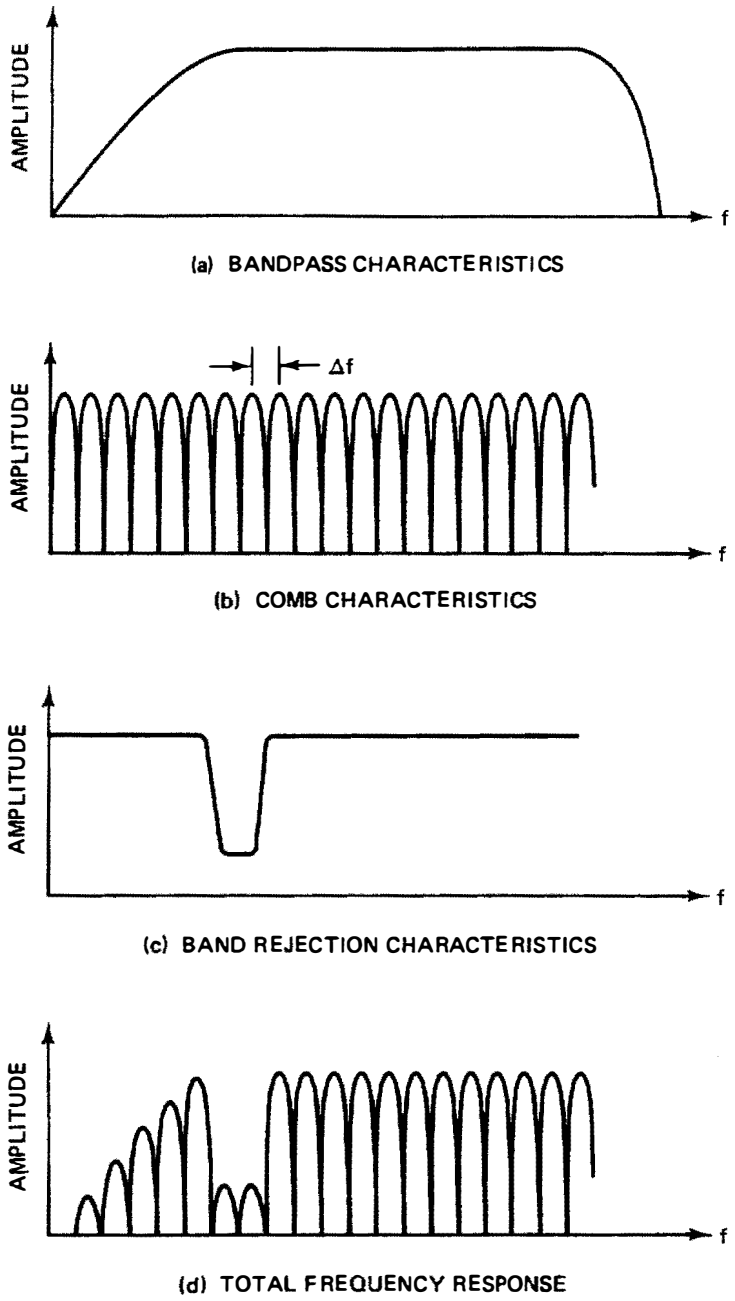
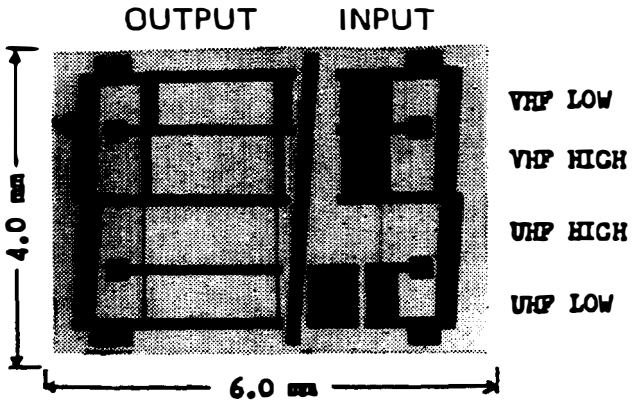


Figure 4.6.21 Frequency response of comb filter: (a) bandpass characteristics; (b) comb characteristics; (c) band rejection characteristics; (d) total frequency response. (From Kishimoto et al. [59].)

VHF low band (150 to 162 MHz), the VHF high band (230 to 276 MHz), and the UHF band (530 to 824 MHz) in current use in Japan. A photograph of the system and its characteristics is shown in Fig. 4.6.22.

A later version of this system has also been used in a feedback circuit, controlling the varactor-tuned oscillator frequency to provide automatic tuning of the oscillator to designated channels.



(a)

		INPUT IDT	GRATING REFLECTOR	OUTPUT IDT
VHF BAND	LOW	APODIZED SPLIT N = 30.5 PAIRS W = 2.0 μm		UNAPODIZED N = 10 PAIRS X 2 W = 5.4 μm
	HIGH	APODIZED SPLIT N = 46.5 PAIRS W = 1.2 μm		UNAPODIZED N = 8 PAIRS X 2 W = 3.4 μm
UHF BAND	LOW	APODIZED N = 61.5 PAIRS W = 1.2 μm	N = 200 W = 1.7 μm	UNAPODIZED N = 7 PAIRS X 2 W = 1.4 μm
	HIGH	UNAPODIZED N = 4 PAIRS W = 1.1 μm		UNAPODIZED N = 4 PAIRS X 2 W = 1.1 μm

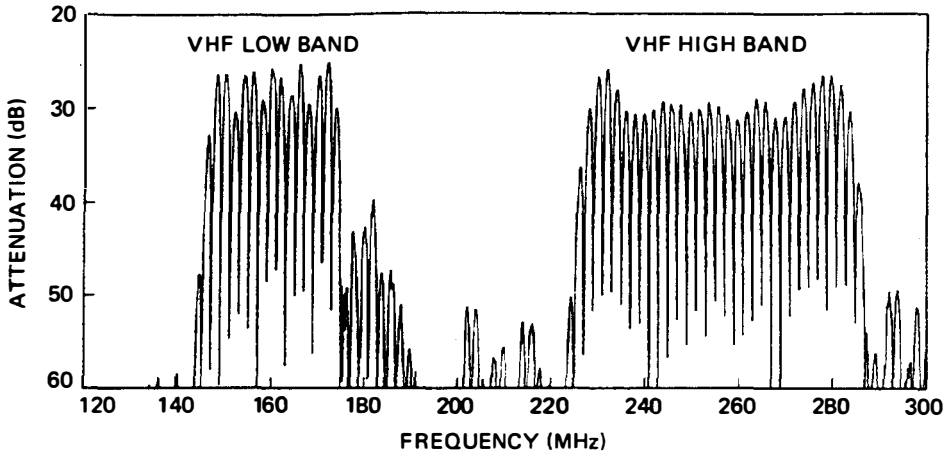
N: NUMBER OF ELECTRODES
W: MINIMUM ELECTRODE WIDTH

(b)

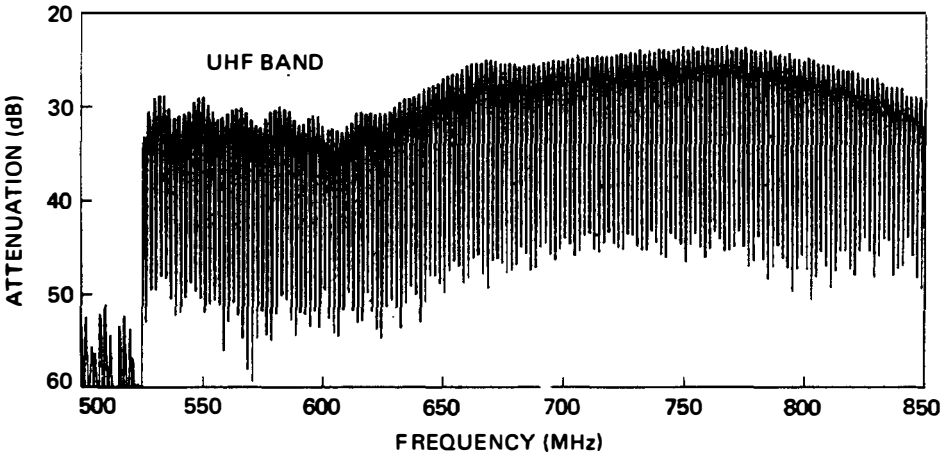
Figure 4.6.22 Frequency characteristics and design features of comb filter, with a photograph of the filter. (After Matsu-ura et al. [60].)

C. Fiber-Optic Recursive Filters

Fiber-optic delay lines also lend themselves well to use as recursive filters for very high frequency signals in the range from several megahertz to several gigahertz. In such cases, we are normally interested in carrying out signal processing on the



(c)



(d)

Figure 4.6.22 Continued

modulation instead of the carrier, which is at optical frequencies. Thus it is important to eliminate interference effects between carriers. For this reason, a relatively incoherent laser beam is used that has considerable phase fluctuation within the shortest modulation time required. This is equivalent to requiring that the laser should have an output bandwidth larger than that of the highest frequency in the modulation.

One basic component used in fiber-optic signal processing devices is a directional coupler. Normally, the light beam is confined to a glass core that, in the case of a *single-mode fiber*, is of the order of 4 to 10 μm in diameter. This glass core is embedded in a glass cladding, with the glass chosen so that the velocity of the wave in the core of the material is slightly slower than that in the cladding. The outer diameter of the cladding is typically of the order of 100 μm. This implies

that a wave approaching the boundary of the core material at a small angle to the axis will be totally reflected at the edge of the core, and that any fields associated with this wave will fall off exponentially into the outer glass cladding. Consequently, the light traveling down the core is inaccessible from outside the fiber.

There are two possible ways to tap a fiber-optic delay line. If, as illustrated in Fig. 4.6.23, the delay line is coiled in a tight coil, some light will leak from the core to the outer radius of the coil because the effective velocity in the cladding will be larger than that in the core, which means that a small amount of light will "leak" outside the fiber. A second technique, illustrated in Fig. 4.6.24, is to lap off the cladding and place the flat surface so formed in contact with or very close to a similarly lapped fiber. This forms a directional coupler, which operates on the same principle as the strip coupler already described, and makes it possible to transfer energy from one fiber to another. By adjusting the length or width of the overlap region, or the spacing between the two fibers, the coupling from one fiber to the other can be changed; thus almost all the power can be transferred from one fiber to the other, or only a small proportion can be transferred. In all cases, the directivity of such couplers is usually better than 60 dB.

A fiber-optic recursive filter can be made using two directional couplers in the configurations shown in Figs. 4.6.25 and 4.6.26, then coupling the signal out of the main fiber into a coiled fiber with a delay time T, and coupling the signal from the delay fiber back into the original fiber. In the simplest case, we can use two 3-dB couplers where half the power is coupled into the delay line and half of that power is coupled back into the original line. At the same time, half the power in the original line is coupled through the second directional coupler into the fiber following the directional coupler, and is not used. In the case illustrated in Figs. 4.6.25 and 4.6.26, a variable delay line is employed so that the output fiber can be translated from one fiber tap to another, to vary the delay T in controllable increments of time.

We consider the operation of this device with a modulated incoherent optical signal. It is not desirable to use purely coherent optical signals, because then there would be interference of the carriers in the two paths, and the interference would

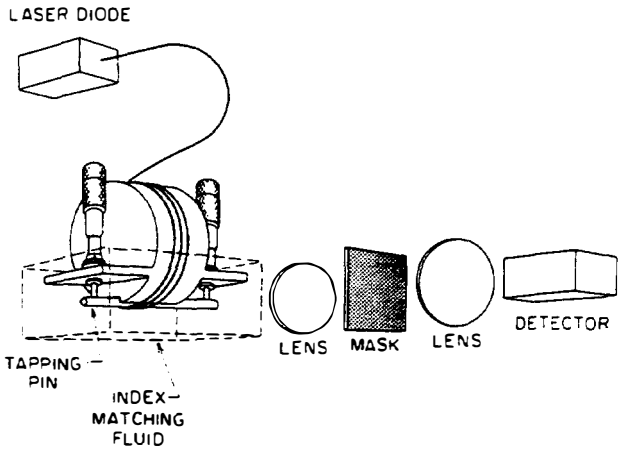


Figure 4.6.23 Macrobend tapped delay line. (After Jackson et al. [10].)

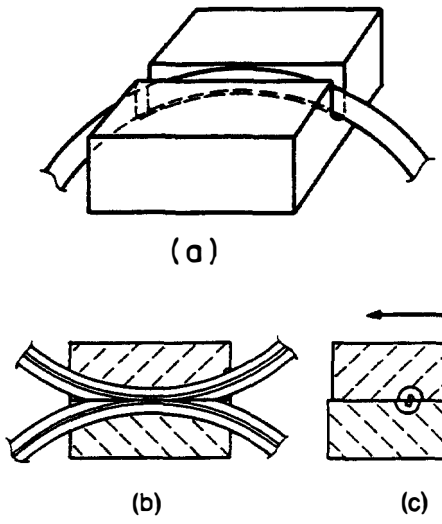


Figure 4.6.24 Construction of a fiber-optic coupler: (a) the fiber is placed in a channel of a solid material, such as silicon or quartz, and bonded in place with epoxy; (b) the two fibers in the coupler are lapped to the core and placed in contact; (c) the coupling may be varied by changing the amount of overlap of the cores. (After Digonnet and Shaw [62].)

depend critically on small delay changes through the fibers due to bending or slight flexing of the fiber. We take the input signal to be of the form

$$u(t) = (1 + A \cos \Omega t)e^{j[\omega t + \phi(t)]} \tag{4.6.38}$$

where ω is the carrier frequency and Ω is the modulation frequency, with A the amplitude of the modulation. The phase of the laser $\phi(t)$ is regarded as fluctuating with time in a random fashion, so that $\langle \phi(t) \rangle = 0$.

After passing through the delay-line path, the signal obtained from the delay line will be of the form

$$u(t - T) = [1 + A \cos \Omega(t - T)]e^{j[\omega(t-T) + \phi(t-T)]} \tag{4.6.39}$$

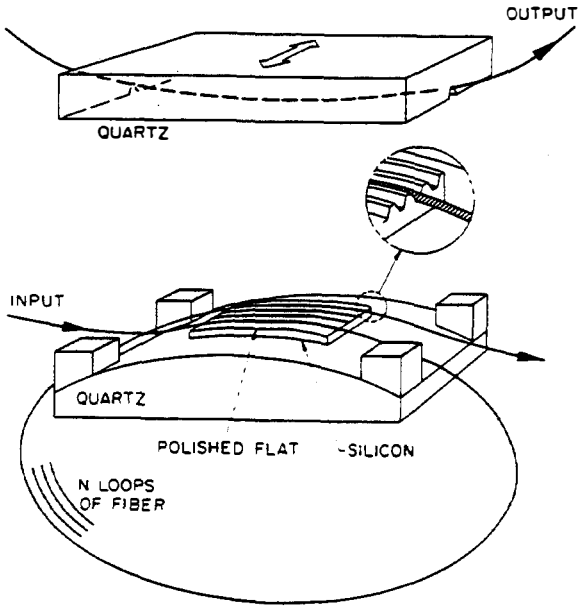
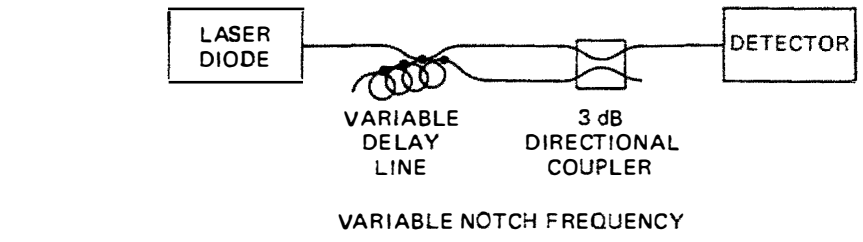
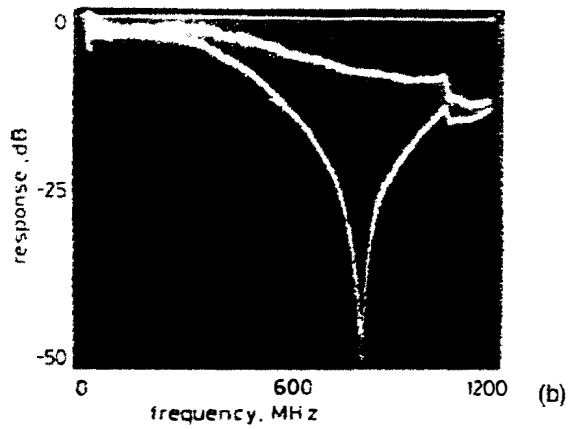


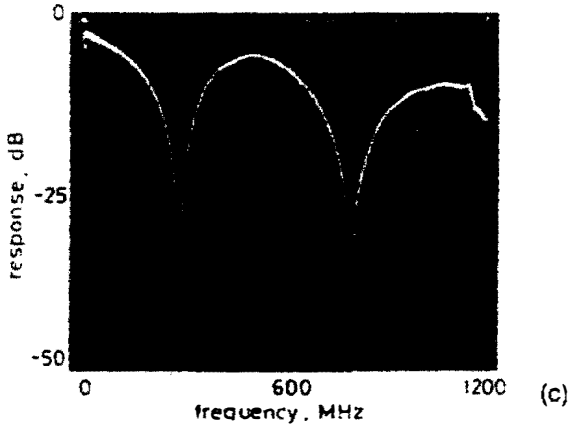
Figure 4.6.25 A variable delay line, in which the output fiber is translated from one input fiber to another. (After Bowers et al. [11].)



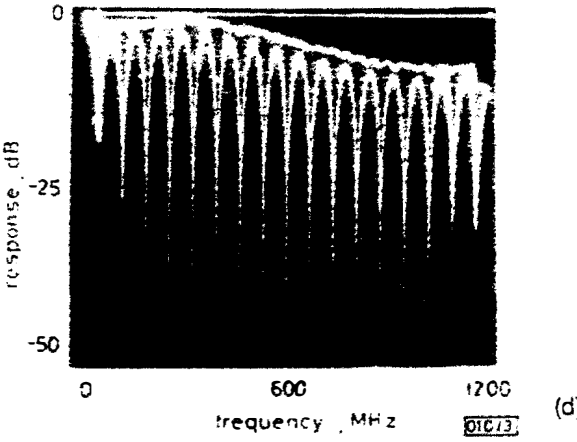
(a)



(b)



(c)



(d)

Figure 4.6.26 Use of variable delay line in a fiber-optic programmable filter: (a) schematic diagram showing source, programmable filter, and detector; (b) frequency response for relative delay of 0.6 ns; (c) frequency response for relative delay of 1.9 ns; (d) frequency response for relative delay of 14.4 ns. Upper trace in parts (b) and (d) shows frequency response of laser diode, amplifier, and detector. (After Bowers et al. [11].)

The sum of two signals $u(t)$ and $u(t - T)$ is inserted into a photodetector that gives an output proportional to the intensity of the light incident on it.

We are concerned with a signal that is the expectation value of the output from the photodetector in a time scale comparable to that of the modulation. This signal is of the form

$$V(t) = \langle [u(t) + u(t - T)]^2 \rangle \tag{4.6.40}$$

Because the laser phase is incoherent, cross-products of the two terms in Eq. (4.6.40) do not give any output. Consequently, the output signal will be of the form

$$V(t) = (1 + A \cos \Omega t)^2 + [1 + A \cos \Omega(t - T)]^2 \tag{4.6.41}$$

If this output signal is passed into a filter so that only signals of frequency Ω can pass through it, we will obtain the result

$$V_\Omega(t) = 4A \cos \Omega\left(t - \frac{T}{2}\right) \cos \frac{\Omega T}{2} \tag{4.6.42}$$

We conclude that the fiber-optic system behaves like a standard recursive delay line, with sharp notches at frequencies corresponding to

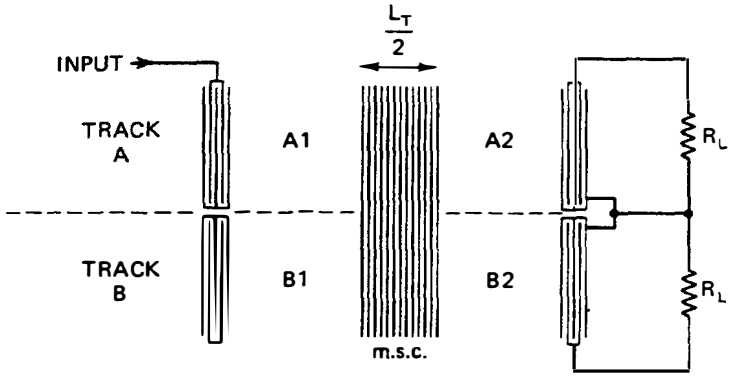
$$f = \frac{2m + 1}{T} \tag{4.6.43}$$

where m is an integer and $f = \Omega/2\pi$. Results obtained by Bowers et al. [11] with a device of this kind are shown in Fig. 4.6.26. Notches approximately 50 dB deep are easily observed, so that the device operates as a recursive filter for the modulation frequency. Coherence of the laser can sometimes be a problem. One way to eliminate this difficulty is to use only one directional coupler, and to take the outputs from the two fibers to separate photodetectors whose outputs are summed electrically. In this case there will be no cross-coupling between the two lasers.

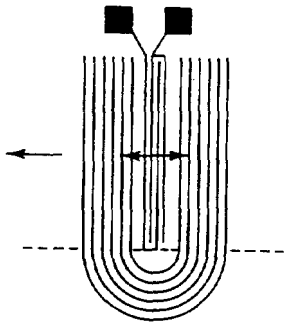
In practice, this type of filter can be very convenient for very high frequency signals. A semiconductor laser source, which is normally fairly incoherent, can be employed, and the laser can be amplitude modulated by modulating the current passing through the laser. Modulation frequencies up to several gigahertz can be obtained, although a very high speed photodetector is required to detect them. Detectors are now being developed for use at modulation frequencies as high as 100 GHz.

PROBLEM SET 4.6

1. Consider a strip coupler with $k_1 \neq k_2$ and coupling coefficient κ. Work out the maximum transfer efficiency of the coupler as a function of $(k_1 - k_2)/\kappa$.
2. Consider a multistrip coupler with an optimum transfer length L_T from track A to track B (i.e., the length for complete power transfer is $L_T = \pi/2\kappa$), as shown [50, 51].



(a)



(b)

(a) Show that if a length $L_T/2$ is employed, the powers in tracks A and B are equal, and the signals are $\pi/2$ out of phase in each track.

(b) Consider the system shown in figure (a). Normally, there is a reflection from an output transducer, which gives rise to the so-called triple transit echo. This is caused by reflections from the output and input transducers in turn, which give rise to a later signal from that desired. In the schematic shown, a strip coupler of length $L_T/2$ is placed across the beam. Two identical output transducers 1 and 2 have identical loads R_L and are placed in tracks A and B, respectively. Show that the signal reflected back into track A is zero, so that the reflections from the output transducers are eliminated.

(c) Consider the system shown in figure (b), where the strip coupler is curved in shape, but still chosen to be of length $L_T/2$. If the input transducer radiates equally in both directions, as it must, show that when it is placed $\lambda/8$ from the centerline between the two strip couplers, all the power radiated from this system will be radaiated in one direction. Thus, by using a "multistrip mirror," we have constructed a unidirectional surface wave coupler with relatively broadband operation. In this problem, ignore the capacity of the curved strip region. Both these systems have been demonstrated as practical configurations.

3. Consider a phase-weighted transducer of the type illustrated in Fig. 4.6.9. Suppose that each transducer is of length $L = VT$, and each is weighted as $w(z) = \cos(\omega_0 z/V + \mu z^2/2V^2)$, where the origin of z is at the center of the transducer. Let the centers of the transducers be separated by a length $H > L$. Write down the time responses $h_1(t)$ and $h_2(t)$ of the two transducers. Then determine an approximate expression for the time response of the complete filter using only the term in the integrand for which square-

law terms cancel out, as in chirp transform theory. From this result, use the method of stationary phase to find the frequency response of the filter.

Answer: The frequency response $H(\omega)$ can be written in the following form:

$$H(\omega) = (1 + j)\sqrt{\frac{\pi}{\mu}}\, e^{j(\omega_0 - \omega)^2/2\mu} \frac{\sin(\omega_0 - \omega)T}{(\omega_0 - \omega)T} \Pi\left(\frac{\omega - \omega_0}{\mu T}\right)$$

4. Consider a strip coupler in which the two acoustic beam paths are of different widths, w_1 and w_2. Let $w_1 = Rw_2$. If a_1 and a_2 are proportional to the voltages at the surface of each beam path, we might expect the coupled mode equations to take the form

$$\frac{da_1}{dz} + jk_1a_1 = j\kappa(a_2 - Ra_1)$$

$$\frac{da_2}{dz} + jk_2a_2 = j\kappa(Ra_1 - a_2)$$

where k_1 and k_2 are the propagation constants in the two beam paths, respectively, and the coupling to the strips is proportional to the width of the beam path. Show that if

$$k_2 - k_1 = \kappa(R - 1)$$

complete power transfer from one beam to the other can be obtained. This process is carried out in practice by keeping $k_1 = k_2 = k_0$ and varying the strip spacing, as illustrated in Fig. 4.7.9(a).

Note: If a_1 and a_2 are taken to be proportional to the square root of the power in each beam, a_1 must be replaced by $a_1/\sqrt{R}$ in these equations. The results obtained will be identical, but the equations will display conservation of power.

5. Consider the problem of designing a recursive filter for a given passband characteristic. Such a filter samples the input signal with a sample spacing T. Using Laplace transform notation $s = j\omega$, a filter can be shown to have a response

$$H(s) = \sum_{p=1}^{P} \frac{A_p}{s - s_p}$$

The corresponding impulse response is of the form

$$h(t) = \sum_{p=1}^{P} A_p e^{s_p t} u(t)$$

where $u(t)$ is a unit step function.

Now suppose that we sample $h(t)$ at times $t = nT$. Then, with the proper normalization,

$$h(nT) = \sum_{p=1}^{P} A_p T e^{s_p nT} u(nT)$$

The Laplace transform of the sampled function is

$$H(s) = \sum_{p=1}^{P} A_p T \sum_{n=0}^{\infty} e^{(s_p - s)nT}$$

$$= \sum_{p=1}^{P} \frac{A_p T}{1 - e^{(s_p - s)T}}$$

Thus we can write $H(z)$ in normalized form as

$$H(z) = \sum_{p=1}^{P} \frac{A_p T}{1 - z^{-1} e^{s_p T}}$$

where $z = e^{sT}$.

Provided that $\omega T << 1$, we expect the sampled filter to give the same response as that of the equivalent analog filter. The result given here is equivalent to that of Eq. (4.6.34), although it is expressed as a sum, not as a product.*

(a) Consider the design of a sampled low-pass Butterworth recursive filter with a response

$$|H(s)|^2 = \frac{1}{1 + (\omega/\omega_c)^{2N}} = \frac{1}{1 + (s/s_c)^{2N}}$$

where $s_c = j\omega_c$. We choose $H(s)$ to have poles at

$$s_p = s_c e^{j\pi/2N} e^{jp\pi/N}$$

where we take only terms for a causal filter for which $\mathcal{R}e(s_p) < 0$. The complex conjugate terms $H^*(s)$ account for the other poles of $|H(s)|^2$.

Taking $\omega_c T = 0.125\pi$ and $N = 2$, find the position of the poles in the s domain and the z domain. Note that the poles occur in complex conjugate pairs.

(b) Express your result in a form similar to Eq. (4.6.34), that is,

$$H(z) = \frac{\alpha_0 + \alpha_1/z}{\beta_0 + \beta_1/z + \beta_2/z^2}$$

and find the values of α_0, α_1, β_0, β_1, β_2.

(c) Following the derivation leading to Fig. 4.6.18, design a recursive Butterworth filter.

(d) Plot the response $|H(z)|^2$ in terms of frequency ω, and compare it with $|H(\omega)|^2$. Plot your results from $\omega = 0$ to $\omega = 16\omega_c$.

(e) Suggest, *without working out numerical values*, the form of the circuit for a recursive filter with $N = 4$. You may place recursive filters in parallel as well as in series.

6. A fiber-optic sensor is used to detect high-frequency vibrations of a surface at a frequency Ω, that is the displacement of the surface is

$$\Delta y = A \cos \Omega t$$

The measuring system illustrated in Fig. 1 is arranged so that light can travel the four possible paths there and back, as shown. In case 3 the light arrives at the substrate earlier than in case 2, but suffers the same total delay T in both cases. In cases 1 and 4 the total delay is either 0 or $2T$, respectively.

(a) Suppose that each directional coupler couples a proportion α of the light power into the neighboring path, leaving $1 - \alpha$ traveling along the original path. What should α be so that light arriving along each path A, B is maximum at the surface being tested?

(b) Suppose that the signal from the laser is incoherent, that is, is of the form $\exp [j(\omega t + \phi)]$, where ϕ changes randomly with time. Then a signal arriving at the substrate with no delay will be phase modulated as

$$e^{j(\omega t + \phi + 2k A \cos \Omega t)}$$

**Reference*: A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing* (Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1975).

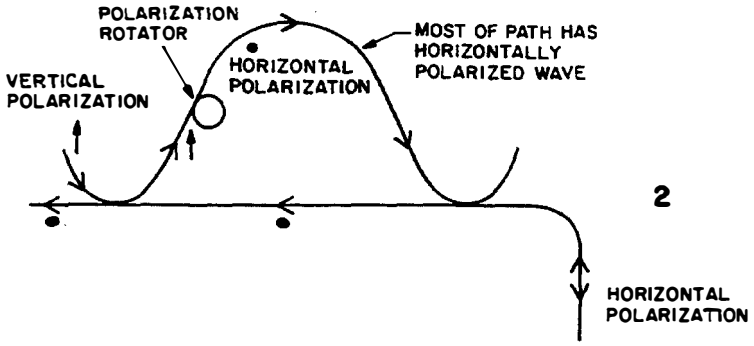
(1)

where $k = 2\pi/\lambda$ is the optical wave number. Show that only for the light traveling on paths 2 and 3, will there be coherent modulation effects in the expected value of the output signal from the square-law detector.

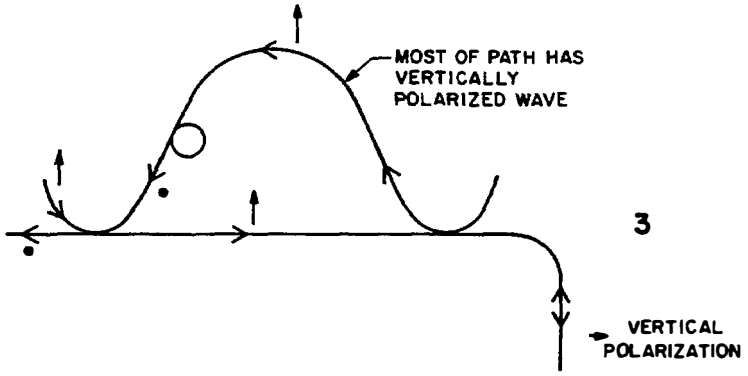
(c) Assume that because light can be polarized, the effective path length changes slightly; thus signals traveling on path A suffer an extra phase shift $\pi/2$ greater than that of path B, in addition to the modulation effects at frequency Ω, as illustrated in Figs. 2 and 3. This can be accomplished by coiling the fiber a few turns around a disk. The stress induced in the fiber changes the velocity of a wave polarized in the plane parallel to the disk and hardly affects the perpendicularly polarized wave. Thus with the correct number of turns and disk radius, the fiber coil will be effectively a quarter-wave polarization plate [9]. The devices are used in the configuration shown as the fiber optic paths 2 and 3. Show that when $kA \ll 1$, the output from the photo-detector at frequency Ω will vary as

$$A \sin \frac{\Omega T}{2}$$

where T is the delay time in the loop. What is the disadvantage of leaving out the extra $\pi/2$ phase shift when A is small?



(2)



(3)

Note: Typically, we might expect A to be 1 Å or less, and the optical wavelength to be 6000 Å. So you can assume that $kA << 1$, and write $\exp(2jkA \cos \Omega t) = 1 + 2jkA \cos \Omega t$.

4.7 CONVOLVERS

4.7.1 Introduction

The transversal filters described so far use fixed tap weights, or tap weights that can be controlled externally at a relatively slow rate that is not comparable to the signal frequencies. It is also useful to have a real-time processing device that can compare an unknown signal to a reference signal and determine the correlation or convolution of the two. Such a device might even change the reference code from pulse to pulse. For example, a spread-spectrum system employing long codes to represent each bit of information could use different codes for each receiver, and these codes could be changed at will. By inserting a reference code, we could feed the tap weights serially into the system instead of adjusting them one by one as we do with an adjustable transversal filter.

One way to convolve two signals, already discussed in Sec. 4.5, is to take the

inverse Fourier transform of the product of the Fourier transforms of two signals; this is mathematically their convolution. Similarly, correlation is obtained by taking the inverse transform of the product of the Fourier transform of one signal and the complex conjugate of the Fourier transform of the other. Such Fourier transform techniques are frequently used in digital processing systems, because they require basically three FFTs, each of order $n \log_2 n$ operations instead of at least n^2 operations, to form a correlation directly.

As we shall see, these considerations do not apply in analog processing because the necessary processes can be carried out in the time domain rather than the frequency domain. Performing the required multiplications and integrations with an analog device is much faster, and certainly requires a less complex system, than with the digital devices needed to take three sets of Fourier transforms. However, an analog system is typically less accurate than a digital system.

In this section we discuss SAW devices that can carry out the direct convolution of two signals by using nonlinear interactions between them; we also describe other kinds of convolvers and correlators based on the use of internal or external nonlinear semiconductor elements. In Sec. 4.8 we discuss closely related nonlinear processing devices in which two tapped CCDs are employed with external multipliers. In Sec. 4.9 we describe acousto-optic devices, which are also used for such processing operations.

4.7.2 Acoustic Convolver

Consider the SAW device illustrated in Fig. 4.7.1. Suppose that complex input signals $f(t)$ exp $(j\omega_1 t)$ and $g(t)$ exp $(j\omega_2 t)$ are supplied to the left- and right-hand sides of the device, respectively. At a position z along a device of length L, the individual signals will be of the forms

$$\hat{f}(t, z) = f\left(t - \frac{z}{V}\right) e^{j\omega_1(t - z/V)} \tag{4.7.1}$$

and

$$\hat{g}(t, z) = g\left(t - \frac{L - z}{V}\right) e^{j\omega_2[t - (L-z)/V]} \tag{4.7.2}$$

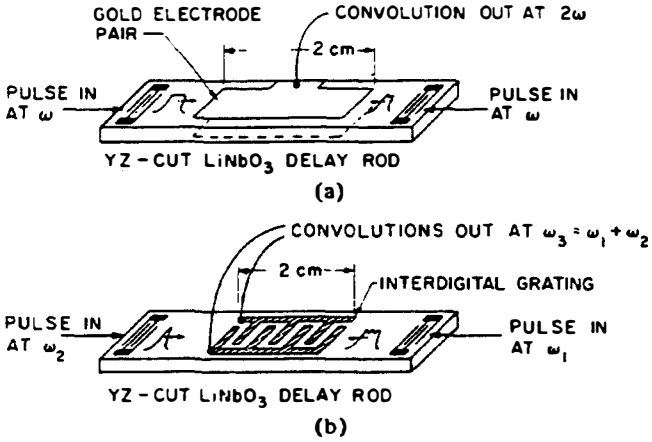


Figure 4.7.1 Transducer configurations used for harmonic generation, frequency mixing, and obtaining convolution between two signals: (a) pair of metal electrodes used as the output transducer for the second harmonic; (b) relatively coarse interdigital transducer used for the sum frequency output at $\omega = \omega_3$. (After Kino [64].)

respectively. We use the notation to indicate a modulated carrier and omit it when describing only the modulation. We assume the medium on which these two waves propagate to be slightly nonlinear. An acoustic wave of amplitude $a(t, z)$ is associated with an electric polarization field in a piezoelectric medium. In turn, a displacement field $\hat{D}(t, z)$ and a displacement current density $\hat{i}(t, z) = \partial\hat{D}(t, z)/\partial t$ will be generated by the applied signal, where $\hat{D}(t, z)$ is of the form

$$\hat{D}(t, z) = K_1\hat{a}(t, z) + K_2\hat{a}^2(t, z) + \cdots \tag{4.7.3}$$

with K_1 and K_2 constants, and the higher-order terms in $\hat{a}(t, z)$, due to the nonlinearity. Because we are concerned with nonlinear interactions, we must take care to express product terms correctly. Thus we use the real part of the waves involved. The interaction between two real waves in a nonlinear medium can be expressed in the form

$$\hat{D}(t, z) = \frac{K_1}{2}[\hat{f} + \hat{f}^* + \hat{g} + \hat{g}^*] + \frac{K_2}{4}[\hat{f} + \hat{f}^* + \hat{g} + \hat{g}^*]^2 \tag{4.7.4}$$

where $\hat{f}^*$ and $\hat{g}^*$ are the complex conjugates of $\hat{f}$ and $\hat{g}$, respectively, and K_1 and K_2 are arbitrary constants. Note that there are product terms due to the second-order nonlinear interaction at frequencies $2\omega_1$, $2\omega_2$, $\omega_1 + \omega_2$, and $\omega_1 - \omega_2$, and at the corresponding negative frequencies.

Suppose that we concentrate on the term of frequency $\omega_3 = \omega_1 + \omega_2$. We can write the displacement field at this frequency in the form

$$\hat{D}_3(t, z) = \alpha f\left(t - \frac{z}{V}\right)g\left(t - \frac{L - z}{V}\right)e^{j\omega_3 t}e^{j(\omega_1 - \omega_2)z/V} \tag{4.7.5}$$

where α is a nonlinear coupling parameter that, for simplicity, includes the phase term $\exp(-j\omega_2 L/V)$.

We see that the phase variation with z of the product signal has the form $\exp[j(\omega_1 - \omega_2)z/V]$. We can therefore detect the product signal with an interdigital transducer of period l, such that

$$\frac{(\omega_1 - \omega_2)l}{V} = 2\pi \tag{4.7.6}$$

When $|\omega_1 - \omega_2|$ is much less than ω_1 or ω_2, the period of output transducer becomes large and a relatively coarse interdigital transducer can be used to detect the output. In the degenerate case in which $\omega_1 = \omega_2 = \omega$, the output is obtained at a frequency $\omega_3 = 2\omega = \omega_1 + \omega_2$, and there is no phase variation with z. In this case, the generated signal in a piezoelectric material will consist of a uniform E field, which can be detected by two metal electrodes on the top and bottom surfaces of the SAW device, as illustrated in Fig. 4.7.1(a). Otherwise, the more general form of the device shown in Fig. 4.7.1(b) must be employed.

We can detect other frequency components by using the correct electrodes. For example, there is a dc term due to the input $\hat{f}$ proportional to $\hat{f}\hat{f}^*$; this has no phase variation with z and can therefore be detected on plate electrodes (see, e.g., Prob. 3). There is also a term of the form $f^2 \exp[2j(\omega t - kz)]$ that can be

detected by an interdigital transducer with half the period of the input transducer. This f^2 term excites an acoustic wave at the second harmonic frequency whose amplitude grows linearly with distance; this second harmonic term can also give rise to saturation effects as it removes power from the fundamental (see Prob. 1).

We shall treat only the parallel-plate output electrode system here. In this case, the output obtained must be proportional to the total charge induced on the plates. If the charge is, in turn, proportional to the displacement field because of the nonlinear interaction, the total output will be of the form

$$y_3(t) = Ke^{2j\omega t} \int f\left(t - \frac{z}{V}\right) g\left(t - \frac{L}{V} + \frac{z}{V}\right) dz \tag{4.7.7}$$

where K is a coupling parameter. By writing $z = V(t - \tau)$, Eq. (4.7.7) becomes

$$y_3(t) = -KVe^{2j\omega t} \int f(\tau) g(2t - \tau - T)\, d\tau \tag{4.7.8}$$

where $T = L/V$ is the time delay between the transducers. Note that we can take the limits of the integral in Eq. (4.7.8) and $+\infty$ and $-\infty$ when each signal consists of a pulse whose length is less than the transit time under the detecting transducer, and when the signals overlap each other while passing under the transducer. The output obtained at a frequency 2ω is the convolution of the two input signals, but it is compressed in time by a factor of 2 because of the $2t$ term in the argument of G. The two signals are passing each other at a velocity $2V$, instead of one signal moving at a velocity V through a stationary filter; the output is therefore a scaled version of the convolution of the two signals. For the same reason, the output is at a carrier frequency of 2ω or, more generally, $\omega_1 + \omega_2$.

We can put this result on a more physical basis by considering what occurs when two square pulses of the type illustrated in Fig. 4.7.2 are injected into the device. As the two pulses pass by each other, we obtain an output that is the integral of the product of the two pulses over their overlap length. This integral increases linearly with time until the two pulses are coincident with each other. After they pass each other, the area of overlap begins to drop, and the amplitude decreases linearly with time until it becomes zero; thus we obtain the convolution of the two signals. If each pulse is of length T, the overlap between the zero output points will also be T, as we would expect from Eq. (4.7.8). Simple results obtained with a device of this type are shown in Fig. 4.7.3. In each case, a symmetric waveform has been convolved with itself.

We often require the correlation between two signals instead of their convolution; for this we need the function

$$y(t) = \int f(\tau) g(t + \tau)\, d\tau \tag{4.7.9}$$

If we compare Eq. (4.7.9) with Eq. (4.7.8), we see that a convolver can be used to obtain correlation only if one of the signals, $f(\tau)$ or $g(\tau)$, is inverted in time. This can usually be done if one of the signals is used as a reference, but generally the device yields only the correlation of two symmetrical signals and the convolution of two arbitrary signals.

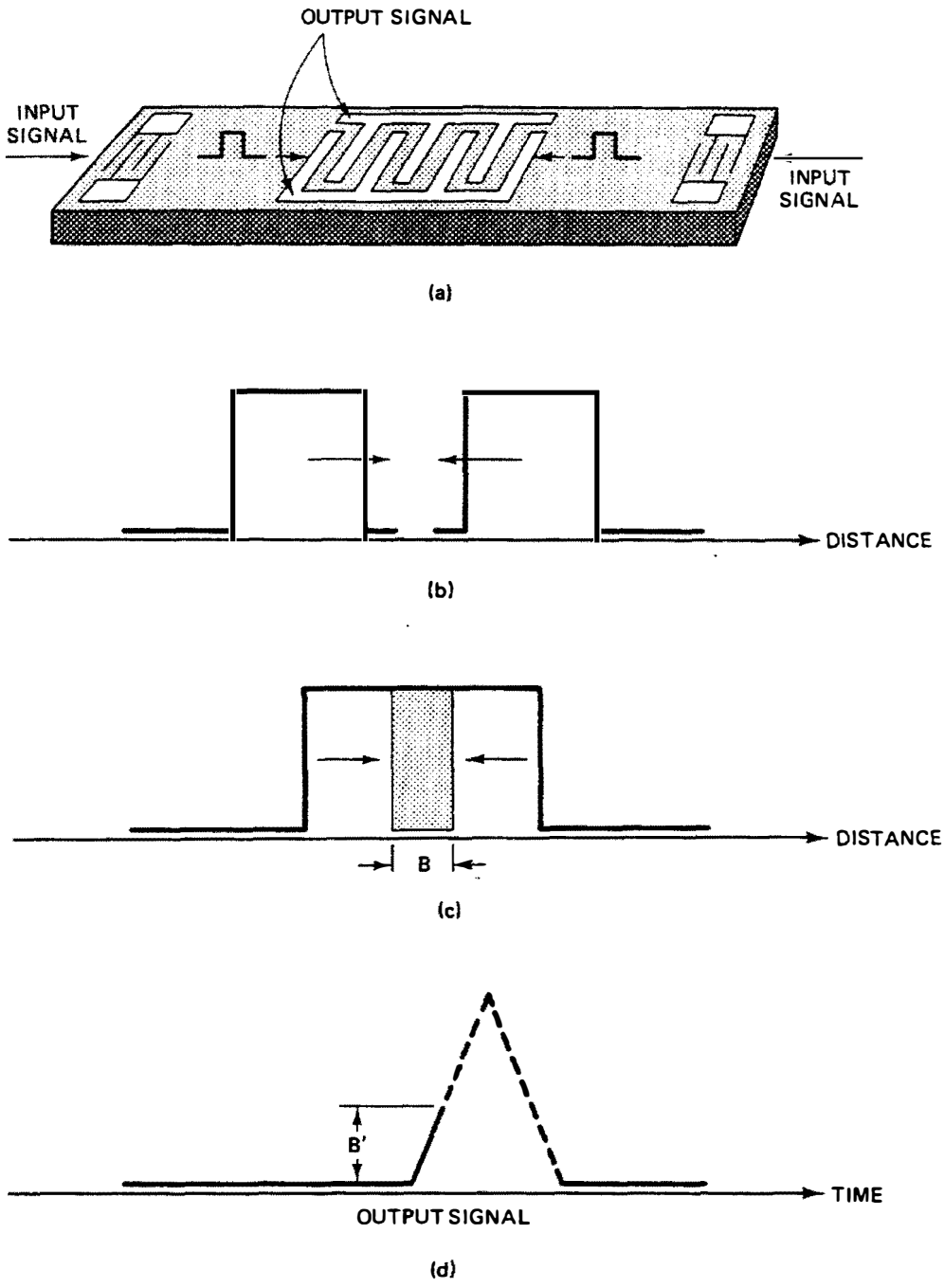


Figure 4.7.2 Nonlinear signal processing is accomplished by introducing two signals at opposite ends of an acoustic wave device (a), and extracting the mathematical result from transducers in the middle. The acoustic wave representations of two rectangular pulses approach each other (b) and begin to overlap (c). When the acoustic signals first touch, the output signal begins to rise (d). When the overlapping signals have proceeded to B, the output signal has risen to B'. The output signal reaches a peak when the two acoustic signals are exactly superposed, and then begins to fall again. Thus the output signal is proportional to the shaded area in part (c). The output is termed the convolution of the two input signals. (After Kino and Shaw [13].)

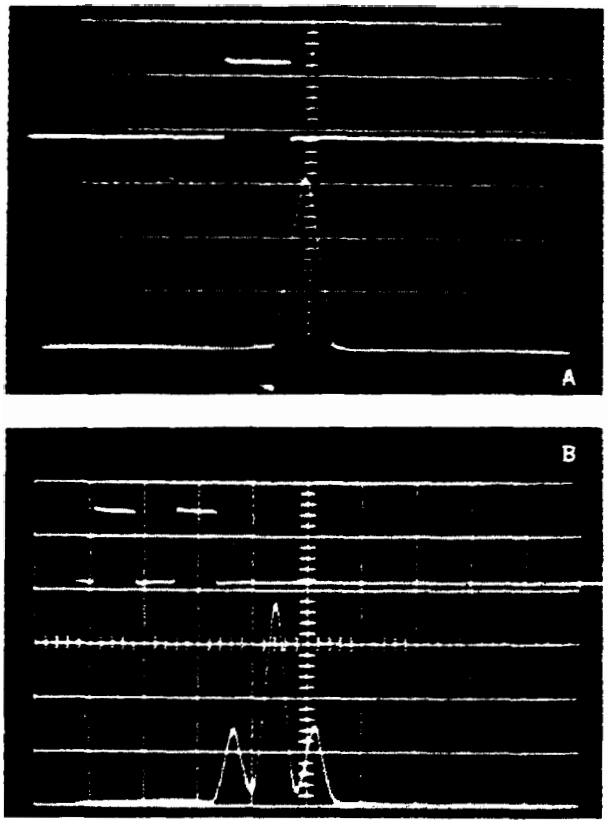


Figure 4.7.3 Autoconvolution of (a) a rectangular pulse observed using a surface wave convolver with an interdigital output transducer; and (b) a double pulse, with the same configuration and frequencies. (After Kino and Matthews [63].)

Devices of this kind can be used to reverse a signal in time to obtain the necessary reference for correlation by using the configuration shown in Fig. 4.7.4. A signal of frequency ω_1 is inserted into the left-hand transducer while a short-pulsed signal of frequency ω_3 is inserted into the center transducer. It can be shown by the process already illustrated that we will obtain an output signal of frequency $\omega_2 = \omega_3 - \omega_1$ traveling in the left-hand direction. When an asymmetrical pulse arrives under the center transducer, a signal will be generated that is the product of the two signals. However, since the regeneration occurs directly under the transducer, the tail end of the original pulse will arrive back at the

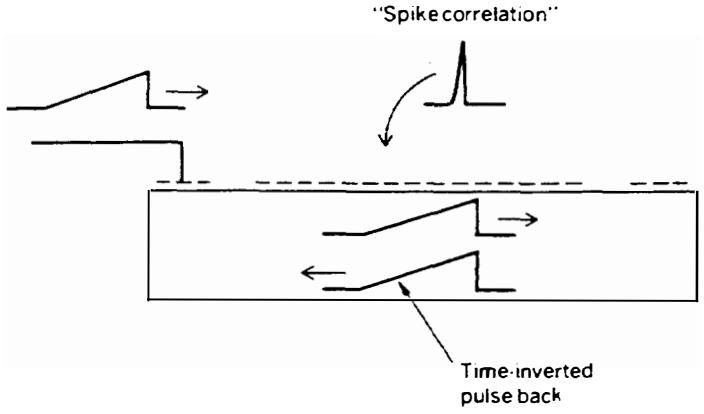


Figure 4.7.4 Configuration used to obtain time reversal of an input RF pulse.

original input transducer first and the front part of the pulse will arrive last; thus the output will be a time-inverted version of the input pulse.

Figure 4.7.5 shows an illustration of results obtained by this process. A short pulse or delta function is applied to the center transducer and a signal is introduced in the left transducer. The time-inverted signal then appears at the original input transducer. This process has been used with two convolvers to generate the correlation between two signals. However, this two-step process is not a practical way to obtain correlation, for it results in a poor signal-to-noise ratio.

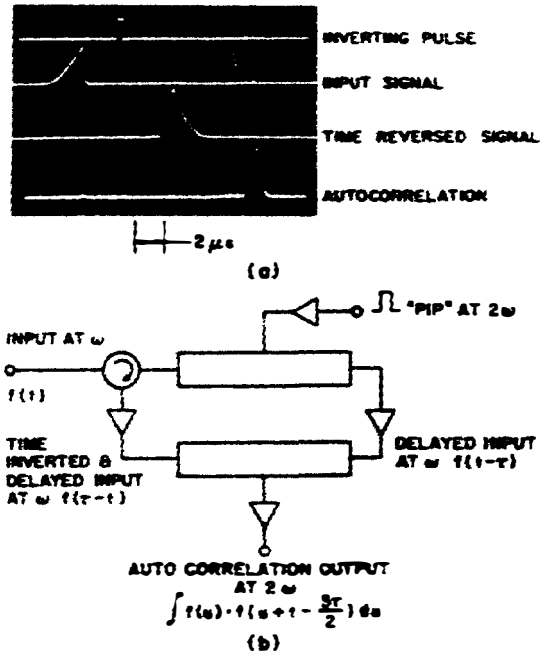


Figure 4.7.5 Inversion of electronic signal can be accomplished by nonlinear signal processing using, for example, the acoustic wave device illustrated in Fig. 4.7.1(a). (a) A signal (top oscilloscope trace) is introduced at the transducer at the left-hand side of the device. A short pip (middle trace) is inserted by way of the transducer in the center of the device. A time-inverted signal (bottom trace) then appears at the original input transducer. The autocorrelation of an asymmetrical triangle with two semiconductor SAW convolvers, is also shown. (b) A schematic of the circuit used. (After Kino [64].)

Nonetheless, the principle of time reversal or phase conjugation is still important (see Probs. 4.7.2 and 4.8.1). When a signal $\exp [j(\omega t - \phi)]$ is injected into a device excited by a signal $\exp (2j\omega t)$ on its plate, the signal generated by the nonlinearity is of the form $\exp [j(\omega t + \phi)]$; thus the phase is reversed. If the phase ϕ contains a term due to distortion in a transmission medium, the convolver will reverse the sign of the phase distortion; this process is called *phase conjugation*. If the phase-conjugated signal is retransmitted through the medium, the frequency component $\exp (j\omega t)$ will arrive without distortion. Thus a two-dimensional optical beam passing through a distorting medium, such as a turbulent atmosphere, and reflected from a target can in principle be regenerated in a crystal excited by a uniform field at a frequency 2ω. If the phase-conjugated beam is reemitted from the crystal and sent through the medium, it will produce an undistorted reflected image (see Prob. 4.8.1) [66].

4.7.3 Spread-Spectrum Communications

The convolver has a sophisticated application to *spread-spectrum communications*. If biphase (0 and π phase) or quadriphase (0, $\pi/2$, π, $3\pi/2$)-coded waveforms are

used to represent one bit of a digital signal, then one bit of a digital signal can consist of several hundred "chips" of a spread-spectrum code. If the bit is 0, the code will not be transmitted; if the bit is 1, it will be. A coded waveform may consist of many "chips" per bit and typically employs a far larger bandwidth than the original digital signal. When the biphase-coded waveform is correlated with itself, it will produce a correlation peak corresponding to the bit of interest, but when the biphase-coded waveform is absent, there will be no correlation peak. A spread-spectrum system of this kind, because it employs a much larger bandwidth than the actual signal itself, is better able to reject unwanted signals and will improve signal-to-noise ratio considerably.

An important application of this system uses different codes for different receivers, so that several codes can use the same wideband communication channel, with the signals addressed to different receivers. In a conventional communication system, the receiver is tuned to the frequency of the signal; in a spread-spectrum system, the receiver code is changed to select the signal required. This gives us two advantages: Because relatively long codes and correlation techniques are employed, the signal-to-noise ratio of the received signal can be very small; and since the types of codes employed are pseudorandom codes, it is very difficult to intercept spread-spectrum signals without prior knowledge of the code.

A spread-spectrum system needs a device that can correlate with the coded waveform over relatively long integration times, of perhaps 1 ms or more. A *PN-coded waveform* (pseudonoise or pseudorandom code) is often used for this purpose. In one example, due to Morgan et al. [67, 68], an SAW convolver is employed to correlate segments of the waveform approximately 30 μs long. This produces correlation peaks, which are subsequently summed coherently in a recirculating loop of the kind illustrated in Fig. 4.7.6. The receiver signal input to the convolver is a biphase PN waveform normally buried in noise. The reference input is a biphase waveform, coded so that segments of length T_d match the time-reversed version of corresponding segments of the signal code. The convolver output is therefore a series of correlation peaks with a spacing T_d. To add these peaks coherently, the recirculation loop has a loop delay equal to T_d. The loop is cleared by opening a gate after a number of circulations designated N_c. The total signal-to-noise ratio ideally is increased by the same ratio as for a matched filter, correlating a length $N_c T_0$ of a waveform, where T_0 is the time delay in the convolver parametric region.

Perhaps the most critical part of this system is the recirculation loop, which consists of a wideband delay line. The gain in this loop must be equal to unity; if it is too large, the signals corresponding to early correlation peaks will continue to increase in amplitude, as will the noise associated with them; if it is too small, the signals from the earlier circulations will become negligible in amplitude. The loop is adding the $(N + 1)^{\text{th}}$ correlation peak to the sum of the previous N peaks. If the bandwidth of the earlier peaks decreases as they circulate, again there will be severe distortions and we will lose the peak amplitude of the correlation pulse. A simple delay line with a $(\sin f/f)^2$ response would have a response after N loops of $(\sin f/f)^{2N}$. Therefore, a delay line must be designed with a very flat response over the bandwidth of interest. In the examples illustrated in Figs. 4.7.6 and 4.7.7,

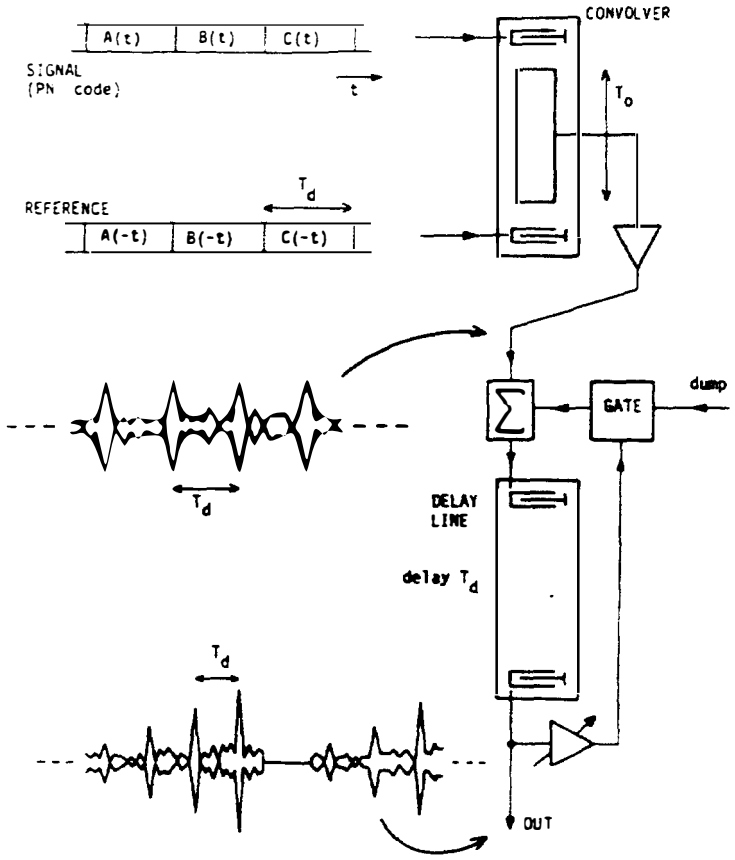


Figure 4.7.6 Principle of spread-spectrum correlator. (After Morgan and Hannah [67].)

two sets of transducers in parallel, with slightly different center frequencies, are used.

Small temperature drifts that, again, would distort the output signals must also be eliminated. This is done by using a quartz delay line with a crystal cut, which has a very small change in transit time with temperature.

Figure 4.7.7 illustrates the buildup of the output in the recirculation unit after 30 circulations with a 9.2-Mbit/s code applied to the input convolver. When the gain is adjusted correctly, the signal grows linearly with time. An automatic-gain-control (AGC) system, with a test pulse passed through the loop, will keep the gain exactly constant, but this precaution is not usually necessary.

Figure 4.7.8 shows results obtained by testing the system with an applied biphase PN signal waveform with a rate of 10 Mbits/s. The output for the seventieth circulation, when noise was added to the signal at the convolver input to give an input signal-to-noise ratio of $-$ 41 dB, is shown. The thermally generated noise was filtered to a bandwidth of 25 MHz, and the time delay T_D was 30 μs. The top trace is a signal without any input noise; the bottom trace is a signal with noise. The measured output signal-to-noise ratio was 4 dB, with an input signal-to-noise ratio of $-$ 41 dB. In theory there should have been an increase of $2BN_cT_D = 2 \times 25 \times 70 \times 30 \approx 100{,}000$ or 50 dB (in a convolver, the output bandwidth is

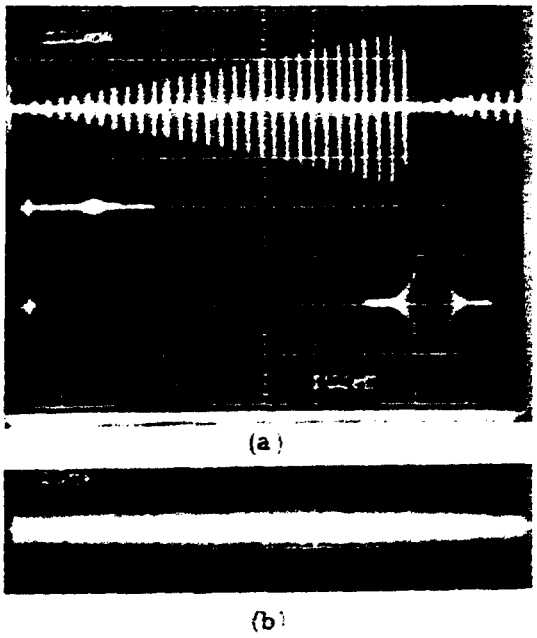


Figure 4.7.7 Loop output waveforms for input waveforms: (a) with no noise added at the convolver input (details of the output pulses are shown at 200 ns/div); (b) twenty-eighth output pulse, for a signal-to-noise ratio (SNR) of −30 dB at the convolver input, at 2 μs/div. (After Morgan et al. [68].)

twice the input bandwidth), which would correspond to a signal-to-noise ratio at the output of 9 dB. But in practice, due to the limited bandwidth of the recirculation loop and its distortions, the gain was 5 dB less than this value. Nevertheless, this technique can detect signals 41 dB below the noise level.

The same system was also demonstrated with a narrower-band input signal, using a 1-Mbit PN code and a noise bandwidth of 2MHz, with the input signal 34 dB below noise. After 70 circulations, the output signal-to-noise ratio was 4.5 dB, which corresponded very closely to the theoretical value of 5.2 dB.

This system demonstrates the great advantage of using spread-spectrum waveforms where a wide bandwidth input signal is employed with a matched filter to

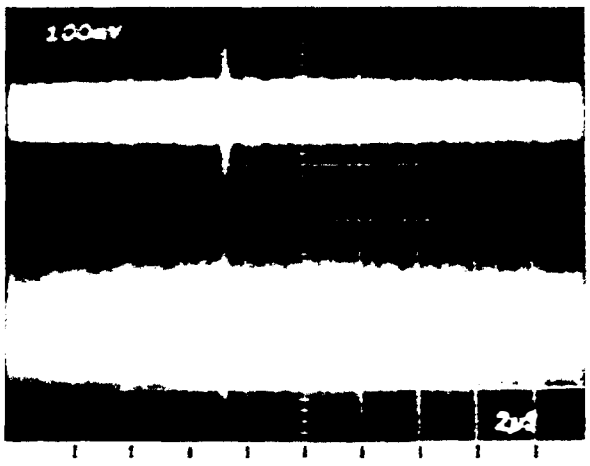


Figure 4.7.8 Output of recirculation unit for seventieth circulation, for 10-Mbit code. Lower trace, with input SNR = −41 dB; upper trace, no noise applied; horizontal, 2 μs/div. (After Morgan and Hannah [67].)

improve the signal-to-noise ratio. Analog systems of this kind are relatively simple to use for side bandwidth signals. The same sort of results have been obtained in digital systems with CCDs and very long codes, but as the need for wider bandwidths increases, SAW devices become particularly suitable.

Spread-spectrum systems for communications. One might think that a spread-spectrum system of this nature makes highly uneconomical use of the radio-frequency spectrum available. Its advantages, of course, are the great improvement it makes in signal-to-noise ratio and its ability to employ many different codes of signals within the same frequency band. This nullifies the disadvantage of using a wide bandwidth for an individual signal. The separation between the different signals is now carried out essentially in the time domain rather than the frequency domain.

The example of spread-spectrum operation given here uses a digital biphase code, but this is not the only possibility. Other techniques employ *frequency hopping modes*, where the frequency rather than the sign of the amplitude is changed from pulse to pulse. Such modes offer the advantages already discussed, and are also extremely difficult to jam or intercept unless the observer knows the exact code.

Systems of this kind require only the correlation peak, not the full correlation function. In principle, such a correlation peak can be obtained by inserting two identical signals into a mixer, taking their product, detecting the output, and integrating the resultant detector output for a time T. The output would then be proportional to $\int_0^T f^2(t)\,dt$, where $f(t)$ is the input signal. An analog system is useful for this purpose, because high-speed multiplication requires high-speed A/D converters and a large, more complex, digital system. But on the other hand, a digital system is well suited to very long narrowband spread-spectrum systems, such as those used for communicating with distant satellites in the neighborhood of the outer planets.

This still leaves us with the problem of finding where the two input codes, the reference and the received signal, are coincident in time, for without coincidence between the codes, no output is obtained. In practice, we find it by altering the time of the reference code several times until we obtain an output. Reference codes can be used at the start of a message for this purpose.

This locking-on process may be relatively slow, because the input signal-to-noise ratio (SNR) is typically -30 to -40 dB. Hence, for a bandwidth of 10 MHz, integration over 1 ms is necessary for each trial to obtain a satisfactory SNR to make a decision. The key advantage of a matched filter, such as a convolver, is that the input signal and the reference signal do not have to be coincident. Provided that the time difference between them is less than the time delay through the convolver, we will obtain an output, and the relative improvement in acquisition time will be equal to the time–bandwidth product of the filter. Thus the locking-on process is greatly simplified and sped up by a factor comparable to the time–bandwidth product of the convolver. Again, this system is far simpler and smaller

than the equivalent digital system, which would have to be extremely large and complex to handle large-bandwidth short-pulse signals.

Range finding. We have described the use of spread-spectrum signals in communications. Another important application is *range finding*. Suppose that a fixed transmitter A emits a signal code with a carrier frequency f_1; the position of a receiver B is required. The received code is reemitted from B with a carrier frequency f_2 and received back at A. The signal received at A is correlated with the original transmitter code. By determining the time of the correlation peak, we can directly measure twice the delay time from A to B, and hence find the range from A to B. We can measure this time, under very noisy conditions, to an accuracy of one chip of the code. Digital versions of this system have been used to measure, with extreme accuracy, the distance of satellites approaching the planets and of course, to communicate with them as well.

This method is also used in the NAVSTAR global navigation system [69]. Each satellite in the system carries an atomic frequency standard and transmits two 1-GHz spread-spectrum signals whose codes are synchronized with the atomic clock. A user with an accurate knowledge of time (i.e., with an atomic clock in his or her receiver), can determine the arrival time of signals from as many as six satellites overhead, and can determine his or her position with an accuracy of 10 to 50 m.

4.7.4 Waveguide Convolver

The convolver we have described above has a major disadvantage: Its operation is based on the use of a weak nonlinearity in the medium. We shall give a heuristic argument to predict how the output signal depends on the input signal amplitudes and the dimensions of the device. The amplitudes of the input signal fields are proportional to the square root of the input signal power intensities of $(P_1/wd)^{1/2}$ and $(P_2/wd)^{1/2}$, respectively, where w is the acoustic beamwidth, the parameter d is proportional to the wavelength and approximately equal to the penetration depth of the surface acoustic wave, and P_1 and P_2 are the acoustic input powers. The dipole polarization field P_D generated by the two input signals can be shown to be proportional to the products of the input fields or power intensities. This polarization is developed over a region of effective thickness d, the generation layer. The voltage V_D across this dipole layer is $V_D = P_D d/E$, where E is the dielectric constant and we have taken P_D to be the average value of polarization in this region. Hence it follows that

$$V_D = \frac{\kappa}{\varepsilon}\left(\frac{P_1}{wd}\right)^{1/2}\left(\frac{P_2}{wd}\right)^{1/2} d \tag{4.7.10}$$

where κ is a coupling constant.

Suppose that the length of the plates is L and that of the generating region is L'. The capacities of these regions are $C = \varepsilon Lw/h$ and $C' = \varepsilon L'w/h$, respec-

tively, where h is the plate spacing. It follows that the open-circuit voltage across the plates, loaded by their self-capacity C, is

$$V_{op} = \frac{V_D C'}{C} = \frac{\kappa L'}{wL\varepsilon} (P_1 P_2)^{1/2} \tag{4.7.11}$$

Thus the open-circuit voltage output of these devices is proportional to $1/w$. If the pulse length is less than that of the output plate, the signal will rise linearly with pulse length, as we would expect from our earlier arguments.

The output power P_3 into a resistive load is proportional to V_{op}^2. Thus P_3 will be proportional to the product of the input powers P_1 and P_2, and we can write the efficiency of the device in the form

$$F_T = \frac{P_3}{P_1 P_2} \tag{4.7.12}$$

For a typical, practical device made of lithium niobate, working at a center frequency of 100 MHz with a beamwidth of 1.25 mm, this number is approximately -85 dBm. This implies that when $P_1 = 20$ dBm (0.1 W), the ratio P_3/P_2 equals -65 dB. The noise level of a receiver is a few decibels worse than kTB, where k is Boltzmann's constant, T is room temperature, and B is the bandwidth. A typical noise figure for a 30-MHz input bandwidth and a 60-MHz output bandwidth is approximately -95 dBm. The maximum input signal level allowed by the transducers before arcing is approximately 20 dBm, so the maximum output level is $(-65 + 20)$ dBm $= -45$ dBm. This result implies that the maximum dynamic range, from the largest input signal to the minimum detectable level, is 50 dB. This dynamic range can severely limit the use of such devices.

Two alternative strategies have been adopted to decrease the problem. One involves the use of a very narrow acoustic beam in an acoustic waveguide; the other makes use of the interaction of a surface acoustic wave with a semiconductor. The first technique uses a thin metal strip, perhaps 50 to 100 μm wide (two to three wavelengths), laid down on the piezoelectric substrate to form an acoustic waveguide. A wider interdigital transducer, perhaps 1 mm wide, is used with a strip coupler to reduce the beamwidth to 100 μm, as illustrated in Fig. 4.7.9(a). The wide input beam is needed so that the impedance of the input transducers will be low enough for matching and the voltage across the fingers for a given input power will be low enough to avoid arcing between the fingers. The metal strip is laid down on the substrate, as illustrated in Fig. 4.7.9(b). The metal short-circuits the electric fields of the wave, causing the acoustic wave propagating beneath it to have a slightly slower velocity than that of the wave in the material outside the metal strip. Consequently, by Snell's law, as illustrated in Fig. 4.7.9(b), a beam approaching the edge of the strip at an angle θ_i will be transmitted at an angle θ_T, such that

$$\frac{\sin \theta_i}{\sin \theta_T} = \frac{V_M}{V} \tag{4.7.13}$$

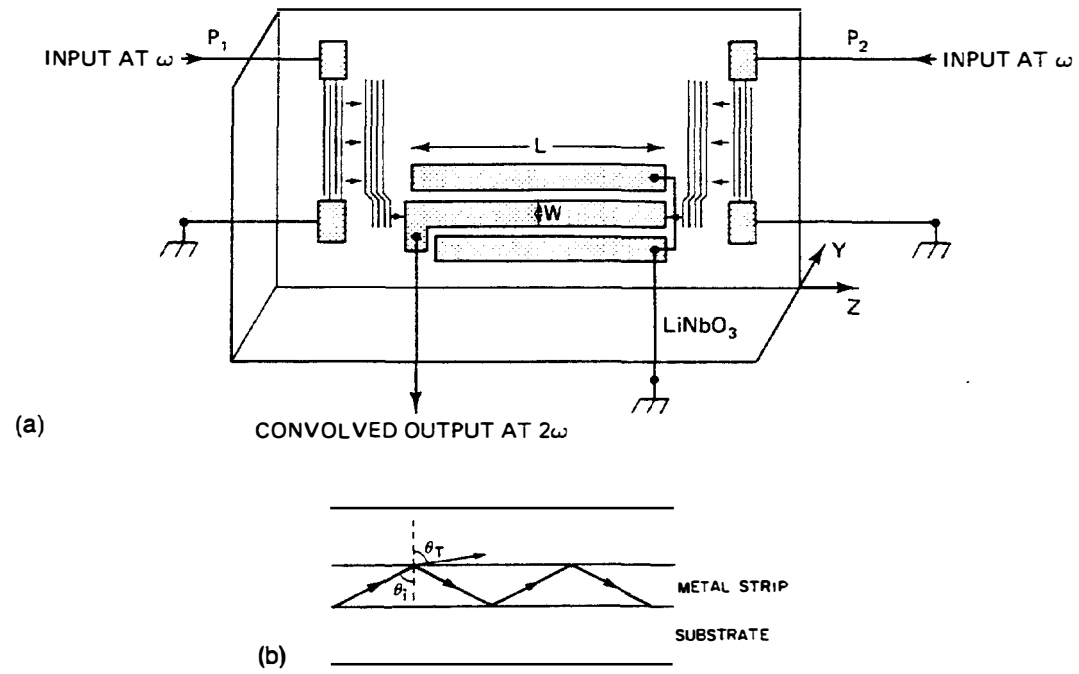


Figure 4.7.9 (a) Waveguide convolver configuration. (After Defranould and Maerfeld [70].) (b) Schematic of the interaction of the waves in the guide. Normally, θ_T is imaginary for the waveguide mode of interest. Unwanted modes may have θ_T real, causing them to radiate sideways and be rapidly attenuated.

where V_M is the wave velocity under the metal. Over a certain range of θ_i, if $V > V_M$, then θ_T will be imaginary and there will be total reflection of the waves at the two edges of the metal. If we arrange the input beam to the strip to be almost parallel to the edges of the strip, we can satisfy this requirement fairly easily, for the strip acts as an SAW waveguide to propagate the narrow beam.

Convolvers of this type have been constructed by Defranould and Maerfeld[70], with the metal stripitself used as one output electrode and two neighboring strips used as the ground electrodes, as illustrated in Fig. 4.7.9(a). By this means, they were able to obtain an improvement in efficiency of approximately 20 dB, an F_T of -65 dBm, and a dynamic range of 70 dB. Such devices are now being employed in military radar systems and developed for use in spread-spectrum communications. Other methods that have been used to compress the beam employ waveguide horns or focused transducers [71, 72].

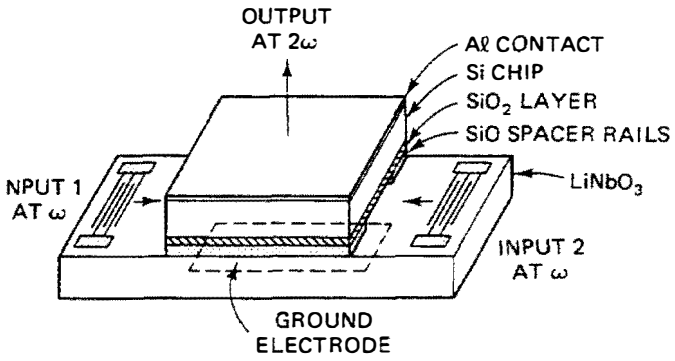
4.7.5 Semiconductor Convolver

An alternative is to employ one of the two configurations shown in Fig. 4.7.10, in which the electric fields associated with the acoustic waves interact with a semiconductor [64, 73–75]. In the first case, a slab of silicon is separated from a piezoelectric substrate by rails, which eliminates the mechanical loading of the

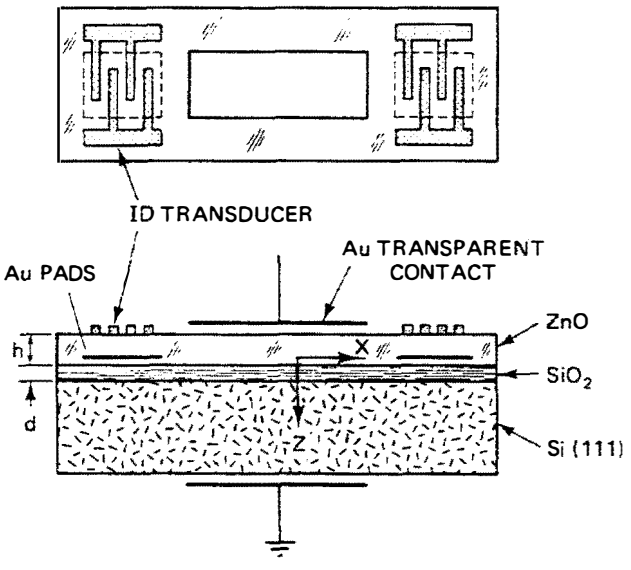
surface acoustic wave. The air gap required is typically a few thousand angstroms. In the second case, a layer of zinc oxide, a piezoelectric material, is laid down on the silicon; this gives piezoelectric coupling to a surface acoustic wave on the silicon substrate. The nonlinearity is now due to the interaction of the electric field with the semiconductor, with the semiconductor behaving like a type of varactor.

Suppose that the doping density of an n-type semiconductor is N_d per unit volume and the semiconductor is depleted at its surface, with the potential at its surface negative in respect to the bulk. If the normal component of the electric field is E_0 at the surface of the semiconductor, the surface will be depleted to a depth h. Integrating the one-dimensional form of Poisson's equation, it follows that

$$\frac{dE_0}{dy} = \frac{\rho}{\varepsilon} = \frac{qN_d}{\varepsilon} \tag{4.7.14}$$



(a)



(b)

Figure 4.7.10 (a) Schematic of the air gap convolver configuration. (After Kino [64].) (b) Schematic of the monolithic zinc-oxide-on-silicon convolver. (After Khuri-Yakub and Kino [74].)

where q is the charge of a donor, y is a coordinate normal to the surface, and ρ is the charge density per unit volume. Therefore, the field at the surface is

$$E_0 = \frac{qN_d h}{\varepsilon} \tag{4.7.15}$$

where it has been assumed that $E_0 = 0$ at the edge of the neutral region. By a further integration, it follows that the potential at the surface is

$$\phi_0 = -\frac{qN_d h^2}{2\varepsilon} = -\frac{E_0^2 \varepsilon}{2qN_d} \tag{4.7.16}$$

We suppose that RF fields $E_1(z, t)$ and $E_2(z, t)$, excited by the input signals at each end, respectively, are applied at the surface in addition to the dc field already present. It is apparent that in this case a voltage proportional to $(E_0 + E_1 + E_2)^2$ is generated at the surface. This yields an output potential with a component $E_1 E_2$ that can be detected by the metal electrode illustrated in the two configurations of Fig. 4.7.10.

This system is far more sensitive, albeit more complex, than the simple lithium niobate convolver. The measured efficiencies F_T of devices of this kind are in the range -42 to -60 dBm. Although the devices saturate at lower power levels than the simple nonlinear acoustic convolvers, the dynamic range is approximately 60 to 70 dB, and devices of this kind have been used in radar systems with time–bandwidth products of up to 2000 (22 μs length, 96 MHz bandwidth) [75–77].

Most of the existing development has been made in the *air gap convolver*. The versions made at Lincoln Laboratories and Texas Instruments are extremely stable, with spacings between the semiconductor and the piezoelectric material of only 3000 Å [75–77]. However, the optical finishing required makes these devices expensive and difficult to reproduce on a large scale. The zinc oxide on silicon monolithic convolver, as developed at Stanford University and elsewhere, shows good efficiency but has a narrower bandwidth than the air gap convolver [74]. Because of its simplicity, the use of the lithium niobate waveguide convolver appears to be a better choice.

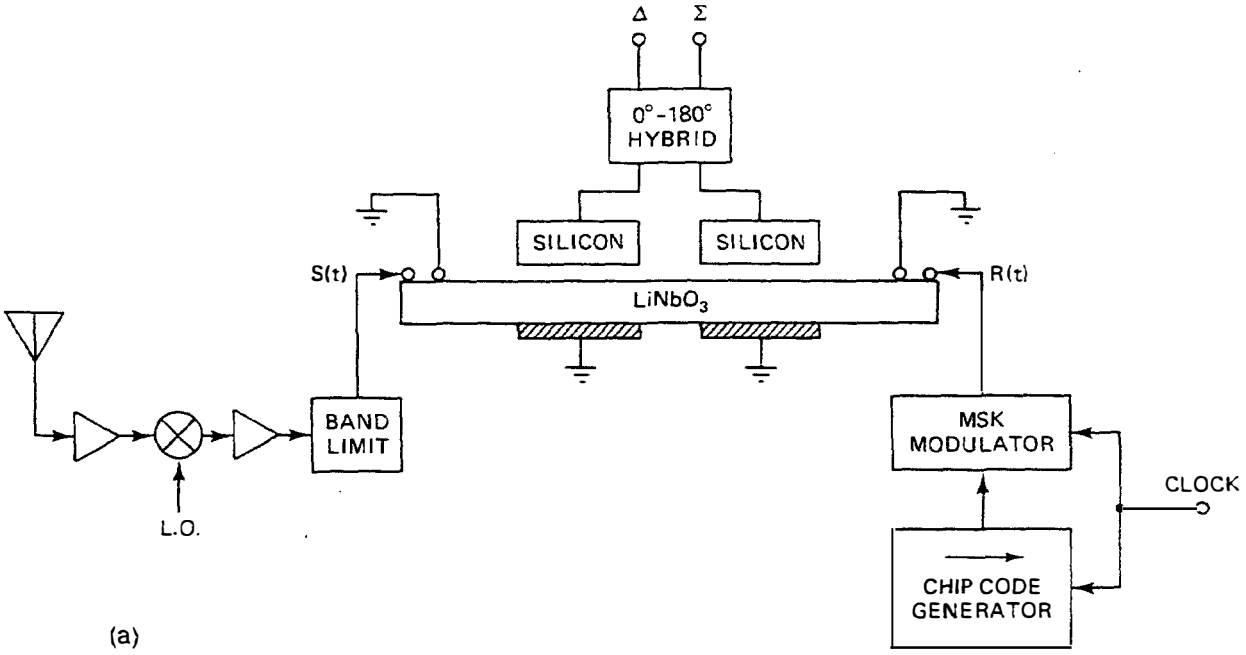
At the present time, perhaps the most important feature of the semiconductor convolver is that it has led to other types of related devices, such as the storage correlator, which will be discussed in Sec. 4.7.8. Furthermore, as surface acoustic waves can propagate directly on silicon, it follows that such devices can be incorporated with other circuit components on the same substrate. The system has even broader implications, because it follows that any type of SAW device could be made on silicon or, for that matter, gallium arsenide, as part of a total integrated circuit.

Where code processing is concerned, the results obtained with semiconductor convolvers of this type are similar to those obtained with the earlier lithium niobate convolvers. Lincoln Laboratories and Texas Instruments have demonstrated semiconductor convolvers with as much as a 240 MHz input bandwidth and a delay time of 10 μs, corresponding to a time-bandwidth product of 2400 [77]. This implies that the analog equivalent of $(2400)^2 = 5.8 \times 10^6$ bits of information can

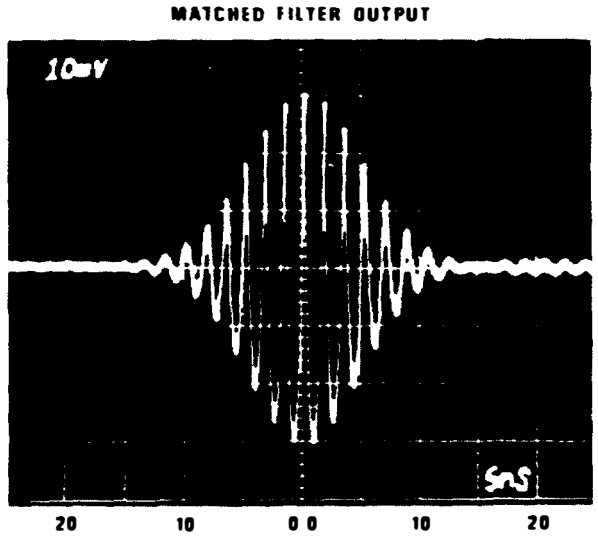
be processed in 10 μs when the convolution is taken between two signals. An example of using a convolver to process a pseudorandom code of 1024 chips at a 96-MHz rate is shown in Fig. 4.7.11.

4.7.6 Acoustic Convolver Using External Mixers

An alternative to the semiconductor convolver, pioneered by Montress and Reeder [78], conducts the mixing process in external nonlinear devices rather than in

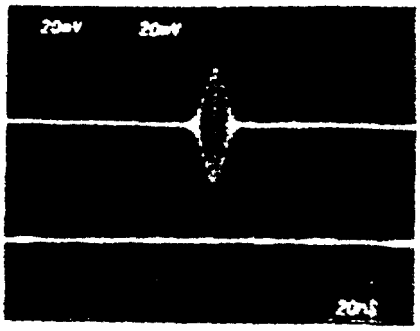


(a)

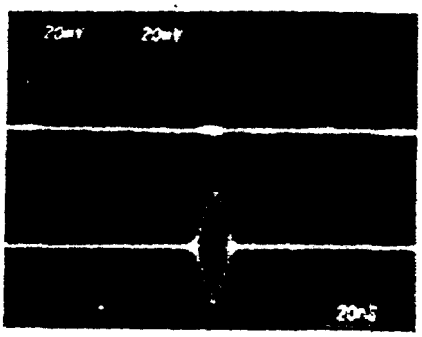


(b)

Figure 4.7.11 (a) Split-output electrode convolver used as part of a spread-spectrum receiver. Depending on the relative phases of the two parts of the reference code and signal code used, the output comes out of the Σ or Δ port. In this first case the bit code is 1; in the second case it is 0. (b) Correlation output with a running pseudorandom code. Each segment contains 1024 chips at a 96-MHz rate. The ideal response would be a width of 22 ns at the nulls of the main response. (c) Response of the device to a bit code 1. The peak Σ output (top trace) is more than 20 dB larger than the Δ output (bottom trace). (d) The output with a bit code 0. The Δ output (lower trace) is more than 20 dB larger than the Σ output. (After Goll and Malocha [76].)



(c)



(d)

4.7.11 (Continued)

internal devices. They employ an SAW delay line with taps along it, each tap connected to an external mixer, as illustrated in Fig. 4.7.12. This device can be optimized for maximum efficiency because the nonlinear properties and impedances of the mixers can be chosen independently of the acoustic properties and piezoelectric coupling coefficient of the delay line material. The mixers themselves can use the nonlinear $I-V$ characteristic of a diode or another semiconductor device, or they can use a nonlinear reactive effect, as does a normal convolver.

The optimum design of these arrays becomes a trade-off between choosing the diode impedance for maximum power sensitivity or for minimum reflection from the individual elements of the array of tapped transducers. The tapped transducers themselves, because there are normally a large number, should be made with split fingers of one-eighth of a wavelength to minimize reflections. The best sensitivity that Montress and Reeder obtained with diode mixers corresponds to $F_T \sim -22$ dBm, with saturation powers of the order of 0 dBm. By adjusting the current, they obtained a saturation power of 10 dBm, with $F_T \sim -33$ dBm. Obviously, a considerable improvement in convolver efficiency can be obtained by this technique, although as it reaches optimum efficiency, the saturation power is correspondingly decreased. External mixers do not necessarily increase the dynamic range of the system, but they can considerably improve the sensitivity of the device.

A disadvantage of tapped devices is that they use only a finite number of sampling points along the delay line, instead of interacting continuously, as does

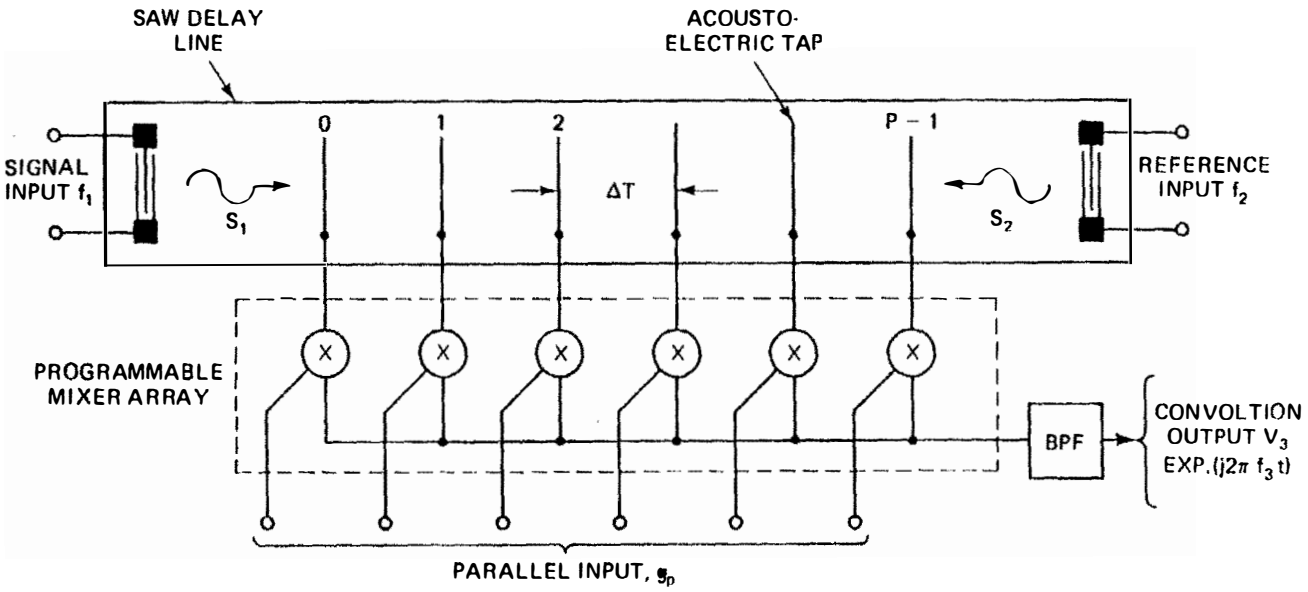


Figure 4.7.12 Fourier transform convolver. (After Montress and Reeder [78].)

a normal convolver. Thus the device can operate as a convolver with a number of possible input frequencies differing by integral multiples of a frequency that depends on the tap spacing L_T. Consider the situation where the two input frequencies are ω_1 and ω_2, respectively, so that the outputs obtained at the nth tap from the center of the device have the forms $\exp [j\omega_1(t - nL_T/V)]$ and $\exp [j\omega_2(t + nL_T/V)]$, respectively. After mixing, the output obtained from the system at the sum frequency will be of the form

$$y(t) = Ae^{j(\omega_1+\omega_2)t} \sum_n \exp\left(\left|\frac{-jnL_T(\omega_1 - \omega_2)}{V}\right|\right) \tag{4.7.17}$$

where A is a constant. It is apparent that the output at the sum frequency, $\omega_3 = \omega_1 + \omega_2$, is strong when the signals from all taps are in phase. Therefore, the mixed output signals from the taps can be summed in phase if

$$\frac{(\omega_1 - \omega_2)L_T}{V} = 2M\pi \tag{4.7.18}$$

where M is an integer. Because the taps are periodically spaced, the convolver can operate at several different input frequencies spaced by V/L_T. This result implies that the maximum bandwidth is limited by this frequency excursion; with larger bandwidths we would find aliasing of the input signals. Such a result is also implied by the sampling theorem, which indicates that at least two samples per RF modulation cycle are required. One advantage, however, is that if the bandwidth required is much less than either of the carrier frequencies, we need only a relatively coarse tap spacing.

The general form of this device is shown in Fig. 4.7.12. Using VHF bipolar transistors as the mixers, the taps themselves can be weighted over a wide dynamic range. We can obtain a significant reduction in spurious output signal levels because the high reverse isolation of the transistors, as compared to diodes, provides greater intertap isolation.

A device of this kind has been operated in a nondegenerate mode. As an example, with a 32-tap system, the two center input frequencies were taken as -83 MHz and 70 MHz. The results obtained, using Hamming weighting for a Fourier transform, are shown in Fig. 4.7.13.

4.7.7 CCD Convolver or Correlator

If the same principles are applied to CCDs, the outputs from two tapped CCDs can be mixed to provide convolution or correlation. A block diagram of a CCD convolver is given in Fig. 4.7.14. It operates by first loading CCD 1 with a reference signal and then clocking a second signal through CCD 2. Since the charge is stored dynamically in CCD 1, it decays because of thermal leakage and must be refreshed after 10 to 100 ms. Each stage of the CCD is tapped. The outputs of the corresponding taps on the two CCDs are taken to corresponding multipliers, using

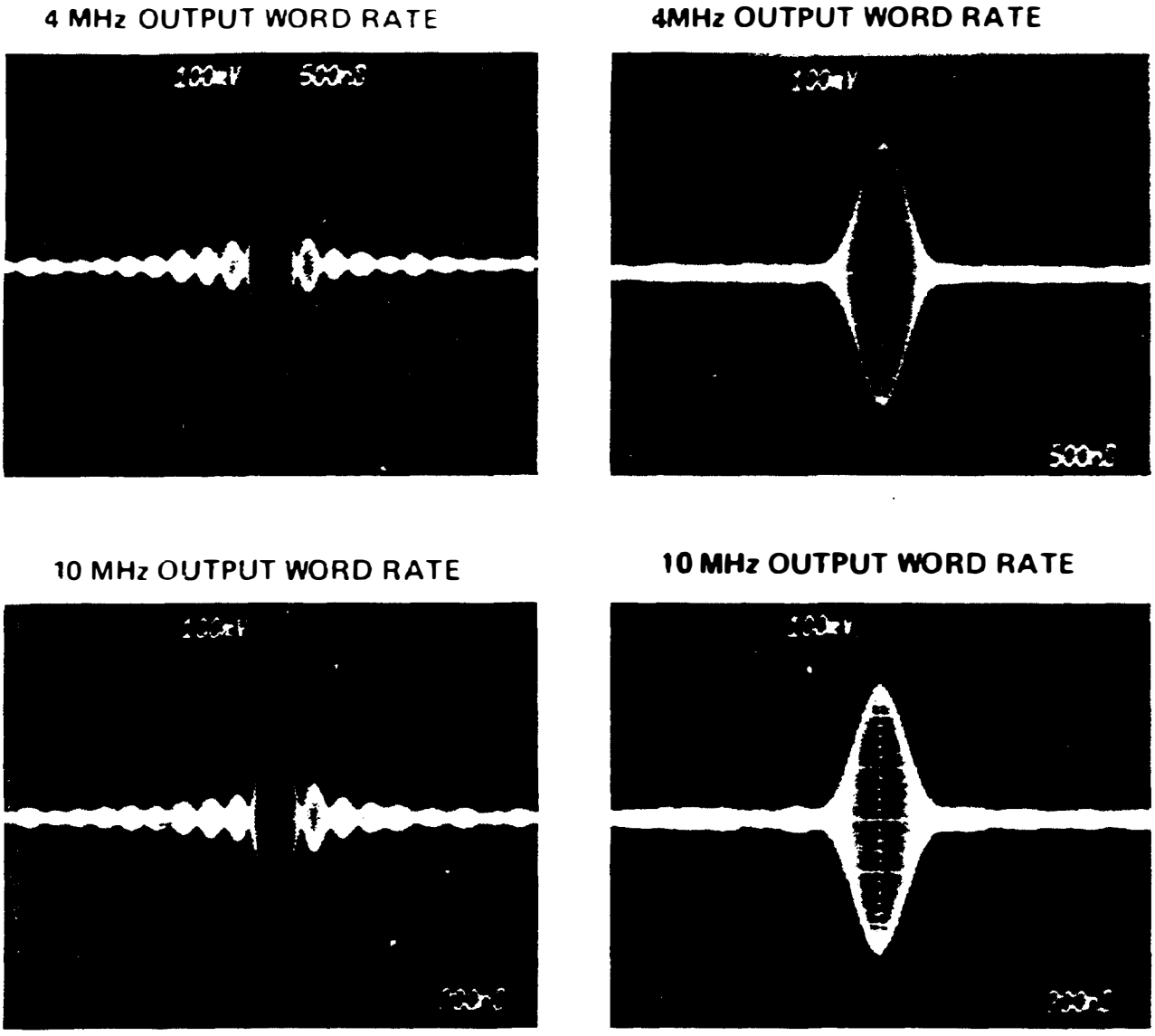


Figure 4.7.13 Fourier transform output: (a) uniform tap weighting; (b) Hamming tap weighting. (After Montress and Reeder [78].)

the circuit of Fig. 4.7.15. When the signal in CCD 1 is stationary in this configuration, it is possible, in principle, to feed the signal in either direction to CCD 2 and obtain either correlation or convolution. In the circuit shown, transistors Q_3 and Q_4 act as voltage-controlled resistors in a source-follower configuration. The balanced configuration overcomes the inherent nonlinearity of MOS transistors used as resistors.

A 100-stage correlator has been designed by Buss et al. [4] that uses floating-gate output taps for high charge-transfer efficiency. The device has been demonstrated to perform 100-point convolutions at clock rates up to 5 MHz. Because floating-gate outputs are used in the device, neither CCD can be stationary. Therefore, a device operated this way will be suitable only for convolution.

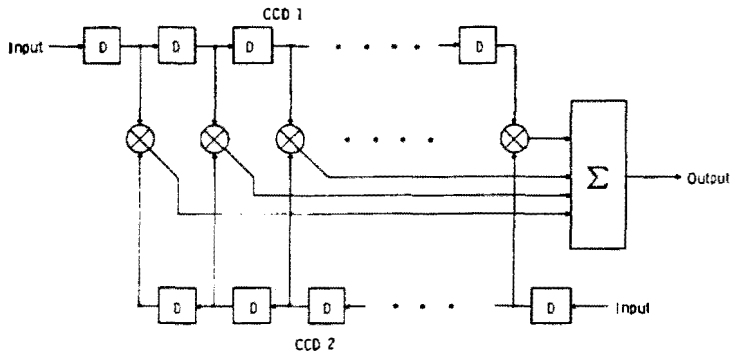


Figure 4.7.14 CCD correlator. (After Buss et al. [4].)

4.7.8 SAW Storage Correlator

The acoustic convolver has three major disadvantages that have limited its use to certain applications of the type already described:

1. The reference signal must be available at the time when the signal being interrogated is entering the device. This effect may be partially eliminated by using a periodically repeated reference and two convolvers, with one convolver in operation at all times. This does not necessarily cause problems in spread-spectrum operation.
2. The processing time in an acoustic convolver is at best only a few tens of a microsecond. In some cases, we may need much larger processing times of several milliseconds. As we have seen in the example of spread-spectrum communication, we can obtain this result to some extent by using an external integrator.
3. Because the device takes only the convolution between two signals, to obtain correlation we must time-reverse one of the signals. If the time-reversed version of one of the signals is not available in the form of a reference, a second convolver must be employed for the time reversal. Such a system has very limited dynamic range.

In this section, after describing a historically important concept, we shall describe a single device that can carry out both the required operations of long-term integration and internal storage of a reference signal. In this way we can

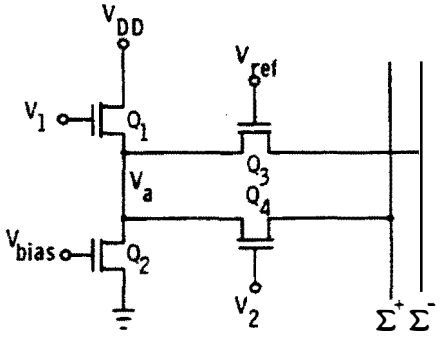


Figure 4.7.15 Circuit for four-quadrant multipliers. (After Buss et al. [4].)

obtain correlation as easily as we can convolution, and we eliminate the need for both signals to be present in the device at the same time. Furthermore, by operating in a different mode, we can correlate two signals, each several milliseconds long, in a device that normally has an acoustic signal delay time of several microseconds.

Tapped-delay-line correlator. One way to eliminate some of the difficulties of the convolver is to adopt the configuration shown in Fig. 4.7.16. A tapped delay line is employed with external mixers using integrators on the output of each mixer. One of the input signals $f(t)$ is fed directly to the mixers, while the other input signal $g(t)$ is fed through the SAW delay line via the taps to the mixers. The output from the nth tap is therefore

$$y(\tau_n) = \int f(t)g(t - \tau_n)\, dt \tag{4.7.19}$$

where $\tau_n = z_n/V$ is the time delay from the input to the nth tap at position z_n. The output is the correlation function at time τ_n. Thus by switching from one output tap to another, we can sample the correlation function over a limited range of time corresponding to the delay time of the delay line.

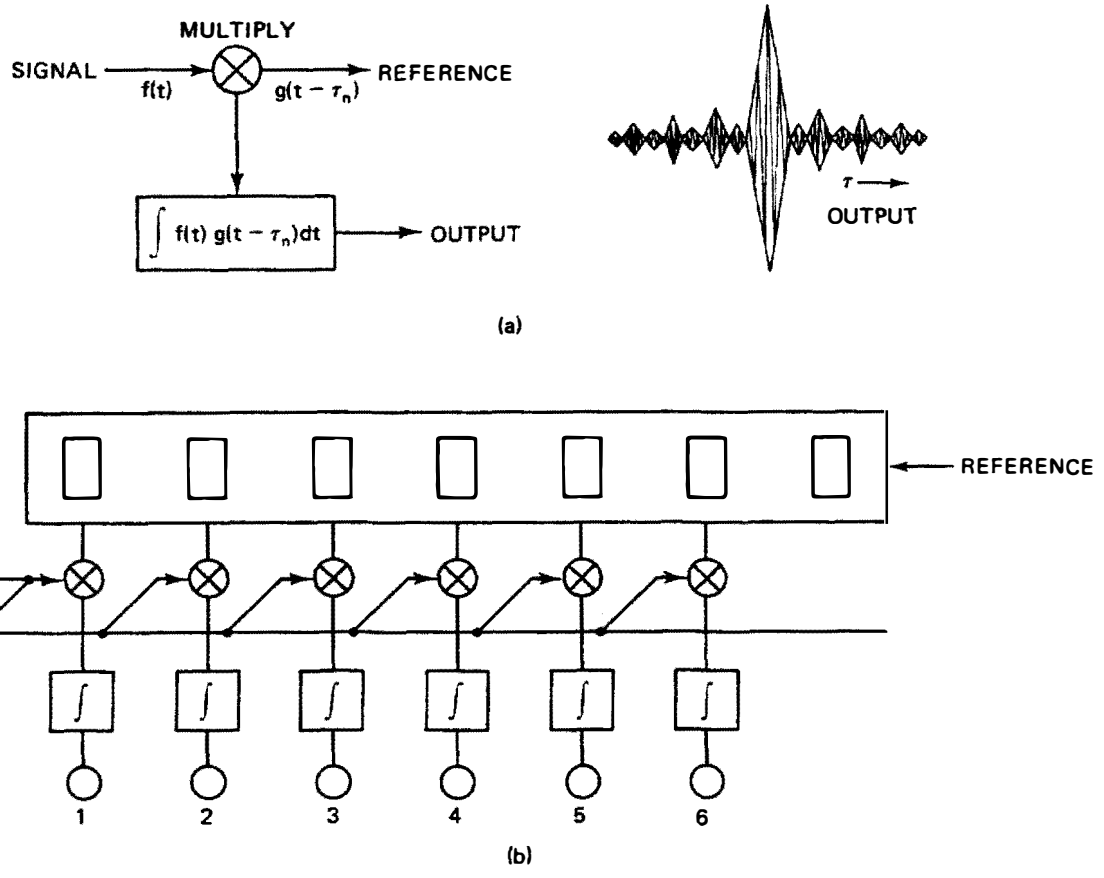


Figure 4.7.16 (a) Common form of active correlator; (b) tapped-delay-line correlator that is less sensitive to synchronization of signal and reference. (From Maines and Paige [44].)

We might imagine other versions of this device with internal mixers (i.e., strips of silicon placed near the substrate). But a technique is needed that will put all these components in one device, including the final multiplexing (i.e., switching to the output taps). The *storage correlator* satisfies these criteria.

Monolithic and air gap storage correlators. Consider the monolithic SAW storage correlator employing zinc oxide on silicon, illustrated in Fig. 4.7.17. In this device, a row of diodes with a spacing of the order of one-fourth of a wavelength of the acoustic wave is laid down in the silicon. SAW transducers are employed at each end of the device, and a top plate electrode is used as an input or output port in much the same manner as with the convolver.

Consider now what occurs when a signal is read into the device. Suppose that a short pulse less than half an RF cycle long, with voltage V_p, is applied to the top plate. This will turn on the individual diodes, as illustrated in Fig. 4.7.18. The capacity between an individual diode and a top plate is therefore charged up to a voltage approximately equal to V_p, with a corresponding charge in the capacitor of $Q_p = C_p V_p$, where C_p is the capacitance between the top plate and an individual diode. The diode itself is now reverse-biased, so that the only current flowing in the circuit comes from the leakage current I_s of the diode, which normally has a very low value. Thus the capacitance remains charged for a time $T_s \approx Q_p/I_s$, typically in the range 10 ms to 5 s, depending on the design of the diode and its temperature.

Now suppose that at the same time the pulse is applied to the top plate, an input signal $f(t) \cos \omega t$ is also applied to the device, exciting a surface acoustic wave so that the signal under a particular diode at position z becomes $f(t - z/V) \cos \omega(t - z/V)$. If the short pulse is applied at a time $t = 0$, we observe that the total charge stored in the diode is

$$Q(z) = C_p V_p + \alpha f\left(\frac{-z}{V}\right) \cos\left(\frac{\omega z}{V}\right) \tag{4.7.20}$$

A spatial pattern of charge, representing the input signal, has been stored in the capacitors, and the device is thereby capable of storing a reference signal.

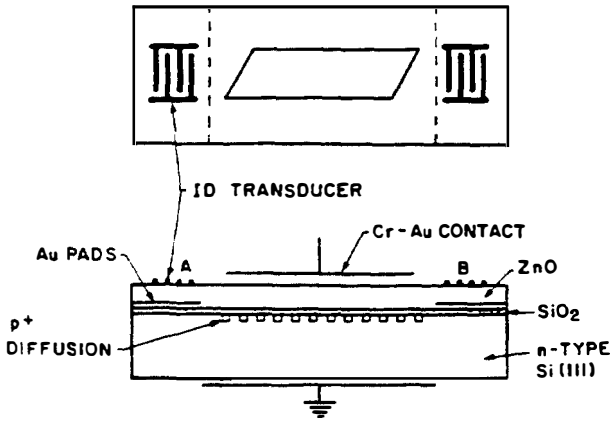


Figure 4.7.17 Surface acoustic wave storage correlator configuration. (After Tuan et al. [79].)

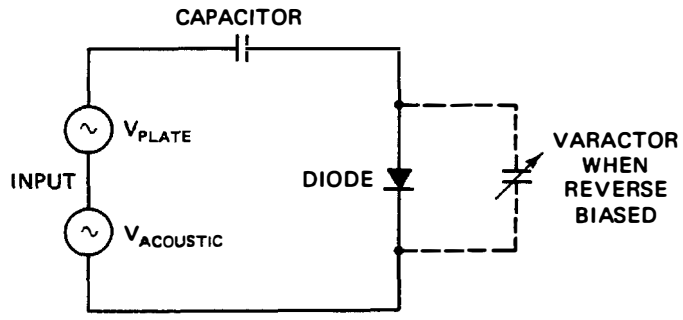


Figure 4.7.18 Equivalent circuit of correlator on read-in and read-out.

The reference charge is actually stored in the top plate, in a combination of the top plate capacitor and the capacitance of the diode itself. The total charge stored in the depletion layer of the diode with a depletion layer length h is

$$Q = \frac{qN_d hA}{\varepsilon} \tag{4.7.21}$$

where A is the area of the diode. We assume that the depleted region is in an n-type semiconductor of doping density N_d and dielectric constant ε. The voltage across the diode will therefore be

$$\phi = -\frac{qN_d h^2 A}{2\varepsilon} = -\frac{Q^2 \varepsilon}{2AN_d q} \tag{4.7.22}$$

and we can write that its effective capacitance is

$$C = \frac{\partial Q}{\partial \phi} = \frac{\varepsilon A}{h} = \frac{qN_d A}{Q\varepsilon} \tag{4.7.23}$$

The capacity of the diodes depends on the stored charge and hence on the initial reference signal. Suppose that the initial reference signal has a variation of the form $\cos \omega(t - z/V)$. There will then be a stored charge component $Q_1 \cos(\omega z/V)$, because of the exciting acoustic wave. The total charge Q will therefore be of the form

$$Q = Q_0 + Q_1 \cos \frac{\omega z}{V} \tag{4.7.24}$$

This implies that the capacity C has a spatially varying component $C_1(z)$ that is proportional in amplitude to $Q_1 \cos(\omega z/V)$, where it is assumed that $Q_1 \ll Q_0$.

Suppose that a "readout" signal $g(t)$ is now applied to the top plate. This readout signal will develop a voltage across the diode capacitance of the form

$$V(t, z) = g(t) \frac{C_p}{C_0 + C_1 + C_p} \cos \omega t \tag{4.7.25}$$

From Eqs. (4.7.24) and (4.7.25), it follows that if $C_p \ll C_0 + C_1$, as it often is, then $V(t, z)$ will have a component proportional to $Q_1\, g(t) \cos(\omega z/V) \cos \omega t$. The individual diodes behave like fingers of an interdigital transducer and can excite acoustic waves propagating in both directions along the device, as though the device had been excited by a voltage with a spatially varying pattern $\cos \omega z/V$.

The input, as we have observed, is stored in the form $f(-z/V)\cos \omega z/V$. The output from a particular point is proportional to the product of the readout signal and the stored signal, so the output obtained from an output transducer at $z = 0$ will be of the form

$$y(t, 0) = \alpha \cos \omega t \int g\left(t - \frac{z}{V}\right) f\left(\frac{-z}{V}\right) dz \tag{4.7.26}$$

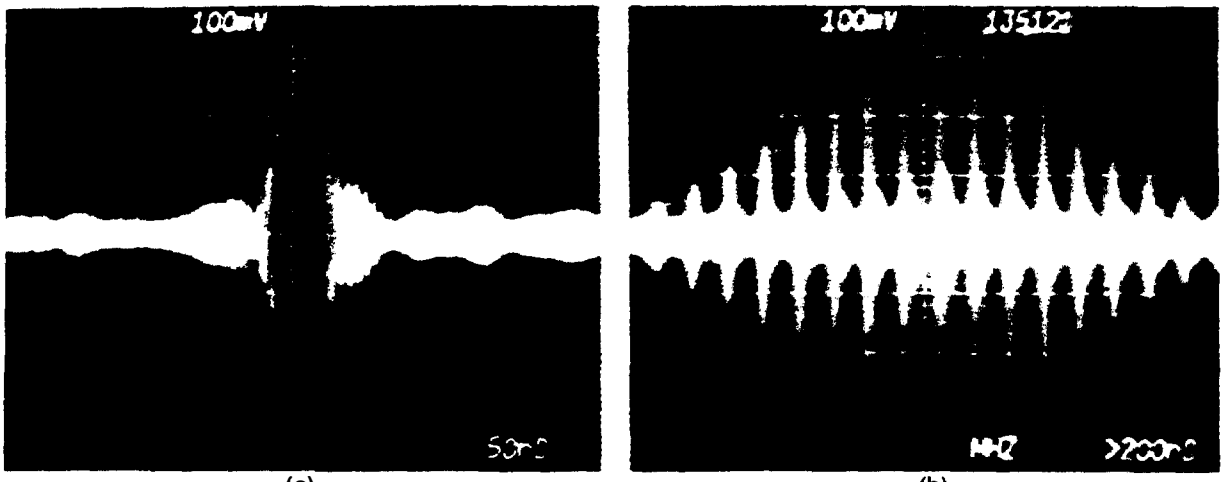
This is the correlation of the stored signal and the readout signal. Similarly, the output obtained from an output transducer at $z = L$ will be

$$y(t, L) = \alpha \cos \omega t \int g\left(t - \frac{L}{V} + \frac{z}{V}\right) f\left(\frac{-z}{V}\right) dz \tag{4.7.27}$$

This is the convolution of the two input signals.

To put it more simply, if we were to use a short pulse $g(t) = \delta(t)$ to read out from the device, we would obtain either the time-reversed waveform $f(-t)$ at $z = 0$, or the stored signal $f(t - T)$ itself at $z = L$, where $T = L/V$.

It follows that the device can be operated as a programmable matched filter for an arbitrary input waveform if a reference signal is read into it at an earlier time. An example of results taken by Ingebrigtsen, who constructed the first Schottky diode correlator in an air gap configuration, is shown in Fig. 4.7.19, with a schematic of his device in Fig. 4.7.20. In this device, a large number of Schottky diodes in the form of small square diodes were constructed in the silicon substrate, and the silicon was separated from a lithium niobate substrate by means of spacer rails. The results shown in Fig. 4.7.19 correspond to a 1.5-μs linear FM chirp, stored for 1 ms and correlated with a similar chirp in the readout. This example



(a) (b)

Figure 4.7.19 Correlation of 30-MHz-bandwidth, 1.5-μs linear FM chirps stored 1 ms: (a) details of single chirp correlation output; (b) nine stored chirps, each delayed by 180 ns, correlated with a similar sequence of nine chirps. Delay-line data: center frequency 145 MHz, 3 dB bandwidth, 30 MHz; Silicon data: 25 Ω-cm *n*-type platinum silicide diodes 4 μm in diameter on a 5.8-μm center distance. (After Ingebrigtsen [80].)

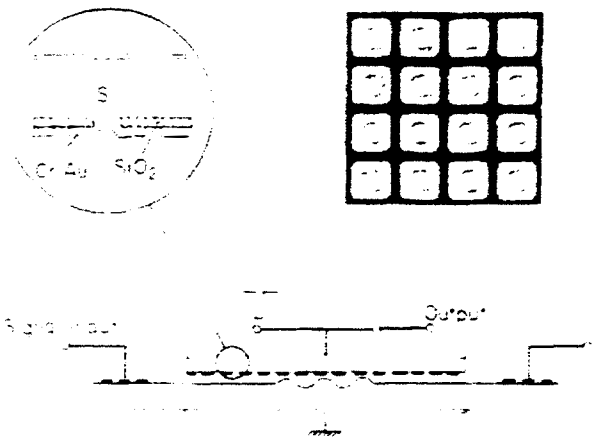


Figure 4.7.20 Diode memory correlator. (After Ingebrigtsen [80].)

corresponds, in fact, to nine stored chirps, each delayed by 180 ns, correlated with a similar sequence of nine chirps. The envelope of the function obtained, of course, is like that of two square pulses correlated with each other, and is hence triangular in form.

Input correlation mode. A second mode of operation is the *input correlation mode*. Suppose that two signals $f(t)$ and $g(t)$ are injected into a transducer and the top plate, respectively. If neither signal is sufficiently large in amplitude to switch on the diodes, no storage will take place unless the sum of the amplitudes of the two signals is high enough to do so. In this case, storage occurs only at the peak of the combined RF waveform. Furthermore, if the diodes are operated relatively slowly, as they would be with a *pn* diode, relatively little charge will be stored on each peak of the RF signal, because the diodes do not turn on completely and do not store all the available charge.

Suppose the current that flows during the peak of the RF cycle is $I(t)$; then the total charge injected into the diode and passing through the capacitor will be

$$\Delta Q = \int_0^{\tau} I(t)\, dt \tag{4.7.28}$$

where τ is the time for which forward current flows in the diode. If the recombination time for carriers in the diode is τ_p, we expect a proportion τ/τ_p of the total charge to recombine in the diode; therefore, the change in charge on the capacitor after one RF cycle will be approximately $\tau\Delta Q/\tau_p$. Because the effect is cumulative, after many RF cycles we would expect to see a gradual buildup of stored charge in the capacitor.

This process is highly nonlinear and depends on the peak amplitude of the two signals present; thus there will always be a stored charge component that depends on the product of the two signals. This stored charge component, in turn, depends on the integral of the product of the two signals over many RF cycles. Therefore, with two input signals $f(t)$ and $g(t)$ on the interdigital transducer and

top plate, respectively, we would expect the stored charge, at any position z, to have a component

$$Q(z) = \alpha \int_0^t f\left(t - \frac{z}{V}\right) g(t)\, dt \tag{4.7.29}$$

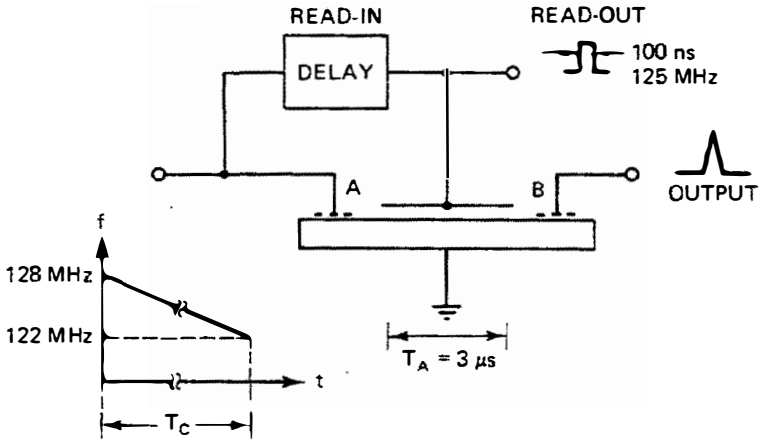
The spatial variation of the stored component of charge corresponds to the correlation of the two input signals.

In this system, the floating variable in the correlation function is time t; this can be relatively long compared to the delay time T of a surface acoustic wave in the device itself, and may be comparable to the storage time of the device. Thus correlation can take place over extremely long time periods, and the device can handle signals with a time–bandwidth product comparable to the product of the storage time and the bandwidth of the device. The storage times can be tens of milliseconds for any bandwidth that can be accepted by the input transducer; thus the possibility arises of correlating signals with time–bandwidth products as large as 10^6. In practice, correlation has been observed in these devices with time–bandwidth products up to 100,000 (in one case corresponding to a bandwidth of 20 MHz with an integration time of 5 ms) [80–82].

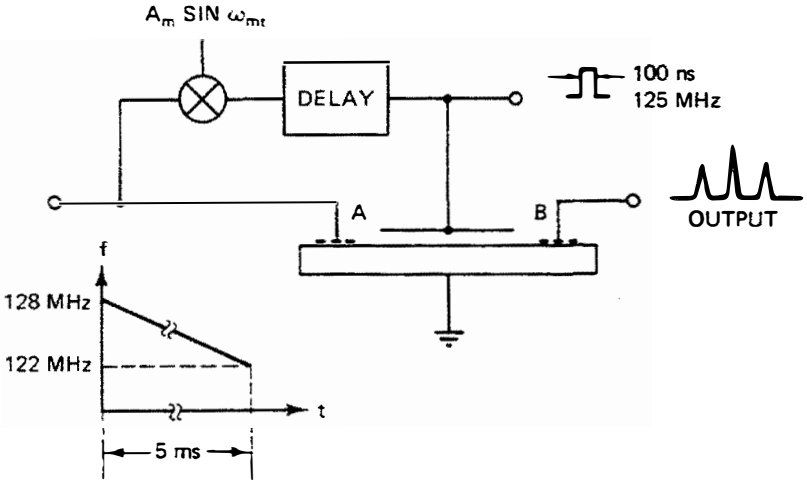
Note that the correlation peak occurs at $z = 0$; this may correspond to a spatial position outside the top plate itself and therefore may not be a readable point. To place the correlation peak at the center of the device, we must use a delay line to delay $g(t)$ by half the time delay in the device, as illustrated in Fig. 4.7.21. The correlation peak itself may be read out using a short interrogating pulse for the readout. If we use a more complicated waveform for the readout, of course, we can obtain correlation or convolution of the readout signal with the stored information. It is apparent that a wide range of signal processing techniques is possible with this configuration.

In one example, shown in Fig. 4.7.22, two chirps, each 5 ms long with a bandwidth of 6 MHz, were read into the device. One of the chirps was modulated with a signal of 186 Hz, using two sidebands on the carrier 372 Hz apart. The correlation of the function therefore consisted of two peaks. In this case, two peaks approximately 350 ns apart were observed, corresponding to a theoretical difference of 420 Hz in frequency. The agreement is reasonable, for it was not known how linear the chirps were, and there was some uncertainty as to their exact bandwidth. By varying the chirp lengths, components in different ranges could be observed, and thus the device could be used as a variable-bandwidth spectrum analyzer. The top trace shows the output signal and the short readout pulse alone as applied to the device. This represents the spurious signal levels of the device, which created considerable difficulties. These spurious signals occurred because there was some direct feedthrough of the signal injected on the top plate into the output transducers.

Spread-spectrum communications. The storage correlator is ideally suited to signal processing of spread-spectrum signals, for it carries out the required integration and correlation functions. Ideally, to correlate a PN code or another



(a)



(b)

Figure 4.7.21 Modes of operation: (a) input correlation; (b) AM modulation of one of the input chirps. (After Tuan and Kino [81].)

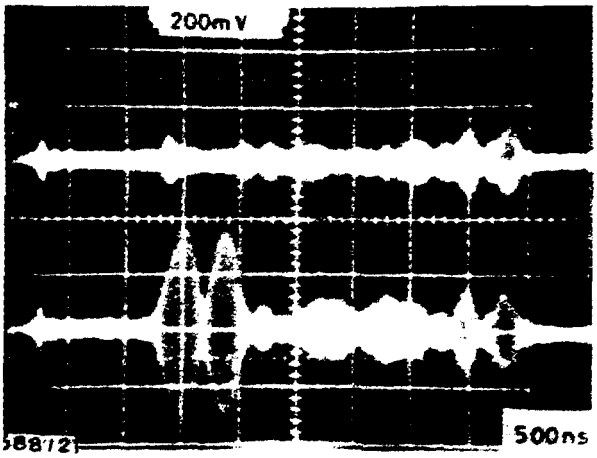


Figure 4.7.22 Correlation output of AM modulation experiment. Top trace, no signal present; bottom trace, a *modulated* chirp modulated at 186 Hz is used as the input signal. (After Tuan and Kino [81].)

type of code suitable for spread-spectrum communications, the device should build up the correlation peak continually with integration time. This experiment has been carried out with a reference code 1 ms long [83]. As shown in Fig. 4.7.23, we see that the correlation peak amplitude increases linearly with integration time, as it also does with the amplitude of the input signal. Figure 4.7.24 shows an example of the correlation peak obtained with a PN code 1 ms long and 5 MHz in bandwidth, corresponding to a time–bandwidth product of 5000. The three traces shown correspond to codes read in with slightly different delays between signal and reference ports. Similar results were obtained with codes as long as 10 ms, corresponding to a time–bandwidth product of 50,000.

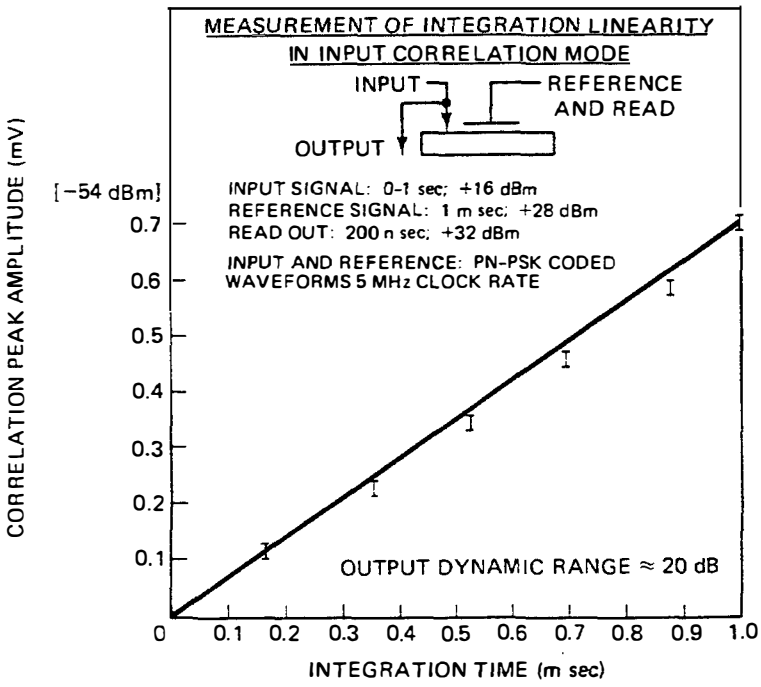


Figure 4.7.23 Measurement of device linearity with input signal duration. (After Tuan et al. [83].)

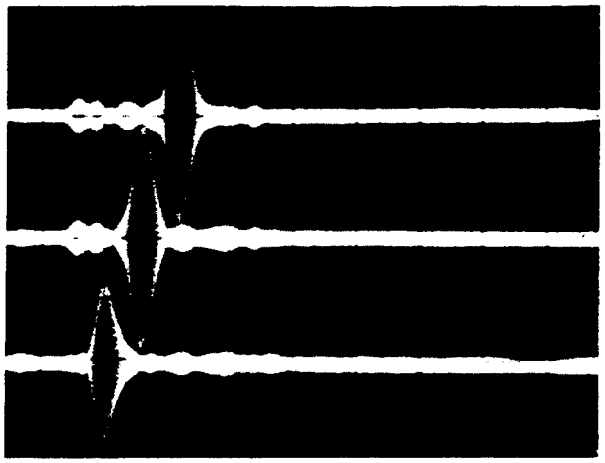


Figure 4.7.24 Output signal from large time–bandwidth product PN-PSK matched filter. (After Tuan et al. [83].)

PROBLEM SET 4.7

1. Consider nonlinear interactions in an SAW structure. Suppose that the propagation constants of the waves at the fundamental frequency are k_1, and at the second harmonic, $k_2 = 2k_1 + \Delta k$. We put

$$\hat{a}_1(z, t) = a_1(z)e^{j(\omega_1 t - k_1 z)}$$

$$\hat{a}_2(z, t) = a_2(z)e^{j(\omega_2 t - k_2 z)}$$

   (a) Assuming that the system possesses a second-order nonlinearity so that there is an excitation at the second harmonic frequency proportional to $\hat{a}_1(z, t)^2$, find a coupled-mode equation at the frequency ω_2 with an arbitrary coupling parameter to $(\hat{a}_1)^2$, which represents how $\hat{a}_2(z, t)$ changes with distance.

   (b) Now, assuming that the power in the second harmonic is small compared to that in the fundamental, we can write $a_1(z) =$ constant. In this case, find how $a_2(z)$ varies with distance from the input when $k_2 = 2k_1$ and when $k_2 \neq 2k_1$. The problem illustrates that slight dispersion will tend to make second harmonic generation very weak. Without dispersion, the second harmonic will build up indefinitely, and the fundamental amplitude will eventually decrease (i.e., there will be saturation with increase of the fundamental input signal).

2. (a) Consider mathematically what occurs when a signal $f(t)$ exp $(j\omega t)$ is injected in one port of a convolver, and a modulated RF signal $g(t)$ exp $(2j\omega t)$ is injected at the center plate such that $g(t)$ is a short pulse approaching a δ function. Show that a time-reversed signal of the form $f(T - t)$ exp $(j\omega t)$ is emitted from the input port.

   (b) What are the forms of the outputs from each end of the device when two signals $f(t)$ exp $(j\omega t)$ and $g(t)$ exp $(2j\omega t)$ are injected at one acoustic transducer and the center plate, respectively?

3. Find the dc potential generated on the surface of a semiconductor placed near a piezoelectric substrate by a surface acoustic wave RF signal field normal to the semiconductor surface of value E_1 exp $[j(\omega t - kz)]$. This phenomenon is known as the transverse acoustoelectric effect. Suggest how measurement of this potential might be used to determine the carrier density in the semiconductor.

4. A zinc-oxide-on-silicon convolver has a dispersive characteristic so that the phase velocity in the device varies with frequency. The phase velocity is defined as $V_p = \omega/k$ and the group velocity is $V_g = d\omega/dk$. The group velocity is the velocity at which the modulation travels through the system. Let us assume that the propagation constant is a parabolic function of frequency. Thus we can write

$$k \approx k_0 + (\omega - \omega_0)\left.\frac{\partial k}{\partial \omega}\right|_{\omega 0} + \tfrac{1}{2}(\omega - \omega_0)^2 \left.\frac{\partial^2 k}{\partial \omega^2}\right|_{\omega 0} \tag{1}$$

or

$$k = k_0 + \frac{\omega - \omega_0}{V_{g0}} - \frac{(\omega - \omega_0)^2}{2}\frac{\tau}{V_{g0}} \tag{2}$$

where

$$\frac{1}{V_{g0}} = \left.\frac{\partial k}{\partial \omega}\right|_{\omega 0} \tag{3}$$

and

$$\tau = \frac{1}{V_{g0}}\left.\frac{\partial V_g}{\partial \omega}\right|_{\omega 0} \tag{4}$$

(a) First, consider physically what occurs when two short pulses of frequency ω_0 are inserted into each end of a convolver of length L at a time $t = 0$. When does the output pulse occur? If the signal frequency is changed to ω_1, when does the output pulse occur? Finally, if a linear up-chirp varying in frequency from ω_0 to ω_1 is convolved with a linear down-chirp whose frequency varies from ω_1 to ω_0, and both chirps are of a time length T_1 less than the phase or group delay time through the convolver, what would be the approximate minimum length of output pulse that you would expect if: (i) there were no dispersion? (ii) there were dispersion? *Use only physical arguments and no detailed mathematical derivations* other than differentiating Eq. (2).

(b) Try the same argument on a dispersive storage correlator of length L, in which the same up-chirp is injected at an acoustic port and stored by putting a very short electrical pulse on the plate. The correlation is then read out with the same chirp as the stored chirp. Would dispersion now be expected to affect the output pulse width?

(c) From physical arguments, what effect would you expect a uniform attenuation per unit length to have on the form of the output of both devices (the convolver and correlator) when they are excited with a pulsed asymmetric signal with a carrier frequency ω_0, assuming that they are nondispersive?

(d) Now treat the dispersive convolver and correlator mathematically. Consider signals $f(t)$ and $g(t)$ inserted at each end of the convolver. Write

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)e^{j\omega t}\,d\omega \tag{5}$$

and

$$g(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} G(\omega)e^{j\omega t}\,d\omega \tag{6}$$

where

$$f(t) = a(t)e^{j\omega_0 t} \tag{7}$$

$$g(t) = b(t)e^{j\omega_0 t}$$

and the signals are narrowband, as [illegible] end of Sec. 4.4.5.

Show that the convolver out[illegible] put-modulated signals of carrier frequency ω_0 is like that of a theor[illegible] nondispersive convolver in which the output is passed into a chirp fil[illegible] domain response

$$Ke^{j(2\omega_0 t - t^2/\tau'^2)} \tag{8}$$

where K is a constant and

$$\tau' = \sqrt{\frac{L\tau}{2v_{g0}}} \tag{9}$$

Taking $L = 2$ cm, $\tau = 5 \times 10^{-10}$ s, and $V_{g0} = 2.8 \times 10^5$ cm/s, estimate the maximum usable input bandwidth of the convolver. Suggest what filter could be used on the output terminal to eliminate the problem with dispersion.

Note: You will find it convenient to use the method of stationary phase in Appendix G to work out what happens to each Fourier component of the input signal, and to find the inverse Fourier transform of the output signal and how it is distorted.

4.8 ADAPTIVE FILTERS

4.8.1 Introduction

We have thus far described various types of fixed transversal filters, and convolvers and correlators whose parameters can be adjusted by using a reference signal as the input to match a required characteristic. In some cases, however, the required filter characteristic is unknown; we know only the output required from a test input signal. When a signal is distorted by transmission through a random medium, such as the atmosphere, or an imperfect communication system, we need a filter that can deconvolve or equalize the distorted signal to produce the original, undistorted signal. Such a filter is termed an *adaptive filter*: it must adapt itself to the distortion present in the system.

In some cases, such as propagation of waves through a turbulent atmosphere, the filter need only remove the phase distortion of a signal; in others, where we need to improve the response characteristics of a sonar transducer, for example, in order to obtain the shortest possible pulse, we may also want the filter to obtain a flatter amplitude response as a function of frequency. To remove unwanted echoes in a television ghost signal, for instance, we need a filter that can change both the amplitude and the phase characteristics of the received signal.

Another common application of an adaptive filter is to eliminate an interfering signal. Consider, for example, the difficulty of observing a fetal heartbeat in the presence of the mother's much stronger heartbeat; an adaptive filter is needed to

eliminate the mother's heartbeat. Similarly, in a communication system, when the presence of an interfering signal masks the required signal, we can use an adaptive filter to eliminate the interfering signal.

We describe here a number of adaptive filter concepts that can be used for these purposes. We consider examples employing SAW filters, CCD filters, and acoustic storage correlators.

4.8.2 Removing Phase Errors with an Adaptive Matched Filter

An adaptive filter can eliminate phase errors that arise from distortion in a propagating medium or a signal processing system. Suppose that the distortion of a test impulse $x(t) = \delta(t)$, which has a uniform amplitude as a function of frequency when it passes through a transmission medium, is $H(\omega)$, where

$$H(\omega) = A(\omega)e^{-j\Phi(\omega)} \tag{4.8.1}$$

and $A(\omega)$ is a real function.

If a compensating matched filter is constructed with a response $H^*(\omega)$, the resultant output, when a signal $X(\omega)$ is passed through the medium, will be

$$Y(\omega) = X(\omega)H(\omega)H^*(\omega) = A^2(\omega)X(\omega) \tag{4.8.2}$$

All phase errors due to the transmission medium are removed by this process, and the output of a test impulse, after passing through the distorting medium and the matched filter, has no phase variation with frequency. The transform of this output signal will be $y(t)$, and the combined response of the transmission system and the equalizer will be the autocorrelation function of $h(t)$ convolved with $x(t)$.

By removing the phase errors, an output at $t = 0$ that corresponds to the peak of the autocorrelation function of $h(t)$ can be obtained from an impulse input $[x(t) = \delta(t)]$. If $H(\omega)H^*(\omega)$ is constant in amplitude over a finite frequency range Ω, so that $H(\omega) = \Pi[(\omega - \omega_0)/\Omega]$, the output with an impulse input will be a sinc function of time. If there are major peaks in the frequency spectrum, the time response will exhibit severe ringing, but provided that the input spectrum is reasonably smooth (ideally, a Gaussian function), the time-domain response corresponding to $H(\omega)H^*(\omega)$ should be reasonably compact with no severe ringing.

Suppose that a reference impulse is sent through a distorting medium. If a distorted pulse $h(t)$ is stored in a storage correlator, phase distortions caused by the medium can be removed from it. Now any other signal $x(t)$, sent along the same path into the correlator where the reference pulse is stored, will be convolved with the stored reference $H^*(\omega)$ or $h(-t)$ and will have its phase distortions removed.

A practical application of this technique is to improve resolution when a transducer with a poor pulse response is used in a pulse echo NDT or sonar system [84]. To demonstrate this application, a 3.25-MHz-center-frequency, 2.5-MHz-bandwidth P ZT transducer was placed 16 cm from a plastic block in a water tank, as illustrated in Fig. 4.8.1. The transducer was excited with a linear FM chirp of constant amplitude. A 6-μs-long, 2.5-MHz segment of the first reflected pulse

from the plastic block was gated, mixed with 98 MHz, and stored in the correlator. The gate to the top plate was then reopened to allow a second echo pulse to correlate with the stored first reflected pulse. This second echo was the triple transit signal returning from the plastic block, reflected from the transducer face, reflected a second time from the plastic block, and received approximately 210 μs after the first one was used as the stored reference. Figure 4.8.2(a) and (b) show the correlation peak obtained and, for comparison, the impulse response of the transducer. The width of the correlation peak corresponded to a compression ratio of 9, compared with the time–bandwidth limit of 15.

The same experiment was tried a second time with a poor-quality transducer that exhibited severe ringing in its impulse response. Figure 4.8.2(c), (d), and (e) show the first reflected pulse, the correlation peak, and the transducer impulse response, respectively. The correlation peak width was essentially the same as that obtained with the original high-quality transducer. The results showed that by using a storage correlator with a poor-quality transducer, we could considerably improve the response of the pulse-echo system.

Similar techniques have been used in optics to remove phase errors. A distorted optical beam is used to store a holographic grating in a photosensitive material. This grating is then used to remove distortions in an image when the optical beam has passed through the same distorting medium [85].

4.8.3 Inverse Filter

The basic principle for constructing an equalizer or deconvolver is to send a reference pulse through the distorting medium and then use a transversal filter to remove the distortion. Consider a general input signal of the form $X(\omega)$. After passing through the distorting medium, whose response is $H(\omega)$, it will be of the

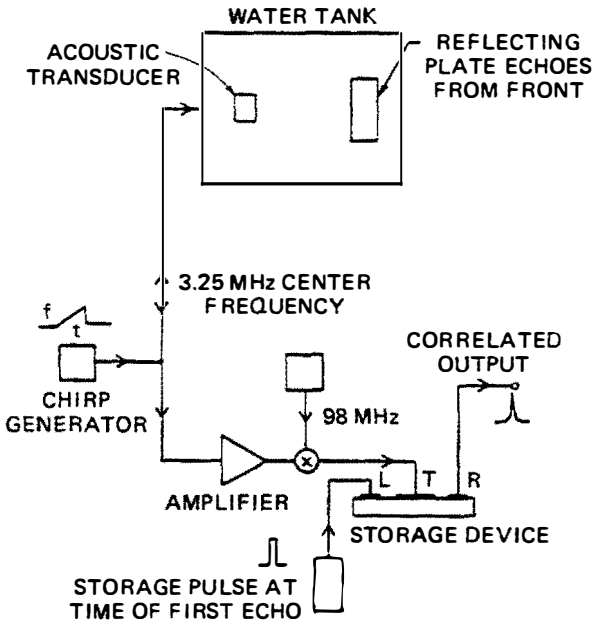


Figure 4.8.1 Acoustic pulse-echo system. (After Borden and Kino [84].)

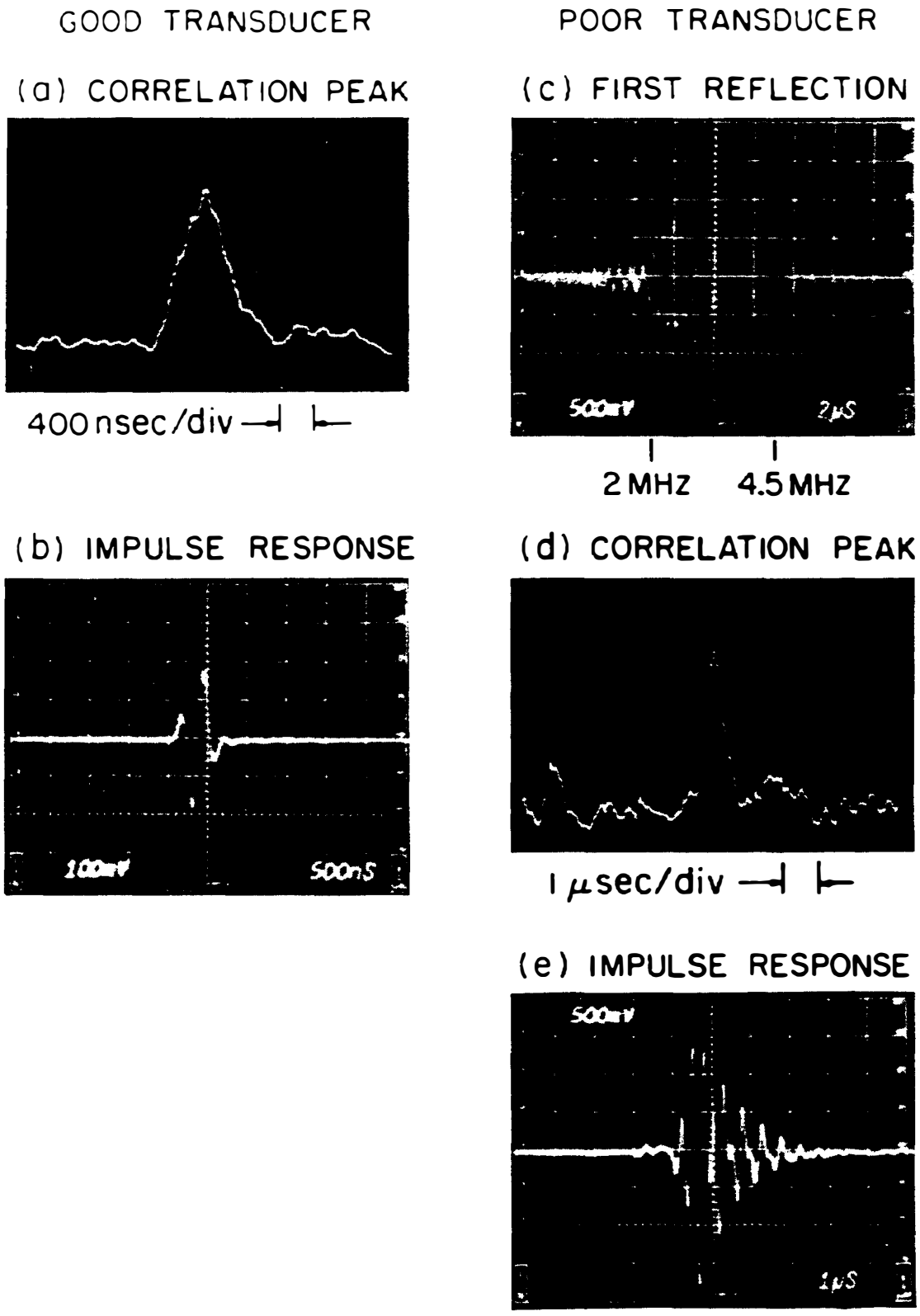


Figure 4.8.2 Pulse-echo experiment results carried out with both good and poor transducers. (After Borden and Kino [84].)

form

$$Y_1(\omega) = X(\omega)H(\omega) \tag{4.8.3}$$

We have already seen that a matched filter of the form $H^*(\omega)$ can remove the phase distortions due to $H(\omega)$. Such a filter will, however, accentuate amplitude variations over the frequency range of interest, for the response will now be $H(\omega)H^*(\omega)$, or essentially the square of the original amplitude response. To recover the original signal $X(\omega)$, we ideally want to construct an *inverse filter*, that is, a transversal filter with a characteristic frequency response $W(\omega) = 1/H(\omega)$. The output after passing the distorted signal through the inverse filter will be

$$Y_2(\omega) = \frac{X(\omega)H(\omega)}{H(\omega)} = X(\omega) \tag{4.8.4}$$

and the effective value of $H(\omega)$ of th d system, which we shall call $H'(\omega)$, will be

$$H'(\omega, - \frac{}{H(\omega)} = 1 \tag{4.8.5}$$

A filter with a response $W(\omega) = 1/H(\omega)$ is known as an *inverse filter*. This filter is obviously impossible to realize because it needs an infinite response at the points where $H(\omega) = 0$. Thus only some approximate form of equalizer, at best, can be constructed. One way to do this, of course, is to measure the characteristic response $H(\omega)$ of the system. If $H(\omega)$ is finite over the frequency range $\omega_1 < \omega < \omega_2$, we can multiply the output by $W(\omega) = 1/H(\omega)$ over this frequency range and put $W(\omega) = 0$ at all other frequencies. This procedure yields a uniform response $H'(\omega) = 1$ in the frequency range $\omega_0 - \Omega/2 < \omega < \omega_0 + \Omega/2$ with zero output outside this frequency range, as illustrated in Fig. 4.8.3. The time response of this filtered system is then

$$h'(t) = \frac{\Omega}{2\pi} e^{j\omega_0 t} \frac{(\Omega t/2)}{\Omega t/2} \tag{4.8.6}$$

Another possibility is to apodize the response of the inverse filter with a Hamming weighting to obtain a smooth frequency response or a smooth and compact time response. In all cases, unless some type of gating or apodizing function is used, the noise output will vary as $1/H(\omega)$ and will be very large in the regions where $H(\omega)\ \omega \to 0$ and $1/H(\omega)\ \omega \to \infty$.

A CCD or an acoustic transducer with the required inverse filter response over this frequency range can be constructed on this principle. This process has been carried out by Kerber et al. [86] using an SAW filter to compensate for the characteristics of an acoustic bulk wave transducer. There are also several examples in the literature that use a computer-based system for this process [87]. Computer-based systems, however, will not easily give large bandwidths and real-time processing.

Consider the system illustrated in Fig. 4.8.4. A reference pulse is sent into a tapped analog delay line. The output taps can be varied in amplitude by a weighting network that is controlled by computer or microprocessor. If the weight

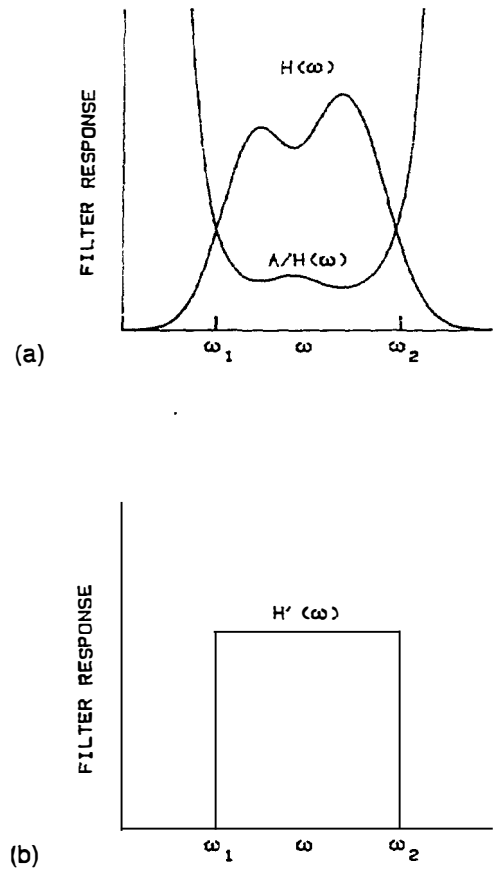


Figure 4.8.3 (a) Response $H(\omega)$ of the system and that of the inverse filter $A/H(\omega)$; (b) inverse filter for which the response of the system would be $H'(\omega)$, where $H'(\omega) = 1$, $\omega_1 < \omega < \omega_2$, $H'(\omega) = 0$, $\omega < \omega_1$, and $\omega > \omega_2$.

of the delay line can be adjusted to compensate for the distortion of the medium, the delay line becomes a real-time inverse filter, which can compensate for the distortion of any signal passing through the same distorting medium. The device is trained relatively slowly, using a repeated distorted input pulse that eventually produces an output pulse of the same form as the known undistorted reference pulse. The signal $x(t)$ entering the delay line is sampled at a rate $f_s = 1/\tau$, which corresponds to the delay time between taps. If the weighting of the nth tap is w_n, the output from the device at a time $t = m\tau$ will be

$$y(m\tau) = \sum_{n=1}^{N} w_n x[(m - n)\tau] \tag{4.8.7}$$

To train the device in the manner prescribed, we therefore require that

$$\begin{aligned} y(m\tau) &= 1; \; m = k \\ y(m\tau) &= 0; \; m \neq k \end{aligned} \tag{4.8.8}$$

where the kth tap is called the reference tap.

The process was carried out in this example with a Reticon TAD-32 tapped analog delay line controlled by a microprocessor. The filter tap weights were adjusted by a process known as a *zero-forcing algorithm*, which is now commonly used for automatic equalization of digital communication systems employed in the

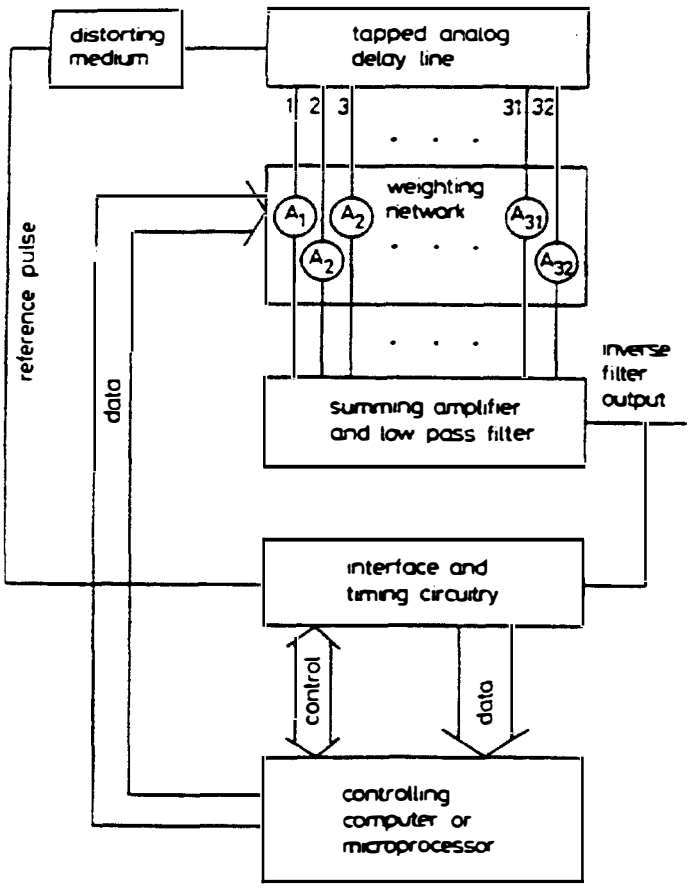


Figure 4.8.4 Adaptive inverse filter processor. (After Corl [88].)

telephone system. In the telephone system, for instance, by adjusting the weights of a tapped electromagnetic delay line, such methods have made it possible to upgrade standard telephone lines, which formerly operated in the frequency range below 3 kHz, to process digital codes with frequencies as high as 64 kHz. Such equalizers automatically come into operation to compensate for distortions over the changing paths used in a switched telephone network; this eliminates errors when telephone lines are used for transmitting digital codes.

The zero-forcing algorithm used in this example was first employed for telephone lines by Lucky [89]. It uses one tap, designated as the reference tap, whose weight is maintained at a constant nonzero value. The algorithm operates by increasing or decreasing each tap weight by one step, according to whether the filter output is positive or negative during the corresponding clock period. This forces the filter output to a maximum during the clock period corresponding to the reference tap, and to zero during the clock periods corresponding to the other taps of the delay line. The algorithm has the advantage of great simplicity and relatively high speed convergence. Its disadvantage is that it does not converge for any arbitrary distortion; however, it does converge for a large class of waveforms of the types associated with digital codes. As it was relatively easy to implement, it was the first one tried with CCD and BBD filters.

In the example shown, an *LC* resonator was constructed to simulate the

response of a distorting medium [88]. This resonator was pulsed and gave a ringing signal, as shown in Fig. 4.8.5, and the CCD adjusted itself until it converged on an output pulse after 300 iterations. We expected the required filter to superimpose a second ringing waveform on the first one, delayed by one-half an RF cycle but reduced in amplitude, which would just cancel out the rest of the ringing waveform. Thus the inverse filter had to consist of a weighting corresponding to two impulses of opposite sign, one following the other. This was basically the form of the filter produced, as shown in Fig. 4.8.6. In another example, illustrated in Fig. 4.8.7., more complicated signals were injected into the device after the filter was adjusted and the distortion was then eliminated from them.

A completely monolithic LSI transversal filter has been constructed on the same principles [90]. This device uses BBD filters in the configuration shown in Fig. 4.8.8. To avoid making a tapped BBD, 16 separate BBDs, with time delays in steps from T to $16T$, are used with their outputs summed. The amplitude of the input signal to each BBD is varied by multiplying D/A converters. These devices behave like variable attenuators whose attenuations can be set by a digital input. In this case the digital inputs are set by a microprocessor. The 16 BBDs are therefore equivalent to a tapped delay line in which the signals are inserted

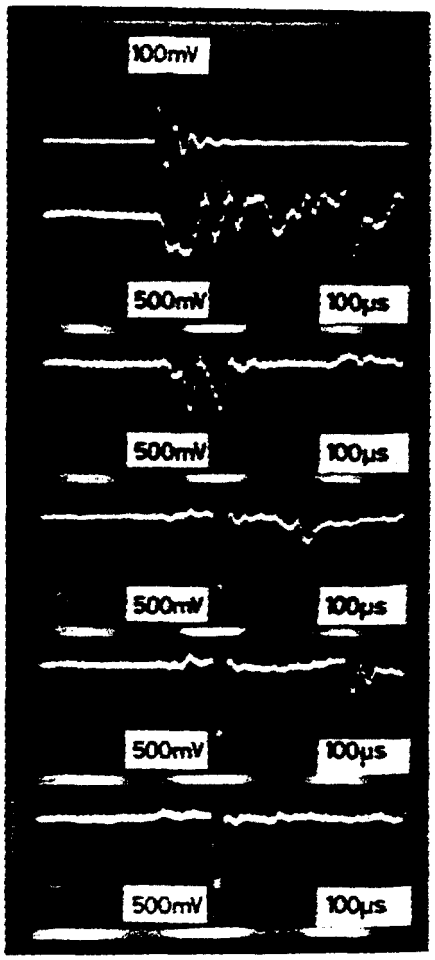


Figure 4.8.5 Convergence of the CCD filter. (After Corl [88].)

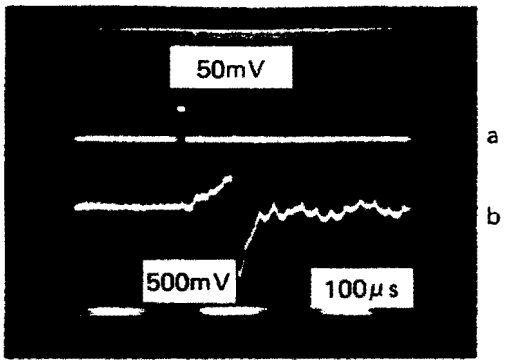


Figure 4.8.6 Inverse filter impulse response. (After Corl [88].)

into the taps. An adaptive inverse filter using the Lucky forcing algorithm was made with this system to obtain the results shown in Fig. 4.8.9.

4.8.4 Wiener Filter

A matched filter can remove phase errors but not amplitude errors. A more general filter, as we have seen, is the inverse filter. Unfortunately, this filter cannot be practically realized, because at certain frequencies it would require an infinite response.

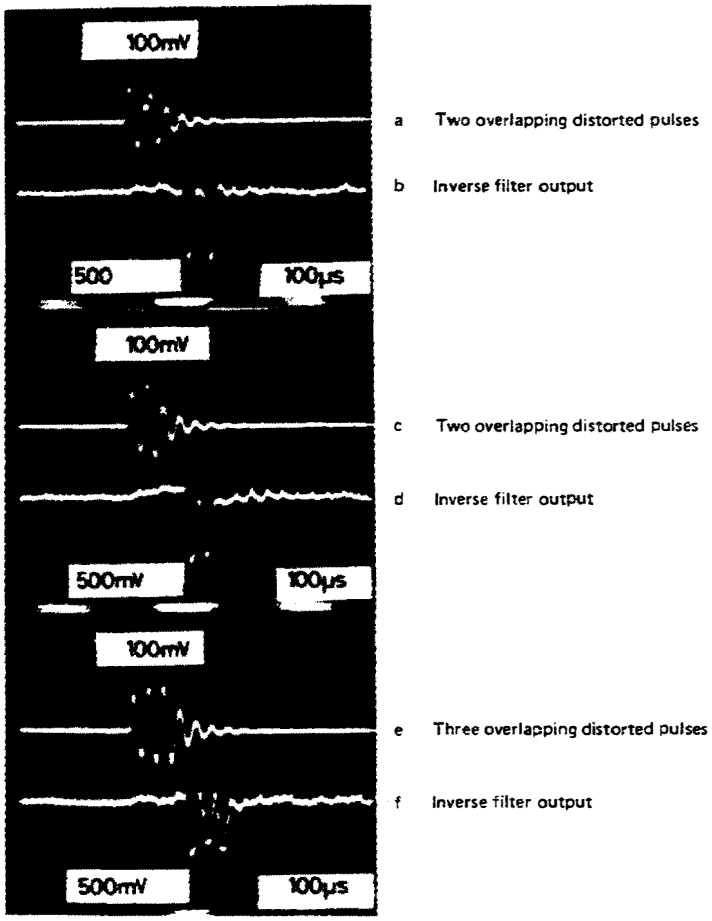


Figure 4.8.7 Deconvolution of overlapping distorted pulses using the inverse filter. (After Corl [88].)

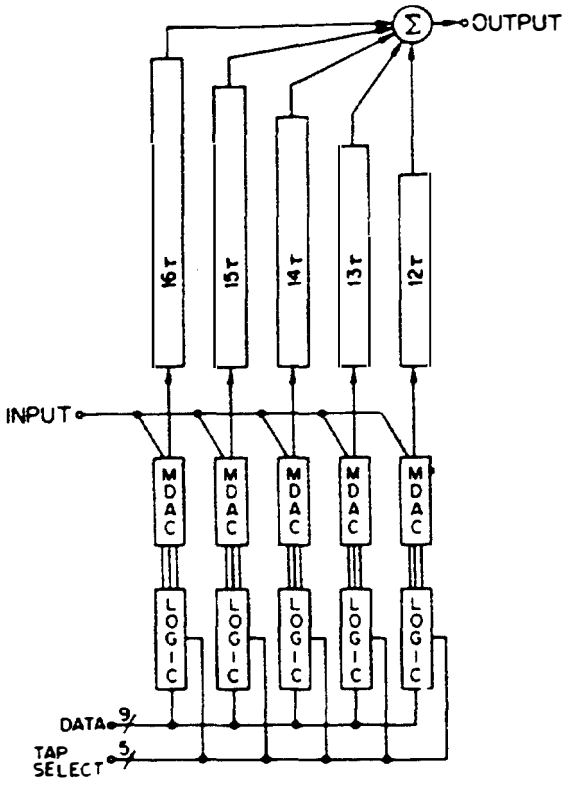


Figure 4.8.8 Tapped filter using separate BBD delay lines. Only five delay lines are shown. (After Tanaka et al. [90].)

The best approximation to such a filter is one with the best mean-square amplitude error over the passband; this is known as a *Wiener filter*. Suppose that the distorted signal is of the form $x(t)$ with a Fourier transform $X(\omega)$. After removing the distortion, the required signal is $d(t)$ and its Fourier transform is $D(\omega)$. Suppose that in addition to the input signal, there is noise present with a spectrum $N(\omega)$. When the input signal plus noise is inserted into a filter with a response $W(\omega)$, the output is

$$Y(\omega) = W(\omega)[X(\omega) + N(\omega)] \tag{4.8.9}$$

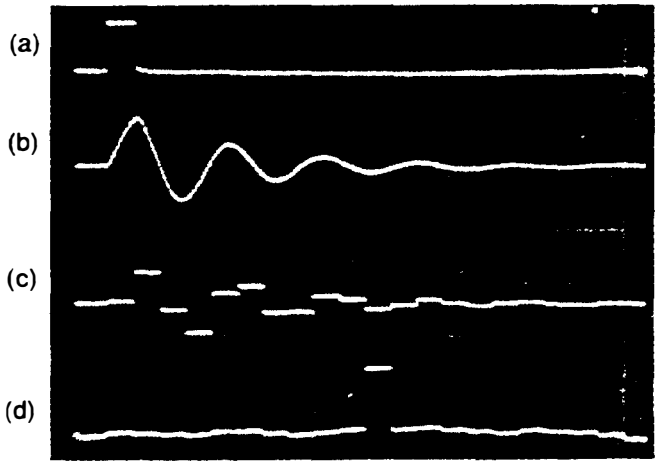


Figure 4.8.9 Adaptive inverse filter waveforms: (a) input impulse; (b) distorted input signal; (c) sampled input signal to the filter; (d) restored impulse output of the filter. (After Tanaka et al. [90].)

The error between the desired signal and the filter output is

$$e(t) = d(t) - y(t) \tag{4.8.10}$$

or

$$E(\omega) = D(\omega) - Y(\omega) \tag{4.8.11}$$

Our criterion in the time domain for an optimum filter will be to choose the parameter ε defined by the relation

$$\varepsilon = \int_0^T \langle e^2(t) \rangle \, dt \tag{4.8.12}$$

to be minimum in the time length T of the signal. This is equivalent to requiring that

$$\varepsilon = \frac{1}{2\pi} \int \langle |E^2(\omega)| \rangle \, d\omega \tag{4.8.13}$$

be minimum over the frequency range of interest. In each case, the symbol $\langle \ \rangle$ indicates the expectation value after a long period or many repeated pulses. Thus we choose the Wiener filter because it gives the least-mean-square error between a desired function and the output of the filter over the time or frequency range of interest.

It will be seen that

$$|E^2(\omega)| = E(\omega)E^*(\omega) = (D - Y)(D^* - Y^*) \tag{4.8.14}$$

so that

$$\begin{aligned} |E^2(\omega)| &= [D - W(X + N)][D^* - W^*(X^* + N^*)] \\ &= DD^* + WW^*(X + N)(X^* + N^*) - WD^*(X + N) \\ &\quad - W^*D(X^* + N^*) \end{aligned} \tag{4.8.15}$$

Because the noise signal can be assumed to be uncorrelated with x or d, there is no phase coherence between N or X and D. Hence, over a long period, $\langle ND^* \rangle = 0$, $\langle NX^* \rangle = 0$, and so on. Thus we can see that

$$\langle |E^2(\omega)| \rangle = DD^* + WW^*[XX^* + \langle NN^* \rangle] - DW^*X^* - D^*WX \tag{4.8.16}$$

The conditions for minimum error are $\partial P/\partial W^* = 0$ and $\partial P/\partial W = 0$. We differentiate inside the integral of Eq. (4.8.13) and equate the differential of the integrand to zero; this process yields the conditions that the differentials with respect to W and W^* of the integrand must be zero. In addition, we have regarded W and W^* as independent quantities, as are their real and imaginary parts. This condition yields the relation

$$\frac{\partial}{\partial W^*} \langle |E^2(\omega)| \rangle = W(XX^* + \langle NN^* \rangle) - DX^* = 0 \tag{4.8.17}$$

Thus the simplest condition for minimum mean-square error between the desired function and the filtered distorted function is that at any frequency within the range

of interest

$$W = \frac{DX^*}{XX^* + \langle NN^* \rangle} \tag{4.8.18}$$

This equation defines the *Wiener filter.*

We note that if $XX^* >> \langle NN^* \rangle$ and the signal-to-noise ratio is large, with D = constant, then $W \rightarrow 1/X$, and the filter behaves like an inverse filter. On the other hand, if $XX^* << \langle NN^* \rangle$, then $W \rightarrow DX^*/\langle NN^* \rangle$. If $D(\omega)$ and $\langle NN^* \rangle$ are constants, the latter response has the form of a matched filter.

It is possible to use Fourier transform techniques to program a computer to set the taps of a delay line transversal filter to the filter of Eq. (4.8.18). This filter will then give the lowest mean-square error between the response desired and the response obtained.

Pseudo-inverse filter. A Wiener filter, as defined in Eq. (4.8.18), is an ideal filter. It is difficult to realize such a filter in practice for two reasons: (1) we must know $N(\omega)$ precisely; and (2) a filter based directly on this principle would not be useful because slight changes in the noise level or the input distortion could change the output signal radically.

There are many approaches to make the design less critical. One method is to construct a filter with a characteristic

$$W(\omega) = \frac{DX^*}{XX^* + \Phi} \tag{4.8.19}$$

where Φ is chosen to be a constant considerably larger than the maximum value of $\langle NN^* \rangle$ in the passband of interest. Such a filter is known as a *pseudo-inverse filter.*

An illustration of the power of this type of filter has been made using a digital computer to process real signals; a schematic of the system is shown in Fig. 4.8.10. An ultrasonic transducer operating at a center frequency of 300 MHz was driven by a 2-ns electric pulse. The transducer excited an acoustic pulse in a solid; this pulse was reflected from the back surface of the solid and used to probe for flaws within it. For good range definition, the shortest possible acoustic pulse is desirable; thus, in this example, a test echo (the reutrn signal from the rear surface of the sample) was made as short as possible after filtering to improve the effective response of the transducer when used as both transmitter and receiver.

The signal received at the transducer was passed into a sampling oscilloscope; this yielded an output that was a slowed-down version of the pulse. An analog-to-digital converter sampled the output and digitized it for insertion into a PDP 11-10 minicomputer. The computer was used to take fast Fourier transforms and to correct the received signal by constructing a pseudo-inverse filter of the form given in Eq. (4.8.19). The value of Φ was fixed at a constant arbitrary value, typically 20 dB below the peak signal value at the center frequency of the transducer. This made the filter less sensitive to slight changes in input conditions. In practice, the smaller the effective value of $\langle |N^2| \rangle$ or Φ used in the design, the more critical

the parameters required. We expect this because we need the pseudo-inverse filter to make far more severe corrections when Φ is effectively very small.

An example of the results obtained are shown in Fig. 4.8.10(a) and the frequency response in Fig. 4.8.10(b). We chose $\langle NN^* \rangle$ or Φ with a value 20 dB below the peak value of $X(\omega)X^*(\omega)$; thus the effective bandwidth of the system now became that of the transducer between its 20-dB points. The output of the Wiener filter is shown in Fig. 4.8.10(c). After gating this output over the frequency range between the original 20-dB points, and retransforming it, we obtained the much improved pulse response shown in Fig. 4.8.10(d). With the pseudo-inverse filter in place, the transducer was successfully used to probe for flaws in solid materials. The transducer could be moved over the surface of the object without readjusting the signal compensation in the computer.

4.8.5 Storage Correlator as an Adaptive Wiener Filter

The storage correlator lends itself well to such applications as an adaptive Wiener or pseudo-inverse filter. One way to construct this type of adaptive filter is to let a computer set the taps of a transversal CCD or SAW filter; this has been done with good results. Another possible application uses a storage correlator in a feedback loop in which the correlator sets itself for minimum error after several iterations. The iteration technique employed is closely related to the *LMS algorithm* (least-mean-square algorithm), due to Widrow and his co-workers [91].

We demonstrate this iterative mechanism with a mathematical derivation carried out in the frequency domain. Suppose that the filter is adjustable and that after the kth iteration, its response is $W_k(\omega)$. Then with a signal $X(\omega)$ inserted, the output is

$$Y_k = W_k(X + N) \tag{4.8.20}$$

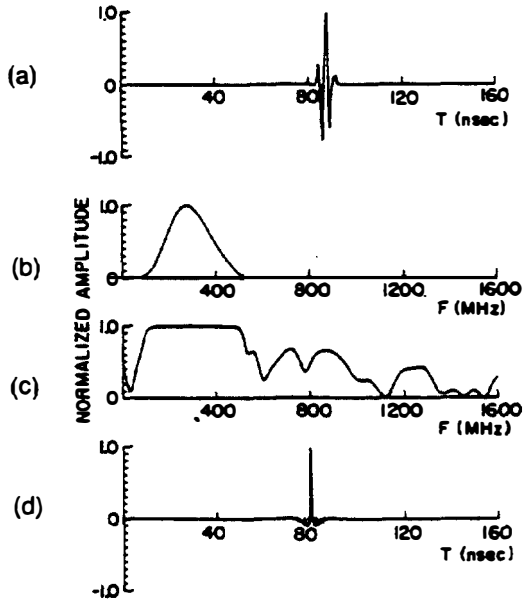


Figure 4.8.10 (a) Impulse response of zinc-oxide-on-sapphire transducer; (b) frequency response of zinc-oxide-on-sapphire transducer; (c) frequency response of the output signal through the filter (note that frequencies above 600 MHz are not used in the final transform; (d) impulse response of the output signal through the Wiener filter. (After Murakami et al. [87].)

where $N(\omega)$ is the noise spectrum. The error between the desired signal and the output is

$$E_k = D - (X + N_k)W_k \tag{4.8.21}$$

Suppose that we try to minimize $|E_k(\omega)|^2$. If we do this, the final value of $W_k(\omega)$ should correspond to a Wiener filter. On the kth iteration $W_k(\omega)$ changes to $W_{k+1}(\omega)$, where

$$W_{k+1} - W_k = \Delta W_k \tag{4.8.22}$$

If we write $S_k = \langle E_k|^2 \rangle$, it follows that

$$\Delta S_k = S_{k+1} - S_k = \langle E_k^* \, \Delta E_k + E_k \, \Delta E_k^* \rangle \tag{4.8.23}$$

Therefore, from Eqs. (4.8.21) through (4.8.23), we see that

$$\Delta S_k = \langle -E_k^*(X + N_k) \, \Delta W_k - E_k(X^* + N_k^*) \, \Delta W_k^* \rangle \tag{4.8.24}$$

We wish to choose ΔW_k so that ΔS_k is always negative and $S_{k+1} < S_k$. This is the condition for the error to decrease on each iteration. A good choice for this purpose is to put

$$\Delta W_k = W_{k+1} - W_k = \mu E_k(X^* + N_k^*) \tag{4.8.25}$$

where μ is a constant. In this case it follows that

$$\Delta S_k = S_{k+1} - S_k = -2\mu S_k \, (XX^* + \langle NN^* \rangle) \tag{4.8.26}$$

For simplicity, we assume that each frequency component is independent and the noise is negligible. We can then write

$$S_{k+1} = S_k[1 - 2\mu(XX^* + \langle NN^* \rangle)] \tag{4.8.27}$$

Thus after k iterations, it follows that

$$S_k = S_0[(1 - 2\mu(XX^* + \langle NN^* \rangle)] \tag{4.8.28}$$

The error converges to zero as $K \to \infty$, provided that $1 - 2\mu(XX^* + \langle NN^* \rangle) < 1$ or $\mu XX^* < 1$ at all frequencies of interest. It also follows that the filter converges to a Wiener filter as $k \to \infty$.

We can explain the operation of the device equally well in the time domain. The output from the filter, when the weighting is $w_k(t)$ and the input is $x(t)$, is

$$y_k(t) = x(t) * w_k(t) \tag{4.8.29}$$

This is the time domain form of Eq. (4.8.20) with the noise term omitted. We subtract the output from the desired signal $d(t)$, and it follows from Eq. (4.8.25) that the weighting must be changed by

$$\Delta w_k(t) = \mu e_k(t) \bigstar x(t) \tag{4.8.30}$$

or by correlating $e_k(t)$, the error in the time domain, with the input signal.

The storage correlator is well suited to carry out these operations. Such a procedure has been carried out by Bowers et al., and the two configurations used

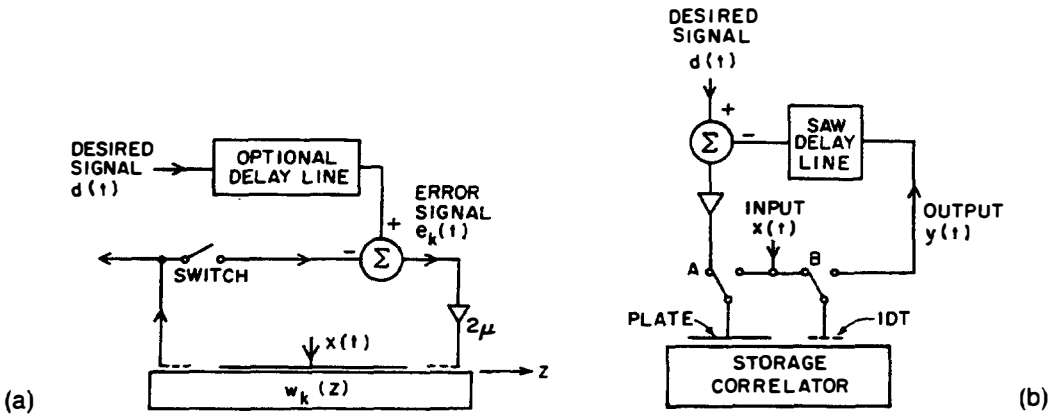


Figure 4.8.11 (a) Adaptive Wiener filter using a storage correlator; (b) second implementation of the adaptive Weiner filter. (From Bowers et al. [92].)

are illustrated in Fig. 4.8.11. The experiment consisted of applying a modulated RF signal $x(t)$ at the acoustic port of a monolithic storage correlator to represent, for instance, a television signal followed by a later echo or its ghost. Using a test signal, such as the line sync pulse, the correlator was required to give a desired output $d(t)$, which ideally would be the signal with the ghost echo removed. The feedback loop carried out the iterative process to make the output $y(t)$ approach the desired single $d(t)$ (a signal pulse) as closely as possible. The received signal $x(t)$ and desired signal $d(t)$ were entered repetitively into the storage correlator for several iteration cycles, where one cycle consisted of two samples of $x(t)$ and one sample of $d(t)$. The whole process involved two basic operations, illustrated in Fig. 4.8.11(a):

1. The convolution between $x(t)$ and the stored impulse response of the correlator $w_k(t)$ was obtained on the kth iteration. The output after the kth iteration was

$$y_k(t) = \int_0^L x\left(t - \frac{z}{V}\right) w_k(z)\, dz \tag{4.8.31}$$

where z is the distance along the correlator, V is the acoustic wave velocity, and L is the length of the device, where $L = VT$. The error $e_k(t) = d(t) - y_k(t)$ was then obtained.

2. The cross-correlation between $x(t)$ and the error signal $e_k^{(t)}$ was taken by inserting $e_{k(t)}$ and $x(t)$ into the device. Note that a delay line was used to give time to clear the device of the original input signal. This process changed the stored weights to $w_{k+1}(z)$, where

$$w_{k+1}(z) = w_k(z) + 2\mu \int_{T_0} x(t) e_k\left(t - \frac{L}{V} + \frac{z}{V}\right) dt \tag{4.8.32}$$

and T_0 is the length of the error signal. The factor 2μ is now a multiplication factor proportional to the gain in the feedback loop and determines the convergence rate. The delays in the system are chosen so that $w_{k+1}(z)$ and

$w_k(z)$ are lined up with each other. This implies that the signals $d(t)$ and $x(t)$ are repeated periodically and that $e_k(t)$ starts at a time $t = T = L/V$. We recognize the integral term as the cross-correlation between $x(t)$ and $e_k(t)$. The process was then repeated. Finally, the output was taken from point A after opening the feedback loop with the switch.

If the $x(t)$ and $d(t)$ are modulated carriers, with the same carrier for both signals, a delay line with the same delay as the storage correlator can be used to delay $d(t)$. In this case the response will be independent of the carrier frequency, provided that this frequency is within the bandwidth of the device.

A somewhat more complicated circuit, like that of Fig. 4.8.11(b), has also been used. The actual configuration uses an RF switch A at the acoustic port, driven synchronously with switch B at the plate. This configuration requires less signal for error correction, using the acoustic port as the input, and does not suffer from reflections at a second interdigital transducer.

An experiment was carried out using this configuration with a zinc-oxide-on-silicon monolithic storage correlator with a bandwidth of 10 MHz, a center frequency of 124 MHz, and a time delay of 3 μs. The signal $x(t)$ consisted of a pulse followed by an echo 6 dB lower in amplitude than the main pulse; each pulse was 0.4 μs long and separated from the next by 0.8 μs. The aim was to remove the echo pulse and to obtain an output consisting of a single pulse.

Figure 4.8.12 shows the signal output before iteration, after one iteration (when the output is the correlation of the distorted signal with itself), and after 22 iterations. The device has been shown to decrease the echo level by approximately 10 dB. This type of filter performs far better even after two iterations than a simple correlation filter, for it is a better approximation of an inverse or a Wiener filter. Other experiments indicate that the time for final convergence is of the order of 200 μs and that it should be possible to improve the performance. The results clearly agree with computer simulation, so the operation of the device is well understood.

4.8.6 Elimination of an Interfering Signal with an SAW Correlator

One technique for eliminating a CW interfering signal has already been described, in Sec. 4.5.4 [46]. This employs a Fourier transform system with a limiter on the Fourier transform output to whiten the spectrum. Another technique uses a storage correlator as an adaptive filter; this forms a *notch filter*, which will eliminate a CW interfering signal. In this case, if we regard the interfering signal as an error signal, we program the filter to make the error zero. One method of doing this is similar to the system already described. Another, which has been demonstrated, is to use the system shown in Fig. 4.8.13. In this system, the device is used to eliminate CW interference from a wideband PN code. Both signals are inserted into the storage correlator at the same time, so that both switches shown are in position 1; the storage correlator then correlates the input signal with itself.

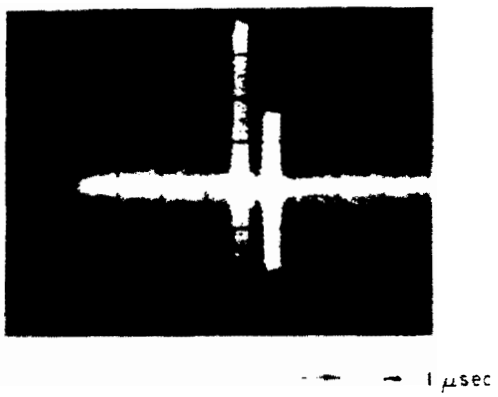


Figure 4.8.12 (a) Input signal. (After Behar et al. [93].) (b) Correlation signal output after one iteration. (c) Output signal after 22 iterations. (After Behar et al. [93].)

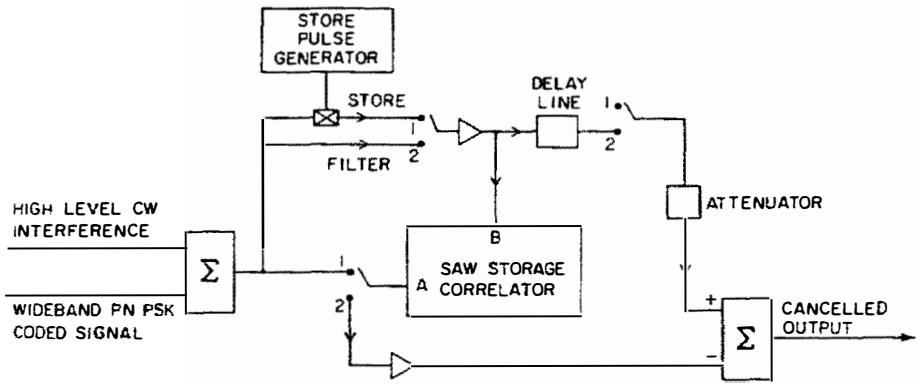


Figure 4.8.13 Surface acoustic wave implementation of narrowband interference canceler. (From Tuan et al. [83], and Grant and Kino [94].)

The PN code gives a correlation peak outside the device, so it has no effect here. With the CW interfering signal, however, the stored weight is of the form

$$w(z) = A \int_0^T \cos \omega_0 t \cos \omega_0 \left(t - \frac{z}{V} \right) dt \approx \tfrac{1}{2} A T \cos \frac{\omega_0 z}{V} \qquad (4.8.33)$$

where A is a constant. If a later signal is read into this device, it will form a narrowband filter. This filter will have a response of the form

$$W(\omega) = \frac{\sin (\omega - \omega_0) L/2V}{(\omega - \omega_0) L/2V} \qquad (4.8.34)$$

Thus its bandwidth will be approximately $\Delta f = 1/T$, where T is the time delay in the storage region of the device.

Returning to Fig. 4.8.13, the switches are switched to position 2 and the storage correlator now functions as a narrowband filter for the CW signal. By adjusting the attenuator, we can use this device to cancel out the direct CW signal passing into the summer. Thus only the PN code is observed at the output. In general, once the attenuator has been adjusted, and provided that the amplitude of response of the device is constant over a given band and the delay in the delay line is equal to that in the correlator, the device will act as a filter for any frequency within this band. Using this technique in the experiment, the 5-MHz rate PN code on a 120-MHz carrier was injected at the same time as an interfering CW signal. By this means, a 30-dB notch with a bandwidth of 330 kHz was introduced into the CW signal, and the CW interference was thus reduced by 30 dB. The correlation peak in the final matched filter for the PN code can be seen clearly in Fig. 4.8.14. The device is capable of simultaneously notching out one or more interfering frequencies anywhere within its 8-MHz bandwidth.

Other and better forms of an interference removing filter are based on the premise that as the error between the filter output and the desired signal decreases on each iteration, the system can be trained with the interfering signal entering each port of Fig. 4.8.11, or with $d(t) = x(t) = i(t)$, where $i(t)$ is the interfering signal. The system makes the error zero. If $d(t)$ is replaced by a signal corresponding to the required signal $s(t)$ and the interfering signal $i(t)$, the output from the operational amplifier will consist of only $s(t)$. Thus the system automatically nulls out interference [95]. If a radar system, for example, is trained on a jamming signal and its own transmitter is then turned on, the jammer will still be canceled out.

Other uses for adaptive filters. Several other types of adaptive filters that use the LMS algorithm to find the optimal set of tap weights have been developed. In one example, a CCD adaptive filter was made with analog tap weights held in sample-and-hold circuits. [96]. Another system uses an analog–digital hybrid approach with MOS-LSI technology. This has the advantage of lowering the power consumption and allowing 32 taps to be used without undue complexity of the external circuit. This technique yielded a large dynamic range, but the bandwidth was limited to less than 1 MHz [97].

An alternative approach for implementing filters with as much as a 500-MHz

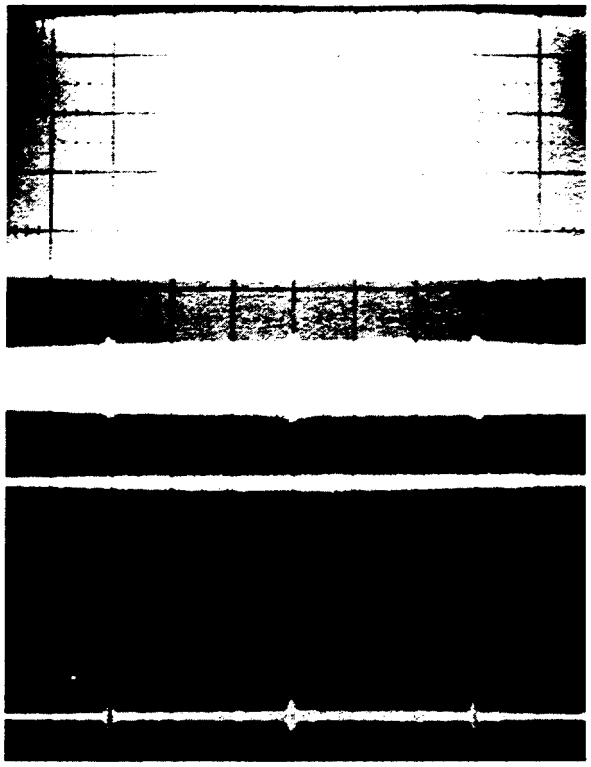
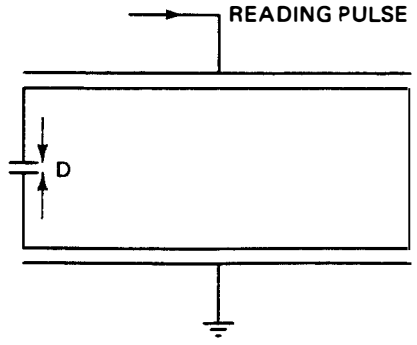


Figure 4.8.14 SAW adaptive filter demonstration. (After Tuan et al. [83].)

bandwidth has been developed by Masenten [98]. It uses tapped SAW filters with a large, complex, computer-controlled system to adjust the tap weights. This system has been used to cancel out interference and to put nulls into an antenna pattern to remove an interfering source.

PROBLEM SET 4.8

1. (a) A modulated RF acoustic pulse (a tone burst) of length τ is injected from a source of very small width D ($D << \lambda$) into a piezoelectric crystal. This pulse passes into the crystal and may be reflected several times from its various surfaces. A very short RF tone burst of frequency 2ω and length τ_0, where $\tau_0 << \tau$ excites an electric field between the two plates, is excited at a time $T > \tau$, as shown. Show that an undistorted but time-reversed pulse should arrive back at the original source after a further time T.

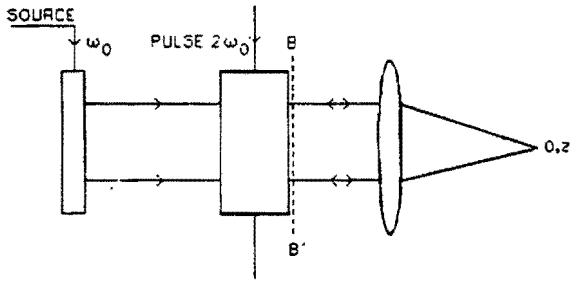


One way to deal with this problem is to take an excited signal of frequency $\omega_0 + \Omega$, amplitude $F(\Omega)$, and phase $\phi(x, y, z)$, together with a reading signal of frequency $2\omega_0$, and consider in general what occurs when a carrier of frequency ω_0 is amplitude-modulated by a signal $f(t)$. The other way is to deal with the problem directly in the time domain and consider the reading pulse to be very short.

(b) Show that if a spatially varying tone burst $f(x, y)$ exp $(j\omega t)$ is injected at the surface $z = 0$ (the front surface of the crystal) and interacts with a short RF E field pulse of frequency 2ω after a time T, there will be a signal received at the input surface at a time $2T$. Suppose that this received signal is retransmitted at this time and a further 2ω pulse excites the crystal at a time $t = 3T$. Show that the main component of the received signal returning to $z = 0$ at a time $t = 4T$ will be $f(x, y)$ exp $(j\omega t)$, so that all spatial and time distortions due to multiple reflections have been removed. You will need to consider carefully the phase of the signals involved, writing phase $\phi(x, y, z)$ along the beam as $\phi(x, y, z) = \omega t - k_x x - k_y y - k_z z$.

Hint: In part (a), the source on the crystal of width D can be regarded as a point source. The problem is set up to avoid difficulties, with the source being a finite number of wavelengths wide. Carry out a simple analysis for a ray propagating at an angle θ to the axis, and derive the rest from physical arguments in parts (a) and (b).

2. (a) Consider the system shown below. A sonar device emits a plane parallel beam from a pulsed source of frequency ω_0 through a plane parallel crystal which has nonlinear characteristics. The parallel beam passes through a lens and is focused in water at the point 0, z. A signal is reflected from an object at this point, returns to the lens and the crystal, and is received on the same transducer, which can only respond to a parallel beam $[\phi(x, y, z) = \text{constant}]$.



Now suppose that the water through which the focused beam passes is turbulent, so that the beam is distorted. The return echo signal from 0, z then arrives with distorted wave fronts after a time T. When the return echoes pass through the short crystal, it is excited with a frequency $2\omega_0$. Show that an acoustic beam will be emitted that will return from 0, z as a parallel beam, whatever the distortion of the medium. You can regard the lens and turbulent medium as a system that introduces a phase change $\phi(x, y, z)$. If there is no turbulence, $\phi(x, y, z) =$ constant at the plane BB'. You do not need to know $\phi(x, y, z)$.

Hint: For simplicity, ignore multiple reflections at the surface of the crystal.

(b) You may wonder if it is possible to obtain a signal large enough to be retransmitted out of the crystal. Consider an equivalent one-dimensional problem by coupled-mode theory. Let the signal entering the crystal be

$$a_1(z, t) = A_1(z)e^{j\omega t} \tag{1}$$

and the signal emitted from the crystal in the backward direction be

$$a_2(z, t) = A_2(z)e^{j\omega t} \tag{2}$$

and the signal (or pump) exciting the crystal between the plates be

$$a_3(t) = A_3 e^{2j\omega t} \tag{3}$$

where A_3 does not vary with z. As the generated signal must depend on the product of the other signals, we can write

$$\frac{dA_1}{dz} + jk_0A_1 = j\alpha A_3A_2^* \tag{4}$$

The second wave is traveling in the backward direction. So we can put

$$\frac{dA_2}{dz} - jk_0A_2 = -j\alpha A_3A_1^* \tag{5}$$

where α is a coupling constant. Note the changes in sign in Eq. (5) from the equivalent forms in Eqs. (4.6.5) and (4.6.6), or Eq. (4).

Beginning at the line before Eq. (6), the problem should read: Prove that it is still true, in this case, that

$$A_1A_1^* + A_2A_2^* = \text{constant} \tag{6}$$

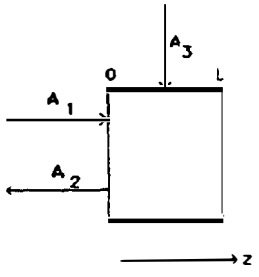
Note that this result is a degenerate form of a more general formula. The power in a transmission line of impedance Z_0, in which there is a voltage V_1e^{-jkz} propagating in the forward direction, is $P_1 = V_1V_1^*/2Z_0$. The power propagating on a transmission line in the backward direction with a voltage V_2e^{jkz} is $P_2 = V_2V_2^*/2Z_0$. Thus, Eq. (6) does, in fact, correspond to $P_1 = P_2 =$ constant if P_2 is defined as being in the backward direction, and $V_1 \alpha A_1$, $V_2 \alpha A_2$.

Power P_3 is supplied to the system, so this is an "active" system, an amplifier which can supply more power output than the input power. In a passive system, in which a forward wave is coupled to a backward wave, conservation of power requires $P_1 + P_2 =$ constant. Here, the situation is more complicated, and the Manley-Rowe relations (conservation of the number of quanta $N = P/\hbar\omega$, or $N_3 = N_1 + N_2$), and conservation of power P must be satisfied.

Now solve for A_1 and A_2^* using the appropriate boundary conditions *at each end* of a crystal of length L. Show that $A_1(z)$ has the form

$$A_1(z) = A_1(0)e^{-jk\alpha z}\frac{\cos \kappa(z - L)}{\cos \kappa L} \tag{7}$$

where $\kappa = \alpha(A_3A_3^*)^{1/2}$. *You will find it convenient to assume that A_3 is real.* Find $A_2(0)/A_1(0)$ and show that the system can self-oscillate if $\kappa L > \pi/2$ (i.e., there will be infinite gain). A device of this kind is known as a *parametric amplifier* or a *parametric oscillator.*



4.9 ACOUSTO-OPTIC FILTERS

4.9.1 Introduction

Light and sound waves interact when they are passed through the same medium. Light is scattered by sound in much the same way as by a diffraction grating, so we can regard the sound wave as a *moving* diffraction grating. Since the wavelength of a very high frequency acoustic wave can be comparable to the wavelength of light, the light beam can be scattered through relatively large angles, thus allowing us to observe a number of important effects. For example, this scattering phenomenon provides an important diagnostic tool for measuring the wavelength, type, and direction of propagation of sound waves and thermal phonons in solid and liquid media. It is also a very useful technique for deflecting a light beam: The amplitude of the deflected beam depends on the amplitude of the sound wave, and the angle of deflection of the beam depends on the acoustic frequency or its wavelength.

The deflection of the optical beam varies almost linearly with the frequency of the acoustic wave, which implies that the light pattern formed by the deflected beam is the Fourier transform of the acoustic beam modulation. If we utilize this type of interaction, a wide range of signal processing methods become available. If we also employ the ability of the optical lens to perform Fourier transforms or use interactions with two acoustic beams, we open many other possibilities for signal processing with acoustic waves in an acousto-optic system, including applications to real-time wideband convolution and correlation of electrical signals.

We can also use the light wave as a "pump" to generate very high frequency sound waves. This technique provides an important means of generating optical phonons in the far-infrared range. Its disadvantage, however, is that laser light of high intensity may unintentionally generate sound waves in certain materials, creating fine damage lines or fractures along its path.

The coupling between light and sound waves results from two effects:

1. The *photoelastic* or piezo-optic effect. Strain due to the presence of a sound wave causes a change in the atomic lattice spacing and hence to the dielectric constant. Thus the relative change in dielectric constant $\Delta\varepsilon/\varepsilon$ is generally proportional to the amplitude of an acoustic wave passing through it, a relation to be described quantitatively by Eq. (4.9.8).
2. The converse *electrostrictive* effect. The presence of an electric field causes a strain essentially proportional to the square of the electric field.

4.9.2 Photoelastic Effect

It is convenient to define the *dielectric impermeability* B as the reciprocal of the relative dielectric constant. In one-dimensional form, we define B by the relation

$$B = \varepsilon_0 \frac{\partial E}{\partial D} \tag{4.9.1}$$

where E is the electric field and D the displacement density. For small signals, the dielectric constant ε would be defined in the same way as $\partial D/\partial E$.

We can write Eq. (4.9.1) in the more general tensor form

$$B_{ij} = \varepsilon_0 \frac{\partial E_i}{\partial D_j} \tag{4.9.2}$$

It follows that the one-dimensional relation between ε and B is

$$B = \frac{\varepsilon_0}{\varepsilon} \tag{4.9.3}$$

More generally, we write

$$\sum_j \frac{\varepsilon_{ij}}{\varepsilon_0} B_{jk} = I \tag{4.9.4}$$

or,

$$\sum_j \frac{\varepsilon_{ij}}{\varepsilon_0} B_{jk} = \delta_{ik} \tag{4.9.5}$$

where $\delta_{ik} = 1, i = k; \delta_{ik} = 0, i \neq k$.

The photoelastic effect relates the change in dielectric impermeability to the strain (the relative change in a length l of the material, defined as $S = \Delta l/l$), as follows:

$$\Delta B = pS \tag{4.9.6}$$

where p is the piezo-optic coefficient. More generally, we can write

$$\Delta B_{ij} = p_{ijkl} S_{kl} \tag{4.9.7}$$

where p_{ijkl} or p_{IJ} is the piezo-optic tensor.

For a cubic crystal, which has an isotropic permittivity, the effective photoelastic constant will depend on the polarization of light relative to the crystal axes and the sound waves. The one-dimensional form of B_{ij} or B_I, which we can use for plane waves, varies with the angle between the direction of sound wave propagation and the direction of the light waves. Different coefficients are required to describe the interaction of light with longitudinal or shear waves, but the relation retains the same form.

It is interesting to note that because of the symmetry relations of the tensors, light diffracted by a longitudinal wave normally gives rise to a diffracted light beam with the same plane of polarization. Light diffracted by a shear wave, however, results in a light wave with a plane of polarization perpendicular to the polarization of the incident light.

It follows that if we keep only first order terms in $\Delta B/B$, $\Delta\varepsilon/\varepsilon$ or $\Delta n/n_0$ in the one-dimensional relations, then

$$\frac{2\Delta n}{n_0} = \frac{\Delta\varepsilon}{\varepsilon} = -\frac{\varepsilon}{\varepsilon_0} pS \tag{4.9.8}$$

where n_0 is the refractive index of the medium and $n_0^2 = \varepsilon/\varepsilon_0$.

We expect the relative change in the dielectric constant to be of the same order as the relative change in density, so that

$$\frac{\Delta\varepsilon}{\varepsilon - \varepsilon_0} \sim -S \tag{4.9.9}$$

Thus we expect $p \sim \varepsilon_0(\varepsilon - \varepsilon_0)/\varepsilon^2$. This is a reasonable approximation to the truth for many materials.

We can carry out a simple analysis, based on the Lorentz–Lorenz law, for a gas or liquid. If the number of molecules per unit volume in the medium is N, the Lorentz–Lorenz law states that

$$\frac{\varepsilon - \varepsilon_0}{\varepsilon + 2\varepsilon_0}\frac{1}{N} = \text{constant} \tag{4.9.10}$$

For a plane wave, where the strain $[S = (\Delta l/l)]$ is assumed to be small, we may write

$$N = N_0(1 - S) \tag{4.9.11}$$

It follows that

$$\frac{\Delta\varepsilon}{\varepsilon} = -\frac{(\varepsilon + 2\varepsilon_0)(\varepsilon - \varepsilon_0)}{3\varepsilon_0\varepsilon}S \tag{4.9.12}$$

or

$$p = \frac{(\varepsilon + 2\varepsilon_0)(\varepsilon - \varepsilon_0)}{3\varepsilon^2} \tag{4.9.13}$$

Example: Piezo-Optic Constant of Water

For water with an index of refraction $n_0 = 1.33$ for light, $\varepsilon/\varepsilon_0 = 1.769$, $\Delta\varepsilon/\varepsilon = -0.546S$, and $p = 0.31$. This compares with a measured value of $p = 0.31$. Values of p for a number of solid materials are given in Appendix B.

4.9.3 Light Diffraction by Sound

An acoustic wave propagating through a solid or a liquid diffracts a light beam much as a diffraction grating does. The sound wave creates a periodic variation of the permittivity of the medium in time and space. Because it travels through the medium with a finite velocity, rather than remaining stationary in space, the acoustic wave Doppler-shifts the frequency of the diffracted light beam, as well as exciting it at an angle to the incident beam.

We first consider a simple physical model for *Raman–Nath diffraction*, when an optical wave of frequency ω_0 and propagation constant k_0 in the material propagates in the z direction. The optical beam interacts with a sound wave of frequency ω_a and propagation constant $k_a = 2\pi/\lambda_a$, propagating in the x direction. We determine the phase shift of the optical beam by using a simple integration of

the phase shift across the acoustic beam. As we shall see later, the results we obtain are valid only when the Raman–Nath parameter Q, defined as

$$Q = \frac{k_a^2 w}{k_0} \tag{4.9.14}$$

is such that $Q << \pi/2$, where w is the acoustic beam width [99]. Several diffraction spots, or orders of diffraction, occur, as illustrated in Fig. 4.9.1. We shall also derive the same results more rigorously and in a convenient normalized form as Eq. (4.9.61) in Sec. 4.9.5.

In Sec. 4.9.5 we shall consider *Bragg interaction*, which occurs when the angle between the direction of propagation of the input optical beam and the z axis is $\pm\theta_B$, where $\sin\theta_B = k_a/2k_0$ and θ_B is called the *Bragg angle*, in analogy to Bragg diffraction of x–rays from crystals. In this case, discussed in detail in Sec. 4.9.5, when $Q >> \pi/2$, only one diffraction spot occurs.

It is convenient to write the variation in the dielectric constant in the form

$$\frac{\Delta\varepsilon}{\varepsilon} = \text{Re}\left[\alpha e^{j(\omega_a t - k_a x)}\right] \tag{4.9.15}$$

or

$$\frac{2\Delta n}{n_0} = \frac{\Delta\varepsilon}{\varepsilon} = \alpha\cos(\omega_a t - k_a x) \tag{4.9.16}$$

where

$$\alpha = \frac{\varepsilon}{\varepsilon_0} pS_0 = n_0^2 pS_0 \tag{4.9.17}$$

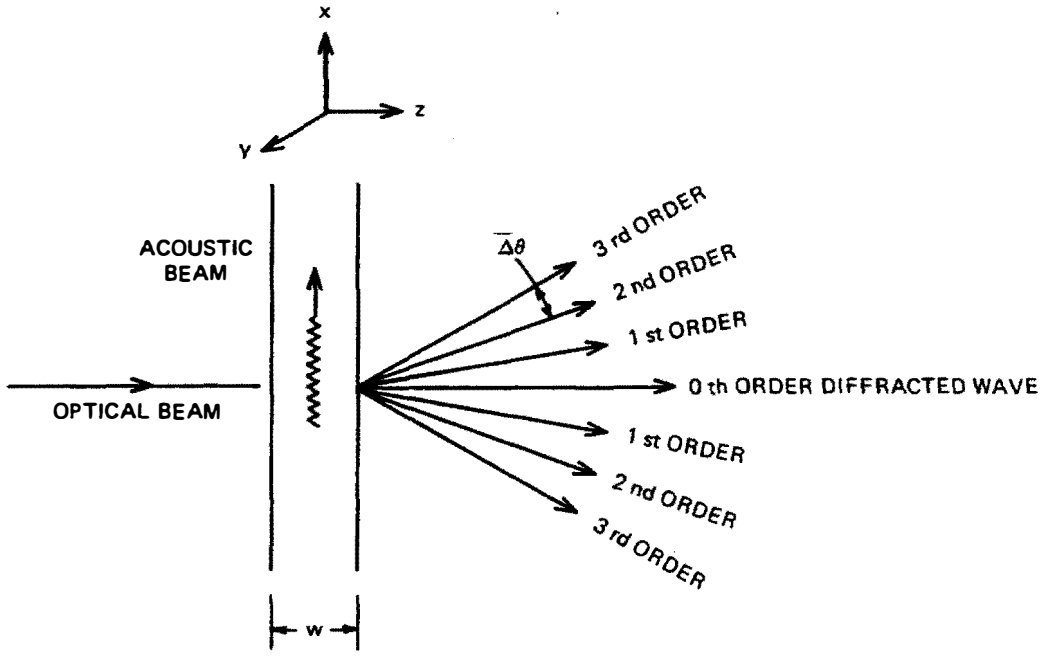


Figure 4.9.1 Raman–Nath diffraction.

and S_0 is the peak strain. As an example, since the acoustic wave velocity in water is 1.5 km/s, the wavelength λ_a at a frequency of 5 MHz is 300 μm or 0.3 mm. This is far larger than the optical wavelength, which is typically less than 1 μm.

We shall assume that the input optical wave in the medium is of frequency ω_0, with propagation constant k_0 in the z direction in the material, which has a refractive index n_0. A simple approximation to the phase change per unit length of the optical wave is therefore

$$\frac{\partial \phi}{\partial z} = -k_0\left(1 + \frac{\Delta\varepsilon}{2\varepsilon}\right) = -k_0\left(1 + \frac{\Delta n}{n_0}\right) \tag{4.9.18}$$

It follows that

$$\frac{\partial \phi}{\partial z} = -k_0\left[1 + \frac{\alpha}{2}\cos(\omega_a t - k_a x)\right] \tag{4.9.19}$$

where $k_0^2 = \omega^2\mu\varepsilon$. The phase change of a wave passing a distance z through the medium is therefore

$$\phi = -k_0 z - \frac{k_0}{2}\int_0^z \alpha \cos(\omega_a t - k_a x)\, dz \tag{4.9.20}$$

We define an amplitude parameter $v(z)$ by the relation

$$v(z) = \tfrac{1}{2}\int_0^z k_0\alpha\, dz \tag{4.9.21}$$

This type of interaction between light and an acoustic beam depends on the integral of the amplitude across the acoustic beam.

If the acoustic beam is uniform over a width w, the phase change in a length w is

$$\phi = -k_0 w\left[1 + \frac{\alpha}{2}\cos(\omega_a t - k_a x)\right] \tag{4.9.22}$$

where

$$v(w) = \frac{k_0\alpha w}{2} \tag{4.9.23}$$

Acousto-optic figure of merit. If A is the area of the acoustic beam, the power in the beam is

$$\begin{aligned} P &= \tfrac{1}{2}AZ_0V_a^2S^2 \\ &= \tfrac{1}{2}A\rho_{m0}V_a^3S_0^2 \end{aligned} \tag{4.9.24}$$

where ρ_{m0} is the density of the material, V_a is the sound wave velocity, and Z_0 is the acoustic impedance.

It follows that we can write $v(w)$ in the form

$$v(w) = \frac{\pi w}{\Lambda}\left(\frac{2PM}{A}\right)^{1/2} = \frac{\pi w\alpha n_0}{\Lambda} \tag{4.9.25}$$

where Λ is the optical wavelength in free space. The parameter M is defined as the *acoustic-optic figure of merit* with

$$M = \frac{p^2 n_0^6}{\rho_{m0} V_a^3} \tag{4.9.26}$$

A table of the parameter M for various materials is given in Appendix B.

For most materials, M is less than the value for water, but for a few exceptional and technically very important materials like tellurium dioxide, M is larger than the value for water.

Example: Calculation of $v(w)$ for Water

For water, $p = 0.32$, $n_0 = 1.33$, $Z_0 = 1.5 \times 10^6$ kg/m^2−s, $V_A = 1.5$ km/s, and $P = 1.69 \times 10^{12} AS_0^2$ watts. With $A = 10^{-5}$ m^2 and $P = 1$ W, it follows that $S_0 = 2.43 \times 10^{-4}$. Thus for $\Lambda = 0.63$ μm and $w = 3 \times 10^{-3}$ m, it follows that $\alpha = 1.38 \times 10^{-4}$, $v(w) = 2.75$.

Returning now to the calculation of phase, it follows that at a position z in the medium, the field $E(x, z)$ is of the form

$$E(x, z) = E_0 e^{j[\omega_0 t - k_0 z - (\alpha k_0 z/2)\cos(\omega_a t - k_a x)]} \tag{4.9.27}$$

where E_0 is the entering plane wave field at $z = 0$. This is an expression for a phase-modulated wave which can be written in the form

$$E(x, z) = E_0 e^{j(\omega_0 t - k_0 z)} \sum_{-\infty}^{\infty} (-j)^n J_{|n|}(v) e^{jn(\omega_a t - k_a x)} \tag{4.9.28}$$

where $J_n(v)$ is an nth-order Bessel function of the first kind and $v = \alpha k_0 z/2$.

We observe that the nth component of the wave has an effective propagation constant

$$\mathbf{k} = \mathbf{a}_z k_0 + \mathbf{a}_x n k_a \tag{4.9.29}$$

where $\mathbf{a}_x$ and $\mathbf{a}_z$ are unit vectors in the x and z directions, respectively. These waves propagate at angles $\pm n\Delta\theta$ to the axis, where $\Delta\theta \approx 2\theta_B$ and θ_B is the Bragg angle, which is defined as

$$\sin \theta_B = \frac{k_a}{2k_0} \tag{4.9.30}$$

Here we have assumed that $\Delta\theta$ is very small.

After a distance w, the intensity of the nth order scattered waves is

$$I_n = I_0 J_n^2(v) \tag{4.9.31}$$

where I_0 is the intensity of the entering optical beam and $v = \alpha k_0 w/2$. The nth-order diffracted beam propagates at an angle $n\Delta\theta$ to the axis. A plot of $[J_n(x)]^2$ for various orders of n is given in Fig. 4.9.2.

From Snell's law, we call the diffraction angle in air

$$\Delta\theta(\text{air}) = n_0 \Delta\theta \text{ (medium)} \tag{4.9.32}$$

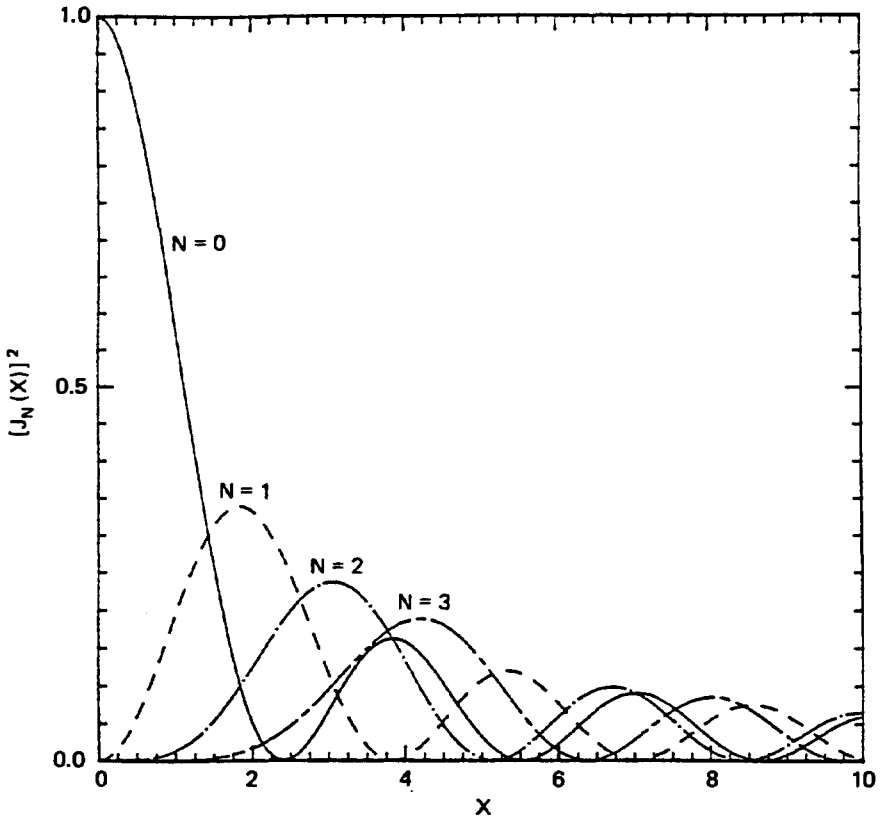


Figure 4.9.2 Plot of the function $[J_n(x)]^2$ for $n = 0, 1, 2$, or 3 as a function of x.

where n_0 is the refractive index of the medium through which the acoustic wave is passing.

We expect to see a series of diffraction spots, the *Raman–Nath diffraction spots*, at angles $\theta_n = n\Delta\theta$ (air). The rays at these angles are Doppler-shifted to frequencies

$$\omega = \omega_0 + n\omega_a \tag{4.9.33}$$

If the parameter v is such that $v \ll 1$, the first-order diffraction spot has an amplitude proportional to v, and the nth-order spot has an amplitude proportional to v^n. If v is larger, the amplitude of the zeroth-order spot decreases to become 0 when $v = 2.405$, while the amplitude of the first-order spot becomes maximum when $v = 1.84$. When v is sufficiently large, we see many diffraction spots, sometimes 20 or more.

We can observe the individual diffraction spots separately only if the beam diameter is somewhat smaller than the separation between spots. Suppose that the beam passes through a hole in a diaphragm of radius a. We can then show, from Fraunhofer diffraction theory (Sec. 3.2), that if the distance from the source is $z \gg a^2/\Lambda$, as illustrated in Fig. 4.9.3, the angle at which the beam amplitude becomes zero is

$$\theta_E(0) = \frac{0.61\Lambda}{a} \tag{4.9.34}$$

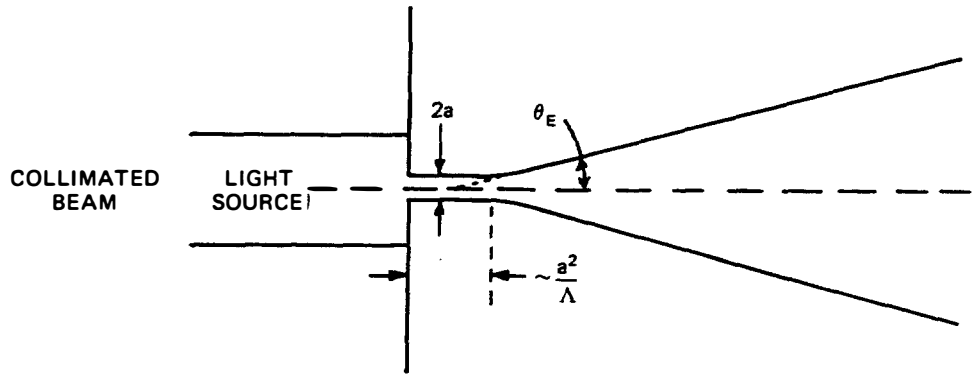


Figure 4.9.3 Collimated beam light source passed through a hole of radius a.

Each acoustically diffracted beam will behave the same way. If we place the maximum amplitude of the diffracted spot at the zero amplitude point of the neighboring spot, each spot should be distinguishable. This condition is the same as the Rayleigh criterion for the definition of a lens, where we require $\Delta\theta > \theta_E(0)$, which is equivalent to saying

$$\frac{k_a}{K} > \frac{0.61\Lambda}{a} \tag{4.9.35}$$

or

$$\frac{a}{\lambda_a} > 0.61 \tag{4.9.36}$$

where $K = 2\pi/\Lambda$. Such a condition is valid only at distances z from the diaphragm, where $z >> a^2/\Lambda$, or the *Fresnel length*.

Suppose that $a = 0.5$ mm and $\Lambda = 6300$ Å. In this case, $a^2/\Lambda = 40$ cm. The Fresnel length varies as a^2, so the condition of Eq. (4.9.35) may be difficult to realize unless we use a lens to refocus the beam. In this case the condition is satisfied at the image plane of the lens.

On the other hand, if we observe the collimated beam at some plane $z < a^2/\Lambda$, the beam radius is approximately a and we require that

$$\frac{k_a}{K} = \Delta\theta \text{ (air)} > \frac{2a}{z} \tag{4.9.37}$$

or

$$a < \frac{\Lambda z}{2\lambda_a} \tag{4.9.38}$$

Limitations of Raman–Nath theory. We have assumed that $k_z \approx k_0$ throughout the analysis. The nth diffracted ray is diffracted at an angle $2n\theta_B$ so its phase change in the z direction becomes approximately $k_0 w \cos 2n\theta_B$. The

Raman–Nath theory neglects the difference between this phase shift and $k_0 w$. Therefore, this theory is valid only if

$$k_0 w(1 - \cos 2n\theta_B) << \frac{\pi}{4} \tag{4.9.39}$$

For $n\theta_B$ small, this relation becomes

$$2n^2 k_0 w \theta_B^2 << \frac{\pi}{4} \tag{4.9.40}$$

where

$$\theta_B \approx \frac{k_a}{2k_0} \approx \frac{\omega_a c}{2V_a \omega_0} \tag{4.9.41}$$

and c is the velocity of light in the photoelastic medium. If $Q = 4k_0 w\theta_B^2$, it follows that the criterion for Raman–Nath diffraction of the first diffraction spot is

$$Q << \frac{\pi}{2} \tag{4.9.42}$$

Example: Raman–Nath Criterion for Water

At $f = 5$ MHz in water, refractive index $n_0 = 1.33$, acoustic velocity $V_a = 1.5$ km/s, we define the maximum value of w for $Q = \pi/2$ as w_{max}. With an optical wavelength $\Lambda = 0.63$ μm and $n = 1$, then $w_{max} = 4.8$ cm. For $n = 2$, w_{max} decreases to 1.2 cm. For $f_a = 50$ MHz and $n = 1$, $w_{max} = 0.5$ mm. For $f_a = 500$ MHz and $n = 1$, $w_{max} = 5$ μm.

We can also see that at $f = 5$ MHz, with $\lambda_a = 0.3$ mm and the wavelength in air $\Lambda = 0.63$ μm, the Bragg angle is $\theta_B(\text{air}) = 0.06°$. At 500 MHz, $\theta_B(\text{air}) = 6°$. Thus, for low acoustic frequencies, θ_B is very small.

4.9.4 Schlieren Effect

The *Schlieren effect* is an important application of Raman–Nath diffraction used in measuring acoustic beam profiles. Consider what occurs when a wide optical beam of width L is passed through a rectangular acoustic beam of width w whose profile is nonuniform in the x and y directions, as illustrated in Fig. 4.9.4. It follows from Eqs. (4.9.28) and (4.9.31) that the acousto-optic interaction yields two first-order diffracted beams, each with an amplitude

$$\begin{aligned} |E(y, z)| &= |E_0 J_1[k_0 w\alpha(y, z)]| \\ &\approx \tfrac{1}{2} |E_0 k_0 w\alpha(y, z)| \qquad (k_0 \alpha w << 1) \end{aligned} \tag{4.9.43}$$

We suppose that after the optical beam passes through the acoustic beam, it enters a convex lens. The undiffracted component of the beam passes through the focal point of the lens. We can remove this component with a stop consisting of a small dot on a glass plate, or with a knife edge, which removes not only the undiffracted beam but one of the first-order diffracted beams as well. It is sometimes convenient to keep both first-order diffracted components to maximize the available light; in this case we use the dot instead of the knife edge.

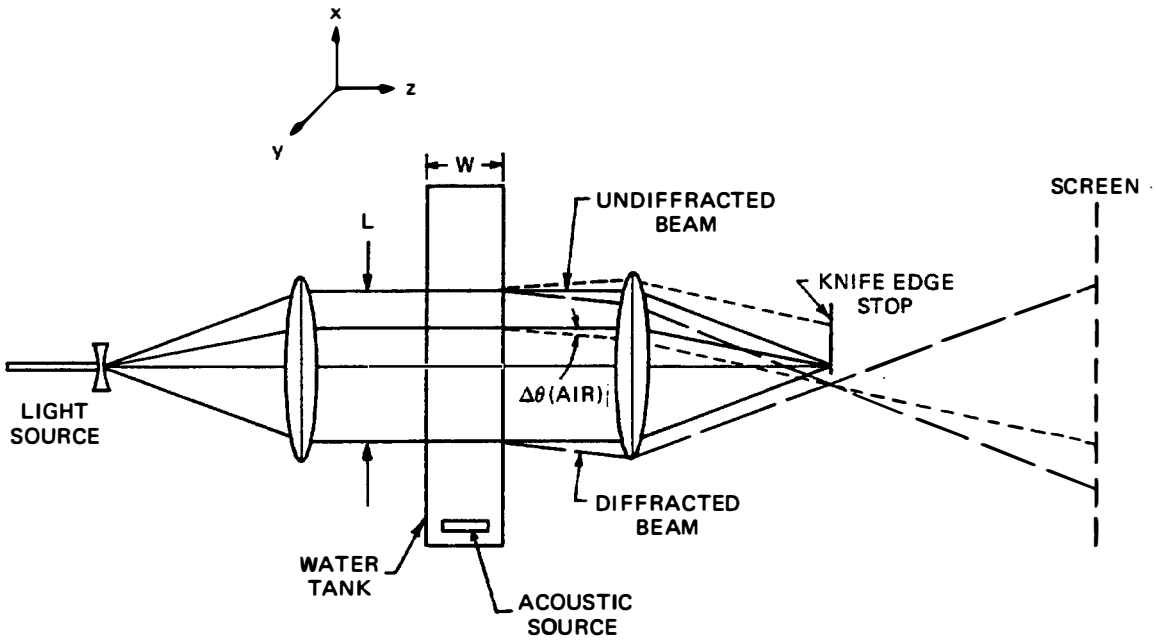


Figure 4.9.4 Schlieren optical system. The first-order diffracted rays are shown as dashed lines. Not all the i-order rays are shown; they may be eliminated by using an opaque dot or a knife edge as a stop.

We then display the diffracted beam on a distant screen. Each point on the screen x_I,y_I has a corresponding point in the acoustic beam x, y, where

$$x_I = x\frac{z - F}{F} \tag{4.9.44}$$

$$y_I = y\frac{z - F}{F} \tag{4.9.45}$$

F is the focal length of the lens and z is the distance of the screen from the focal plane of the lens.

This diffracted beam has each ray displayed by an angle $\Delta\theta$(air) from the corresponding undiffracted ray. If $\Delta\theta$(air) is small, we can display both diffracted rays without significantly blurring the intensity image of the acoustic beam. But if we want a highly defined image, we use a knife edge stop with only one diffracted beam, which yields an image corresponding to the intensity variation $\alpha^2(x, z)$ of the acoustic beam. An image of a square acoustic beam taken with a Schlieren optical system is shown in Fig. 4.9.5(a) and compared to a theoretical diffraction calculation for the same beam in Fig. 4.9.5(b).

If we use a high-intensity laser, an argon laser beam for instance, and pulse both the optical and acoustic source, we can observe the way in which an acoustic pulse propagates through a medium and even how longitudinal, shear wave, and Rayleigh wave modes are generated as a wave is reflected from a solid surface. (Other examples of the use of such a Schlieren optical system are shown in Figs. 2.5.8 and 4.9.15.)

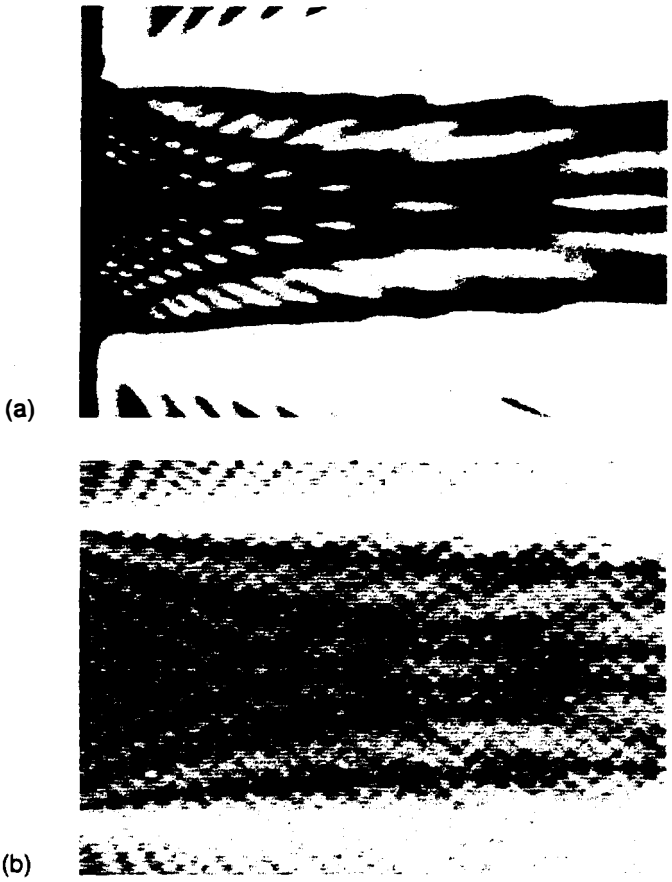


Figure 4.9.5 (a) Schlieren image of the near field of a square transducer with a ratio of width to acoustic wavelength of 14/37. (After Osterhammel, as noted by Cook et al. [100, 101].) (b) Schlieren image numerically calculated by Cook for the same transducer. (After Cook et al. [101].)

4.9.5 Bragg Diffraction by Sound

We now consider *Bragg diffraction*, a condition that occurs when the width of the acoustic beam is large, $Q >> \pi/2$, and there are cumulative effects as the incident optical beam is injected at an angle θ_B to the wavefront of the acoustic beam. In this case, the wave is diffracted at an angle $-\theta_B$ to the acoustic wavefront and there is only one strong diffracted beam. The mathematical formalism we shall use is a rigorous one, suitable for dealing with Raman–Nath as well as Bragg diffraction.

We consider the configuration shown in Fig. 4.9.6, in which a sound wave passes in the x direction through an isotropic solid or liquid. Suppose that a light wave is incident on the material at an angle θ_i to the z direction, and diffracted by the sound wave at an angle $-\theta_R$ to the z direction, as shown in this example. We assume that the sound wave has frequency ω_a and propagation constant k_a, and we take the vector $\mathbf{k}_a$ to be in the x direction. If we want to see continuous and cumulative interaction over a finite length of the crystal, resulting from nonlinear interaction with the sound wave, all contributions to the diffracted wave must be added in phase. This implies that

$$\omega_0 \pm \omega_a = \omega_1 \tag{4.9.46}$$

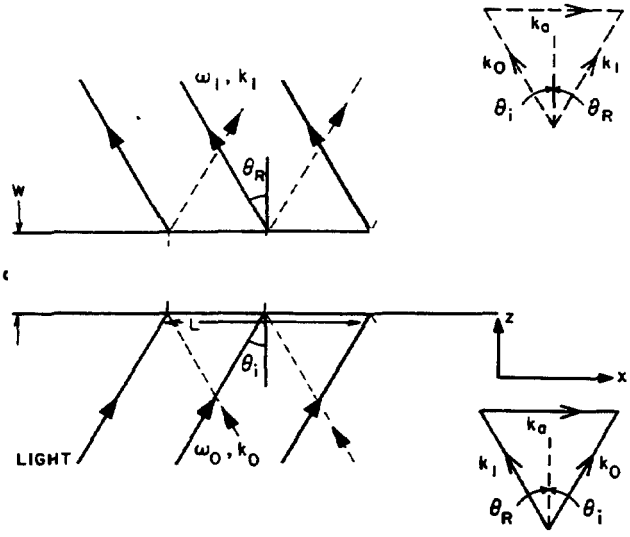


Figure 4.9.6 Bragg diffraction by sound. The solid-line vector diagram shows θ_i positive and $\mathbf{k}_a$ in the $+x$ direction; therefore, $\mathbf{k}_1 = \mathbf{k}_0 - \mathbf{k}_a$ and a diffracted optical beam travels in the $-z$ direction. The dashed lines illustrate Bragg diffraction with θ_i negative and the entering optical beam traveling in the $-x$ direction; $\mathbf{k}_1 = \mathbf{k}_0 + \mathbf{k}_a$ and a diffracted optical beam travels in the $+x$ direction. In the first case the frequency of the diffracted beam is downshifted; in the latter it is upshifted.

and

$$\mathbf{k}_0 \pm \mathbf{k}_a = \mathbf{k}_1 \tag{4.9.47}$$

where all parameters and angles are defined in the material where the interactions take place.

If we take the negative sign and use the angle convention given in the figure, it follows that

$$\omega_0 - \omega_a = \omega_1 \tag{4.9.48}$$

$$k_0 \sin \theta_i - k_a = -k_1 \sin \theta_R \tag{4.9.49}$$

and

$$k_0 \cos \theta_i = k_1 \cos \theta_R \tag{4.9.50}$$

When we consider diffraction by a sound wave in the UHF range, where $\omega_a \ll \omega_0$ or ω_1, we also see that $k_1 \approx k_0$, at least in an isotropic material, in which all light waves of the same frequency have the same propagation constant.† It follows that as $k_1 \approx k_0$, Eqs. (4.9.49) and (4.9.50) yield the results

$$\theta_i = \theta_R \tag{4.9.51}$$

and

$$\sin \theta_i = \sin \theta_R = \frac{k_a}{2k_0} = \frac{\lambda_0}{2\lambda_a} = \sin \theta_B \tag{4.9.52}$$

respectively. The latter condition is the familiar Bragg scattering relation, with the spacing d of the crystal lattice spacing replaced by the acoustic wavelength λ_a.

Note that because λ_0, the light-wave wavelength is typically much smaller than the acoustic wavelength λ_a and the diffraction angles are normally very small.

†The reader is referred to Appendix J for a discussion of Bragg diffraction in anisotropic materials.

We can also consider the solution using positive signs in Eqs. (4.9.46) and (4.9.47). In this case the frequency ω_1 is the upper, rather than the lower, sideband. Now $|\theta_i| = |\theta_R|$ as before, but the incident and diffracted rays, shown by the dashed lines, have positions on the other side of the z axis for those given in the first case.

Diffraction theory. We can now estimate how the amplitude of the diffracted wave increases with distance. The wave equation for the optical waves has the approximate form

$$\frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial z^2} - \frac{\partial^2\phi}{\partial t^2}\mu\varepsilon[1 + \alpha\cos(\omega_0 t - k_a x)] = 0 \tag{4.9.53}$$

where ϕ is any component of the E field.

We look for a solution of the form

$$\phi = e^{j(\omega_0 t - k_0 z\cos\theta_i - k_0 x\sin\theta_i)} \sum_{n=-\infty}^{\infty} e^{jn(\omega_a t - k_a x)}\phi_n(z) \tag{4.9.54}$$

We assume ϕ_n to be a relatively slowly varying function. We write

$$\frac{\partial^2}{\partial z^2}\left(\phi_n(z)e^{-jkz}\right) \approx -\left(k^2\phi_n + 2jk\frac{\partial\phi_n}{\partial z} + \frac{\partial^2\phi_n}{\partial z^2}\right)e^{-jkz} \tag{4.9.55}$$

Since ϕ_n is a slowly varying function, we neglect the last term, $\partial^2\phi_n/\partial z^2$, in the parentheses on the right-hand side of Eq. (4.9.55) compared to $2jk(\partial\phi_n/\partial z)$.

We now consider the term that varies as

$$e^{j[(\omega_0 + n\omega_a)t - (k\sin\theta_i + nk_a)x - k_0 z\cos\theta_i]}$$

We substitute Eqs. (4.9.54) and (4.9.55) into Eq. (4.9.53) to obtain, after some algebra, the following result:

$$\frac{\partial\phi_n}{\partial z} - j\beta_n\phi_n = -\frac{j\alpha k_0}{4\cos\theta_i}(\phi_{n+1} + \phi_{n-1}) \tag{4.9.56}$$

where

$$\beta_n = \frac{2nk_0k_a\sin\theta_i + n^2k_a^2}{2k_0\cos\theta_i}. \tag{4.9.57}$$

and we have assumed that $|n\omega_a| \ll \omega_0$.

Raman–Nath interaction: $Q \ll \pi/2$. Here we put $\theta_i = 0$ and assume that $wn^2k_a^2/(2k_0\cos\theta_i) \to 0$, or $Q \to 0$. Thus we can also assume that $\beta_n = 0$. Now consider the recurrence relation for the Bessel function:

$$\frac{\partial J_n}{\partial z} = -\tfrac{1}{2}[J_{n+1}(z) - J_{n-1}(z)] \tag{4.9.58}$$

and compare it to the simplified form of Eq. (4.9.56) with $\beta_n = 0$:

$$\frac{\partial\phi_n}{\partial z} = -\frac{j\alpha k_0}{4\cos\theta_i}(\phi_{n+1} + \phi_{n-1}) \tag{4.9.59}$$

If we write $\phi_n = (-j)^n \phi_n'$, and set $\theta_i = 0$, we find that

$$\frac{\partial \phi_n'}{\partial z} = \frac{-\alpha k_0}{4} (\phi_{n+}' \quad \phi_{n-1}') \tag{4.9.60}$$

Thus the solution to Eq. (4.9.60), if $\phi_0(0) = 1$, is

$$\phi_n' = J_n\left(\frac{k_0 \alpha z}{2}\right) = J_n\left[\frac{\pi z}{\Lambda}\left(\frac{PM}{2A}\right)^{1/2}\right] \tag{4.9.61}$$

This leads us to the Raman–Nath solution described by Eq. (4.9.28) and (4.9.31), with the same limitation, that $Q << \pi/2$.

Bragg regime: $Q >> \pi/2$. We now assume that the only waves of importance are the $n = 0$, and either the $n = 1$ or $n = -1$ components. When only the $n = 0$ and $n = 1$ terms matter, Eq. (4.9.56) leads to the results

$$\frac{\partial \phi_1}{\partial z} = \frac{-j\alpha k_0}{4 \cos \theta_i} \phi_0 \tag{4.9.62}$$

and

$$\frac{\partial \phi_0}{\partial z} = \frac{-j\alpha k_0}{4 \cos \theta_i} \phi_1 \tag{4.9.63}$$

If $\beta_1 = 0$, it follows from Eq. (4.9.57) that

$$\sin \theta_i = -\frac{k_a}{2k_0} = -\sin \theta_B \tag{4.9.64}$$

The similar condition for $n = -1$ is exactly the same as the Bragg condition of Eq. (4.9.52), while Eq. (4.9.64) simply corresponds to the Bragg condition when θ_i is negative.

When the Bragg condition is satisfied, the interaction is cumulative. When α is constant, ϕ_1 initially increases linearly with distance. Equations (4.9.62) and (4.9.63) take the form of two simple coupled-mode equations. For α constant, Eqs. (4.9.62) and (4.9.63) give the solution

$$\phi_1(z) = -j\phi_0(0) \sin\left(\frac{\alpha k_0 z}{\cos \theta_i}\right) \tag{4.9.65}$$

So the ratio of the diffracted beam intensity to the incident beam intensity I_0 is

$$\frac{I_1}{I_0} = \sin^2\left(\frac{\pi z}{\Lambda}\sqrt{\frac{PM}{2A}}\right) \tag{4.9.66}$$

We observe that the intensity of the diffracted beam initially increases as the square of the distance, reaches a maximum value, and then decreases.

Response as a function of acoustic beam profile. In the general case when the Bragg condition is not exactly satisfied and we assume α to be nonuniform, we can obtain the solution to Eqs. (4.9.62) and (4.9.63). For simplicity, we shall

assume from now on that the interaction length is short enough so that $\phi_0(z)$ is constant. We can then solve Eq. (4.9.62) by writing it in the form

$$e^{-j\beta_1 z}\frac{\partial}{\partial z}(\phi_1 e^{j\beta_1 z}) = -\frac{j\alpha k_0}{4\cos\theta_i}\phi_0 \tag{4.9.67}$$

with the result

$$\phi_1 = -\frac{jk_0 e^{-j\beta_1 z}\phi_0}{4\cos\theta_i}\int_0^z \alpha(z)e^{-j\beta_1 z}\,dz \tag{4.9.68}$$

It follows that, in this case, ϕ_1 is the Fourier transform of the acoustic beam profile.

Response as a function of input angle θ_i. When β_1 is finite and $\theta_1 \neq -\theta_B$, then it is convenient to write, for θ_B small,

$$\theta_i = -\theta_B + \Delta\theta \tag{4.9.69}$$

It follows from Eqs. (4.9.57) and (4.9.69) that $\beta_1 \approx k_a\,\Delta\theta$ and

$$\phi_1(\Delta\theta) = -\frac{jk_0 e^{-jk_a z\Delta\theta}\phi_0}{4\cos\theta_B}\int_0^z \alpha(z)e^{-jk_a z\Delta\theta}dz \tag{4.9.70}$$

If we rotate the crystal through an angle $\Delta\theta$ from the correct incident angle θ_B, Eq. (4.9.70) implies that the Fourier transform of the acoustic beam profile will be $\phi_1(\Delta\theta)$.

Response as a function of input frequency. We can find a similar relation for the response as a function of input frequency. From Eq. (4.9.54) we see that if the $n = 0$ term has a constant amplitude, the $n = 1$ diffracted wave varies as

$$\phi(x, z) = \phi_1 e^{j(\omega_0+\omega_a)t}e^{-jk_0 z\cos\theta_i}e^{-j(k_0\sin\theta_i + k_a)x} \tag{4.9.71}$$

If θ_i is kept constant and ω_a is varied, the output angle is θ_R, where the propagation constant of the diffracted beam in the x direction is $k_x = k_0 \sin\theta_R$, and

$$\begin{aligned}\sin\theta_R &= \frac{k_a + k_0\sin\theta_i}{k_0}\\ &= \sin\theta_i + \frac{k_a}{k_0}\end{aligned} \tag{4.9.72}$$

If we take the center frequency to be ω_{a0}, corresponding to a value of k_{a0}, where

$$-\sin\theta_{i0} = \frac{k_{a0}}{2k_0} = \sin\theta_{R0} = \sin\theta_B \tag{4.9.73}$$

and we assume that θ_R and θ_i are small, we find that

$$\theta_R \approx \frac{k_a - k_{a0}}{k_0} + \theta_{R0} \tag{4.9.74}$$

The angle of the emerging beam therefore varies linearly with the frequency of the acoustic wave. For general values of θ_R and θ_{R0}, the x position of the diffracted beam spot at the plane z is

$$x = z \sin \theta_R = z \sin \theta_{R0} + \frac{k_a - k_{a0}}{k_{a0}} z \tag{4.9.75}$$

Therefore, the x position of the diffracted spot varies linearly with frequency, and the amplitude variation of the diffracted beam with x position is the Fourier transform of the time response of the input acoustic signal.

When we vary the frequency, it follows from Eq. (4.9.57) that as $\beta_1 = 0$ when $k_a = k_{a0}$, then to first order in $(k_a - k_{a0})$,

$$\beta_1 = -(k_a - k_{a0}) \tan \theta_{i0} \tag{4.9.76}$$

If we substitute Eqs. (4.9.73) and (4.9.74) in Eq. (4.9.76), we see that for θ_{R0} and $|\theta_R - \theta_{R0}|$ small,

$$\beta_1 = \frac{(\theta_R - \theta_{R0})k_{a0}}{2} \tag{4.9.77}$$

We can then show from Eq. (4.9.68) that for a beam of width w, the variation of the diffracted beam amplitude with β_1 is

$$\left|\frac{\phi_1}{\phi_{10}}\right| = \left|\frac{\sin \beta_1 w/2}{\beta_1 w/2}\right| \tag{4.9.78}$$

where ϕ_{10} is the amplitude of the output beam when $\theta_i = \theta_{i0}$ and $\omega_a = \omega_{a0}$.

The angular range of the output beam to the 3-dB points is

$$\Delta\theta_R(3 \text{ dB}) = \frac{0.89\lambda_{a0}}{w} \tag{4.9.79}$$

Similarly, if we observe the output beam at a constant angle θ_{R0}, the bandwidth B between 3-dB points is

$$B = 0.89 \frac{V_a}{w \tan \theta_{i0}} \tag{4.9.80}$$

If the length of the acoustic beam is L, however, we can show that the diffracted optical beam has an angular extent to the 3-dB points of

$$\Delta\theta_E(3 \text{ dB}) = \frac{0.89\lambda_0}{L} \tag{4.9.81}$$

It follows that the number of resolvable spots N in the beam is

$$N = \frac{\Delta\theta_R}{\Delta\theta_E} = \frac{L}{w}\frac{\lambda_{a0}}{\lambda_0} \tag{4.9.82}$$

or

$$N = 1.12BT \tag{4.9.83}$$

where $T = L/V_a$ is the time delay of the acoustic beam in a length L.

The basic physical principles governing each of the transversal filters we have examined in this chapter lead us to the same conclusion: *The number of resolvable points in any signal processing system is approximately equal to the time–bandwidth product of the device.*

4.9.6 Application of Bragg Deflection Systems

The first demonstrations of Bragg interaction used bulk waves and gave only limited angular and frequency bandwidths, as Eqs. (4.9.80) and (4.9.81) predicted. To circumvent this limitation in practical systems, several separate transducers, each with a different center frequency, are used. The individual transducers are placed at the optimum angle to the incident optical beam for efficient acousto-optic interaction. By this means, we can divide the total bandwidth into several frequency ranges to obtain a flat confined response over bandwidths of several hundred megahertz.

Surface acoustic waves are often used to deflect the optical beam. This makes it possible to employ a guided optical wave in place of the large optical components, attached to a stable table, that are usually required in a system using bulk acoustic waves. A layer of material is typically deposited on an SAW substrate, as illustrated in Fig. 4.9.7, and this waveguide layer is then arranged to propagate light waves with lower velocity than the substrate. As we saw earlier in our discussions of acoustic waveguides (Chapter 2 and Sec. 4.7), this implies that the total internal reflection of the light occurs at the top and bottom edges of the layer. Thus this layer acts as a waveguide, one that can consist, for instance, of a zinc oxide layer, like those used in the SAW devices described in Secs. 4.7 and 4.8. We can also form a high-quality, low-loss guide by diffusing titanium into lithium niobate.

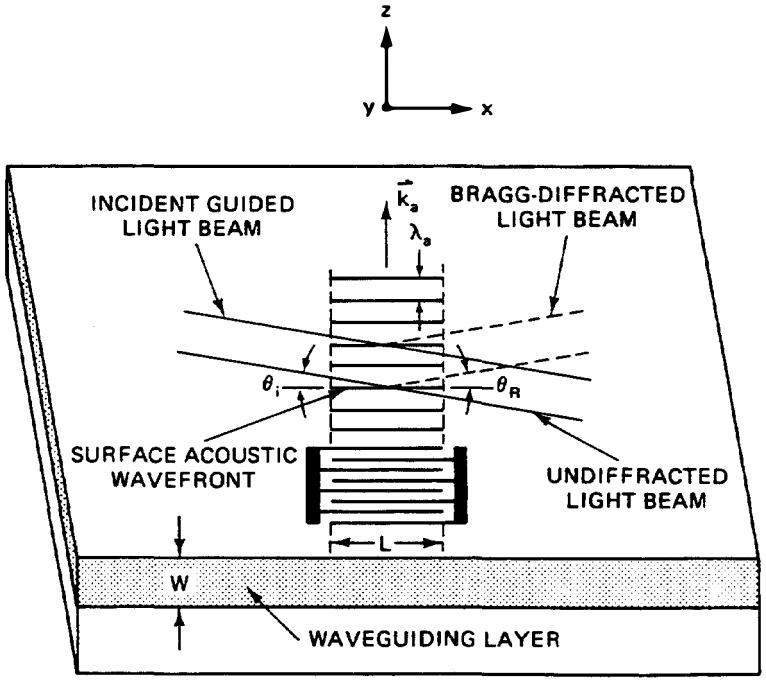


Figure 4.9.7 Guided-wave Bragg diffraction from single-surface acoustic wave. (After Tsai [102].)

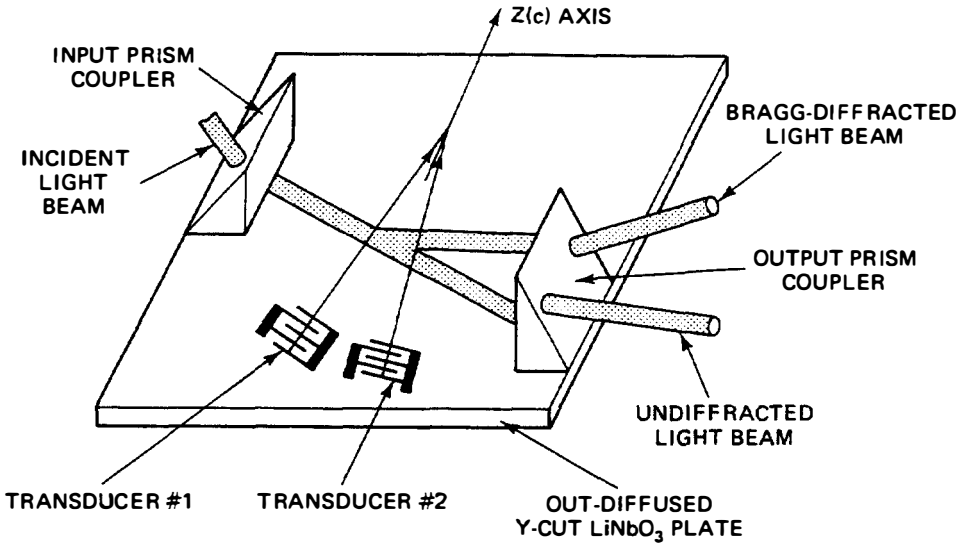


Figure 4.9.8 Experimental configuration for guided-wave acousto-optic Bragg deflection. (After Tsai [102].)

We introduce the optical beam into the guide by means of a *prism coupler*, illustrated in Fig. 4.9.8, a device operated on the same principle as the wedge coupler used with surface acoustic waves, which was described in Sec. 2.5.2. The prism is normally constructed of a material with a larger refractive index than the waveguide material. Thus if the propagation constant of a wave in the prism is k_p and the angle of the prism is θ, the effective propagation constant along the surface of the guide is $k_p \sin \theta$. If k_0 is the propagation constant of the guided wave and

$$k_p \sin \theta = k_0 \tag{4.9.84}$$

the two propagation constants will match and we can excite a waveguide mode. Using this method, optimum optical conversion efficiencies of the order of $\sim 80\%$ have been observed.

The Bragg deflector can be used for broadband operation if several SAW transducers, tilted with respect to each other, are employed, as illustrated in Fig. 4.9.9. Using three transducers in this arrangement, shown in Fig. 4.9.10, Tsai obtained a bandwidth of 385 MHz for a Bragg deflector.

Devices of this kind have been employed as wideband spectrum analyzers, operating in real time as acousto-optic modulators and acousto-optic phase shifters. Thus they provide a very useful component for optical waveguide systems.†

4.9.7 Acousto-Optic Convolver

An important type of signal processing device can be constructed using two acousto-optic deflectors in succession, as illustrated in Fig. 4.9.11. In such a device, known as the *acousto-optic convolver*, the optical beam enters at an angle θ_i to the normal

†An alternative technique which uses anisotropic materials to obtain broadband operation is discussed in Appendix J.

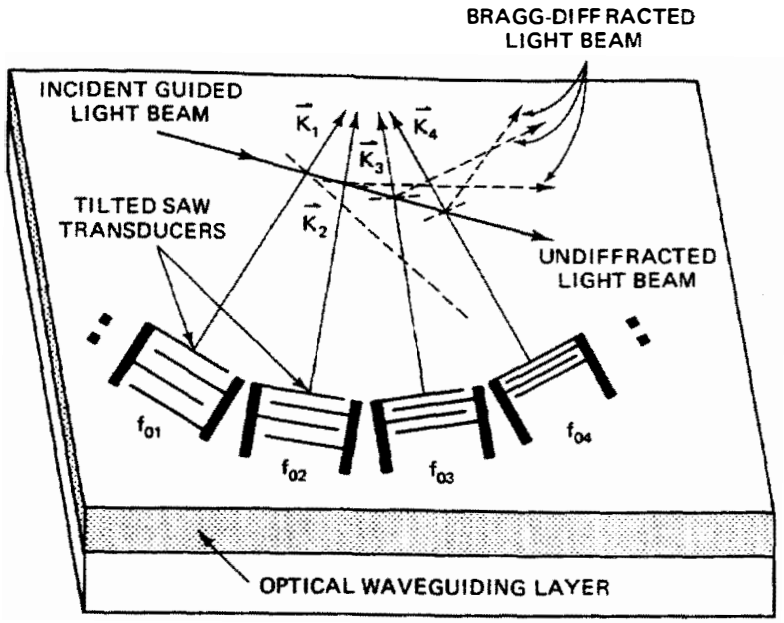


Figure 4.9.9 Guided-wave Bragg diffraction from multiple tilted surface acoustic waves. (After Tsai [102].)

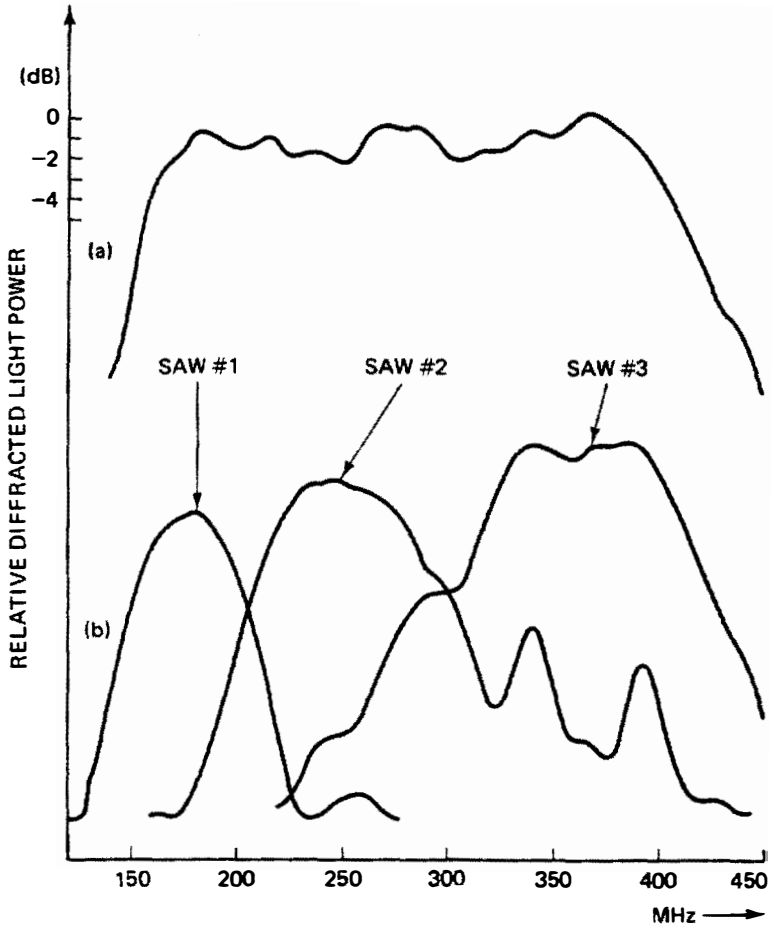


Figure 4.9.10 Frequency response of the Bragg diffracted light power from tilted surface acoustic waves (experimental): (a) three combined-surface acoustic waves; (b) individual-surface acoustic waves. (After Tsai [102].)

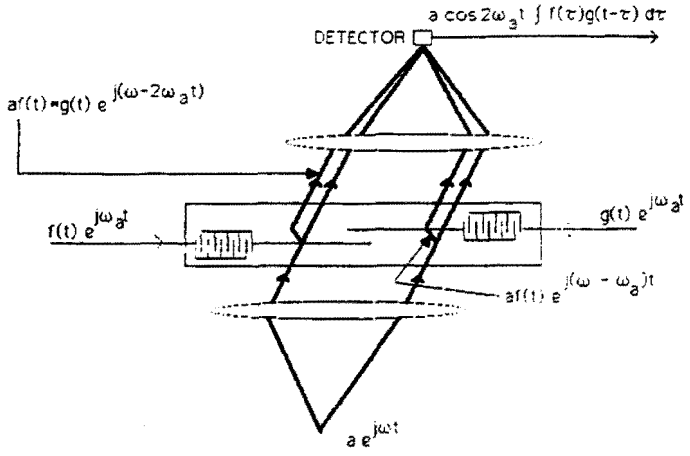


Figure 4.9.11 Waveform convolution using acousto-optic Bragg interaction. (From Lee et al. [103].)

acoustic beam axis. After diffraction by the first acoustic beam, the diffracted beam exits at an angle $\theta_R = -\theta_i$. This diffracted beam is diffracted once more by an acoustic beam traveling in the opposite direction and the rediffracted beam leaves the system at an angle θ_i. This beam, twice deflected by the acoustic waves, is modulated at a frequency $2\omega_a$. If we can separate this modulation component from the carrier and the component of frequency ω_a, we can detect the twice-deflected beam.

This system was first demonstrated by Lee et al. [103]. They employed a thin flat laser beam, focused by a cylindrical lens instead of an optical waveguide mode because the length of the acoustic beam (the aperture was 15 cm) and the lack of suitable lenses made it easier to produce a flat sheet beam from a 400-mW krypton laser ($\lambda = 0.657$ μm) than from a waveguide mode. A lens was used on the output to focus the diffracted beam to a small spot on a high-speed photodetector.

Suppose that the two acoustic wave signals entering the system in opposite directions are $f(t)\exp(j\omega_a t)$ and $g(t)\exp(j\omega_a t)$, respectively. The twice-diffracted beam is focused onto a photodetector. The signal received at the photodetector will be of the form

$$S(t) = e^{j(\omega+2\omega_a t)} \int f\left(t - \frac{z}{v}\right) g\left(t + \frac{z}{v}\right) dz + \text{c.c.} \tag{4.9.85}$$

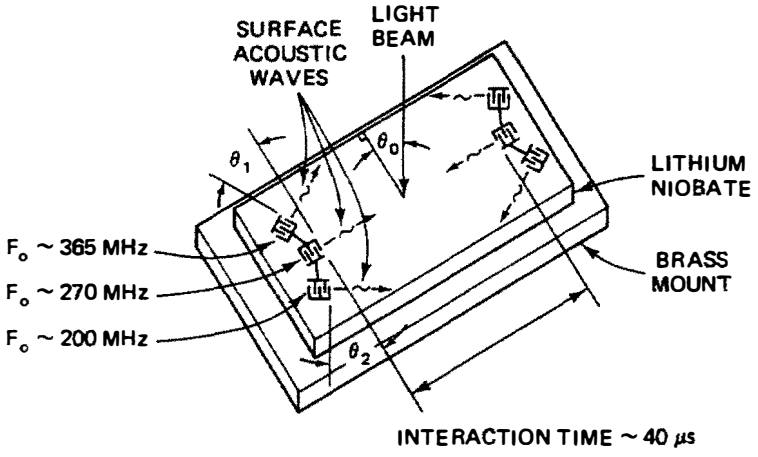


Figure 4.9.12 Large time-bandwidth, acousto-optic convolver. (After Berg and Udelson [104].)

where c.c. denotes the complex conjugate. In addition to this diffracted beam, there will be an undiffracted beam of frequency ω. When these two signals are mixed at the photodetector, the output at frequency $2\omega_a$ will be the convolution of the two input signals.

Another practical system, demonstrated by Berg and Udelson and illustrated in Fig. 4.9.12, employed a configuration with several input transducers of the type already described. Using three transducers in this manner, they obtained an input bandwidth of approximately 200 MHz and therefore a 400-MHz output bandwidth. Because the time delay in their device was of the order of 36 μs, we might expect it to be able to process signals with a time–bandwidth product of the order of 7000. Berg and Udelson obtained a dynamic range of 50 to 60 dB. They have used the system to carry out fast Fourier transforms and to process wide-bandwidth digital codes. Figures 4.9.13 and 4.9.14 show examples of phase-coded input and the corresponding outputs from the acousto-optic convolver. It is apparent that the acousto-optic convolver is a very powerful device for processing large-bandwidth signals.

4.9.8 Time-Integrating Correlator

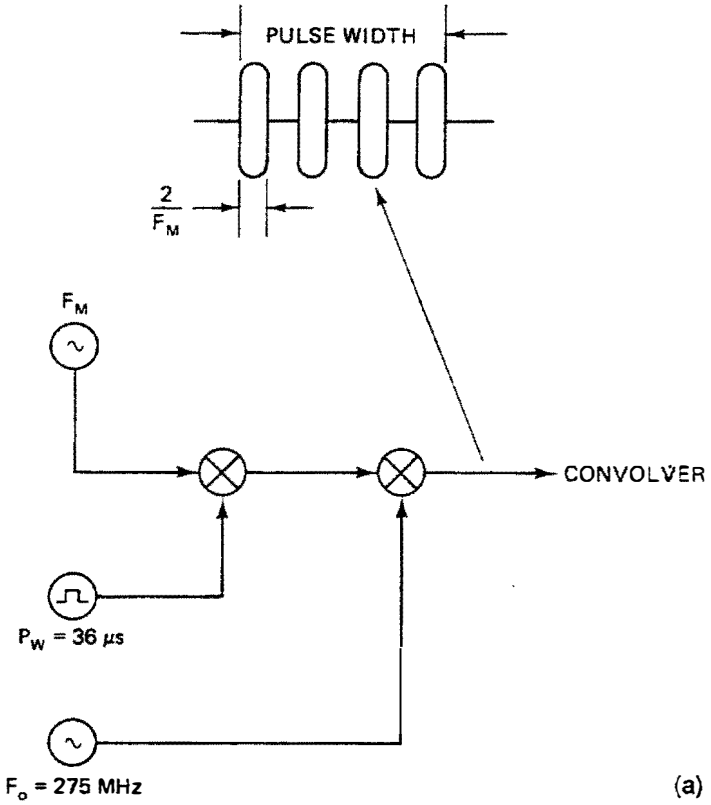
Another interesting example of the many types of acousto-optic processors which have been developed is the time-integrating optical signal processor. This can make use of an incoherent optical source and, like the storage correlator, uses integration in time to obtain correlation between two signals. Because it employs optical signal processing, it is also possible to construct a two-dimensional correlator which can directly display real-time ambiguity functions.

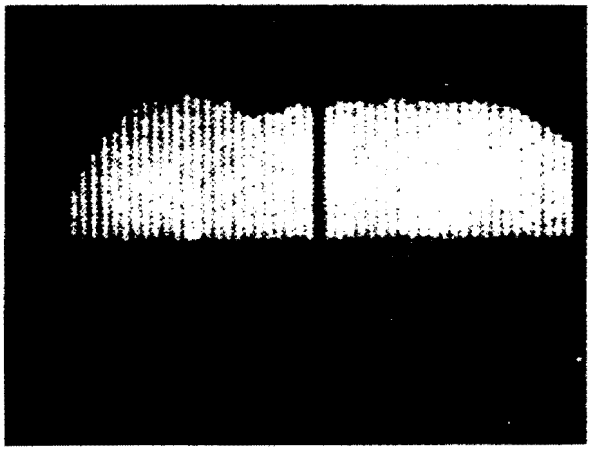
We consider the system illustrated in Fig. 4.9.15. In this device, a laser diode or an LED is modulated by an input signal $f(t)$. A second signal $g(t)$ is inserted into an acousto-optic modulator and the output light beam is then passed through a Schlieren optical system of the type described in Sec. 4.9.4. A second lens is added to the system after the knife-edge stop and the output beam is imaged onto a CCD array. The output received at a point x on the array contains a term proportional to $f(t)\, g(t - x/V)$, where V is the effective velocity of a signal along the array. As we saw above, a Schlieren optical system basically images the amplitude of the acoustic wave at any point in the acoustic beam to a corresponding point on the image plane of the Schlieren optical system. By integrating the output of the photodetector as a function of time, we obtain the correlation function of the two input signals

$$R(x) = \int f(t)\, g\left(t - \frac{x}{V}\right) dt \tag{4.9.86}$$

By reading out the image on the CCD array, we can obtain the correlation function directly.

The time–bandwidth product of the system, much as with the storage correlator, is dictated by the bandwidth of the input system and the integration time of the CCD array elements. The portion of the correlation function that can be read out is dictated by the delay time in the acousto-optic modulator. Thus the





(b)

Figure 4.9.13 (a) Circuit generating phase-coded inputs for the acousto-optic convolver; (b) frequency-domain output from the convolver. (After Berg and Udelson [104].)

basic system has the same kinds of limitations on time bandwidth as the storage correlator, as well as a portion of the correlation function that can be read. The system itself has major advantages over many other kinds of acousto-optic correlators utilizing an optical reference to convert the acoustic phase modulation to amplitude modulation, because it does not require coherent optics with their attendant problems of speckle and noise.

We now analyze this system in more detail. The output of the photodetector is proportional to the image intensity $I(t, x)$, which can be written in the form

$$I(t, x) = I_1(t)I_2\left(t - \frac{x}{V}\right) \tag{4.9.87}$$

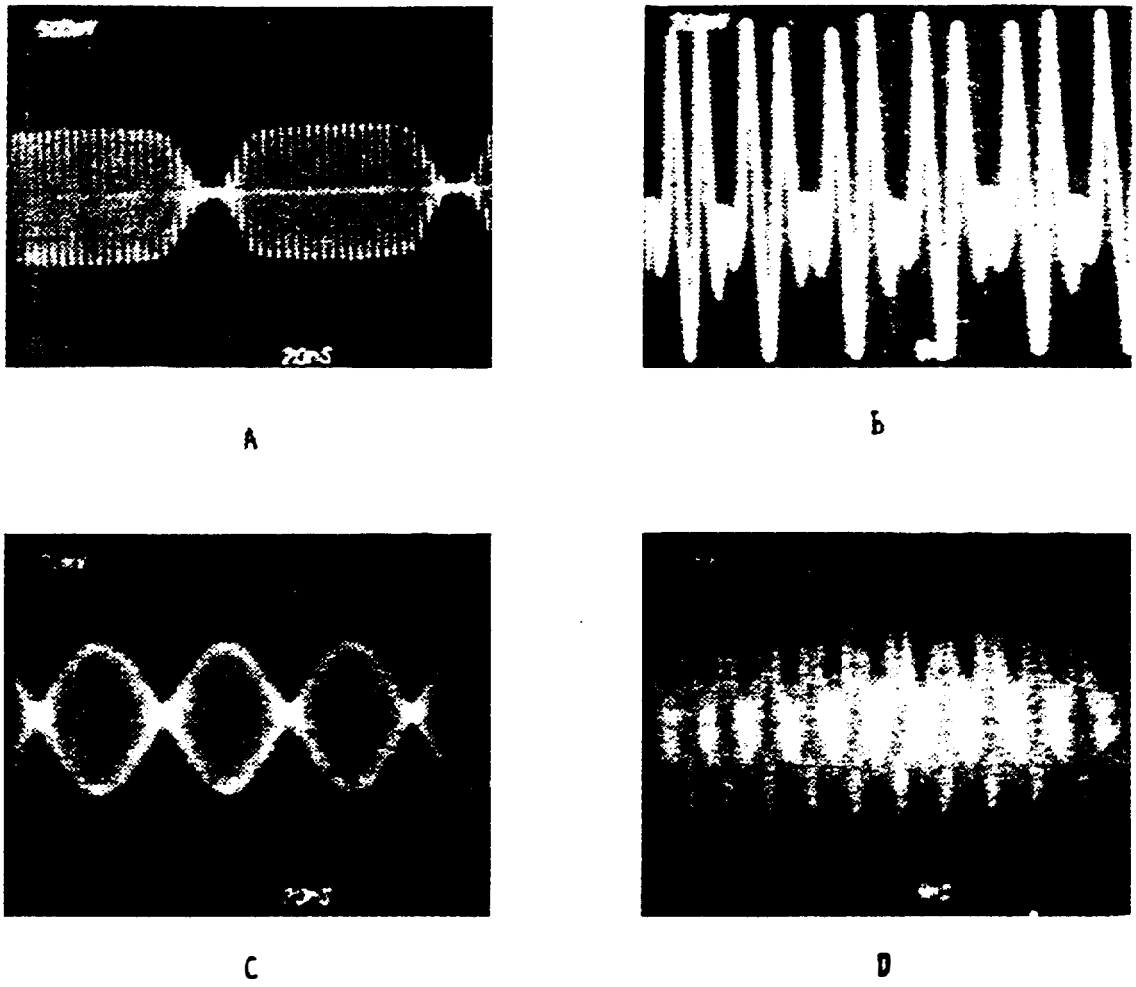


Figure 4.9.14 (a) Phase-coded inputs to the acousto-optic convolver; (b) corresponding time-domain outputs from the convolver. (After Berg and Udelson [104].)

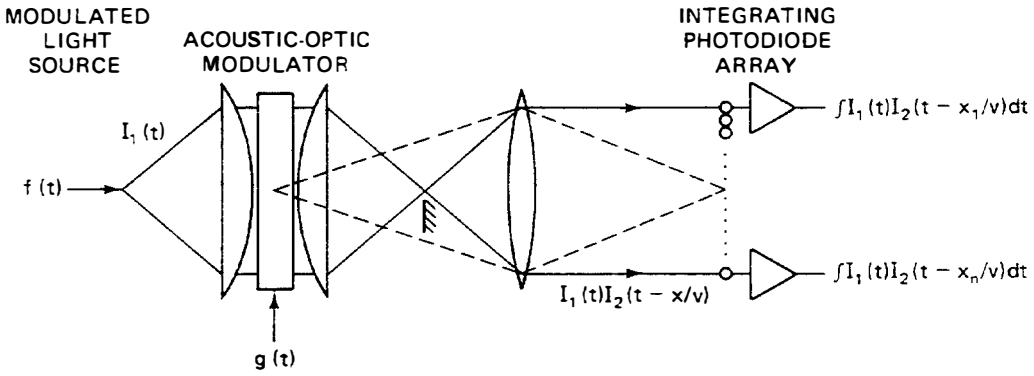


Figure 4.9.15 Time-integrating correlator. (From Kellman [105].)

For double-sideband suppressed carrier modulation of the laser by a signal $f(t)$ cos $(2\pi f_0 t)$, it follows that

$$I_1(t) = A_1[1 + f(t) \cos (2\omega_0 t)] \tag{4.9.88}$$

where ω_0 is the carrier frequency of the signal and A_1 is a constant.

The complex electric field modulation due to the acousto-optic modulator is $E_2(t - x/V)$, where $I_2 = |E_2|^2$. For first-order diffraction,

$$E_2\left(t - \frac{x}{V}\right) = A_2^{1/2}\left[1 + g\left(t - \frac{x}{V}\right)e^{j\omega_0(t - x/V)}\right]e^{j\omega_c(t - x/V)} \tag{4.9.89}$$

where A_2 is a constant. In this case, the input acoustic signal is at a frequency $\omega_c \pm \omega_0$, but as the detector receives nothing but the first-order diffraction term, only the terms of frequencies ω_c and $\omega_c + \omega_0$ arrive at the detector. It follows that

$$I_2\left(t - \frac{x}{V}\right) = A_2\left[1 + g^2\left(t - \frac{x}{V}\right) + 2g\left(t - \frac{x}{V}\right) \cos \omega_0\left(t - \frac{x}{V}\right)\right] \tag{4.9.90}$$

If B is the bandwidth of the signals $f(t)$, $g(t)$, and $f_0 > 3B$, several cross-terms effectively integrate to zero and it can be shown that the total output from the detector is of the form

$$R\left(\frac{x}{V}\right) = TA_1A_2\left[1 + \frac{1}{T}\int_0^T g^2\left(t - \frac{x}{V}\right) + \cdot \int_0^T f(t)g\left(t - \frac{x}{V}\right) dt\right] \tag{4.9.91}$$

where T is the integration time of the system. The first two terms are essentially dc outputs; the last term is the correlation function we require. In practice, these unwanted terms can be eliminated by filtering.

One-dimensional transforms. One example of the use of this time-integrating system is for spectral analysis. In this case, the two input signals are FM chirps of $S_1(t)$ cos $(\omega_0 t + \mu t^2/2)$ and cos $(\omega_0 t + \mu t^2/2)$, respectively. The input signal to the laser is a modulated chirp and the output is the Fourier transform of the modulation. The output is of the form

$$Y_{\text{out}}(x) = \text{Re}\left[e^{-j(\omega_0 x/V + \mu x^2/2V^2)}\int_0^T S_1(t)e^{-j\mu tx/V} dt\right] \tag{4.9.92}$$

Since the output occurs as a function of time $\tau = x/V$, we can multiply the output, as discussed in Sec. 4.7 by a chirp exp $(j\mu\tau^2/2)$ to eliminate the chirp carrier and obtain phase information. The output with an input signal modulation $s_1(t) =$ exp $(j\Omega t)$ is

$$|Y_{\text{out}}(x)| = T\left|\frac{\mu x/V)T/2]}{}\right| \tag{4.9.93}$$

If the time delay in the acousto-optic delay line is $\tau_0 = L/V$, where L is the physical

length of the array and there are N elements in the array, the minimum resolution is

$$\Delta f = \frac{\mu\tau_0}{2N\pi} \tag{4.9.94}$$

The bandwidth of the chirp, however, is

$$B = \frac{\mu T}{2\pi} \tag{4.9.95}$$

and the bandwidth of the chirp in a time τ_0 is

$$B' = \frac{\tau_0}{T}B \tag{4.9.96}$$

Thus

$$\Delta f = \frac{B\tau_0}{NT} = \frac{B'}{N} \tag{4.9.97}$$

The number of resolvable frequencies within the array is $B'/\Delta f = N$. It follows from Eq. (4.9.96) that the best possible resolution we could hope for is

$$\Delta f \approx \frac{1}{T} \tag{4.9.98}$$

Thus for optimum resolution,

$$B\tau_0 = N \tag{4.9.99}$$

which leads to an optimum choice of the chirp rate μ.

If we wish to examine a range of frequencies equal to the chirp bandwidth, we find, just as for the earlier cases examined in Sec. 4.7, that

$$B\tau_0 < \frac{N}{4} \tag{4.9.100}$$

An example of a time-integrating spectrum analyzer output is shown in Fig. 4.9.16. In this example, $N = 1000$, $\tau_0 = 40$ μs, $B = 2.5$ MHz, $T = 2$ ms, and

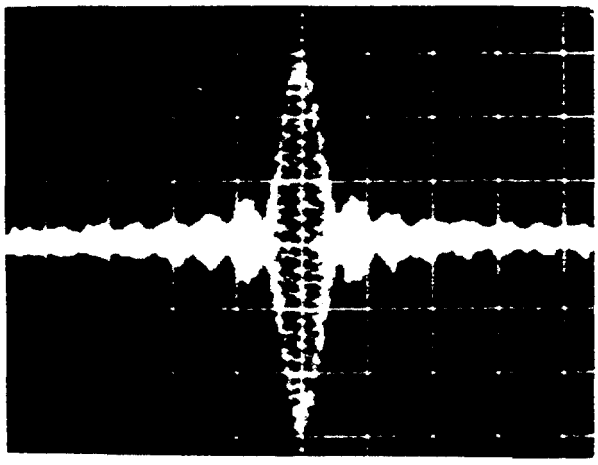


Figure 4.9.16 Time-integrating spectrum analyzer output. (After Kellman [105].)

$B\tau_0 = 100$. The possible resolution is 50 Hz. The dynamic range of the device was 40 dB.

Two-dimensional transforms. Two-dimensional transforms can also be carried out with another version of this system, as illustrated in Fig. 4.9.17. Two acousto-optic deflectors at right angles to each other are employed together with a modulated LED source. The output is now of the form

$$R(x, y) = \int f(t)g\left(t - \frac{x}{V}\right)h\left(t - \frac{y}{V}\right) dt \tag{4.9.101}$$

where the input signals are $f(t)$, $g(t)$, and $h(t)$, respectively.

Suppose that we wish to find the ambiguity function

$$\chi(\tau, f) = \int S_1(t)S_2^*(t - \tau)e^{-2\pi ft}\, dt \tag{4.9.102}$$

Then if we make the x dimension correspond to f and the y dimension to τ, we require the Fourier transform in the y direction of the function $S_1(t)S_2(t - \tau)$. To do this we must have complex input signals of the form

$$\begin{aligned} f(t) &= S_1(t)e^{j\mu t^2/2} \\ g(t) &= S_2^*(t)e^{-j\omega_2 t} \\ h(t) &= e^{j\mu t^2/2}e^{j\omega_1 t} \end{aligned} \tag{4.9.103}$$

The functions $f(t)$ and $g(t)$ without the $\exp(-j\mu t^2/2)$ term provide the required convolutions of $S_1(t)$ and $S_2(t)$. The combination of $h(t)$ with the $\exp(-j\mu t^2/2)$ term then provides the required Fourier transform. By substituting Eq. (4.9.103) in Eq. (4.9.101), it can be shown more directly that the output from the system is

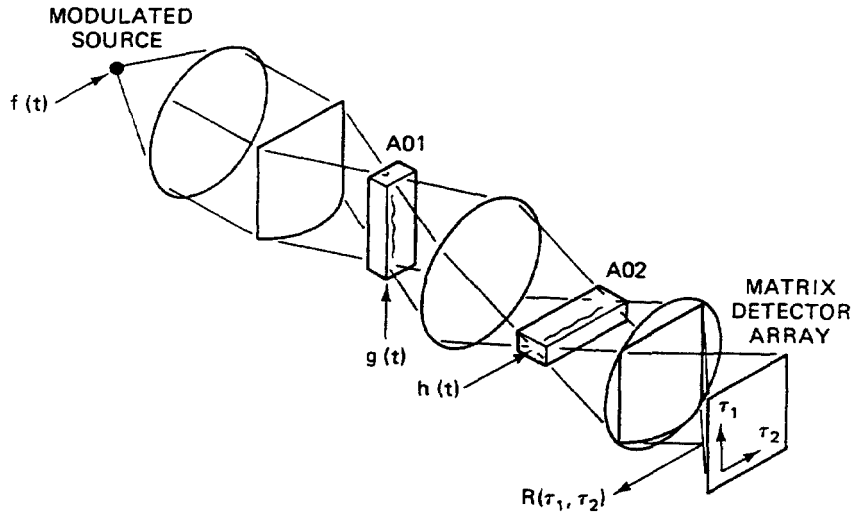


Figure 4.9.17 Two-dimensional time-integrating optical processor. (From Kellman [105].)

of the form

$$R(x, y) = \mathrm{Re}\left[e^{-j(\omega_1 x/V - \omega_2 y/V)} e^{j\mu y^2/V^2} \int S_1(t)\, S_2^*\left(t - \frac{y}{V}\right) e^{j\mu t x/V}\, dt \right] \qquad (4.9.104)$$

where the input signals to the two acousto-optic modulators have carrier frequencies of ω_1 and ω_2, respectively, and the integration time is the frame time of the system. The chirp itself must be of this length.

The modulation of the ambiguity function by the quadratic phase term can be canceled out through post-detection weighting. The variables x and y now correspond to Doppler shift and delay, respectively. The Doppler resolution is commensurate with the integration time and the time resolution is determined by the number of elements in the x dimension.

Kellman [105] has illustrated ambiguity function processing using a Hitachi HLP-20 light-emitting diode, a Fairchild SL62926 charge-coupled-device image sensor, and Isomet acousto-optic devices. The diode had a 30-MHz, 3-dB bandwidth and was biased at an average optical power of approximately 10 mW. The image sensor had 380 × 488 elements and was operated with an integration period of 33.34 ms. The acousto-optic devices had a 30-MHz, 1-dB bandwidth and a 50-μs delay. Chirp waveforms with a very large time–bandwidth product BT = 524,288 were synthesized digitally. The delay range was limited to 36 μs in the y direction and 27 μs in the x direction, to increase the bandwidth. The image plane sampling was therefore 380/36 $\mu s^{-1} \approx 10$ MHz in one delay dimension and 18.1 MHz in the other.

A short pseudorandom code is shown in Fig. 4.9.18. The Doppler range was 1 to 3.8 kHz, determined by the chirp rate. The code-repetition period was 6.2 μs, so that four correlation peaks are evident in the ambiguity function. The Doppler shift in this example was 2 kHz. The time–bandwidth product was 5 MHz × 33.34 ms = 166,700. The sensor dynamic range was approximately 60 dB and the resultant input signal dynamic range was approximately 48 dB.

A similar system has been used for spectral analysis. This made it possible to obtain fine spectral information in one direction and coarse spectral information in the other. Thus the two-dimensional acousto-optic signal processor is an ex-

Figure 4.9.18 Ambiguity function for short code. (After Kellman [105].)

tremely powerful device and can provide a remarkable amount of real-time information. Demonstrations have been made of various other systems that can perform two-dimensional transforms. Examples of these are two-dimensional acoustic or acousto-optic convolvers. At the present time, the system described here seems to be the most practical.

PROBLEM SET 4.9

1. An isotropic material can be used for an acousto-optic deflector. Prove that interaction of a polarized optical beam with a longitudinal acoustic wave gives rise to an optical beam polarized in the same direction as the entering beam. Then prove that interaction of a polarized optical beam with an acoustic shear wave produces an optical beam polarized perpendicular to the polarization of the entering beam.

   Set up the problem by assuming that the perturbation of the dielectric constant due to the acoustic beam is very small, so that only first-order effects due to the acoustic wave need be considered. For a solid isotropic material, in a reduced coordinate system (see Sec. 1.5 and Appendix A), $p_{11} = p_{22} = p_{33}$, $p_{12} = p_{13} = p_{23} = p_{21} = p_{31} = p_{32}$, and $p_{44} = p_{55} = p_{66} = (p_{11} - p_{12})/2$; thus there are only two independent constants. For a liquid, in addition, $p_{11} = p_{12}$.

   We can write

$$\varepsilon_0 E_i = B_{ij} D_j + \Delta B_{ij} D_j$$

   with

$$\Delta B_{ij} = p_{ijkl} S_{kl}$$

   or, in reduced coordinates,

$$\Delta B_I = p_{IJ} S_J$$

   (a) Consider a plane-polarized optical wave, polarized with the D field in the x direction and propagating in the z direction. Suppose that it is perturbed by a plane longitudinal acoustic wave propagating in the x direction (i.e., a wave with S_{xx} or S_1 finite). In the reduced coordinate system, what components of ΔB_I are finite? What components of E_i are finite? What is the polarization of the wave generated by the acousto-optic interaction?

   (b) What components of ΔB_I and E_i are finite for an incident optical beam, with the same intensity and interaction length as in part (a), when it interacts with a longitudinal acoustic wave propagating in the y direction?

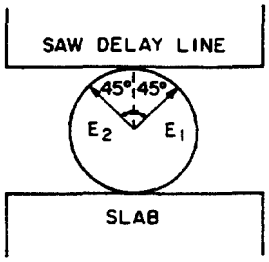
   (c) If the interactions of parts (a) and (b) take place in a liquid instead of a solid, show that the respective output beam amplitudes are indentical.

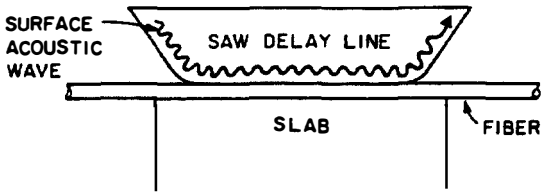
   (d) What components of ΔB_I and E_i are finite, and what is the polarization of the generated optical beam, for interaction of an incident optical beam with a shear acoustic wave, propagating in the y direction, whose particle motion is in the x direction?

2. A Schlieren optical system (Fig. 4.9.4) is used to image a square acoustic beam in water, 1 cm on each side. The acoustic power in the beam is 50 mW and the operating frequency is 5 MHz. A knife edge is employed to pass only the first-order diffraction image.

A 5-mW helium–neon laser (Λ = 6300 Å in air) is used to produce the image, and the acoustic beam amplitude is uniform (although in practice, diffraction causes it to be nonuniform). Optical lenses of 4-cm diameter and focal length 50 cm are used to produce a uniform 4 cm diameter parallel optical beam and to focus it to a diffraction limited spot (see Sec. 3.3). Estimate the intensity of the Schlieren image and the width of the image of the acoustic beam at 3 m from the second lens.

3. A single-sideband acoustic modulator for a bimode fiber-optic delay line is constructed as shown below. The fiber's core is shaped slightly elliptically; this causes the fiber to have two possible modes of propagation, with fields E_1 and E_2 polarized at right angles. The fiber is bonded to an SAW delay line and a flat slab, and placed between them. The fiber is rotated so that the polarizations of the two orthogonal modes are at 45° to the surface of the slab. The displacement normal to the surface of the delay line, due to the surface acoustic wave, applies a stress component at 45° to the planes of polarization of the two optical waves, and couples them together.





For optimum coupling of the two waves we know that

$$\omega_2 = \omega_1 \pm \omega_a \tag{1}$$

and

$$k_2 = k_1 \pm k_a \tag{2}$$

where ω_1 and ω_2 are the frequencies and k_1 and k_2 are the effective propagation constants of the two optical waves, and ω_a and k_a are the corresponding parameters of the acoustic wave. Typically, fibers can be obtained for which the beat wavelength λ_b is of the order of 1 mm, where $\lambda_b = 2\pi/(k_2 - k_1)$.

We may use a coupled mode theory, like that in Sec. 4.6.2, to write

$$\frac{\partial A_2}{\partial z} + jk_2A_2 = j\kappa(z)A_1 \tag{3}$$

$$\frac{\partial A_1}{\partial z} + jk_1A_1 = j\kappa(z)A_2 \tag{4}$$

where $\kappa(z)$ is the coupling coefficient between the optical waves, due to the presence of the acoustic wave, and $A_1(z, t)$ and $A_2(z, t)$ are the amplitudes of the two optical waves.

We may write

$$\kappa(z, t) = \kappa_0 \cos(\omega_a t - k_a z) = \frac{\kappa_0}{2}\left[e^{j(\omega_a t - k_a z)} + e^{-j(\omega_a t - k_a z)}\right] \tag{5}$$

where

$$A_1(z, t) = a_1(z)e^{j(\omega_1 t - k_1 z)} \tag{6}$$

$$A_2(z, t) = a_2(z)e^{j(\omega_2 t - k_2 z)} \tag{7}$$

and in all cases Eq. (1) is satisfied.

(a) By putting $k_2 = k_a + k_1$, neglecting one of the terms in the brackets, and putting $A_2(0) = 0$ with $A_1(0)$ finite, find an expression determining the amplitudes of $A_2(z)$ and $A_1(z)$ along the fiber. Show that when k_a is positive (the acoustic wave is effectively propagating in the $+z$ direction), the output signal in mode 2 is upshifted in frequency.

(b) Show that if the effective propagation direction of the acoustic wave is reversed (k_a negative, $\theta > \pi/2$), the output signal is downshifted in frequency.

Note: A polarizer can easily separate out one signal from the other.

4. Following the assumptions and analysis of Prob. 3, assume that the coupling κ is weak ($\kappa \ll k_1$, $\kappa \ll k_2$) and, therefore, the amplitude of $a_1(z)$ is constant.

(a) Using the first of the coupled-mode equations [Eq. (3)], determine the 3-dB bandwidth for a coupler of length L, as ω_a is varied. Consider coupling from mode 1 to mode 2 and use only the first term in the brackets of Eq. (5). Put $k_a = \omega_a/V_a$ and assume V_a is constant. Then find $\Delta k_a/k_{a0}$ where k_{a0} is the acoustic propagation constant for optimum coupling.

(b) Using the second term in the brackets of Eq. (5), find how much signal is transferred into the lower sideband.

(c) Show that if the coupling coefficient κ is decreased and the length L is increased to obtain a given transfer efficiency to the upper sideband, both the bandwidth and the spurious signal level in the lower sideband are decreased.

5. An optical wave of the form $\phi_0(x) \exp[j(\omega_0 t - k_0 x)]$, traveling forward in the $+x$ direction, interacts with an acoustic wave traveling in the opposite direction that perturbs the effective dielectric constant of the medium by

$$\Delta\varepsilon = \alpha\varepsilon \cos(\omega_a t + k_a x)$$

The interaction of the two waves gives rise to an optical wave traveling backward in the $-x$ direction, of the form $\phi_1(x) \exp[j(\omega_1 t + k_1 x)]$, such that

$$\omega_1 = \omega_0 + \omega_a$$

and

$$k_1 = k_a - k_0$$

For strong interaction ($k_1 \approx k_0$), $k_a \approx 2k_0$ and the acoustic wave must have a frequency of the order of 25 GHz.

Write a wave equation similar to Eq. (4.9.53) for this interaction. Simplify this result in the same way as for Eq. (4.9.56), neglecting all terms but ϕ_0 and ϕ_1, to obtain two differential equations relating ϕ_0 and ϕ_1. Solve these equations for a system of length L with boundary conditions $\phi_0(0)$ = constant, $\phi_0(L) = 0$, and $k_1 = k_0 = k_a/2$, to find the amplitude of the excited wave at $x = 0$.*

Note: This effect has been used to detect thermal phonons. In addition, if two thin optical beams interact with an acoustic beam at the Bragg angle, the frequencies and propagation constants satisfy Eqs. (4.9.46) and (4.9.47); however, this type of analysis is better suited to thin optical beams than the one employed for broad beams in the text.

REFERENCES

1. C. E. Cook and M. Bernfeld, *Radar Signals: An Introduction to Theory and Application*. New York: Academic Press, Inc., 1967.
2. H. Matthews, ed., *Surface Wave Filters: Design, Construction, and Use*. New York: John Wiley & Sons, Inc., 1977.
3. "Special Issue on Surface Acoustic Wave Devices and Applications," *Proc. IEEE*, 64, No. 5 (May 1976), 577–832.
4. D. D. Buss, A. F. Tasch, Jr., and J. B. Barton, "Applications to Signal Processing," in *Charge-Coupled Devices and Systems*, M. J. Howes and D. V. Morgan, eds. Chichester, West Sussex, England: John Wiley & Sons Ltd., 1979, Chapter 3, pp. 103–75.
5. C. H. Séquin and M. F. Tompsett, *Charge Transfer Devices*, Supplement 8 for *Advances in Electronics and Electron Physics*, L. Marton, series ed. New York: Academic Press, Inc., 1975, Chapter VI, pp. 201–35.
6. R. Melen and D. Buss, eds., *Charge-Coupled Devices: Technology and Applications*. New York: IEEE Press, 1977.
7. R. W. Broderson, P. R. Gray, and D. A. Hodges, "MOS Switched-Capacitor Filters," *Proc. IEEE*, 67, No. 1 (Jan. 1979), 61–75.
8. F. L. J. Sangster and K. Teer, "Bucket-Brigade Electronics—New Possibilities for Delay, Time-Axis Conversion, and Scanning," *IEEE J. Solid-State Circuits*, SC-4, No. 3 (June 1969), 131–36.
9. S. A. Newton, K. P. Jackson, J. E. Bowers, C. C. Cutler, and H. J. Shaw, "Fiber-Optic Delay Line Devices for Gigahertz Signal Processing," *IEEE Int. Conf. Acoustics Speech Signal Process.*, Vol. 3 (Apr. 1983), 1204–7.
10. K. P. Jackson, S. A. Newton, and H. J. Shaw, "1-Gbit/s Code Generator and Matched Filter Using an Optical Fiber Tapped Delay Line," *Appl. Phys. Lett.*, 42, No. 7 (Apr. 1983), 556–58.
11. J. E. Bowers, S. A. Newton, and H. J. Shaw, "Fibre-Optic Variable Delay Lines," *Electron. Lett.*, 18, No. 23 (Nov. 11, 1982), 999–1000.
12. S. A. Reible, "Wideband Analog Signal Processing with Superconductive Circuits," *1982 Ultrason. Symp. Proc.* (IEEE), 82CH1823-4, Vol. 1, 190–201.

*Reference: C. F. Quate, C. D. W. Wilkinson, and D. K. Winslow, *Proc. IEEE*, 53, No. 10 (Oct. 1965), 1604–23.

13. G. S. Kino and H. J. Shaw, "Acoustic Surface Waves," *Sci. Am.*, 227, No. 4 (Oct. 1972), 50–68.

14. M. G. Unkauf, "Surface Wave Devices in Spread Spectrum Systems," in *Surface Wave Filters: Design, Construction, and Use*, H. Matthews, ed. New York: John Wiley & Sons, Inc., 1977, Chapter 11, pp. 477–509.

15. C. S. Hartmann, D. T. Bell, Jr., and R. C. Rosenfeld, "Impulse Model Design of Acoustic Surface-Wave Filters," *IEEE Trans. Microwave Theory Tech.*, MTT-21, No. 4 (Apr. 1973), 162–75.

16 D. T. Bell, Jr., and R. C. M. Li, "Surface-Acoustic-Wave Resonators," *Proc. IEEE*, 64, No. 5 (May 1976), 711–21.

17. R. L. Rosenberg and L. A. Coldren, "Scattering Analysis and Design of SAW Resonator Filters," *IEEE Trans. Sonics Ultrason.*, SU-26, No. 3 (May 1979), 205–30.

18. R. C. Williamson, "Reflection Grating Filters," in *Surface Wave Filters: Design, Construction, and Use*, H. Matthews, ed. New York: John Wiley & Sons, Inc., 1977, Chapter 9, pp. 381–442.

19. C. Lardat, "Improved SAW Chirp Spectrum Analyser with 80 dB Dynamic Range," *1978 Ultrason. Symp. Proc.* (IEEE), 78 CH 1344-1SU, 518–21.

20. H. M. Gerard, O. W. Otto, and R. D. Weglein, "Development of a Broadband Reflective Array 10,000:1 Pulse Compression Filter," *1974 Ultrason. Symp. Proc.* (IEEE), 74 CHO 896-1SU, 197–201.

21. J. Crabb, M. F. Lewis, and J. D. Maines, "Surface-Acoustic-Wave Oscillators: Mode Selection and Frequency Modulation," *Electron. Lett.*, 9, No. 10 (May 17, 1973), 195-97.

22. A. J. Budreau, P. H. Carr, and K. R. Laker, "Frequency Synthesizer Using Acoustic Surface-Wave Filters," *Microwave J.*, 17, No. 3 (Mar. 1974), 65–69.

23. R. D. Weglein and O. W. Otto, "Microwave SAW Oscillators" *1977 Ultrason. Symp. Proc.* (IEEE), 77 CH 1264-1SU, 913–22.

24. Reticon Corporation, "Product Summary: Discrete Time Analog Signal Processing Devices." Sunnyvale, Calif.: Reticon Corp., 1977.

25. J. T. Walker and J. D. Meindl, "A Digitally Controlled CCD Dynamically Focused Phased Array," *1975 Ultrason. Symp. Proc.* (IEEE), 75 CHO 994-4SU, 80–83.

26. M. A. Jack, P. M. Grant, and J. H. Collins, "The Theory, Design, and Applications of Surface Acoustic Wave Fourier-Transform Processors," *Proc. IEEE*, 68, No. 4 (Apr. 1980), 450–68.

27. Reticon Corporation, "Preliminary Data Sheet: R5501-32 Parallel In/Serial Out (PISO)." Sunnyvale, Calif.: Reticon Corp., 1978.

28. Reticon Corporation, "S-Series Solid State Line Scanners, 512 and 1024 Elements." Sunnyvale, Calif.: Reticon Corp., 1978.

29. I. Deyhimy, R. C. Eden, R. J. Anderson, and J. S. Harris, Jr., "A 500-MHz GaAs Charge-Coupled Device," *Appl. Phys. Lett.*, 34, No. 2, (Jan. 15, 1980), 151–53.

30. R. D. Baertsch, W. E. Engeler, H. S. Goldberg, C. M. Puckette, and J. J. Tiemann, "Two Classes of Charge-Transfer Devices for Signal Processing," *Int. Conf. Technol. Applications Charge Coupled Devices* (Proc.), (Sept. 1974), 229–36.

31. R. D. Baertsch, W. E. Engeler, H. S. Goldberg, C. M. Puckette IV, and J. J. Tiemann, "The Design and Operation of Practical Charge-Transfer Transversal Filters," *IEEE J. Solid-State Circuits*, SC-11, No. 1 (Feb. 1976), 65–74.

32. R. Bracewell, *The Fourier Transform and Its Applications*, 2nd ed. New York: McGraw-Hill Book Company, 1978.

33. A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing.* Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1975.

34. P. M. Woodward, *Probability and Information Theory with Applications to Radar*, 2nd ed. Oxford: Pergamon Press Ltd., 1953.

35. D. Gabor, "Theory of Communication," *J. IEE* (Brit.), 93, Pt. III, No. 26 (Nov. 1946), 429–57.

36. H. Stark and F. B. Tuteur, *Modern Electrical Communications: Theory and Systems.* Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1979.

37. M. Bernfeld, C. E. Cook, J. Paolillo, and C. A. Palmieri, "Matched Filtering, Pulse Compression and Waveform Design," Parts I-IV, *Microwave J.*, 7, No. 10 (Oct. 1964), 57–64, No. 11 (Nov. 1964), 81–90, No. 12 (Dec. 1964), 70–76, and 8, No.1 (Jan. 1965), 73–81.

38. J. R. Klauder, "The Design of Radar Signals Having Both High Range Resolution and High Velocity Resolution," *Bell Sys. Tech. J.*, XXXIX, No. 4 (July 1960), 809-20.

39. F. J. Harris, "On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform," *Proc. IEEE*, 66, No. 1 (Jan. 1978), 51–83.

40. J. F. Kaiser, "Digital Filters," in *System Analysis by Digital Computer*, F. F. Kuo and J. F. Kaiser, eds. New York: John Wiley & Sons, Inc., 1966, Chapter 7, pp. 218-85.

41. R. M. Hays, W. R. Shreve, D. T. Bell, Jr., L. T. Claiborne, and C. S. Hartmann, "Surface Wave Transform Adaptable Processor System," *1975 Ultrason. Symp. Proc.* (IEEE), 75 CHO 994-4SU, 363–70.

42. J. D. Maines, G. L. Moule, C. O. Newton, and E. G. S. Paige, "A Novel SAW Variable-Frequency Filter," *1975 Ultrason. Symp. Proc.* (IEEE), 75 CHO 994-4SU, 355–58.

43. C. Atzeni, G. Manes, and L. Masotti, "Programmable Signal Processing by Analog Chirp Transformation Using SAW Devices," *1975 Ultrason. Symp. Proc.* (IEEE), 75 CHO 994-4SU, 371–76.

44. J. D. Maines and E. G. S. Paige, "Surface-Acoustic-Wave Devices for Signal Processing Applications," *Proc. IEEE*, 64, No. 5 (May 1976), 639–52.

45. H. M. Gerard, P. S. Yao, and O. W. Otto, "Performance of a Programmable Radar Pulse Compression Filter Based on a Chirp Transformation with RAC Filters," *1977 Ultrason. Symp. Proc.* (IEEE), 77 CH 1264-1SU, 947–51.

46. G. R. Nudd and O. W. Otto, "Chirp Signal Processing Using Acoustic Surface Wave Filters," *1975 Ultrason. Symp. Proc.* (IEEE), 75 CHO 994-4SU, 350–54.

47. M. A. Jack and J. H. Collins, "Fast Fourier Transform Processor Based on the SAW Chirp Transform Algorithm," *1978 Ultrason. Symp. Proc.* (IEEE), 78CH 1344-1SU, 533–37.

48. M. A. Jack and E. G. S. Paige, "Fourier Transformation Processors Based on Surface Acoustic Wave Chirp Filters," *Wave Electron.*, 3, No. 3 (Nov. 1978), 229–47.

49. Reticon Corporation, "Preliminary Data Sheet: Quad Chirped Transversal Filter R5601." Sunnyvale, Calif.: Reticon Corp., 3/28/78.

50. F. G. Marshall, C. O. Newton, and E. G. S. Paige, "Theory and Design of the Surface

Acoustic Wave Multistrip Coupler," *IEEE Trans. Microwave Theory Tech.*, MTT-21, No. 4 (Apr. 1973), 206–15.

51. F. G. Marshall, C. O. Newton, and E. G. S. Paige, "Surface Acoustic Wave Multistrip Components and Their Applications," *IEEE Trans. Microwave Theory Tech.*, MTT-21, No. 4 (Apr. 1973), 216–25.

2. R. H. Tancrell and H. Engan, "Design Considerations for SAW Filters," *1973 Ultrason. Symp. Proc.* (IEEE), 73 CHO 807-8SU, 419–22.

53. C. Atzeni and L. Masotti, "Weighted Interdigital Transducers for Smoothing of Ripples in Acoustic-Surface-Wave Bandpass Filters," *Electron. Lett.*, 8, No. 19 (Sept. 21, 1972), 485–86.

54. T. Kodama, "Optimization Techniques for SAW Filter Design," *1979 Ultrason. Symp. Proc.* (IEEE), 79CH1482-9, 522–26.

55. C. S. Hartmann, W. S. Jones, and H. Vollers, "Wideband Unidirectional Interdigital Surface Wave Transducers," *IEEE Trans. Sonics Ultrason.*, SU-19, No. 3 (July 1972), 378–81.

56. A. J. DeVries, "A Design Method for Surface-Wave Filters Using Simple Structures as Building Blocks," *1973 Ultrason. Symp. Proc.* (IEEE), 73 CHO 807-8SU, 441–44.

57. A. J. DeVries and R. Adler, "Case History of a Surface-Wave TV IF Filter for Color Television Receivers," *Proc. IEEE*, 64, No. 5 (May 1976), 671–76.

58. A. J. DeVries, T. Sreenivasan, S. Subramanian, and T. J. Wojcik, "Detailed Description of a Commercial Surface-Wave TV IF Filter," *1974 Ultrason. Symp. Proc.* (IEEE), 74 CHO 896-1SU, 147–52.

59. K. Kishimoto, M. Ishigaki, K. Hazama, and S. Matsu-ura, "SAW Comb Filter for TV Frequency Synthesizing Tuning System," *1980 Ultrason. Symp. Proc.* (IEEE), 80CH1602-2, Vol. 1, 377–81.

60. S. Matsu-ura, K. Hazama, and T. Murata, "TV Tuning Systems with SAW Comb Filter," *IEEE Trans. Microwave Theory Tech.*, MTT-29, No. 5 (May 1981), 434–39.

61. R. A. Bergh, G. Kotler, and H. J. Shaw, "Single-Mode Fibre Optic Directional Coupler," *Electron. Lett.*, 16, No. 7 (Mar. 27, 1980), 260–61.

62. M. J. F. Digonnet and H. J. Shaw, "Analysis of a Tunable Single Mode Optical Fiber Coupler," *IEEE J. Quantum Electron.*, QE-18, No. 4 (Apr. 1982), 746–54.

63. G. S. Kino and H. Matthews, "Signal Processing in Acoustic Surface-Wave Devices," *IEEE Spectrum*, 8, No. 8 (Aug. 1971), 22–35.

64. G. S. Kino, "Acoustoelectric Interactions in Acoustic-Surface-Wave Devices," *Proc. IEEE*, 64, No. 5 (May 1976), 724–48.

65. G. S. Kino, S. Ludvik, J. H. Shaw, W. R. Shreve, J. M. White, and D. K. Winslow, "Signal Processing by Parametric Interactions in Delay-Line Devices," *IEEE Trans. Microwave Theory Tech.*, MTT-21, No. 4 (Apr. 1973), 244–55.

66. D. M. Bloom and G. C. Bjorklund, "Conjugate Wave-Front Generation and Image Reconstruction by Four-Wave Mixing," *Appl. Phys. Lett.*, 31, No. 9 (Nov. 1, 1977), 592–96.

67. D. P. Morgan and J. M. Hannah, "Correlation of Long Spread-Spectrum Wave-forms Using a SAW Convolver/Recirculation Loop Sub-system," *1976 Ultrason. Symp. Proc.* (IEEE), 76 CH 1120-5SU, 436–40.

68. D. P. Morgan, J. H. Hannah, and J. H. Collins, "Spread-Spectrum Synchronizer Using a SAW Convolver and Recirculation Loop," *Proc. IEEE*, 64, No. 5 (May 1976), 751-53.

69. P. K. Blair, "Receivers for the NAVSTAR Global Positioning System," *IEE Proc.* (Brit.), 127, Pt. F, No. 2 (Apr. 1980), 163–67.

70. P. Defranould and C. Maerfeld, "A SAW Planar Piezoelectric Convolver," *Proc. IEEE*, 64, No. 5 (May 1976), 748–51.

71. I. Yao, "High Performance Elastic Convolver with Parabolic Horns," *1980 Ultrason. Symp. Proc.* (IEEE), 80CH1602-2, Vol. 1, 37–42.

72. J. B. Green and G. S. Kino, "SAW Convolvers Using Focused Interdigital Transducers," *IEEE Trans. Sonics Ultrason.*, SU-30, No. 1 (Jan. 1983), 43–50.

73. G. S. Kino and H. Gautier, "Convolution and Parametric Interaction with Semiconductors," *J. Appl. Phys.*, 44, No. 12 (Dec. 1973), 5219–21.

74. B. T. Khuri-Yakub and G. S. Kino, "A Detailed Theory of the Monolithic Zinc Oxide on Silicon Convolver," *IEEE Trans. Sonics Ultrason.*, SU-24, No. 1 (Jan. 1977), 34-43.

75. J. H. Cafarella, W. M. Brown, Jr., E. Stern, and J. A. Alusow, "Acousto-electric Convolvers for Programmable Matched Filtering in Spread-Spectrum Systems," *Proc. IEEE*, 64, No. 5 (May 1976), 756–59.

76. J. H. Goll and D. C. Malocha, "An Application of SAW Convolvers to High Bandwidth Spread Spectrum Communications," *IEEE Trans. Microwave Theory Tech.*, MTT-29, No. 5 (May 1981), 473–83.

77. S. A. Reible, "Acoustoelectric Convolver Technology for Spread-Spectrum Communications," *IEEE Trans. Sonics Ultrason.*, Su-28, No. 3 (May 1981), 185–95.

78. G. K. Montress and T. M. Reeder, "A High Performance SAW/Hybrid Coponent Fourier Transform Convolver," *1978 Ultrason. Symp. Proc.* (IEEE), 78CH 1344-1SU, 538–42.

79. H. C. Tuan, J. E. Bowers, and G. S. Kino, "Theoretical and Experimental Results for Monolithic SAW Memory Correlators," *IEEE Trans. Sonics Ultrason.*, SU-27, No. 6 (Nov. 1980), 360–69.

80. K. A. Ingebrigtsen, "The Schottky Diode Acoustoelectric Memory and Correlator—A Novel Programmable Signal Processor," *Proc. IEEE*, 64, No. 5 (May 1976), 764-69.

81. H. C. Tuan and G. S. Kino, "Large-Time-Bandwidth-Product Correlation and Holographic Storage with an S.A.W. Storage Correlator," *Electron. Lett.*, 13, No. 24 (Nov. 24, 1977), 709–10.

82. S. A. Reible and I. Yao, "An Acoustoelectric Burst Waveform Processor," *1980 Ultrasonics Symp. Proc.* (IEEE), 80CH1602-2, Vol. 1, 133–38.

83. H. C. Tuan, P. M. Grant, and G. S. Kino, "Theory and Application of Zinc-Oxide-on-Silicon Monolithic Storage Correlators," *1978 Ultrason. Symp. Proc.* (IEEE), 78 CH 1344-1SU, 38–43.

84. P. G. Borden and G. S. Kino, "Correlation with the Storage Convolver," *Appl. Phys. Lett.*, 29, No. 9 (Nov. 1, 1976), 527–29.

85. M. Cronin-Golomb, B. Fischer, J. O. White, and A. Yariv, "Passive (Self-Pumped) Phase Conjugate Mirror: Theoretical and Experimental Investigation," *Appl. Phys. Lett.*, 41, No. 8 (Oct. 15, 1982), 689–91.

86. G. L. Kerber, R. M. White, and J. R. Wright, "Surface-Wave Inverse Filter for Nondestructive Testing," *1976 Ultrason. Symp. Proc.* (IEEE), 76 CH 1120-5SU, 577–81.

87. Y. Murakami, B. T. Khuri-Yakub, G. S. Kino, J. M. Richardson, and A. G. Evans,

"An Application of Wiener Filtering to Nondestructive Evaluation," *Appl. Phys. Lett.*, 33, No. 8 (Oct. 15, 1978), 685–87.

88. D. Corl, "A C.T.D. Adaptive Inverse Filter," *Electron. Lett.*, 14, No. 3 (Feb. 2, 1978), 60–62.

89. R. W. Lucky, "Automatic Equalization for Digital Communication," *Bell Syst. Tech. J.*, XLIV, No. 4 (Apr. 1965), 547–88.

90. S. C. Tanaka, H. Tseng, L. T. Lin, and P. Chen, "An Integrated Real-Time Programmable Transversal Filter," *IEEE J. Solid-State Circuits*, SC-15, No. 6 (Dec. 1980), 978–83.

91. B. Widrow, "Adaptive Filters," in *Aspects of Network and System Theory*, R. E. Kalman and N. DeClaris, eds. New York: Holt, Rinehart and Winston, 1970, pp. 563–87.

92. J. E. Bowers, G. S. Kino, D. Behar, and H. Olaisen, "Adaptive Deconvolution Using a SAW Storage Correlator," *IEEE Trans. Microwave Theory Tech.*, MTT-29, No. 5 (May 1981), 491–98.

93. D. Behar, G. S. Kino, J. E. Bowers, and H. Olaisen, "Storage Correlator as an Adaptive Inverse Filter," *Electron. Lett.*, 16, No. 4 (Feb. 14, 1980), 130–31.

94. P. M. Grant and G. S. Kino, "Adaptive Filter Based on S.A.W. Monolithic Storage Correlator," *Electron. Lett.*, 14, No. 17 (Aug. 17, 1978), 562–64.

95. J. E. Bowers and G. S. Kino, "Adaptive Noise Cancellation with a SAW Storage Correlator," *Electron. Lett.*, 17, No. 13 (June 25, 1981), 460–61.

96. D. F. Barbe, W. D. Baker, and K. L. Davis, "Signal Processing with Charge-Coupled Devices," *IEEE J. Solid-State Circuits*, SC-13, No. 1 (Feb. 1978), 34–51.

97. B. K. Ahuja, M. A. Copeland, and C. H. Chan, "A Sampled Analog MOS LSI Adaptive Filter," *IEEE J. Solid-State Circuits*, SC-14, No. 1 (Feb. 1979), 148–54.

98. W. K. Masenten, "Adaptive Processing for Spread Spectrum Communication Systems," *Hughes Aircraft Company Report*, TP 77-14-22 (Sept. 27, 1977).

99. W. R. Klein and B. D. Cook, "Unified Approach to Ultrasonic Light Diffraction," *IEEE Trans. Sonics Ultrason.*, SU-14, No. 3 (July 1967), 123–34.

100. K. Osterhammel, "Optische Untersuchung des Schallfeldes Kolbenformig Schwinger Quarze," *Akust. Zh. Bd.*, 6 (1941), 73–86.

101. B. D. Cook, E. Cavanagh, and H. D. Dardy, "A Numerical Procedure for Calculating the Integrated Acoustooptic Effect," *IEEE Trans. Sonics Ultrason.*, SU-27, No. 4 (July 1980), 202–7.

102. C. S. Tsai, "Wideband Guided-Wave Acoustooptic Bragg-Devices and Applications," *1975 Ultrason. Symp. Proc.* (IEEE), 75 CHO 994-4SU, 120–25.

103. J. N. Lee, N. J. Berg, and B. J. Udelson, "Large Time-Bandwidth Acousto-Optic Signal Processors," *1977 Ultrason. Symp. Proc.* (IEEE), 77 CHU 1264-1SU, 451–55.

104. N. J. Berg and B. J. Udelson, "Large Time-Bandwidth Acousto-Optic Convolver," *1976 Ultrason. Symp. Proc.* (IEEE), 76 CH 1120-5SU, 183–88.

105. P. Kellman, "Time Integrating Optical Signal Processing," in *Acousto-Optic Bulk Wave Devices*, J. B. Houston, Jr., ed., *Proc. Soc. Photo-Opt. Instrum. Eng.*, 214 (1979), 63–73.