Chapter 9

Chirp Filters and Their Applications

A chirp filter is a linear device designed to give a delay varying substantially with frequency. An equivalent definition is that the impulse response is a frequency-modulated pulse, often known as a "chirp" pulse. We have seen in Chapter 1 that such devices are used in pulse-compression radar systems, where the transmitter generates a chirp pulse and a matching chirp filter is used in the receiver; the chirp filter compresses the pulse to a much shorter duration and also increases the signal-to-noise ratio, thus increasing the sensitivity of the receiver. The pulse compression principle was first put forward in the 1940's, but depended for its practical implementation on the development of a suitable technology for the chirp filters. A variety of technologies have been developed, including surface-wave devices which first emerged in the late 1960's. Today, pulse compression is commonly used in radar systems, and surface-wave technology is usually chosen for the chirp filters because of the exceptional performance capabilities and high precision that it offers. As already mentioned in Chapter 1, there are two main types of surface-wave chirp filter — the interdigital device, using an interdigital transducer with graded periodicity, and the Reflective Array Compressor (RAC), using arrays of grooves.

In addition to the use of chirp filters in pulse-compression systems, a number of other applications will be considered in this Chapter. Most of these make use of linear-chirp waveforms, in which the frequency varies linearly with time. The concept of a time-dependent "frequency" will be justified in Section 9.2; strictly, we should refer to the "instantaneous frequency". A basic property of a linear-chirp waveform is that if it is amplitude modulated its spectral amplitude, in the frequency domain, has the same shape as the time-domain envelope. This property is exploited in systems that use linear-chirp filters to perform Fourier transformation, giving an output waveform proportional to the spectrum of the input waveform. In particular the "compressive receiver", which uses two chirp filters, gives the magnitude of the spectrum of the input waveform, and gives this information much faster than a conventional spectrum analyser. There are also several other applications for linear-chirp filters, including for example a technique giving a variable delay.

This chapter commences with a summary of the principles of pulse-compression radar in Section 9.1. It will be seen in particular that there are several types of chirp

waveform suitable for use as the transmitted pulse, and as the impulse response of the chirp filter in the receiver, and the properties and design of these waveforms are considered in Section 9.2. This material is not simply incidental to the subject; it is usually the case that the starting point for development of a surface-wave device is the specification of some basic performance requirements for the radar system, and hence the first task is the design of chirp waveforms satisfying these requirements. In addition, Section 9.2 introduces the very useful "stationary-phase approximation", which is also used in later sections in connection with device analysis. Interdigital devices are considered in the following two sections: the analysis and design of chirp transducers are described in Section 9.3, and Section 9.4 describes the design and performance of interdigital devices. Section 9.5 is concerned with a number of second-order effects, particularly the phase errors associated with velocity errors and temperature changes, and includes the consequent degradation of the radar system performance which must be considered when establishing tolerances. This topic is conveniently placed here since most of it is relevant not only to interdigital devices, considered in the previous sections, but also to RAC's which follow in Section 9.6. Finally, Section 9.7 describes a variety of system applications other than pulse-compression radar, including the compressive receiver and Fourier transformation.

9.1 PRINCIPLES OF PULSE-COMPRESSION RADAR

In a radar system, the transmitter emits a short electromagnetic pulse which is reflected by a target and then arrives at the receiver after a delay proportional to the target range. As in all electronic receivers the input signal will be accompanied by random noise, originating in the receiver itself or elsewhere, and the ability of the receiver to detect the signal is determined by the relative powers of the signal and noise. To optimise the performance of the system the linear portion of the receiver, prior to the detector, should be designed such that the signal-to-noise ratio at the detector input is maximised. A conventional radar system is illustrated in Figure 9.1(a), omitting amplifiers and frequency-conversion stages that are necessary

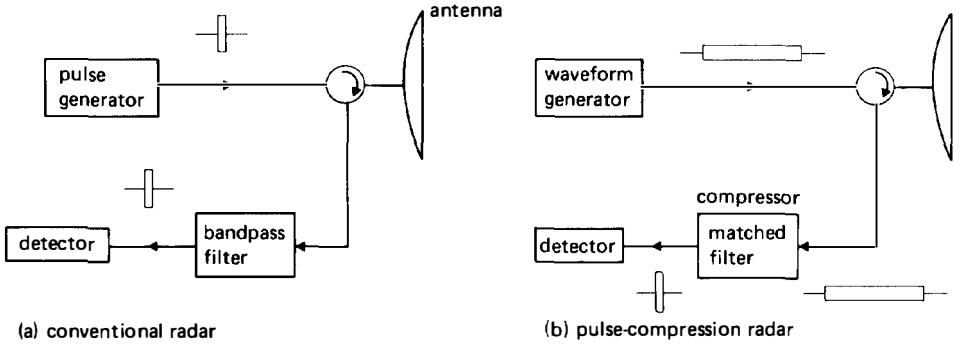


FIGURE 9.1. Comparison between conventional and pulse-compression radar systems. Output waveforms are shown as they appear for a point target.

in practice. The transmitter here generates a simple pulse of carrier with a rectangular envelope, and the system performance is close to optimal if the linear part of the receiver has a simple bandpass response, passing most of the signal spectral components and rejecting noise outside this band.

Suppose now that we wish to increase the sensitivity of such a radar system by increasing the signal-to-noise ratio at the detector input; this will enable the system to detect targets that are more distant, or have lower reflectivity. Clearly the power of the transmitter could be increased, but in practice this would generally be very expensive since microwave frequencies are involved. Alternatively, the length of the transmitted pulse could be increased, thus reducing its bandwidth; the receiver could then have a smaller bandwidth, thus reducing the output noise power. However, this would give a longer output pulse and would therefore degrade the spatial resolution of the system, that is, its ability to distinguish two targets close together.

The pulse-compression concept [277–281] resolves this dilemma, enabling the signal-to-noise ratio to be increased without increasing the transmitted power or degrading the system resolution. The principle is illustrated in Figure 9.1(b). Compared with a conventional system, a pulse-compression system transmits a pulse that is longer and also has a more complex form; the use of a complex waveform enables the pulse to be longer without reducing its bandwidth. In the linear part of the receiver the signal is applied to a special type of filter known as a *matched filter* [277–279, 282, 283], and this is usually a surface-wave device. The term "matched filter" refers to a linear filter designed such that, for a specified input waveform accompanied by noise, the signal-to-noise ratio at the filter output is maximised. The requirement is analysed in Appendix A, Section A.3, which shows that the required impulse response of the matched filter is essentially the time-reverse of the input waveform. It will be seen later that there are two key features associated with the use of a matched filter. Firstly, the output pulse produced has a width determined by the *bandwidth* of the input pulse, and not its duration; thus the latter can be increased without degrading the radar resolution. Secondly, the output signal-to-noise ratio is determined by the *energy* of the input pulse, and can therefore be enhanced by increasing the length of the transmitted pulse without changing its power level. Since the transmitter generates pulses repetitively, the length increase will in fact increase the average power, but this is generally acceptable because most radar transmitters are not limited in average power. On the other hand the *peak* power is often limited by breakdown considerations. Thus the pulse-compression principle enables the receiver signal-to-noise ratio to be increased without making radical changes to the transmitter, and without degrading the radar resolution.

For a pulse-compression system, the transmitted pulse is required to have a length much greater than the reciprocal of the bandwidth, that is, the time-bandwidth product, TB, must be large. This implies the use of some form of modulation. Most commonly, a *chirp* waveform is used, in which the frequency (strictly, the "instantaneous" frequency, Section 9.2) is swept monotonically with time. This is illustrated in Figure 9.2. The waveform $s(t)$ at the left represents the receiver input waveform due to a point target, and is therefore the same as the transmitted waveform except for a delay and a reduction of amplitude. In the receiver, the matched filter is

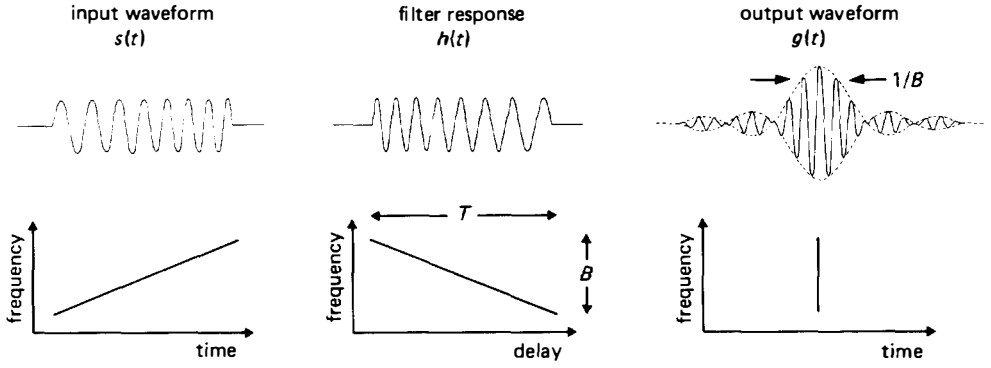


FIGURE 9.2. Pulse compression using linear-chirp waveforms.

matched to the transmitted waveform so that its impulse response is the time-reverse, a chirp waveform $h(t)$ with frequency sweeping in the opposite direction. A filter whose impulse response is a chirp waveform is known as a *chirp filter*. Here the chirp filter acts as a particular type of matched filter, though a number of other applications for chirp filters will be described in this chapter. The output waveform, shown in Figure 9.2, typically has a narrow peak called the *correlation peak*, with small *time-sidelobes* on either side. It will be shown later that the width of the correlation peak is approximately $1/B$, where B is the bandwidth (in Hz) of the input waveform. This is approximately equal to the range of frequencies swept by the waveform. Generally the width of the output peak is much less than the length T of the input waveform, and hence the process is often called *pulse compression*, and the matched filter is also called a *compressor*. The ratio of the input and output pulse widths is called the *compression ratio*. This is approximately equal to TB, the time-bandwidth product of the input waveform, and of the matched filter.

The lower part of Figure 9.2 shows a simplified interpretation of the pulse compression process. The matched filter, whose impulse response has a frequency decreasing with time, can be regarded as a dispersive delay line. Its group delay decreases with frequency, and has a slope opposite to that of the frequency–time characteristic of the input waveform. The frequency components of the input waveform are thus delayed such that they all arrive at the output at the same time, so that a large narrow correlation peak is produced. Although this view is over-simplified, some justification for it is provided by the stationary phase principle, considered in Section 9.2 below.

If there are two point targets, two similar chirp waveforms will arrive at the receiver input, with delays corresponding to the target ranges. Since the matched filter is linear, the output will be simply the sum of the outputs due to each target taken individually. This is true even if the two input chirps overlap in time; two output correlation peaks are produced, and these will be resolved if the delay difference is greater than the peak width $1/B$. The range resolution, expressed in terms of delay, is therefore $1/B$, which

is quite different from the length T of the input waveform. In a conventional radar, using a simple rectangular pulse, the resolution is approximately equal to the pulse width, and this in turn is approximately the reciprocal of the bandwidth. Thus, in both cases the resolution is approximately $1/B$. A typical bandwidth is 20 MHz.

Quantitatively, the advantage to be gained from pulse compression can be assessed by considering the signal-to-noise ratio (SNR) at the output of the matched filter. From Section A.3, the output signal-to-noise power ratio is $2E_s/N_i$, where E_s is the signal energy, and N_i the noise spectral density, at the filter input. In a radar system the input signal has a flat envelope, so its energy can be written $E_s = P^sT$, where P^s is the average power during the pulse. The output SNR is therefore

$$\mathrm{SNR}_0 = 2P^sT/N_i. \tag{9.1}$$

This is valid for any waveform with a flat envelope, applied to an appropriate matched filter. Consider now the output SNR for a pulse-compression radar and for a conventional radar, assuming both systems use a matched filter. If the pulse length is T for the pulse compression system, and T_0 for the conventional system, the output SNR's will be in the ratio T/T_0, assuming P^s and N_i to be the same for both. If the two radars have the same resolution, the bandwidth B of the pulse-compression system must be approximately $1/T_0$, and hence the ratio of the SNR's is

$$\frac{\mathrm{SNR}_0(\text{chirp})}{\mathrm{SNR}_0(\text{conventional})} \approx TB. \tag{9.2}$$

This ratio is called the *processing gain*, and is seen to be approximately equal to the compression ratio. In practice, conventional radars do not generally use matched filters, so the output SNR is somewhat less than the value given by equation (9.1); however, the difference is quite small because the time-bandwidth product is close to unity [278, p. 413]. The benefit obtained can be substantial: surface-wave devices give TB-products typically in the range 50 to 1000, so the use of pulse compression gives an SNR increase equivalent to increasing the peak power of a conventional radar by a factor of 50 to 1000.

In a practical system the linear part of the receiver includes components for amplification and frequency conversion in addition to the surface-wave matched filter, which is usually at an I.F. stage. Ideally, the entire pre-detection system should be matched to the input waveform. Usually, the surface-wave device is designed to match an I.F. version of the waveform, and the other components are specified to be adequately free of distortion. The tolerances required are considered in Section 9.5 below.

Waveform Considerations. Although only chirp waveforms have been considered so far, several other types of waveform are applicable to pulse compression radar. There are several basic requirements. Nearly always, the envelope of the transmitted waveform must be flat, because practical transmitters are most efficient when such waveforms are used. The TB-product should be large, so that significant processing gain can be obtained [equation (9.2)]. These requirements imply that some

form of frequency or phase modulation is required. In addition, the waveform should be such that the matched filter output gives a narrow peak with relatively small time-sidelobes. The sidelobes appear as false targets, and can cause a weakly-reflecting target to be obscured by a nearby strongly-reflecting target. Thus it is usually important to ensure that the sidelobes are adequately suppressed.

The commonest waveform types satisfying these criteria are chirp waveforms and phase-shift-keyed (PSK) waveforms. The latter are often used in communication systems, and are considered in Chapter 10. In radar systems, chirp waveforms are usually preferred because there are some very effective techniques available for minimising the time-sidelobe levels. In addition, chirp waveforms are in general less affected by doppler shifts due to motion of the target relative to the radar, and this is often significant when high-velocity targets must be catered for. A variety of waveforms with large TB-products, including chirp and PSK waveforms, is discussed by Cook and Bernfeld [279].

In the simplest form of chirp radar the transmitted waveform is a *linear chirp*, that is, its instantaneous frequency (Section 9.2) varies linearly with time. When this waveform is applied to its matched filter, the output waveform has an envelope of the form $(\sin x)/x$, as illustrated in Figure 9.2. The time-sidelobes of this waveform are too large for most radar applications. Often, the sidelobes are reduced by incorporating amplitude weighting in the filter, giving sidelobes typically 35 dB below the main peak. This incurs an SNR penalty of about 1 dB, due to the fact that the receiver is no longer matched to the input waveform. Alternatively, a *non-linear chirp* waveform, in which the frequency is a prescribed non-linear function of time, enables low sidelobe levels to be obtained without the SNR penalty. However, non-linear chirps are not always applicable, because they are more sensitive to doppler shifts than linear chirps.

Generation of Chirp Waveforms. A variety of techniques have been developed for generation of chirp waveforms in the radar transmitter [279, 280]. Techniques involving the use of active circuits are generally called *active generation*. Conceptually, the simplest of these is a voltage-controlled oscillator, with the control voltage having the form of a ramp. This method is relatively straightforward, but since the frequency is required to change rapidly the required accuracy can be difficult to achieve; usually, it is necessary to distort the applied ramp to compensate for distortion in the response of the oscillator.

An alternative approach is to impulse a filter whose impulse response is essentially the waveform required. This is known as *passive generation*. Clearly, the device required is essentially the same as the matched filter in the receiver, except that its impulse response has the frequency sweeping in the opposite direction. A device used in this way is known as an *expander*. This method has the attraction that a highly accurate waveform can be obtained. Moreover some forms of error can be significantly reduced by this method. For example, if a surface-wave expander and compressor are used, errors due to temperature changes are much reduced if the devices are mounted close together, so that their temperatures are similar. Many systems therefore use a "matched pair" of devices. The main disadvantage of passive

generation is that the expansion process reduces the power level of the signal, particularly if the TB-product is large. For this reason, the SNR of the expanded signal may be inadequate, and then passive generation cannot be used.

Technologies for Chirp Filtering. In this chapter we are mainly concerned with surface-wave chirp filters, which are widely used in modern radar systems. Interdigital surface-wave devices first emerged in the late 1960's [284, 285], and rapidly demonstrated the ability to give highly accurate and reproducible performance. Shortly afterwards, reflective array compressors (RAC's) using arrays of grooves were developed [286], and these enable very large time-bandwidth products, up to several thousand, to be used. The RAC is very similar in principle to an earlier device called the IMCON which also uses groove arrays, but in this case the wave motion involved is a bulk acoustic wave rather than a surface wave. In all three of these devices, the wave motion is non-dispersive, though the device response is dispersive. Interdigital devices are described in Sections 9.3 to 9.5, while RAC's and IMCON's are considered in Section 9.6.

A variety of other techniques have been developed [279, 280, 287], though most of these have been superseded by surface-wave devices and IMCON's. The required dispersion can be obtained by using an acoustic wave with dispersive propagation. For example, in a layered half-space Rayleigh waves and Love waves are both dispersive, as described in Section 2.2.4. Both have been considered for pulse compression, particularly Love waves [288]. Alternatively, a parallel-sided plate supports dispersive waves, as shown in Section 2.2.5. Chirp devices using these modes take the form of thin metal strips with a bulk-wave transducer at each end, and are known as "strip delay lines". Other acoustic devices use non-dispersive bulk waves, arranging for the path length to vary with frequency. This principle is used in perpendicular diffraction delay lines and wedge delay lines. Several optical techniques are also applicable, including acousto-optic devices in which a light beam is diffracted by an acoustic wave.

Digital processing [283] has considerable attractions because of the flexibility obtainable. In comparison with surface-wave techniques, digital methods are generally limited to lower bandwidths, and result in devices that are more bulky and consume more power. However current developments in integrated circuit technology, increasing the operating speed and the number of components per chip, can be expected to make digital methods increasingly attractive in future.

9.2. WAVEFORM CHARACTERISTICS AND DESIGN

In Section 9.1 we have seen that chirp waveforms are commonly used in pulse-compression radar systems, both for the transmitted pulse and as the required impulse response of the compressor in the receiver. Some of the basic performance criteria for the system – the range resolution, processing gain and the level of the time-sidelobes — are determined by the design of the chirp waveforms and by the accuracy with which they are implemented in practical devices. In the design of surface-wave chirp filters it is usually the case that the first task is to design chirp

waveforms satisfying the basic system requirements, and this is the topic for this section.

The stationary-phase approximation, introduced in Section 9.2.1, is widely used in this context, and will also be used later in connection with analysis of surface-wave devices. Properties of linear-chirp waveforms are given in Section 9.2.2, and the amplitude weighting of linear chirps to reduce the time-sidelobes of the compressor output waveform is explained in Section 9.2.3. The use of non-linear chirp waveforms to reduce time-sidelobes is described in Section 9.2.4. The effects of a doppler shift, due to motion of the target relative to the radar, must usually be considered when designing the waveforms, and these effects are dependent on the waveform design. However, it is convenient to defer this topic to Section 9.5.3, because the analysis makes use of some results in Section 9.5.1 concerning the effects of phase errors.

We first define some basic parameters. A chirp waveform $v(t)$, representing either the transmitted radar waveform or the impulse response of the chirp filter in the receiver, can be written

$$v(t) = a(t) \cos [\theta(t)]. \tag{9.3}$$

Here $\theta(t)$ is the time-domain phase, and will be a non-linear function of t. The envelope $a(t)$ is zero outside a time interval of length T, the length of the waveform. Clearly, $\theta(t)$ must be a monotonic function of time. We define $\theta(t)$ such that it *increases* with time; this can be done without loss of generality, because reversing the sign of $\theta(t)$ does not affect the waveform $v(t)$. The differential of $\theta(t)$ is defined to be the *instantaneous frequency* $\Omega_i(t)$, so that

$$\Omega_i(t) = \dot{\theta}(t). \tag{9.4}$$

Thus, $\Omega_i(t)$ will always be positive. It will be shown later than $\Omega_i(t)$ can be regarded approximately as the "frequency" at time t. We also define the *chirp rate* $\mu(t)$ as the rate of change of instantaneous frequency. It is convenient to express this in units of Hz/sec, so that

$$2\pi\mu(t) = \dot{\Omega}_i(t) = \ddot{\theta}(t). \tag{9.5}$$

For the waveforms considered in this chapter, the instantaneous frequency is a monotonic function, either increasing or decreasing with time. Thus $\mu(t)$ is either positive throughout the waveform, or negative throughout. Waveforms with $\mu(t) > 0$ are called *up-chirps*, while those with $\mu(t) < 0$ are called *down-chirps*. For a linear chirp $\mu(t)$ is a constant, and hence $\Omega_i(t)$ is a linear function of t and $\theta(t)$ is a quadratic function.

If $v(t)$ in equation (9.3) represents the trasmitted radar waveform, the required impulse response for the matched filter is $h(t) = v(-t)$, apart from an arbitrary delay and a multiplying constant. The phase of $h(t)$ is taken to be $-\theta(-t)$ rather than $\theta(-t)$, because here it is assumed to be increasing with time. In practice there is usually an unknown additional constant α, so that $h(t) = a(-t) \cos [\alpha - \theta(-t)]$. Although this is not in fact matched to the input waveform, the constant α has little

effect on the output waveform: it leaves the envelope unchanged, but affects the phase of the carrier. This change is nearly always of no consequence.

9.2.1. Stationary-phase Approximation

This approximation [279, p. 34] gives a very convenient method for estimating the spectrum of a chirp waveform, and will be used later to solve a wide variety of problems. We consider the chirp waveform of equation (9.3), in the complex form

$$v(t) = \tfrac{1}{2}a(t)\exp[j\theta(t)] + \tfrac{1}{2}a(t)\exp[-j\theta(t)] \tag{9.6}$$

It can be assumed that $v(t)$ is a bandpass waveform, so that its spectrum is negligible for frequencies near zero. Since $\theta(t)$ is taken to be increasing with time, it follows that the Fourier transform of the first term in equation (9.6) gives the positive-frequency part of the spectrum, which is denoted $V_+(\omega)$. We thus have

$$V_+(\omega) = \frac{1}{2}\int_{-\infty}^{\infty} a(t)\exp[j\theta(t) - j\omega t]\,\mathrm{d}t, \tag{9.7}$$

which is the spectrum of $v(t)$ for $\omega > 0$. The negative-frequency part can be deduced from this by noting that $v(t)$ is real. Since $\theta(t)$ is a non-linear function of t, the phase $\theta(t) - \omega t$ in the above equation generally varies rapidly with t, so that the exponential gives rapid oscillations which contribute little to the integral. However, the phase varies slowly for times close to the point where the differential is zero, that is, where $\dot{\theta}(t) = \omega$. The main contribution to the integral therefore arises from times near this point. The time satisfying $\dot{\theta}(t) = \omega$ is known as the *stationary phase point*, denoted by $T_s(\omega)$, so that

$$\dot{\theta}[T_s(\omega)] = \omega. \tag{9.8}$$

For any given ω, this equation has only one solution for $T_s(\omega)$, because $\dot{\theta}(t)$ is assumed to be monotonic.

Since the main contribution to the integral in equation (9.7) arises from times near $T_s(\omega)$, it is a good approximation to replace $\theta(t)$ by its Taylor expansion about this point, neglecting cubic and higher order terms. It is also assumed that the envelope $a(t)$ varies slowly with t, and may therefore be approximated by its value $a[T_s(\omega)]$ at the stationary-phase point. Using also equation (9.8), we find

$$V_+(\omega) \approx \tfrac{1}{2}a(T_s)\exp[-j\omega T_s + j\theta(T_s)]\int_{-\infty}^{\infty}\exp[\tfrac{1}{2}j(t - T_s)^2\ddot{\theta}(T_s)]\,\mathrm{d}t, \tag{9.9}$$

where the frequency argument of $T_s(\omega)$ is omitted for brevity. Here the integral is a form of the standard integral [289, p. 301]

$$\int_{-\infty}^{\infty}\exp(jKt^2)\,\mathrm{d}t = \sqrt{\frac{\pi}{|K|}}\,\mathrm{e}^{\pm j\pi/4}, \tag{9.10}$$

where $K \neq 0$ is a real constant and the sign in the exponential on the right is the same as the sign of K. We define $A(\omega)$ and $\phi(\omega)$ as the amplitude and phase of the spectrum $V_+(\omega)$, so that

$$V_+(\omega) = A(\omega)\exp[j\phi(\omega)]. \tag{9.11}$$

Using equation (9.10), equation (9.9) gives

$$A(\omega) \approx \frac{a(T_s)}{2}\sqrt{\frac{2\pi}{|\ddot{\theta}(T_s)|}} = \frac{1}{2}\frac{a(T_s)}{\sqrt{|\mu(T_s)|}} \tag{9.12a}$$

and

$$\phi(\omega) \approx \theta(T_s) - \omega T_s \pm \pi/4. \tag{9.12b}$$

These results assume that $\ddot{\theta}(T_s) \neq 0$, which is valid here because $\dot{\theta}(t)$ is monotonic. The sign in equation (9.12b) is the same as the sign of $\mu(T_s)$. Note that $a(t)$ is zero outside a specified interval of length T, and hence $A(\omega)$ is zero outside a corresponding frequency band. If $v(t)$ is the impulse response of a filter, the group delay at frequency ω is, from equations (9.12b) and (9.8),

$$\tau_g(\omega) \equiv -\mathrm{d}\phi(\omega)/\mathrm{d}\omega \approx T_s(\omega). \tag{9.13}$$

The stationary-phase approximation shows that, for a particular frequency ω, the spectrum is determined mainly by the part of the waveform in the vicinity of the stationary phase point $T_s(\omega)$, defined as the solution of equation (9.8). Conversely, it is clear that the waveform at some time t is determined mainly by the part of the spectrum in the vicinity of the frequency $\omega = \dot{\theta}(t)$, which by definition is equal to the instantaneous frequency $\Omega_i(t)$. Thus $\Omega_i(t)$ can be regarded as the time-dependent "frequency" of the waveform. The stationary-phase approximation therefore gives a formal meaning to the statement that the frequency varies with time. This statement has no precise meaning otherwise, since the term "frequency" generally refers to periodic waveforms, or to the independent parameter in the Fourier transform. Although $\Omega_i(t)$ is precisely defined, its interpretation as the time-dependent "frequency" is meaningful only when the stationary-phase approximation is valid, and this depends on the nature of the waveform as shown in Section 9.2.2 below.

9.2.2. Linear Chirp Waveforms

For a linear chirp waveform, the instantaneous frequency varies linearly with time so that the chirp rate $\mu(t)$ is a constant. The phase $\theta(t)$ is a quadratic function of t. The waveform can be written as

$$v(t) = a(t)\cos[\theta(t)] = a(t)\cos(\omega_c t + \pi\mu t^2 + \phi_0) \tag{9.14}$$

where ϕ_0 is a constant. We take the waveform to be centred at $t = 0$, so that $a(t) = 0$ for $|t| > T/2$, where T is the length. Thus, ω_c is the centre frequency. If the envelope is flat, as in Figure 9.2, then $a(t)$ will be a constant for $|t| \leqslant T/2$. The instantaneous frequency is

$$\Omega_i(t) = \omega_c + 2\pi\mu t$$

and the stationary-phase point is, from equation (9.8),

$$T_s(\omega) = (\omega - \omega_c)/(2\pi\mu). \tag{9.15}$$

If $v(t)$ is the impulse response of a filter, this expression is approximately equal to the group delay, as shown by equation (9.13). The number of cycles in the waveform is readily found to be $T\omega_c/(2\pi)$, which is independent of the chirp rate μ.

In the stationary-phase approximation, the spectrum of the waveform is given by equations (9.12). Thus the spectral amplitude is

$$A(\omega) \approx \frac{a[T_s(\omega)]}{2\sqrt{|\mu|}}. \tag{9.16a}$$

The spectral phase is

$$\phi(\omega) \approx -\frac{(\omega - \omega_c)^2}{4\pi\mu} + \phi_0 \pm \pi/4, \tag{9.16b}$$

where the sign is the same as the sign of μ. Note that, since $T_s(\omega)$ is a linear function of ω, the spectral amplitude $A(\omega)$ has the same shape as the time-domain envelope $a(t)$.

Since the envelope $a(t)$ falls to zero at the ends of the waveform, where $t = \pm T/2$, $A(\omega)$ is finite only between the corresponding instantaneous frequencies $\dot{\theta}(\pm T/2) = \omega_c \pm \pi\mu T$. It is convenient to define the *bandwidth* B (in Hz) as the range of instantaneous frequencies, so that

$$B = |\mu|T \tag{9.17}$$

and the band edges are at $\omega_c \pm \pi B$. Note that the 3 dB bandwidth of the spectrum is dependent on the form of $a(t)$, and can be quite different from B.

Linear Chirp with Flat Envelope. In a radar system the transmitted signal has a flat envelope, as illustrated in Figure 9.2. In this case the envelope $a(t)$ is a constant for the interval $|t| < T/2$ occupied by the pulse. The spectral amplitude of equation (9.16a) is therefore flat between the points $\omega = \omega_c \pm \pi B$, and is zero elsewhere, according to the stationary-phase approximation.

Some actual spectra of flat-envelope linear chirp waveforms are shown in Figure 9.3, for several values of the time-bandwidth product TB. These were obtained by taking the Fourier transform of the waveform, without approximation. For the phase curves, the quadratic term predicted by the stationary-phase approximation, equation (9.16b), has been subtracted from the data. For large TB-products, the amplitude is quite flat within a band of width $\Delta\omega = 2\pi B$ and falls off rapidly outside this band. However, ripples are present in both the amplitude and the phase, and they become more significant as TB is reduced. Thus the stationary-phase approximation is valid only for large TB. For most purposes the approximation is found to give adequate accuracy if $TB \geqslant 100$. Figure 9.3 also shows that B is approximately the 6 dB bandwidth, in Hz, for the flat-envelope case.

The amplitude curves in Figure 9.3 are closely related to the isotropic diffraction plots shown earlier on the right of Figure 6.6. Mathematically, both cases refer to the Fourier transform of a function with rectangular amplitude and quadratic phase. However, the appearance is rather different because the "TB-products" in Figure 6.6 are very small.

Suppose now that a flat-envelope linear chirp waveform is applied to its matched filter, as in Figure 9.2. The input waveform $s(t)$ is taken to be equal to $v(t)$, equation (9.14), with $a(t)$ set equal to unity for $|t| < T/2$. Taking the filter response

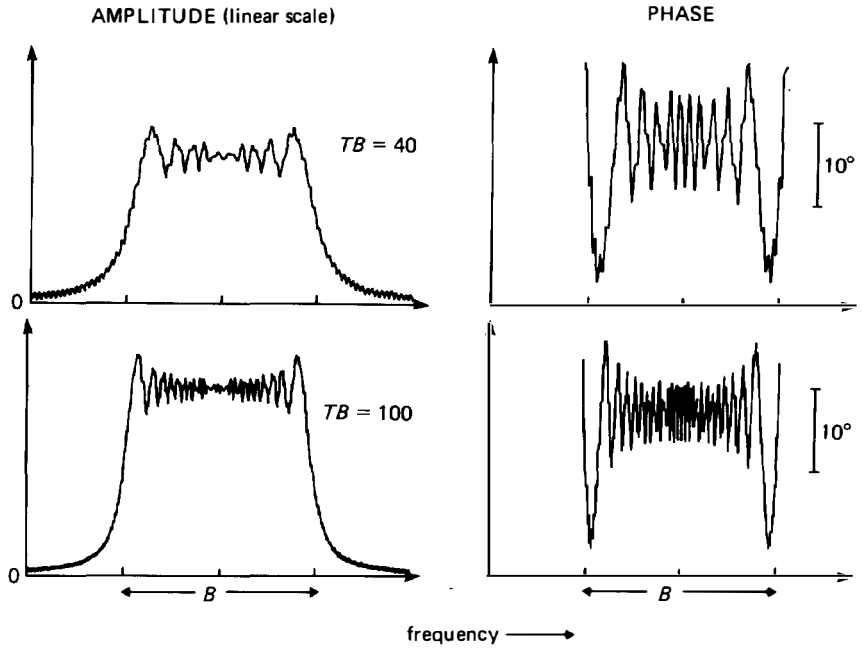


FIGURE 9.3. Spectra of flat-envelope linear chirp waveforms.

to be the time-reverse of this, the spectrum of the output waveform is $G(\omega) = |S(\omega)|^2$, where $S(\omega)$ is the spectrum of $s(t)$. This result follows from equation (A.66) of Appendix A. For positive frequencies the magnitude of $S(\omega)$ is equal to $A(\omega)$, given by equation (9.16a) in the stationary-phase approximation. Transforming $G(\omega)$ to the time-domain, the output waveform is found to be

$$g(t) \approx \tfrac{1}{2} T \cos(\omega_c t) \operatorname{sinc}(\pi B t), \tag{9.18}$$

where $\operatorname{sinc}(x) = (\sin x)/x$. This is illustrated in Figure 9.2. The envelope has its maximum at $t = 0$, and the first zeros are at $t = \pm 1/B$. The 4 dB points are almost exactly at $t = \pm 1/(2B)$.

An exact analysis [279, p. 133] shows that equation (9.18) is a good approximation for times close to the main peak. The main differences are that the exact expression has a finite length $2T$, and, for times remote from the main peak, the locations of the zeros are different.

9.2.3. Weighting of Linear-chirp Filters

It was seen above that, for a flat-envelope linear chirp waveform applied to its matched filter, the output waveform has an envelope of the form $\operatorname{sinc}(\pi B t)$. The largest time-sidelobes of this waveform are only 13 dB below the main peak, and this is unacceptable for most radar applications, which typically require a sidelobe rejection of 30 dB.

A common method of reducing the sidelobes is to apply amplitude weighting in the filter [277, 279, 280], as illustrated in Figure 9.4. The filter impulse response is weighted to give a maximum amplitude at the centre, reducing towards the ends. In addition to reducing the sidelobe levels this also increases the width of the main peak somewhat, though this can be compensated for by increasing the bandwidth B. There is also a small reduction of output signal-to-noise ratio, since the filter is no longer exactly matched to the input waveform. The use of non-linear chirps to avoid this loss of SNR will be considered later, in Section 9.2.4.

In the analysis here it is assumed initially that $TB \geqslant 100$, so that the stationary-phase approximation can be used. The limitations imposed by this assumption are discussed later.

As before, we consider a waveform $s(t)$ applied to a filter with impulse response $h(t)$, giving an output waveform $g(t)$. The input waveform is a flat-envelope linear chirp

$$s(t) = \cos(\omega_c t - \pi\mu t^2 - \phi_0), \quad \text{for } |t| \leqslant T/2, \tag{9.19}$$

with $s(t) = 0$ for other t. The matched filter for this waveform has an impulse response $s(-t)$, but here the filter is taken to have amplitude weighting with an envelope $a(t)$, so that

$$h(t) = a(t)\cos(\omega_c t + \pi\mu t^2 + \phi_0), \quad \text{for } |t| \leqslant T/2. \tag{9.20}$$

The transforms of $s(t)$ and $h(t)$ are respectively denoted $S(\omega)$ and $H(\omega)$, and the spectrum of the output waveform is $G(\omega) = S(\omega)H(\omega)$. The waveforms $s(t)$ and $h(t)$ have the same form as equation (9.14), so their transforms are given by equations (9.16) in the stationary-phase approximation. The spectrum of the output waveform, for $\omega > 0$, is thus found to be the real function

$$\begin{aligned} G(\omega) &\approx \frac{a[T_s(\omega)]}{4|\mu|}, \quad \text{for } |\omega - \omega_c| \leqslant \pi B, \\ &\approx 0, \qquad\quad \text{for } |\omega - \omega_c| > \pi B, \end{aligned} \tag{9.21}$$

where

$$T_s(\omega) = (\omega - \omega_c)/(2\pi\mu) \tag{9.22}$$

is the stationary-phase point for the filter.

It is convenient to express this result in terms of a real frequency-domain *weighting function* denoted by $\bar{W}(\omega)$, which is considered to be centred at zero frequency and

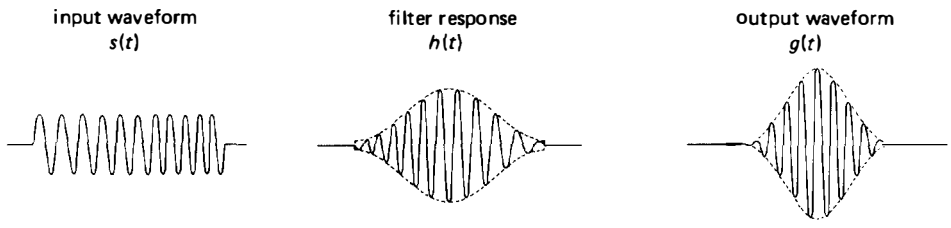


FIGURE 9.4. Pulse compression using linear chirps with amplitude weighting.

is zero for $|\omega| > \pi B$. The output spectrum $G(\omega)$ is thus proportional to $\bar{W}(\omega - \omega_c)$ for positive ω, and can be written

$$G(\omega) = K\bar{W}(\omega - \omega_c), \quad \text{for } \omega > 0, \tag{9.23}$$

where K is an arbitrary real constant. If $W(t)$ is the inverse Fourier transform of $\bar{W}(\omega)$, the output waveform $g(t)$ is given by the shifting theorem, equation (A.11), so that

$$g(t) = 2KW(t) \cos \omega_c t. \tag{9.24}$$

The envelope of the output waveform is therefore proportional to $W(t)$. We also have, from equations (9.21)–(9.23),

$$a(t) = 4K|\mu|\bar{W}(2\pi\mu t). \tag{9.25}$$

Thus the time-domain weighting $a(t)$ is, with a change of scale, proportional to the frequency-domain weighting function. Both sides of equation (9.25) are, by definition, zero for $|t| > \frac{1}{2}T$.

Types of Weighting Function. To design the filter impulse response, we need to find a weighting function $\bar{W}(\omega)$ which is zero for $|\omega| > \pi B$ and has an inverse Fourier transform $W(t)$ giving well-suppressed time-sidelobes. A similar problem was encountered in Section 8.2.1, which discussed the use of window functions for bandpass filter design. This is the dual of the chirp filter case; for a bandpass filter the window function has finite length in the time domain, while for a chirp filter the weighting function has finite length in the frequency domain. Thus a window function suitable for bandpass filters can also be applied to chirp filters if the time and frequency variables are interchanged, with a scaling factor such that the required bandwidth is obtained.

In practice, the weighting functions used for chirp filters have usually been functions developed earlier for antenna weighting, where the polar diagram is required to have well-suppressed sidelobes. While a variety of weighting functions have been considered [277, 279, 280], the usual choice for a chirp filter is either Hamming weighting or Taylor weighting. The Hamming weighting function $\bar{W}_H(\omega)$ is defined by

$$\begin{aligned} \bar{W}_H(\omega) &= 0.54 + 0.46 \cos \omega/B, \quad \text{for } |\omega| \leqslant \pi B, \\ &= 0, \qquad\qquad\qquad\quad \text{for } |\omega| > \pi B. \end{aligned} \tag{9.26}$$

This function, and its transform $W_H(t)$, are shown in Figure 9.5. The output envelope $W_H(t)$ gives a series of sidelobes, the largest of which is the fourth one, peaking at $Bt = 4.5$; this sidelobe is 42.8 dB below the main peak. The width of the main peak, at the 3 dB points, is about 50% larger than the width obtained for an unweighted filter, for the same bandwidth, B. There is therefore some loss of radar resolution, in exchange for the suppression of sidelobes. This weighting is also shown in Figure 9.4, where the envelope of the filter response is proportional to $\bar{W}_H(2\pi\mu t)$ and the envelope of the output waveform is proportional to $W_H(t)$.

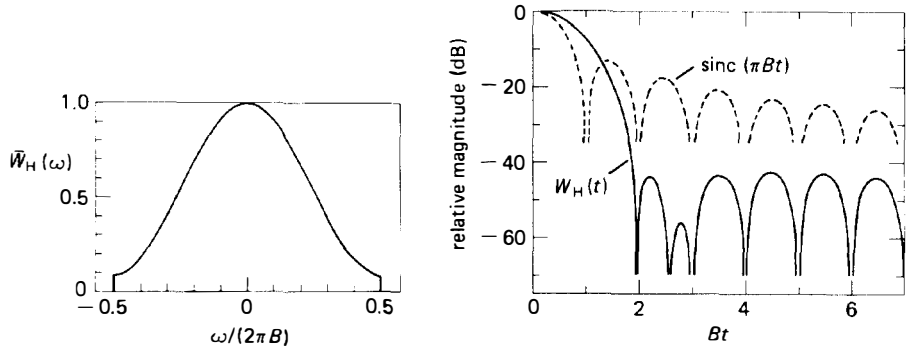


FIGURE 9.5. Hamming weighting function (left) and its transform. The transform is symmetrical about $t = 0$.

The Taylor weighting function [277, 279, 280, 290] enables the time-sidelobe level to be set by the designer, and is given by

$$\begin{aligned} \bar{W}_T(\omega) &= 1 + 2\sum_{m=1}^{\bar{n}-1} F_m \cos m\omega/B, \quad \text{for } |\omega| \leqslant \pi B, \\ &= 0, \qquad\qquad\qquad\qquad \text{for } |\omega| > \pi B. \end{aligned} \tag{9.27}$$

For this function, the largest sidelobe is the one adjacent to the main peak. The coefficients F_m and the number of terms $\bar{n}$ are chosen in accordance with the performance required. To evaluate the function, the maximum sidelobe level is first chosen, and a value for $\bar{n}$ is selected. These values are then inserted into a formula to give the values of the F_m. For a given sidelobe level, increasing the value of $\bar{n}$ reduces the width of the main peak. For large $\bar{n}$ the peak width approaches that given by the Dolph–Chebyshev function, which gives the minimum possible peak width for a given sidelobe level. In practice, the value of $\bar{n}$ is not critical, and typical values are $\bar{n} = 5$ for 30dB sidelobes or $\bar{n} = 10$ for 50dB sidelobes. Generally, $\bar{W}_T(\omega)$ has an appearance similar to $\bar{W}_H(\omega)$, shown in Figure 9.5, though for large $\bar{n}$ it can increase at the band edges.

Figure 9.6 shows, on the left, the width of the main output peak as a function of the sidelobe level relative to the peak. The peak width is measured at the 3dB points of the output waveform, and is normalised by multiplying by B. It can be seen that lower sidelobe levels give wider peaks. The point marked "unweighted" refers to a filter whose impulse response has a flat envelope, that is, the filter matched to the input waveform. In this case the output waveform has an envelope proportional to sinc (πBt), as shown by equation (9.18).

Figure 9.6 also shows the *mis-match loss* for a Taylor-weighted device. This is the reduction of output signal-to-noise power ratio (SNR) due to the fact that the filter is not matched to the input signal. The output SNR for the weighted filter is given by equations (A.55)–(A.57) in Appendix A. For a matched filter the output SNR is $2E_s/N_i$, as shown by equation (A.61). With these two SNR's expressed in dB, the difference is the mis-match loss plotted in Figure 9.6.This figure also shows the mis-match loss and output peak width for a Hamming weighted device.

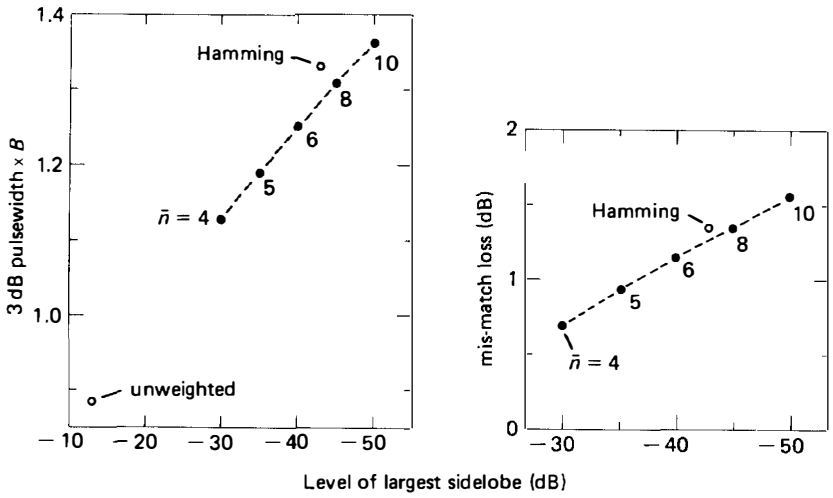


FIGURE 9.6. Output pulsewidth and mis-match loss for linear chirps with Taylor weighting (solid circles) and Hamming weighting. From data given by Farnett *et al.* [280].

Limitations of the Stationary-phase Approach. In the above description, the stationary-phase approximation was used both to establish the design of the filter response and to deduce the nature of the output waveform. In practice, the design method, giving the amplitude weighting of equation (9.25), is usually acceptable if TB exceeds 100. According to the stationary-phase analysis, the output waveform, equation (9.24), has an envelope proportional to $W(t)$. This gives a sequence of sidelobes which decrease progressively for times more remote from the main peak, except for the first three sidelobes in the case of Hamming weighting.

For a device designed as above, accurate calculations show that, for $TB > 100$, the output waveform predicted by the stationary-phase method is a good approximation for times close to the main peak. The main discrepancy is that the more remote sidelobes can be somewhat larger than those predicted by the stationary-phase method. The additional sidelobes are associated with the abrupt truncation of $s(t)$ and $h(t)$, and are known as "gating sidelobes". They are maximised at $t \approx \pm \frac{1}{2}T$. It is common practice to reduce the gating sidelobes by adding short "extensions" to $h(t)$, so that the envelope falls smoothly to zero instead of the abrupt truncation shown in Figure 9.4. When this is done, the maximum level of the gating sidelobes is typically about 20 log (TB) + 3 dB below the main peak, and this is usually acceptable.

For small TB, the stationary-phase design method generally gives unacceptable sidelobe levels. For this case a more sophisticated design method, called the "reciprocal-ripple" method, was introduced by Judd [291], and is also described by Armstrong [292, 293]. The spectrum of the output waveform is $G(\omega) = S(\omega)H(\omega)$, and is required to be proportional to the weighting function $\bar{W}(\omega - \omega_c)$. This requirement can be satisfied by taking the filter response to have the form

$$H(\omega) = \frac{\bar{W}(\omega - \omega_c)}{S(\omega)}. \tag{9.28}$$

For large TB the amplitude of $S(\omega)$ is almost flat within the band, so that $1/S(\omega)$ is approximately proportional to $S^*(\omega)$, and hence the form of $H(\omega)$ given by equation (9.28) is almost the same as that given by the stationary-phase method. However, the ripples of $S(\omega)$, evident in Figure 9.3, are cancelled when $H(\omega)$ is given by equation (9.28). This enables low time-sidelobe levels to be obtained from waveforms with low TB-products, which give large ripples. In practice, the weighting function $\bar{W}(\omega - \omega_c)$ is extended beyond the band edges to avoid abrupt discontinuities. The resulting function $H(\omega)$ is transformed to the time domain, and since this gives a waveform of infinite duration it is necessary to truncate and add short extensions decaying smoothly to zero. It is usually found that the filter impulse response must be somewhat longer than the input waveform. An experimental device [291] gave a time-sidelobe rejection of 34 dB, with a TB-product as low as 8; for this case, the sidelobe rejection for a stationary-phase design is 20 dB.

The main limitation of the reciprocal-ripple method is that it increases the sensitivity to doppler shifts of the input waveform, which cause the sidelobe levels to increase. The doppler sensitivity becomes worse as the TB-product increases, and hence the reciprocal ripple method is usually applied only when $TB < 100$.

9.2.4. Weighting Using Non-linear Chirps

It was seen above that, for linear chirps, amplitude weighting in the filter causes a mis-match loss of typically 1 to 1.5 dB. If weighting could be incorporated in such a way that the filter were still matched to the input waveform, then low time-sidelobe levels would be obtainable without this loss of SNR. An obvious way of doing this is to use linear chirps, with amplitude weighting in both the input waveform and the filter impulse response, but this is usually precluded by the requirement that the waveform transmitted by the radar should have a flat envelope. However, a viable solution is to employ non-linear chirps, which use a form of phase weighting. These enable the time-sidelobes to be suppressed without significant loss of SNR, while still retaining a flat envelope for the input waveform. The method was introduced by Cook and Paolillo [294], and is also described by Cook and Bernfeld [279, p. 222] and by Newton [295].

The method is illustrated by Figure 9.7, which shows the frequency-time curves for the input waveform and for the matched filter. The frequency varies slowly with time at the centre, and more rapidly at the ends. This causes the spectrum of the waveform to have a power density decreasing toward the band edges, even though the time-domain envelope is flat. The input waveform and the filter impulse response appear much as in Figure 9.2, but the spectral weighting reduces the time-sidelobes so that the output waveform is typically as in Figure 9.4.

As in the previous section, the stationary-phase approximation is used in the analysis. We consider a finite-length waveform $s(t)$ applied to its matched filter, whose impulse response is $h(t)$. These waveforms are both flat-envelope chirps, and are given by the expressions

$$\begin{aligned} h(t) &= \cos[\theta(t)], \quad \text{for } |t| \leqslant \tfrac{1}{2}T, \\ &= 0, \qquad\quad \text{for } |t| > \tfrac{1}{2}T, \end{aligned} \tag{9.29}$$

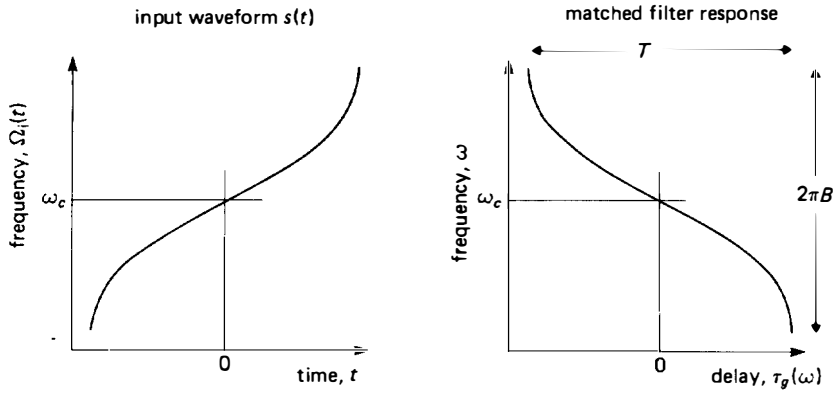


FIGURE 9.7. Dispersion curves for non-linear chirps with Hamming weighting, in the stationary-phase approximation.

and

$$s(t) = h(-t).$$

For convenience, the waveforms are both centred at $t = 0$, and have the same amplitude. According to the matched filter analysis of Section A.3, the waveform spectrum $S(\omega)$ and the filter frequency response $H(\omega)$ are related by $S(\omega) = H^*(\omega)$, and the output waveform $g(t)$ has a spectrum $G(\omega) = |H(\omega)|^2$. In the stationary-phase approximation the frequency response of the filter is given by equations (9.12), and we thus find

$$\begin{aligned} G(\omega) &\approx \frac{\pi}{2|\ddot{\theta}(T_s)|}, \quad \text{for } |\omega - \omega_c| \leqslant \pi B, \\ &\approx 0, \qquad\quad \text{for } |\omega - \omega_c| > \pi B, \end{aligned} \tag{9.30}$$

where $T_s(\omega)$ is the stationary-phase point, defined as the solution of $\dot{\theta}(T_s) = \omega$. The band-edge frequencies $\omega = \omega_c \pm \pi B$ are defined as the values of the instantaneous frequency $\Omega_i(t) = \dot{\theta}(t)$ at the two ends of the waveform, that is, at $t = \pm \frac{1}{2}T$. The centre frequency ω_c is midway between these points.

If $\dot{\theta}(t)$ is a non-linear function of t, so that the chirps are non-linear, the factor $\ddot{\theta}(T_s)$ in equation (9.30) will vary with frequency, thus modifying the amplitude of the output spectrum. This enables the sidelobes of the output waveform to be suppressed by designing an appropriate time-domain phase $\theta(t)$. Since $G(\omega)$ is zero outside a band of width $2\pi B$, it is appropriate to design the device such that

$$G(\omega) = C\bar{W}(\omega - \omega_c), \quad \text{for } \omega > 0, \tag{9.31}$$

where C is a constant. Here $\bar{W}(\omega)$ is a weighting function which is zero for $|\omega| > \pi B$, and has an inverse Fourier transform $W(t)$ with well-suppressed sidelobes. The output waveform $g(t)$ has its envelope proportional to $W(t)$, as can be seen by comparison with equations (9.23) and (9.24). As discussed in the previous section, suitable choices for $\bar{W}(\omega)$ include the Hamming function $\bar{W}_H(\omega)$ and the Taylor function $\bar{W}_T(\omega)$.

From equations (9.30) and (9.31) we have

$$\ddot{\theta}[T_s(\omega)] = \pm \frac{\pi}{2C\bar{W}(\omega - \omega_c)}. \tag{9.32}$$

Now, since $\dot{\theta}[T_s(\omega)] = \omega$ we have $\ddot{\theta}(T_s) \cdot \mathrm{d}T_s/\mathrm{d}\omega = 1$, and hence

$$\mathrm{d}T_s/\mathrm{d}\omega = \pm 2C\bar{W}(\omega - \omega_c)/\pi. \tag{9.33}$$

Thus, $\bar{W}(\omega - \omega_c)$ can be integrated to give $T_s(\omega)$, which in the stationary-phase approximation is equal to the group delay of the filter. The constant of integration determines the value of $T_s(\omega)$ at the centre frequency ω_c. The constant C determines the length T of the waveform, and is chosen such that the values of $T_s(\omega)$ at the band edges $\omega_c \pm \pi B$ differ by T.

Once $T_s(\omega)$ has been found, the instantaneous frequency $\Omega_i(t) = \dot{\theta}(t)$ follows by solving the equation $\dot{\theta}[T_s(\omega)] = \omega$, so that values of $T_s(\omega)$ give $\dot{\theta}(t)$ directly if T_s is interpreted as t and ω is interpreted as $\dot{\theta}(t)$. From $\dot{\theta}(t)$, the phase $\theta(t)$ is obtained by integration, and here the constant of integration is immaterial.

To illustrate the method, consider Hamming weighting, so that the function $\bar{W}(\omega)$ is the Hamming function $\bar{W}_H(\omega)$ of equation (9.26). In this case the stationary-phase point $T_s(\omega)$ of the filter is, from equation (9.33),

$$T_s(\omega) = \pm \frac{T}{2\pi}\left[\frac{\omega - \omega_c}{B} + \frac{0.46}{0.54} \sin\left(\frac{\omega - \omega_c}{B}\right)\right], \quad \text{for } |\omega - \omega_c| \leqslant \pi B \tag{9.34}$$

and this is approximately equal to the group delay. The constant of integration has been chosen such that $T_s(\omega) = 0$ at the centre frequency ω_c. The function is shown on the right of Figure 9.7, where the lower sign has been chosen in equation (9.34) so that the filter impulse response is a down-chirp. Since $\Omega_i(T_s) = \omega$, the curve can also be regarded as the instantaneous frequency, as a function of time. The input waveform $s(t)$ is the time-reverse of the filter impulse response $h(t)$ and therefore has its instantaneous frequency reversed in time, as shown on the left in Figure 9.7.

For a given weighting function, the output envelope is the same as that obtained for an amplitude-weighted linear chirp, considered above. For example, if the Taylor weighting function $\bar{W}_T(\omega)$ is used, the width of the main peak can be traded-off against the sidelobe rejection, as in the left part of Figure 9.6. However, the mis-match loss is zero, since the filter is matched.

The limitations of the stationary-phase approach, mentioned in the previous section, apply here also. Thus, it is usual to add extensions to $h(t)$ to avoid abrupt truncation, and for small TB-products an adaptation of the reciprocal-ripple design method can be used [295]. With these modifications, the filter is no longer exactly matched to the input waveform, but the resulting mis-match loss is very small. For example [295], a reciprocal-ripple design with $TB = 60$ gave 36 dB sidelobe rejection and a mis-match loss of 0.04 dB, which is typical.

9.3. CHIRP TRANSDUCERS

Having considered the design of chirp waveforms, we now turn to the analysis and design of surface-wave interdigital chirp filters, which are used as matched filters for

these waveforms and for passive waveform generation. The present section is concerned with analysis and design of chirp transducers, while the design and performance of devices using these transducers is described in Section 9.4 below. In both of these sections, no restriction is placed on the nature of the chirp waveform involved. Thus the filter impulse response may be any of the types of chirp mentioned in Section 9.2, including weighted and unweighted linear chirps, and non-linear chirps.

In this section we are concerned with chirp transducers such as that shown on the right of the chirp filter of Figure 9.8. The main transducer properties of interest are the frequency response $H_t(\omega)$ and the admittance. The analysis here is based on the quasi-static approach of Chapter 4, assuming that the only acoustic wave present is a piezoelectric Rayleigh wave, and ignoring propagation loss and surface-wave diffraction. Electrode interactions are also ignored, and the reader is reminded that the analysis predicts that the surface-wave velocity within the transducer is the free-surface velocity v_0. In practice the electrodes reduce the velocity a little even when reflections due to interactions are negligible (Section D.2), and this can be allowed for in the analysis below by adjusting the value of the wavenumber k_0. End effects are assumed to be negligible, and this condition can be obtained by adding a few guard electrodes at each end, as explained in Section 4.5.1. However, chirp transducers are usually many wavelengths long, and consequently end effects are usually negligible even if the guards are omitted. For apodised transducers the analysis assumes that dummy electrodes are included. This is usually done in practice, for reasons mentioned in Section 4.7.2. It is assumed that all the electrodes connected to either bus-bar have a uniform potential, and this implies that the metallisation has negligible resistivity, and that the transmission-line effect (Section 9.5.4) is neglected. Second-order effects are considered in Section 9.5. below, which also considers the resulting distortion of the output waveform when the device is used for pulse compression.

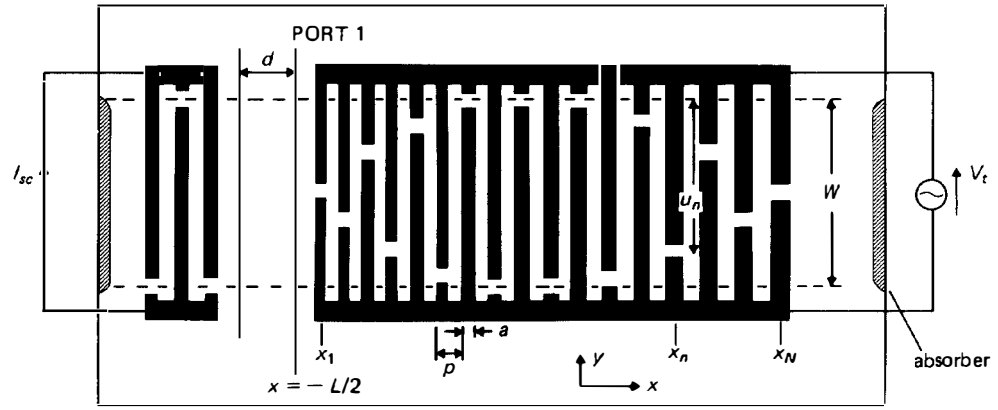


FIGURE 9.8. Single-dispersive interdigital chirp filter. The chirp transducer is the single-electrode ($S_e = 2$) type.

The chirp transducer of Figure 9.8 is typical of designs obtained by synchronous sampling, that is, a chirp waveform with phase $\theta(t)$ is sampled at times corresponding to equal increments of $\theta(t)$. In the case shown there are two samples for period of the waveform, and hence the transducer is a "single-electrode" type, with $S_e = 2$. Multi-electrode designs, with $S_e > 2$, can be obtained by using more than two samples per period, as discussed for constant-pitch transducers in Sections 4.2 and 8.1.3; such designs are often used in order to minimise electrode interaction effects. Synchronous sampling is nearly always used in chirp filters as it generally leads to relatively slow changes of apodisation, reducing second-order effects due to diffraction. In Section 9.3.1 it will be shown that, if synchronous sampling is used, the transducer response can be expressed in terms of an array factor and an element factor, in a manner very similar to the delta-function model of Section 4.1.

9.3.1. Transducer Analysis

We consider first an *unapodised* chirp transducer, such as that shown in Figure 1.5. The electrode pitch p and width a vary along the length. It will be assumed that p and a vary slowly with distance, and that the ratio a/p is constant throughout the transducer. We define $\varrho_e(x)$ as the electrostatic charge density on the electrodes when unit voltage is applied across the transducer, and $\bar{\varrho}_e(\beta)$ as the Fourier transform of $\varrho_e(x)$. The transducer has acoustic ports at $x = \pm \frac{1}{2}L$, and when a voltage V_t is applied, at frequency ω, a surface-wave with potential $\phi_{s1}(\omega)$ emerges at port 1, that is, at $x = -\frac{1}{2}L$. According to the quasi-static analysis of Chapter 4 this potential is given by equation (4.50), so that

$$\phi_{s1}(\omega) = j\Gamma_s V_t \bar{\varrho}_e(k_0) \exp(-\tfrac{1}{2}jk_0 L), \tag{9.35}$$

where $k_0 = \omega/v_0$ is the wavenumber for surface waves on a free surface, and Γ_s is a piezoelectric coupling parameter such that $(\varepsilon_0 + \varepsilon_p^T)\Gamma_s = \Delta v/v$, as in equation (4.21).

In Section 4.5 it was shown that the response can be written in a more convenient form if the electrodes are regular, so that p and a are constants. For a chirp transducer, p and a vary along the length. However, a development similar to that of Section 4.5 can be used here. We define $\varrho_{en}(x)$ as the electrostatic charge density on the electrodes when electrode n has unit voltage and all the other electrodes are grounded. Then, according to the charge superposition principle, equation (4.33), the charge density $\varrho_e(x)$ is given by

$$\varrho_e(x) = \sum_{n=1}^{N} \hat{P}_n \varrho_{en}(x), \tag{9.36}$$

where N is the number of electrodes and $\hat{P}_n$ are the electrode polarities, taking $\hat{P}_n = 1$ if the electrode n is connected to the upper bus and $\hat{P}_n = 0$ if it is connected to the lower bus. In Section 4.5.1 it was shown that, for regular electrodes, $\varrho_{en}(x)$ can be equated with $\varrho_f(x - x_n)$ provided end effects are negligible. Here x_n is the centre location of electrode n, and $\varrho_f(x)$ is the elemental charge density shown in Figure 4.10. In the present case, the electrodes are not regular. However, $\varrho_f(x - x_n)$ gives significant charge densities only in the immediate vicinity of electrode n. Since the pitch p usually varies very little from one electrode to the next, $\varrho_{en}(x)$ will not be substantially affected by the pitch variation, and is therefore given by

$$\varrho_{en}(x) \approx \varrho_f(x - x_n). \tag{9.37}$$

In practice, this equation is usually an excellent approximation. Note that $\varrho_f(x - x_n)$ depends on p, which varies from one electrode to another.

The Fourier transform of $\varrho_f(x)$ is $\bar{\varrho}_f(\beta)$, given by equation (4.88) and shown in Figure 4.11. Equation (4.88) shows that $\bar{\varrho}_f(\beta)$ is a function of the normalised variable βp. Since p varies with position, it is convenient here to change the notation slightly. The pitch in the vicinity of electrode n is denoted by p_n, and the function $\bar{\varrho}_f(\beta)$ for the electrodes in this vicinity is written as $\bar{\varrho}_f(\beta p_n)$. Using equations (9.37) and (9.36), we thus have

$$\bar{\varrho}_e(\beta) = \sum_{n=1}^{N} \hat{P}_n \bar{\varrho}_f(\beta p_n) \exp(-j\beta x_n). \tag{9.38}$$

Evaluating this at $\beta = k_0$ gives $\phi_{s1}(\omega)$, as shown by equation (9.35).

We now consider an *apodised* transducer, such as that shown in Figure 9.8. In this case the surface-wave potential $\phi_{s1}(\omega)$ will vary across the width of the transducer. The general analysis for apodised transducers is given in Section 4.7.2. The frequency response $H_t(\omega)$ of the transducer is defined such that

$$H_t(\omega) = -j\sqrt{\frac{\omega W}{\Gamma_s}} \frac{\langle \phi_{s1}(\omega) \rangle}{V_t}. \tag{9.39}$$

Here it is assumed that a voltage V_t is applied, and $\langle \phi_{s1}(\omega) \rangle$ is the average potential of the surface waves at port 1, taking the average over the aperture W of the transducer. As in Section 4.7.2, the transducer is imagined to be divided into a number of parallel channels, each of which can be regarded as an unapodised transducer. For channel j, the surface-wave potential at port 1 is, from equations (9.35) and (9.38),

$$\phi'_{s1}(\omega) = j\Gamma_s V_t \exp(-\tfrac{1}{2}jk_0L) \sum_{n=1}^{N} \hat{P}_n \bar{\varrho}_f(k_0 p_n) \exp(-jk_0 x_n). \tag{9.40}$$

Here the polarity sequence $\hat{P}_n$ will vary from channel to channel. To find the average potential, we integrate with respect to y and divide by W. This is easily done by integrating individual terms in equation (9.40). For electrode n, the break is at a location given by u_n, as shown in Figure 9.8. Thus the nth term in equation (9.40) is uniform over a distance u_n, and zero for the remaining distance $W - u_n$. The average potential $\langle \phi_{s1}(\omega) \rangle$ is therefore obtained by replacing $\hat{P}_n$ by u_n/W in equation (9.40), and substitution into equation (9.39) gives

$$H_t(\omega) = (\omega W \Gamma_s)^{1/2} \exp(-\tfrac{1}{2}jk_0L) \sum_{n=1}^{N} \frac{u_n}{W} \bar{\varrho}_f(k_0 p_n) \exp(-jk_0 x_n). \tag{9.41}$$

An alternative formulation in terms of gap elements is obtained by the method of Section 4.5.3, and is found to give

$$H_t(\omega) = (\omega W \Gamma_s)^{1/2} e^{-\frac{1}{4}jk_0 L} \sum_{n=1}^{N-1} \frac{u_{n+1} - u_n}{W} \bar{\varrho}_g(k_0 p_n) e^{-jk_0(x_n + p_n/2)} \quad (9.42)$$

where $\bar{\varrho}_g(k_0 p_n)$ is the transform of the elemental charge density for gaps, given by equation (4.96). In both of the above equations, p_n is the pitch in the vicinity of electrode n.

The above results can be simplified if the transducer is designed using synchronous sampling, that is, if it is designed to given an impulse response of the form $a(t) \cos [\theta(t)]$ and the electrode locations correspond to equal increments of $\theta(t)$. This is nearly always the case in practice. Consider a small region of the transducer, containing several periods, and assume that the pitch does not vary much over this region. For synchronous sampling, the electrodes in this region generate surface waves most effectively at a frequency such that the local period $p_n S_e$ equals the wavelength, that is, when $k_0 = 2\pi/(p_n S_e)$. At other frequencies the waves generated by individual periods are not in phase, so the surface-wave amplitude is generally much smaller. Thus, in the function $\bar{\varrho}_f(k_0 p_n)$ we can replace $k_0 p_n$ by $2\pi/S_e$, and this is usually a very good approximation because $\bar{\varrho}_f(k_0 p_n)$ varies slowly with frequency. Furthermore, $\bar{\varrho}_f(2\pi/S_e)$ is a constant, and can therefore be taken outside the summation in equation (9.41). Similarly, $\bar{\varrho}_g(k_0 p_n)$ can be replaced by $\bar{\varrho}_g(2\pi/S_e)$, and thus equation (9.42) becomes

$$H_t(\omega) = (\omega W \Gamma_s)^{1/2} e^{-\frac{1}{4}jk_0 L} \bar{\varrho}_g(2\pi/S_e) \sum_{n=1}^{N-1} \frac{u_{n+1} - u_n}{W} e^{-jk_0(x_n + p_n/2)}. \quad (9.43)$$

This is very similar to the delta-function analysis of Section 4.1, though the element factor, that is, the expression outside the summation, is proportional to $\omega^{1/2}$. The term $(x_n + p_n/2)$ is the location of the nth gap element, midway between the centres of electrode n and electrode $n + 1$.

The above approximation is nearly always adequate for practical purposes. It should be noted that the above refers to the fundamental response. For a harmonic, $k_0 p$ is replaced by $2\pi M/S_e$, where M is the harmonic number.

The transducer response is defined such that the short-circuit response of a two-transducer device is, from equation (4.127),

$$H_{sc}(\omega) \equiv I_{sc}/V_t = H_t^a(\omega) H_t^b(\omega) \exp(-jk_0 d), \quad (9.44)$$

where $H_t^a(\omega)$ and $H_t^b(\omega)$ are the transducer responses, d is the separation between their acoustic ports, and I_{sc} and V_t are shown in Figure 9.8. It is assumed that at least one of the two transducers is unapodised. As in Chapter 4, the transducer admittance can be written $Y_t(\omega) = G_a(\omega) + jB_a(\omega) + j\omega C_t$. For an unapodised transducer the parallel conductance is $G_a(\omega) = |H_t(\omega)|^2$, as shown by equation (4.124), but for an apodised transducer it is necessary to sum the conductances of individual channels to obtain the overall conductance. The capacitance C_t can be obtained by summing the electrostatic charges on all the electrodes connected to one bus. The scattering coefficients for an unapodised transducer are given in Section 4.4.4, and the device response allowing for terminating circuits is given in Section 4.8.

9.3.2. Admittance in the Stationary-phase Approximation

In this section we give an approximate analysis for the parallel conductance $G_a(\omega)$ of a chirp transducer, making use of the stationary-phase approximation introduced in Section 9.2.1. Although the result does not predict the ripples which are given by an accurate analysis, it is very useful in practice and conveniently relates the performance of a chirp transducer to that of a uniform transducer with the same basic electrode structure.

We first consider a uniform transducer with aperture W and with N_{pu} periods. At the fundamental centre frequency ω_c, the parallel conductance $G_{au}(\omega_c)$ can be written

$$G_{au}(\omega_c) = \omega_c N_{pu}^2 W \Gamma_s G_0, \tag{9.45}$$

where G_0 is a constant dependent only on the properties of the substrate and the structure of the electrodes in each period, that is, it depends on S_e and on a/p. This form for $G_{au}(\omega_c)$ is derived in Section 4.6. Comparison with equation (4.118) shows that $G_0 = (\varepsilon_0 + \varepsilon_p^T)^2 \tilde{G}_{a1}$, where $\tilde{G}_{a1}$ is given in Table 4.1. When a voltage V_t is applied to an unapodised transducer, the potential $\phi_s(\omega)$ of the surface wave radiated in either direction is related to the conductance $G_a(\omega)$ by

$$|\phi_s(\omega)|^2 = V_t^2 \Gamma_s G_a(\omega)/(\omega W), \tag{9.46}$$

which follows from power conservation. Thus, in the particular case of a uniform transducer at the centre frequency ω_c we have

$$|\phi_s(\omega_c)| = V_t N_{pu} \Gamma_s G_0^{1/2}. \tag{9.47}$$

We now consider an *unapodised* chirp transducer. For each "period" of the transducer, the surface-wave amplitude generated is obtained by regarding the electrodes as a uniform transducer with one period. At any frequency ω in the band, the surface waves are generated primarily in the region where the periodicity corresponds to the wavelength, so the surface-wave potential can be obtained from equation (9.47). Surface waves generated elsewhere contribute little to the total amplitude, so equation (9.47) may be used for all the periods in the transducer. We define x_m as the location of period m, such that the wave potential ϕ_s due to this period is real if the phase is referred to x_m. The total wave potential generated by the chirp transducer can then be written

$$\phi_{s1}(\omega) \approx V_t \Gamma_s G_0^{1/2} \exp(-\tfrac{1}{2} j k_0 L) \sum_{m=1}^{N_p} \exp(-j k_0 x_m), \tag{9.48}$$

where N_p is the number of periods, so that the number of electrodes is $N = N_p S_e$. The potential is evaluated at port 1, that is, at $x = -\frac{1}{2}L$.

To evaluate $\phi_{s1}(\omega)$, it is assumed that the transducer is designed by sampling a non-linear function $\theta(t)$ at points t_m such that $\theta(t_m) = 2m\pi$. The points t_m therefore have unequal spacing. The locations x_m are taken to be $x_m = v_0 t_m$, and hence $k_0 x_m = \omega t_m$. The summation in equation (9.48) can be written

$$S = \sum_{m=1}^{N_p} \exp(-j\omega t_m) = \sum_{m=1}^{N_p} \exp[j\theta(t_m) - j\omega t_m]. \tag{9.49}$$

Here the spacing of the points is given by

$$t_{m+1} - t_m \approx 2\pi/\dot{\theta}(t_m)$$

and the summation can therefore be replaced by the integral

$$S \approx \frac{1}{2\pi}\int_{t_1}^{t_2} \dot{\theta}(t) \exp [j\theta(t) - j\omega t]\mathrm{d}t, \tag{9.50}$$

where the limits are the times corresponding to the two ends of the transducer. Equation (9.50) can be evaluated by the stationary-phase method. Comparing with equations (9.7) and (9.12a), the magnitude of S is

$$|S| \approx \frac{\omega}{2\pi\sqrt{|\mu(T_s)|}}, \tag{9.51}$$

where $\mu(t)$ is the chirp rate, so that $2\pi\mu(t) = \ddot{\theta}(t)$, and $T_s(\omega)$ is the stationary-phase point, defined as the solution of $\dot{\theta}(T_s) = \omega$. From equation (9.48) the magnitude of the surface-wave potential is $|\phi_{s1}(\omega)| = V_t\Gamma_s|S|G_0^{1/2}$ and, using equation (9.46), the conductance of the unapodised transducer is

$$G_a(\omega) \approx \frac{\omega^3 W G_0 \Gamma_s}{4\pi^2|\mu(T_s)|}. \tag{9.52}$$

For an *apodised* chirp transducer, a simple modification can be made. The apodisation is taken to be given by a function $U(t)$, which gives the electrode overlaps at the corresponding position, so that $U(x_n/v_0) = |u_{n+1} - u_n|$. At frequency ω, the surface wave excitation arises primarily from the region corresponding to the stationary-phase point $T_s(\omega)$. Thus, if $U(t)$ varies slowly enough, the conductance is given by equation (9.52) with the aperture taken as $U[T_s(\omega)]$, so that

$$G_a(\omega) \approx \frac{\omega^3 G_0 \Gamma_s U(T_s)}{4\pi^2|\mu(T_s)|}. \tag{9.53}$$

Equations (9.52) and (9.53) are very convenient for estimating the conductance, and hence the insertion loss of a device.

The acoustic susceptance $B_a(\omega)$ is the Hilbert transform of $G_a(\omega)$, as shown by equation (4.62). It is usually the case that $G_a(\omega)$ varies slowly with ω and hence $B_a(\omega)$ is small, and can therefore be neglected.

Effective number of periods. It is helpful to compare the conductance $G_a(\omega)$ of a chirp transducer, equation (9.53), with the centre-frequency conductance $G_{au}(\omega_c)$ of a uniform transducer, equation (9.45). To clarify this, we introduce an effective number of periods, $N_{\text{eff}}(\omega)$. This is defined as the number of periods in a uniform transducer, with centre frequency ω, such that its conductance equals that of the chirp transducer at frequency ω. Both transducers are taken to have the same aperture. We

thus set equations (9.45) and (9.53) equal, and in equation (9.45) put $\omega_c = \omega$, $N_{pu} = N_{\text{eff}}(\omega)$ and $W = U(T_s)$. This gives

$$N_{\text{eff}}(\omega) = \frac{\omega}{2\pi\sqrt{|\mu(T_s)|}}. \tag{9.54}$$

With this formula, the conductance of a chirp transducer can be found straightforwardly from the centre-frequency conductance of a uniform transducer with the same value of S_e. We can interpret $N_{\text{eff}}(\omega)$ roughly as the number of periods in the chirp transducer that are actively generating surface waves at frequency ω; the remaining periods on either side of the active region behave essentially as a capacitive shunt.

In the case of a linear chirp, $|\mu|$ is a constant equal to B/T, and we find $N_{\text{eff}}(\omega)$ is related to the total number of periods N_p by

$$\frac{N_{\text{eff}}(\omega)}{N_p} = \frac{\omega}{\omega_c\sqrt{TB}}, \tag{9.55}$$

where ω_c is the centre frequency of the chirp. Thus, for $TB = 100$, say, about 10% of the electrodes are active at any particular frequency.

Capacitance and Q-factor. For a uniform transducer, the capacitance C_t is given by equation (4.117), that is, by

$$C_{tu} = WN_{pu}(\varepsilon_0 + \varepsilon_p^T)\tilde{C}_t, \tag{9.56}$$

where N_{pu} is the number of periods and $\tilde{C}_t$ is a constant depending on S_e and a/p, given in Table 4.1 and Figure 4.15. Since this is proportional to N_{pu}, we can regard the contribution of each period as being given by equation (9.56), with $N_{pu} = 1$. The same formula can be applied to each period of a chirp transducer, if the apodisation $U(t)$ is varying slowly. For period m the aperture is $U(t_m)$, where t_m is the solution of $\theta(t_m) = 2m\pi$. The capacitance of the chirp transducer is thus

$$C_t \approx (\varepsilon_0 + \varepsilon_p^T)\tilde{C}_t \sum_{m=1}^{N_p} U(t_m). \tag{9.57}$$

For an unapodised chirp transducer, the capacitance is the same as that of a uniform transducer, equation (9.56), with the same number of periods.

An important result relates the Q-factor of a chirp transducer to that of a uniform transducer with the same acoustic bandwidth. The chirp transducer is taken to be unapodised and to have a linear chirp, so that its centre-frequency conductance $G_a(\omega_c)$ is given by equation (9.52) with $|\mu| = B/T$. The centre-frequency conductance of the uniform transducer is $G_{au}(\omega_c)$, given by equation (9.45). Here N_{pu} is replaced by $\omega_c/(2\pi B)$, so that the acoustic bandwidths of the two transducers are approximately the same. Taking the apertures to be equal, the conductances are found to be in the ratio

$$\frac{G_a(\omega_c)}{G_{au}(\omega_c)} \approx TB. \tag{9.58}$$

The capacitances are given by equations (9.56) and (9.57), and since the chirp transducer has $N_p = T\omega_c/(2\pi)$ periods we have

$$C_t/C_{tu} \approx TB. \tag{9.59}$$

The Q-factor of a transducer is, by definition, $\omega_c C_t/G_a(\omega_c)$. Thus the Q-factors of the chirp and uniform transducers are the same if they have the same acoustic bandwidth. This is of course still true if the apertures are different, since the aperture does not affect the Q-factor. This important result was first noted by Hartmann *et al.* [296]. It enables the Q-factor of a linear chirp transducer to be estimated directly from the analysis for uniform transducers in Section 4.6. The Q-factor is particularly relevant when the transducer is tuned in order to minimise the conversion loss, as discussed in Section 7.1.1.

9.3.3. Transducer Design

It is assumed here that the transducer is to be designed such that its impulse response is a good approximation to some required response, given by the chirp waveform

$$v(t) = a(t) \cos [\theta(t)], \tag{9.60}$$

where $\theta(t)$ is a non-linear function of time. This waveform will be of finite length, so that $a(t)$ is zero outside the interval occupied by the waveform. If the required response is specified in the frequency domain, $a(t)$ and $\theta(t)$ can be found by transforming the spectrum to the time domain. It may then be necessary to truncate in order to obtain a finite length, and if necessary short extensions can be added to avoid abrupt discontinuities. Owing to the sampled nature of the transducer, the frequency response will include a fundamental and a series of harmonics. It is assumed here that the fundamental component is to correspond with the required response of equation (9.60). In most practical cases the harmonics are irrelevant, or are eliminated by filtering elsewhere.

The design procedure is based on some results for non-uniform sampling, derived in Section A.4 and repeated here for convenience. Suppose that the waveform $v(t)$ is sampled at times t_n, such that

$$\theta(t_n) = 2\pi n/S_e + \theta_0, \tag{9.61}$$

where θ_0 is a constant and n is an integer. Thus, synchronous sampling is assumed, and the integer S_e is the number of samples per period of the waveform. The sampled waveform is

$$v_s(t) = \sum_{n=1}^{N} v(t_n)\delta(t - t_n), \tag{9.62}$$

where N is the number of samples. It is assumed that θ_0 and N are chosen such that $v(t) = 0$ for $t < t_1$ and for $t > t_N$. Section A.4 shows that the *fundamental* componential of $v_s(t)$ is

$$\tilde{v}_s(t) = Ca(t)\,\dot{\theta}(t)\cos[\theta(t)], \tag{9.63}$$

where C is a constant. This is valid for $S_e > 2$, and for $S_e = 2$ it is valid provided θ_0 is a multiple of π.

Now consider the design of a transversal filter to given an impulse response $v(t)$. The impulse response of the transversal filter has the form

$$h_s(t) = \sum_{n=1}^{N} h_n \delta(t - t_n) \tag{9.64}$$

with t_n given by equation (9.61). The transversal filter has the form shown in Figure 8.3, though here the tap delays t_n are unequally spaced. Comparing with equations (9.62) and (9.63), it can be seen that if the coefficients h_n were equated with $v(t_n)$ the fundamental component of $h_s(t)$ would be proportional to $v(t)$ except for the amplitude distortion due to the term $\dot{\theta}(t)$. To compensate for this distortion, the tap weights must be given by

$$h_n = v(t_n)/\dot{\theta}(t_n) \tag{9.65}$$

so that $h_s(t)$ is a sampled version of a waveform $v(t)/\dot{\theta}(t)$. The fundamental component of $h_s(t)$ will then be proportional to $v(t)$, equation (9.60).

For a *transducer* required to give an impulse response $v(t)$, we consider the design in terms of gap elements, though an alternative method based on electrode elements is very similar. The frequency response is given by equation (9.43), which has the form

$$H_t(\omega) = K\omega^{1/2} \sum_{n=1}^{N} (u_{n+1} - u_n) \exp(-jk_0 x_n'), \tag{9.66}$$

where K is a constant and x_n' is the location of gap n relative to the acoustic output port. Here N is the number of gaps and u_n are the locations of the electrode breaks, as in Figure 9.8. For comparison, the frequency response $H_s(\omega)$ of the transversal filter considered above is the transform of equation (9.64), so that

$$H_s(\omega) = \sum_{n=1}^{N} h_n \exp(-j\omega t_n). \tag{9.67}$$

Comparing with equation (9.66), it is clear that the required gap locations are given by $x_n' = v_0 t_n$, where $v_0 = \omega/k_0$ is the free-surface velocity. The transducer response of equation (9.66) differs from equation (9.67) in that a term $\omega^{1/2}$ is present. In the design, this is easily allowed for by invoking the stationary-phase approximation. The response at frequency ω is mainly determined by the electrodes in the vicinity of gap n, such that $\dot{\theta}(t_n) = \omega$. Thus the $\omega^{1/2}$ term can be compensated by making the gap strength proportional to $h_n/[\dot{\theta}(t_n)]^{1/2}$, so that

$$u_{n+1} - u_n \propto v(t_n)/[\dot{\theta}(t_n)]^{3/2}. \tag{9.68}$$

A more accurate method of compensation is to transform $v(t)$ to the frequency domain, divide by $\omega^{1/2}$, and then return to the time domain. However, equation (9.68) is usually accurate enough in practice.

To give an example, consider the linear chirp waveform

$$\begin{aligned} v(t) &= a(t)\cos[\theta(t)] = a(t)\cos(\omega_1 t + \pi\mu t^2), \quad \text{for } 0 \leqslant t \leqslant T, \\ &= 0, \qquad\qquad\qquad\qquad\qquad\qquad \text{for other } t. \end{aligned} \tag{9.69}$$

This gives instantaneous frequency $\dot{\theta}(t) = \omega_1 + 2\pi\mu t$ and has bandwidth $B = |\mu|T$. From equation (9.61) the sampling times t_n are given by

$$\mu t_n = [f_1^2 + \mu\theta_0/\pi + 2\mu n/S_e]^{1/2} - f_1, \tag{9.70}$$

where $f_1 = \omega_1/(2\pi)$. The gap strengths are given by equation (9.68), so that

$$u_{n+1} - u_n \propto \frac{a(t_n)\cos[\theta(t_n)]}{(\omega_1 + 2\pi\mu t_n)^{3/2}}. \tag{9.71}$$

For $S_e = 2$ the terms $\cos[\theta(t_n)]$ all have the same magnitude, and simply cause an alternation in sign.

It is often necessary to modify the design somewhat to compensate for second-order effects, but this topic is deferred to Section 9.5.

9.4. DESIGN AND PERFORMANCE OF INTERDIGITAL DEVICES

The analysis and design of chirp transducers have been considered in Section 9.3 above. Here we consider the design of a two-transducer chirp filter. This of course pre-supposes that the required response of the device has already been determined. In practice, the starting point is usually the relevant performance requirements of the radar system, for example the output pulse width, sidelobe levels and the receiver mis-match loss. In this case the first task is to design the *waveforms*, that is, the required transmitter waveform and the impulse response of the filter in the receiver. These are designed by the methods of Section 9.2. The devices considered here may have impulse responses that are any of the types of chirp waveform considered in Section 9.2.

Single-dispersive Devices The simplest form of chirp filter is illustrated in Figure 9.8, and comprises one chirp transducer and one uniform transducer. This is called a *single-dispersive* device to distinguish if from devices using two chirp transducers, considered later. Since the single-dispersive device has one transducer unapodised, its short-circuit response is essentially the product of the two transducer responses, as given by equation (9.44). To design the device, the number of electrodes in the uniform transducer is first chosen, such that its response does not vary by more than a few dB's over the required device bandwidth. The chirp transducer is then designed by the method of Section 9.3.3, with a modification to allow for the roll-off

given by the uniform transducer. In practice it is often necessary to compensate also for a variety of second-order effects, discussed in Section 9.5. The modifications involved are usually small and can be implemented straightforwardly using the stationary-phase approximation, that is, by modifying the time-domain amplitude $a(t)$ and phase $\theta(t)$. Often, the transducers are tuned in order to reduce the conversion losses, and in this case it is usually necessary to allow for the circuit effect when designing the chirp transducer. However, for chirp transducers it is usually difficult to obtain low conversion losses, because the radiation resistance $R_a(\omega)$ is usually much less than $50\,\Omega$. This is because, at any ω, most of the electrodes act as a capacitive shunt across the active region, as seen in Section 9.3.2. This reduces the radiation resistanee, as shown in Section 7.1.2.

The substrate material is usually quartz, chosen because it gives good temperature stability and because electrode interactions are relatively weak. However, for small time-bandwidth products a strongly-piezoelectric material such as lithium niobate is often feasible, and gives lower insertion losses than quartz.

The triple-transit spurious signal can be reduced if necessary by adjusting the terminal impedances "seen" by the transducers. However, it is unusual for the triple-transit signal to cause problems in chirp devices. From the basic nature of the device it is easily seen that, in the *time* domain, the triple-transit response gives a chirp waveform similar to the main response, but stretched in time by a factor of 3. When the device is used as a matched filter the triple-transit response is not matched to the input waveform, and this factor discriminates against the triple-transit response.

Although the device shown in Figure 9.8 is a down-chirp, the device may equally well be an up-chirp, so that the frequency of its impulse response increases with time. This makes no difference to the basic principles. However, for wide-band devices there can be difficulties due to the electrodes scattering surface waves into bulk waves, as explained in Appendix F. For this reason, up-chirp devices are usually constrained to fractional bandwidths less than about 25%, and the bulk wave scattering is then negligible.

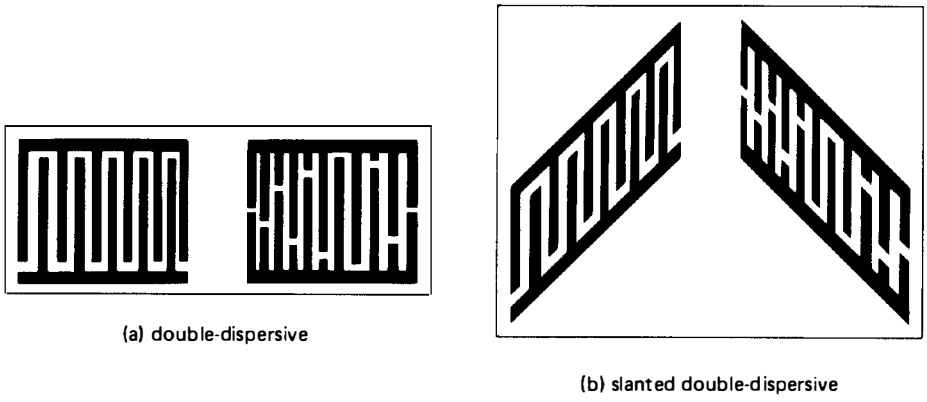
In many radar systems a matched pair of chirp devices is required, one to serve as an expander in the transmitter, generating the chirp waveform, and the other to serve as the compressor in the receiver. Usually, one device will be an up-chirp and the other will be a down-chirp. In fact, the chirp polarity can be reversed in the receiver by using a local oscillator with frequency above the band occupied by the R.F. input signal; for this case the two filters will have the same chirp polarity. The expansion process is most efficient if the expander is excited by a short rectangular pulse of carrier, whose length τ is such that its bandwidth is compatible with the device bandwidth. It is usual to design the expander to allow for the roll-off in the spectral amplitude of this input pulse. An important factor is the loss of power due to the expansion process. For a flat-envelope linear chirp it can be shown that, in comparison with C.W. excitation, the expansion process reduces the output power in the ratio of approximately $\tau^2 B/T$, provided $\tau \leqslant 1/B$. Typically, $\tau \approx 1/B$, giving an "expansion loss" of 10 log (TB) dB. This must be added to the C.W. insertion loss to give the total loss. For large TB the total loss can be prohibitive, and then the transmitter must use active rather than passive generation.

Double-dispersive Devices. For weakly-piezoelectric substrates, such as quartz, the single-dispersive device is generally suitable only if the fractional bandwidth is small. This is due to the difficulty of matching uniform transducers over a wide bandwidth. For example, on *ST*, *X* quartz a matched uniform transducer cannot give more than about 5% bandwidth, as noted in Section 7.1.1; for wider bandwidths the loss is substantial unless a complex matching network is used.

In this situation, a *double-dispersive* design is usually preferrable. As shown in Figure 9.9(a), this design has two chirp transducers. Usually, each transducer contributes about half of the dispersion required of the overall device. One of the transducers is unapodised, so that the device response is given by the product of the transducer responses. Most of the earlier remarks on second-order effects and choice of materials, referring to single-dispersive devices, apply here also.

To design a double-dispersive device, the unapodised transducer is designed first. Typically, the dispersion required for the overall device is halved, and the corresponding phase $\theta_u(t)$ for the unapodised transducer is deduced. The sampling points are then obtained from $\theta_u(t)$ in the usual way. The form of $\theta_u(t)$ is not at all critical, though it is necessary to ensure that the transducer has adequate bandwidth. Once this transducer has been designed, its frequency response can be calculated and divided into the required device response in order to obtain the response required of the apodised transducer. The latter can then be designed by the method of Section 9.3.3.

Figure 9.9(b) shows a *slanted* double-dispersive device [297–299]. Here, waves generated at any point in one transducer travel through relatively few electrodes before emerging on the free surface. Consequently, second-order effects are reduced and, for an up-chirp device, the bulk-wave scattering problem mentioned earlier does not occur. Also, since waves of different frequencies traverse the substrate at different locations, there is the possibility of adding a "phase plate" to correct for phase errors, as done in RAC's (Section 9.6). However, compared with other chirp devices, the



(a) double-dispersive

(b) slanted double-dispersive

FIGURE 9.9. Other types of interdigital chirp filter.

slanted device makes less effective use of the substrate area and is more difficult to fabricate.

Another interesting variant is obtained if one transducer of the double-dispersive device, Figure 9.9(a), is reversed end to end. In this case the device will have one down-chirp transducer and one up-chirp. If the frequency-time curves for the two transducers are the same except for the time-reversal, the overall device will be non-dispersive even though the transducers are dispersive. This configuration can be used effectively to realise a bandpass filter with linear phase, as already mentioned in Section 8.4.5, and for this application the design techniques of Chapter 8 are applicable.

Performance. Interdigital devices typically have bandwidths up to about 500 MHz and dispersion T up to 50 μsec, the latter being limited by the length of substrate required. For large time-bandwidth products the insertion loss can be excessive and errors due to second-order effects can be unacceptable, and consequently the TB-product is generally limited to 1000. Insertion losses (for C.W. signals) are typically 30 to 60 dB, larger values being obtained for larger TB-products and for substrates with weaker piezoelectric coupling.

The accuracy achievable using surface-wave devices is well illustrated by the fact that the time-sidelobes of the compressed output waveform can be as low as 45 dB below the main peak. It will be seen in Section 9.5.1 that this implies a phase accuracy, in the frequency domain, of only a few degrees, while the phase itself varies by typically several thousand degrees over the device bandwidth. The accuracy required becomes more stringent as the required sidelobe level is reduced, but even at low sidelobe levels the performance can be quite close to the theoretical ideal given by the analysis in Section 9.2: sidelobe levels are typically within a few dB's of the ideal, and the output pulse width and the processing gain are also close to ideal. In practice the sidelobe level is not usually minimised because a low sidelobe level implies, even in the ideal case, a relatively wide output pulse and both factors are relevant to the system performance. Typically, the sidelobe suppression is 25 to 40 dB.

A typical example is shown in Figure 9.10, which refers to a matched pair of devices. Here the chirps are linear, with 11 MHz bandwidth and 14 μsec duration, and the compressor has amplitude weighting to suppress the time-sidelobes. The upper photograph shows the two devices, mounted in one package to minimise errors due to temperature differences. Both devices are of the double-dispersive type, with ST, X quartz substrates. The lower photograph on the left shows the expanded pulse, obtained by applying a short rectangular pulse to the expander. The lower right photograph, with the same time scale, is the compressor output when the expanded pulse is applied to it, simulating the output in a radar system for a point target. The width of the output pulse, at the 3 dB points, is 120 nsec, and the largest sidelobe (not visible in the figure) is 35 dB below the peak.

Table 9.1 shows performance data for a variety of chirp devices, and illustrates the wide range of parameters obtainable. For example, bandwidths in the table range from 2.4 to 500 MHz, while the dispersion T ranges from 0.5 to 34 μsec. The first column refers to the devices of Figure 9.10.

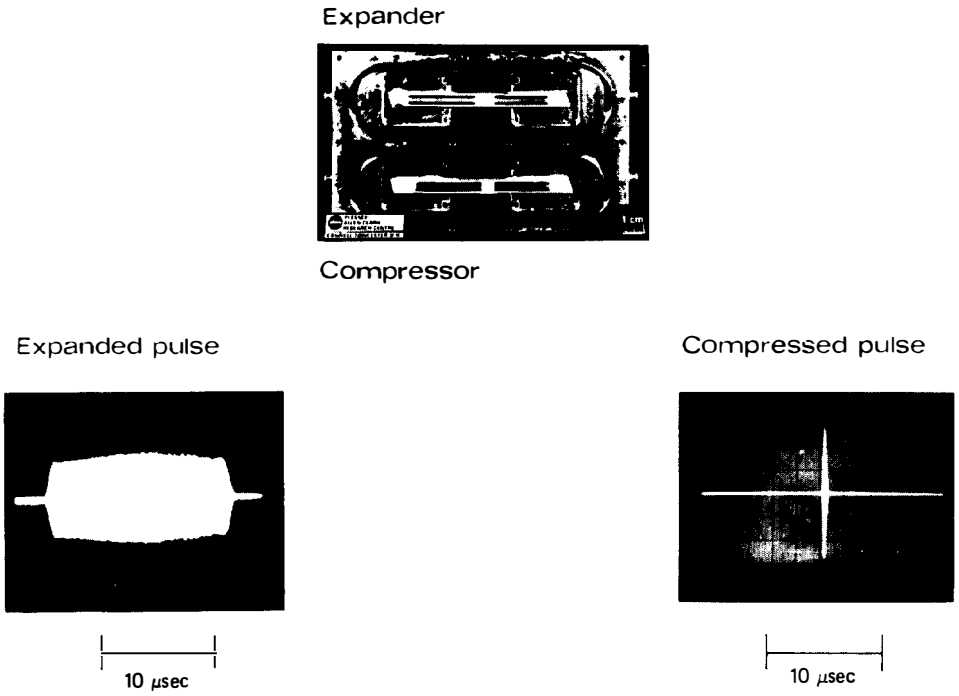


FIGURE 9.10. Performance of matched pair of interdigital chirp filters. (Courtesy Plessey Research)

Devices for Frequency-scanning Radar. The last column on Table 9.1 refers to a chirp filter for a frequency-scanning radar, which uses pulse compression in novel way [303–305]. The radar antenna here is a phased array, with the elements driven from a slow-wave structure with the input port at one end. Owing to the slow-wave structure, the relative phases of the signal at the array elements vary quite rapidly with frequency, and this feature is used to form a narrow transmitted beam whose direction depends on frequency. Thus the beam can be scanned rapidly by electronic means. In the particular example here, the elevation scan is achieved by applying a single chirp pulse. For a point target, the returned echo has a centre frequency indicating its elevation. The echo is essentially a short segment of the transmitted waveform, with duration given by the width and sweep rate of the beam at the target elevation. Since this signal is a chirp waveform, processing gain can be obtained by applying it to a matching chirp filter in the receiver. However, the chirp filter must cover the entire bandwidth of the transmitted waveform in order to process echos from targets at all elevations, and in the present case this implies a very large time-bandwidth product of several thousand. The chirp filter also compensates for range errors due to the fact that beams with different elevation are radiated at different times.

TABLE 9.1

Performance examples for interdigital chirp filters. Data refer to a matched pair of devices, unless noted otherwise

Centre frequency (MHz)	70	60	30	60	300	30	300	750	75
Bandwidth B (MHz)*	11	12	8	4	100	2.4	100	500	27
Duration T (μs)	14	8	4	12	5	34	10	0.5	28
TB	153	96	20	49	500	81	1000	250	737
Chirp type†	L	L	L	NL	L	L	L§	L§	NL§
Time-sidelobe rejection (dB)	35	40	20	45	32	32	20	13	–
Substrate‡	Q	Q	LN	Q	Q	Q	LN	LN	Q
Reference	See text	[300]	[30?]	[300]	[293]	[293]	[301, 302]	[338]	See text

* Range of instantaneous frequencies.

† L = linear; NL = non-linear.

‡ Q = quartz; LN = lithium niobate

§ Data refer to single device, sidelobe rejection is for ideal input waveform.

Owing to the large time-bandwidth product required the receiver used seven interdigital chirp filters, multiplexed in frequency. The data in the last column of Table 9.1 refer to the device with the largest time-bandwidth product. The instantaneous frequency for this device was a very non-linear function of time, but the device nevertheless gave a frequency response with phase accurate to within $\pm 10^\circ$ of the response required.

9.5. SECOND-ORDER EFFECTS IN CHIRP FILTERS

A variety of second-order effects are significant in surface-wave chirp filters, and it is important to consider not only the resulting distortion of the device response but also the effect of this distortion on the output waveform of a pulse-compression radar system. The main practical consequences are degradations in the time-sidelobe levels, the pulse width and the signal-to-noise ratio. In many cases, these degradations are primarily associated with slowly-varying phase errors in the frequency response of the compressor or expander, or both. Such errors arise, for example, if there is a temperature change, or if the surface-wave velocity differs from its expected value due to misalignment during fabrication. Similar effects are caused by doppler shifts.

In order to assess these effects, we first consider the effect of a slowly-varying phase error on the output waveform, without considering the physical origin of the error. This is done in Section 9.5.1. The error is expressed as a sum of Legendre polynomials, and it is shown that individual polynomials have quite distinctive effects on the output waveform. It is often the case that errors due to specific second-order effects can be conveniently expressed in terms of these polynomials, and hence the corresponding distortion of the output waveform is readily assessed. This is illustrated in Section 9.5.2, which describes the effects of velocity errors and temperature changes. Doppler shifts are considered in Section 9.5.3, which uses the analysis of Section 9.5.1 for the case of non-linear chirps.

Nearly all of the material in Sections 9.5.1 to 9.5.3 is applicable to reflective array compressors, described later in Section 9.6, as well as to interdigital chirp filters. However, some further second-order effects, specific to interdigital devices, are considered in Section 9.5.4.

9.5.1 Effect of Phase Errors on Compressor Output Waveform

It is assumed here that an ideal chirp waveform is applied to a compressor designed to give an output waveform with well-suppressed time-sidelobes, as described in Sections 9.2.3 and 9.2.4. However, here the compressor is taken to be imperfect, so that its frequency response is $H'(\omega)$ instead of the ideal response $H(\omega)$. If $S(\omega)$ is the spectrum of the input waveform, the ideal output waveform has a spectrum $G(\omega) = S(\omega)H(\omega)$, while the actual output waveform has a spectrum $G'(\omega) = S(\omega)H'(\omega)$. Thus

$$G'(\omega) = G(\omega)H'(\omega)/H(\omega). \tag{9.72}$$

In practice, phase errors in the compressor response are often much more significant than amplitude errors, and the latter will therefore be ignored here. The phase error will be denoted by $\Delta\phi(\omega)$, so that $H'(\omega)/H(\omega) = \exp[j\Delta\phi(\omega)]$, and the distorted output spectrum becomes

$$G'(\omega) = G(\omega) \exp[j\Delta\phi(\omega)]. \tag{9.73}$$

Here the ideal output spectrum $G(\omega)$ is proportional to $\bar{W}(\omega - \omega_c)$, where $\bar{W}(\omega)$ is the weighting function used to suppress the time-sidelobes, as explained in Sections 9.2.3 and 9.2.4. Note that the input waveform and the ideal filter response need not be considered explicitly. For example, for a given $G(\omega)$ equation (9.73) gives the output distortion irrespective of whether the chirps are linear or non-linear.

Surface-wave devices often exhibit a slowly-varying phase error, considered below, with a rapidly-varying error superimposed. The latter is usually due to either surface wave reflections or inaccuracy in the locations of electrodes or grooves. An indication of the effect of rapid errors is given by "paired-echo" theory [277, 279]. This shows that a small sinusoidal error of the form $\Delta\phi(\omega) = k \sin \omega\tau$, where k and τ are constants, causes spurious time-sidelobes to appear at times $\pm\tau$ relative to the main peak, with relative amplitude $k/2$. However, in practice the rapidly-varying error usually has an irregular form, and hence the paired-echo theory is of limited value. In addition, rapidly-varying errors do not generally cause significant degradations of the output waveform, so they will not be considered further here.

Slowly-varying Phase Errors. Slowly-varying errors can be expressed as simple polynomial functions. In order to distinguish the distortions arising in the compressed output waveform it is important to use orthogonal polynomials, and a suitable choice is the Legendre polynomials $P_n(x)$ defined in Appendix C. These are orthogonal over the interval $-1 \leqslant x \leqslant 1$. The phase error is therefore written as

$$\Delta\phi(\omega) = \sum_{n=0}^{\infty} \Phi_n P_n(x), \tag{9.74}$$

where Φ_n are constant coefficients and

$$x = (\omega - \omega_c)/(\pi B). \tag{9.75}$$

This choice for x makes the polynomials orthogonal over the compressor band, that is, over the interval $|\omega - \omega_c| \leqslant \pi B$. The first four polynomials are $P_0(x) = 1$, $P_1(x) = x$, $P_2(x) = (3x^2 - 1)/2$, and $P_3(x) = (5x^3 - 3x)/2$. The orthogonality relation is

$$\begin{aligned} \int_{-1}^{1} P_n(x)P_m(x)\,dx &= (n + \tfrac{1}{2})^{-1}, \quad \text{for } m = n, \\ &= 0, \qquad\qquad \text{for } m \neq n. \end{aligned} \tag{9.76}$$

Given some phase error $\Delta\phi(\omega)$, this can be used to evaluate the coefficients Φ_n in equation (9.74); $\Delta\phi(\omega)$ is multiplied by $(n + \frac{1}{2})P_n(x)$, and the integral with respect to x from -1 to 1 gives Φ_n. Alternatively, least-squares fitting gives the same results.

Generally, the terms with $n > 3$ in equation (9.74) are found to be of little significance, so here we shall only consider the first four terms. These give rise to distortions of the output waveform in the vicinity of the main peak, and the forms of distortion are quite distinct. The first term, $\Phi_0 P_0(x)$, is a constant, and is rarely of any consequence. The second term, $\Phi_1 P_1(x)$ is proportional to $(\omega - \omega_c)$, and corresponds to a delay $\tau = -\Phi_1/(\pi B)$ of the output waveform. It should be noted that this delay is valid irrespective of the presence of any other terms in the expansion.

To see the effect of a *quadratic* error, suppose that only the $n = 2$ term is present in equation (9.74), so that $\Delta\phi(\omega) = \Phi_2 P_2(x)$. This function is shown on the left in Figure 9.11. The coefficient Φ_2 is the phase error at the band edges. The consequent distortion of the output waveform can be found by using equation (9.73) and transforming to the time domain. The effect of the error depends on the nature of the ideal output waveform, and has been calculated by Klauder *et al.* [277] for a typical case with an ideal time-sidelobe suppression of 40 dB. The error is found to broaden the main peak and reduce its amplitude, though the levels of the sidelobes are hardly affected. The right part of Figure 9.11 shows the 6 dB width of the peak as a function of Φ_2 and also the reduction of output signal-to-noise ratio; the latter follows directly from the peak amplitude since the output noise level is not affected by the phase error.

For a *cubic* error, suppose that only the $n = 3$ term is present in equation (9.74), so that $\Delta\phi(\omega) = \Phi_3 P_3(x)$. As shown in the upper part of Figure 9.12, this gives a phase error $\pm\Phi_3$ at the band edges. The effect of this on the output waveform is shown in the lower left part of the figure, which shows the envelope of the output for $\Phi_3 = 0$ and for $\Phi_3 = 22.4°$. In this case Hamming weighting has been chosen as an example, so that the ideal output spectrum $G(\omega)$ in equation (9.73) is taken as $\bar{W}_H(\omega - \omega_c)$, equation (9.26). It can be seen that the error causes an unwanted sidelobe to appear before the main peak, and also distorts the shape of the peak. However, the peak broadening is negligible, and there is no significant delay or loss of amplitude. The plot on the lower right in Figure 9.12 shows the spurious sidelobe level as a function of Φ_3.

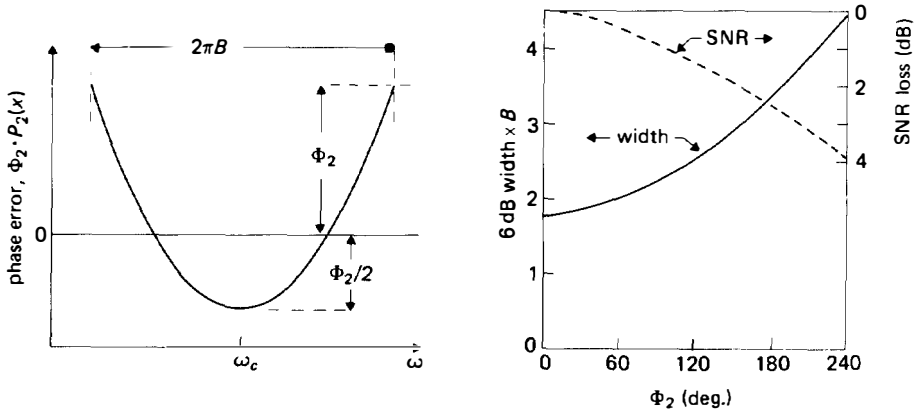


FIGURE 9.11. Effect of quadratic phase error on compressed output waveform. Ideal output has −40 dB sidelobes. From data given by Klauder *et al.* [277].

Note that quadratic and cubic errors have been considered individually, assuming other terms to be absent. In practice it is often the case that one of these two terms is dominant, and the above conclusions are then valid. If both terms are present with comparable magnitude, the above conclusions will not in general be quantitatively correct.

9.5.2. Velocity Errors and Temperature Effects

These factors cause slowly-varying phase errors in the frequency response of the compressor, and the resulting distortion of the compressed output waveform is readily assessed using the method of Section 9.5.1 above. Velocity errors are caused by, for example, misalignment during fabrication (Section 6.4) or the presence of electrodes (Appendix D). It is assumed here that the velocity error is independent of frequency and is uniform over the surface of the substrate, though these assumptions are not always valid.

It was shown in Section 6.4 that a velocity error or temperature change scales the phase of the device frequency response, so that

$$\phi'(\omega) = \phi[\omega(1 + \varepsilon)], \tag{9.77}$$

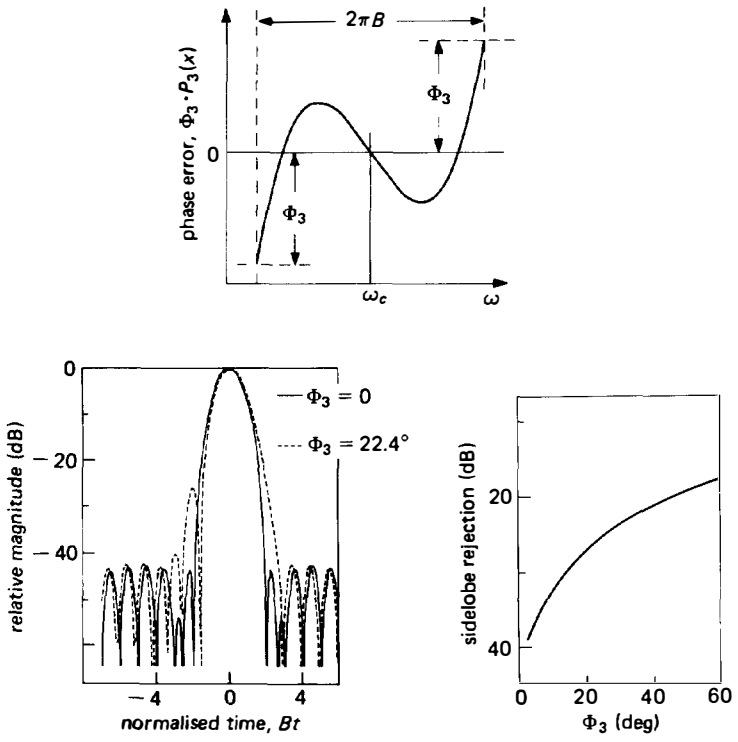


FIGURE 9.12. Effect of cubic phase error on compressed output waveform. Ideal output waveform is Hamming weighted.

where $\phi(\omega)$ and $\phi'(\omega)$ are respectively the ideal and actual phase. There is also a small amplitude change, but this is usually insignificant and is neglected here. The quantity ε is small. For a velocity error, $\varepsilon = -(v' - v)/v$, where v and v' are the ideal and actual velocities. For a temperature change, $\varepsilon = (T' - T)/T$, where T and T' are the delays between two points at the "ideal" temperature and at the actual temperature

As in Section 9.5.1, the spectrum of the compressed output waveform is $G'(\omega) = G(\omega) \exp [j\Delta\phi(\omega)]$, and here $\Delta\phi(\omega) = \phi'(\omega) - \phi(\omega)$. Since $\varepsilon \ll 1$, we obtain from equation (9.77)

$$\Delta\phi(\omega) \approx \varepsilon\omega \frac{d\phi(\omega)}{d\omega} = -\varepsilon\omega\tau_g(\omega), \tag{9.78}$$

where $\tau_g(\omega)$ is the ideal group delay of the compressor. The phase error therefore depends on the dispersion curve of the device.

(a) Linear Chirps. If the chirp is linear, the ideal group delay of the compressor is approximately

$$\tau_g(\omega) \approx (\omega - \omega_c)/(2\pi\mu) + \tau_c. \tag{9.79}$$

This equation is obtained from the stationary-phase approximation, neglecting the ripples in the response. The parameter τ_c is the delay at the centre frequency ω_c. Clearly, the phase error $\Delta\phi(\omega)$ of equation (9.78) is quadratic; it corresponds to a change of chirp rate from μ to μ', where $\mu' = \mu/(1 + 2\varepsilon)$, and also to a change of the delay pedestal. If the phase error is written as a sum of Legendre polynomials, as in equation (9.74), the coefficients Φ_1 and Φ_2 are found to be

$$\Phi_1 = -\varepsilon B[\pi\tau_c + \tfrac{1}{2}\omega_c/\mu]. \tag{9.80}$$

and

$$\Phi_2 = \mp \tfrac{1}{3}\pi\varepsilon TB, \tag{9.81}$$

where the upper sign applies for $\mu > 0$ and the lower sign for $\mu < 0$. The linear term, with coefficient Φ_1, causes a delay equal to $\varepsilon[\tau_c + \omega_c/(2\pi\mu)]$, while the quadratic term broadens the output pulse and reduces the SNR, as shown for example in Figure 9.11. Usually the quadratic term is more significant, and a typical requirement would be that Φ_2 should not exceed $\pi/2$. This implies that $|\varepsilon| \leqslant 3/(2TB)$, and hence the tolerances on velocity and temperature errors become more stringent as TB is increased.

Generally, these tolerances are not difficult to meet. For devices on ST, X quartz substrates, the temperature stability is generally sufficient for operation over a range of 100°C or more. For Y, Z lithium niobate the range is more restricted; for example, if $TB = 200$, temperature changes need to be restricted to about $\pm$ 20°C. Note that, since velocity changes and temperature changes cause similar effects, a

velocity change can be compensated for by changing the device temperature, provided the response is sufficiently sensitive to temperature. This is often feasible for lithium niobate substrates.

For interdigital devices on lithium niobate substrates, difficulties can occur due to the relatively large velocity perturbation, typically 1%, due to electrical loading. A repeatable velocity error can be compensated for by scaling the transducer geometry; however, the velocity depends on the electrode widths, and for large *TB*-products the required tolerance on the widths can be difficult to achieve. In addition, the variation of electrode pitch along the transducer causes the fractional error in the widths to vary, since narrower electrodes are generally more subject to fabrication errors than wider ones. This causes the device phase error to have a cubic component, and hence the output waveform has a spurious sidelobe which may be unacceptable [302]. This type of error cannot be compensated for by changing the temperature.

(*b*) *Non-linear Chirps.* The methods described above can be applied to non-linear chirps, but for brevity the details will not be given here. For a temperature change or a non-dispersive velocity change, the phase error is given by equation (9.78), and for Hamming weighting the group delay $\tau_g(\omega)$ is given by equation (9.34). The phase error is found to have a cubic component, giving rise to a spurious sidelobe. In consequence, a non-linear chirp is found to be much less tolerant of errors than a linear chirp.

A study of phase errors in non-linear chirps, including surface-wave dispersion due to mechanical loading, is reported by Morgan and Deacon [306].

(*c*) *Matched Pairs of Devices.* In many radar systems a matched pair of chirp filters is used, with one device serving as the expander and the other as the compressor, as explained in Section 9.1. In this case the output waveform is affected by errors in both devices. Usually one device will be an up-chirp and the other a down-chirp, and this has the useful feature that much of the phase error is cancelled out. If the ideal group delays of the expander and compressor are respectively $\tau_e(\omega)$ and $\tau_c(\omega)$, then $\tau_e(\omega) + \tau_c(\omega) = \tau_0$, where τ_0 is a constant equal to the sum of the group delays at the centre frequency. If there is a temperature change or a non-dispersive velocity change, and if the change is the same for both devices, the phase error in either device is given by equation (9.78); thus the total phase error is $\Delta\phi(\omega) = -\varepsilon\omega\tau_0$. This simply gives a delay $\varepsilon\tau_0$, without distorting the output waveform. The velocities or temperatures of the two devices may of course be different, in which case the methods described above give the differences allowable if the distortion of the output waveform is to be kept within prescribed limits. Temperature differences can be minimised by locating the devices close to each other, as illustrated in Figure 9.10, and this usually eliminates the need for temperature regulation, even when lithium niobate substrates are used.

9.5.3. Doppler Shifts

A doppler shift, due to motion of the target relative to the radar system, causes phase errors rather similar to those discussed above and thus generally tends to increase the

time-sidelobe levels and to reduce the output SNR. These effects are often significant in practical systems, and must therefore be considered when the transmitted chirp waveform is designed. Here we consider the effects for linear and non-linear chirps, showing that the latter are generally more sensitive to doppler shifts.

It is usual to analyse doppler effects by assuming that each frequency component of the receiver input waveform is shifted upwards in frequency by an amount ω_d, the doppler frequency. Thus, if $S(\omega)$ is the spectrum of the input waveform for zero doppler, the spectrum for non-zero doppler is $S'(\omega)$, given by

$$S'(\omega) = S(\omega - \omega_d), \quad \text{for } \omega > 0, \tag{9.82}$$

where the doppler frequency is

$$\omega_d = 2V_r\omega_0/c. \tag{9.83}$$

Here ω_0 is the centre frequency of the transmitted waveform, at R.F., and c is the velocity of light. V_r is the radial velocity of the target relative to the radar, taken to be positive if the range is decreasing. This representation is actually an approximation. Strictly [279, p. 64], the doppler effect contracts or expands the waveform along the time axis, causing the spectrum to expand or contract along the ω-axis. However, the above equations usually give a very good approximation because the waveform bandwidth is much less than the R.F. centre frequency ω_0. Moreover, equation (9.82) is convenient because it is valid at the filter input, in the I.F. section of the receiver, as well as at the receiver input, which is at R.F. As a numerical example, if the R.F. centre frequency is $\omega_0/(2\pi) = 10\,\text{GHz}$, an aircraft target with a radial velocity of Mach 1, about 330 m/sec, gives a doppler frequency ω_d of about $2\pi \times 22\,\text{kHz}$.

For zero doppler, the filter output waveform $g(t)$ has a spectrum $G(\omega) = S(\omega)H(\omega)$, where $H(\omega)$ is the filter frequency response. For non-zero doppler, the output spectrum is $G'(\omega) = S'(\omega)H(\omega)$. Thus, using equation (9.82) we have

$$G'(\omega) = G(\omega)\frac{S(\omega - \omega_d)}{S(\omega)}, \quad \text{for } \omega > 0. \tag{9.84}$$

Thus the actual output spectrum is the ideal spectrum multiplied by a complex function $S(\omega - \omega_d)/S(\omega)$. We define $A(\omega)$ and $\phi(\omega)$ as the amplitude and phase of $S(\omega)$, so that $S(\omega) = A(\omega)\exp[j\phi(\omega)]$. Equation (9.84) thus becomes

$$G'(\omega) = G(\omega)\frac{A(\omega - \omega_d)}{A(\omega)}\exp[j\Delta\phi(\omega)], \tag{9.85}$$

where $\Delta\phi(\omega) = \phi(\omega - \omega_d) - \phi(\omega)$. In practice the distortion usually arises primarily from the phase error $\Delta\phi(\omega)$, and since the doppler frequency ω_d is small for practical cases the phase error can be well approximated by

$$\Delta\phi(\omega) = -\omega_d\frac{d\phi(\omega)}{d\omega} \approx \omega_d T_s(\omega), \tag{9.86}$$

where $\phi(\omega)$ is the ideal phase. Here $T_s(\omega)$ is the stationary-phase point of the ideal input waveform, which can be equated with $-\mathrm{d}\phi(\omega)/\mathrm{d}\omega$ in the stationary-phase approximation, as shown by equation (9.13). If the input waveform is generated by impulsing an expander, the term $-\mathrm{d}\phi(\omega)/\mathrm{d}\omega$ is equal to the group delay $\tau_g(\omega)$ of the expander.

Doppler Shift for Linear Chirps. For a linear chirp, the spectral phase $\phi(\omega)$ is, in the stationary phase approximation, a quadratic function of ω. The phase $\phi(\omega)$ is given by equation (9.16b) and $T_s(\omega)$ is given by equation (9.15). Using either of these in equation (9.86), equation (9.85) gives

$$G'(\omega) \approx G(\omega) \frac{A(\omega - \omega_d)}{A(\omega)} \exp\left[\frac{j\omega_d(\omega - \omega_c)}{2\pi\mu}\right], \tag{9.87}$$

where μ is the chirp slope of the ideal waveform. For small doppler shifts the amplitude term here is insignificant, while the phase error is equivalent to a delay

$$\tau \approx -\omega_d/(2\pi\mu). \tag{9.88}$$

Thus the doppler shift simply causes a delay of the output waveform, without appreciably distorting it. This is valid irrespective of whether the input waveform or the filter response have amplitude weighting.

If the delay error due to doppler is acceptable it is necessary to consider the amplitude distortion in equation (9.87), and the effect of this depends on the nature of any amplitude weighting in the input waveform or filter response. For example, if the waveform and the filter impulse response both have flat envelopes it is found [279, p. 134] that the amplitude of the output peak is proportional to $1 - |\omega_d|/(2\pi B)$. This result is intuitively reasonable, since the linear phase error does not affect the peak amplitude, while the bands occupied by the input waveform and the filter response overlap by an amount $2\pi B - |\omega_d|$. If, for example, a 1 dB reduction of output amplitude is acceptable, this corresponds to a doppler frequency of about 11% of the bandwidth. In this case, the output SNR is also reduced by 1 dB, since the noise output power is not affected by the doppler shift.

Doppler Shift for Non-linear Chirps. For a non-linear chirp the phase error $\Delta\phi(\omega)$ of equation (9.86) is more complicated because $T_s(\omega)$ is a non-linear function of ω. To illustrate the effect, we consider the representative case of a Hamming-weighted chirp. For this case the stationary-phase point of the compressor response is given by equation (9.34), and the negative of this is equal to $T_s(\omega)$ in equation (9.86). Equation (9.34) includes a term $\sin(\pi x)$, where $x = (\omega - \omega_c)/(\pi B)$ as before. This is expressed as a sum of Legendre polynomials, so that the phase error of equation (9.86) has the form $\Delta\phi(\omega) = \sum \Phi_n P_n(x)$, as in equation (9.74). Apart from the constant term, the values of the first few coefficients Φ_n are found to be

$$\Phi_1 = \mp 0.6295\omega_d T,$$
$$\Phi_2 = 0,$$
$$\Phi_3 = \pm 0.1570\omega_d T, \tag{9.89}$$

where the upper signs apply when the compressor is an up-chirp, and the lower signs for a down-chirp.

Since $\Phi_2 = 0$, the doppler shift causes no appreciable pulse broadening or loss of SNR. However, the linear term causes a delay $\tau = -\Phi_1/(\pi B)$, which is proportional to ω_d. The cubic term gives a spurious sidelobe, whose level can be obtained from the plot in Figure 9.12. The delay and the sidelobe rejection are shown in Figure 9.13, as functions of ω_d. The delay τ is normalised by expressing it as the product τB. Typically, the sidelobe rejection required would be, say, 30 dB or more, requiring $|\omega_d|T < 1.5$. Thus, if $TB = 100$ for example, the doppler shift must not exceed about $\frac{1}{4}\%$ of the bandwidth B. This is much more stringent than the requirement for linear chirps which, as seen above, can tolerate a doppler shift of typically 10% of the bandwidth. The delay error for a non-linear chirp is usually of no consequence because it is less than the resolution of the system, which is typically $1.5/B$.

Accurate simulations of doppler effects are reported by Newton [295] and by Butler [300]. For these simulations the TB-products were less than 100 and the weighting functions were not the same as the Hamming function, so the analysis give above is, strictly speaking, not applicable. Nevertheless, there is quite reasonable agreement with the results of Figure 9.13.

9.5.4. Other Second-order Effects in Interdigital Devices

As in the case of bandpass filters, chirp filters are usually designed taking account of second-order effects, and it is usual to include a design stage compensating for errors observed experimentally. Several types of error are conveniently compensated for by using the stationary phase approximation, that is, amplitude and phase errors at a particular frequency are compensated by changing the apodisation and position of the electrodes at the corresponding location. This applies to velocity errors, considered above, and to errors due to the circuit effect and surface-wave attenuation.

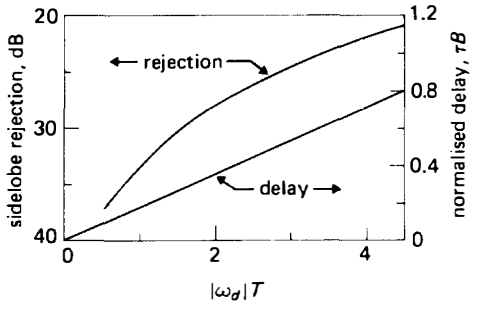


FIGURE 9.13. Effect of doppler shift for Hamming weighted non-linear chirps, in the stationary-phase approximation.

Surface-wave diffraction is often significant, though not if the substrate material is *Y*, *Z* lithium niobate, which is a minimal-diffraction orientation. Here again, the stationary-phase approximation gives a useful simplification; for a particular region of the transducer, corrections can be made by considering diffraction only at the frequency equal to the local instantaneous frequency [307]. Diffraction is described in Section 6.2, and design techniques for compensation are discussed in Section 8.4.3.

In long chirp devices, transmission-line effects can be significant [308]. As in a conventional two-wire transmission line, the transducer bus-bars have a distributed inductance L and capacitance C per unit length, and the velocity for electromagnetic waves is $(LC)^{-1/2}$. Typically, L is about $0.6\,\mu$H/m. The capacitance C is primarily associated with the electrodes, and can be large enough to significantly reduce the velocity. The transmission-line effect causes the voltage across the bus-bars to vary along the length of the transducer, and this is sometimes significant even at relatively low frequencies, below 100 MHz. Usually, the transmission-line effect can be adequately suppressed by limiting the transducer aperture, thus minimising the capacitance, though diffraction effects become unacceptable if the aperture is made too small.

9.6. REFLECTIVE ARRAY COMPRESSORS

The Reflective Array Compressor, or RAC, is an alternative type of surface-wave chirp filter, introduced by Williamson and Smith [286, 309–311]. The device relies for its operation on reflection of surface waves by large arrays of grooves, and thus uses physical principles quite different from the interdigital devices described above. In comparison with interdigital devices, RAC's enable much larger time-bandwidth products, up to about 10,000, to be obtained, though the fabrication procedures needed are somewhat more complex and time-consuming.

Section 9.6.1 below discusses the basic nature of the RAC, and makes a more detailed comparison with interdigital devices. Section 9.6.2 considers the analysis and design, including second-order effects, and concludes with a summary of performance. In Section 9.6.3 some modification of the basic form of RAC are discussed, including the use of different types of reflector and different geometries.

9.6.1. Basic Principles

Figure 9.14 shows schematically the commonest type of RAC, using arrays of grooves. Surface waves are generated by a uniform interdigital transducer at one end, and are then reflected by two identical arrays of grooves. Each array reflects the waves through 90°, so that they arrive at a uniform output transducer at the same end as the input. The groove depth is quite small, typically 1% of the wavelength, and consequently the reflection coefficient of any one groove is small. However, at any frequency in the band of the device the overall reflection coefficient is much larger than that of one groove, because reflected waves from many grooves are added coherently. This condition is obtained when the pitch p of the grooves, measured in the propagation direction of the incident waves, is equal or nearly equal to the surface-wave wavelength. In the RAC, the pitch is varied along the length of the

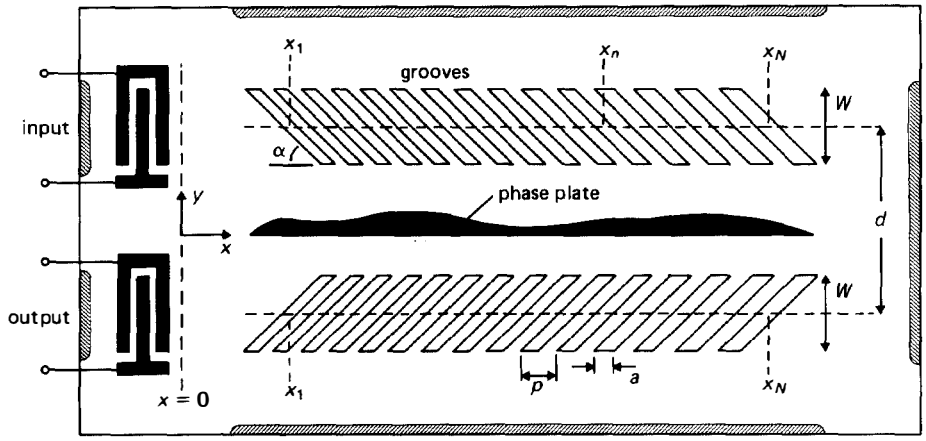


FIGURE 9.14. Reflective Array Compressor (RAC).

device, and consequently waves with different frequencies are reflected at different locations; thus the path length, and hence the delay, varies with frequency. The substrate material is usually *Y*, *Z* lithium niobate, where the orientation notation specifies that the input transducer generates waves directed along the *Z*-axis. This material is favoured because RAC's are particularly suited to large time-bandwidth requirements, where interdigital devices are inapplicable, and for these cases low diffraction and low propagation loss are particularly desirable.

The use of graded-pitch reflective arrays to form dispersive devices was proposed by Sittig and Coquin [312], though not in quite the same configuration as the RAC. A closely related device is the IMCON described by Martin [313], which emerged a few years before the RAC. In this case the wave involved is a non-dispersive bulk shear wave propagating in a parallel-sided steel strip. The two groove arrays are etched on one side of the strip and have the same basic configuration as in the RAC, so that each array reflects the waves through 90°. The waves are launched and received by means of parallel-sided piezoelectric plate transducers bonded on to the end of the strip. The thickness of the strip must be less than about one acoustic wavelength in order to avoid excitation of higher modes, and consequently fabrication considerations limit the IMCON to relatively low frequencies and bandwidths. However, the IMCON can be substantially longer than the RAC, giving larger time dispersion, and large *TB*-products can be obtained.

As for interdigital devices, amplitude weighting is often necessary in order to suppress the time-sidelobes of the compressed waveform, and to compensate for amplitude variations arising from several causes, for example, the varying groove spacing. In the RAC, amplitude weighting is obtained by varying the depth of the grooves, exploiting the fact that shallow grooves give a reflection coefficient approximately proportional to the depth. This somewhat complicates the fabrication procedure. The grooves are made by exposing the substrate to an ion beam, which

slowly etches the surface; to define the geometry, a metal film is deposited first, and the metal is removed at the required groove locations by conventional photolithography. To vary the depths, the ion beam is confined by means of a narrow aperture, and the substrate is drawn past at a varying rate so that different regions are subject to different exposure times.

A useful feature of the RAC is that the phase of the waves propagating between the two groove arrays can be modified by means of a metal film, called the "phase plate", which exploits the velocity reduction due to electrical loading. Since waves of different frequencies traverse the device at different locations, the phase change introduced can be made frequency-dependent by varying the width of the phase plate along its length, as illustrated in Figure 9.14. The phase plate can therefore be used to compensate for phase errors in the frequency response of the device The metallisation is usually of aluminium. Quite commonly, an initial device without a phase plate is fabricated and tested, and a phase plate compensating for the observed phase errors is then designed and fabricated on the device. This enables non-repeatable errors to be compensated, though the fabrication process becomes more complicated.

To ensure that the grooves reflect the waves through 90°, it is necessary to take account of the anisotropy of the substrate. Thus the inclination of the grooves is not generally equal to 45°. From straightforward phase-matching considerations it can be seen that the required inclination, α, is given by

$$\tan \alpha = v_y/v_x. \tag{9.90}$$

Here the x-axis gives the propagation direction of surface waves launched by the input transducer, and α is the groove inclination relative to this axis, as in Figure 9.14; v_x and v_y are the surface-wave phase velocities in the x and y directions. In practical devices the accuracy of the groove inclination is often quite critical. An angular error causes the wavefronts incident on the output transducer to be non-parallel to the transducer electrodes, and hence the output signal level is reduced. Often, an accuracy of 0.1° or better is needed. It is found that velocities calculated from the bulk constants of the material, as in Section 2.3.2, do not give sufficient accuracy, so α-values determined from experimental tests are used for device design. Williamson [310] quotes experimental values for several materials. For example, for Y, Z lithium niobate, $\alpha = 46.82°$ at a temperature of 25°C.

An additional consideration is that the requirement on groove inclination, equation (9.90), must be satisfied with adequate accuracy over a reasonable range of temperatures. Generally, v_x and v_y have different temperature coefficients and, for a particular device, α also varies with temperature because the expansion coefficients in the two directions differ. To satisfy equation (9.90) over a range of temperatures, it is found that the temperature coefficients of delay, $(1/T)\, dT/d\Theta$, should be the same in the x and y directions [310]. These coefficients are generally different, and this applies in particular for ST, X quartz substrates. Thus RAC's on ST, X quartz are quite temperature-sensitive, and so this material is usually avoided. This is in marked contrast to interdigital devices, which give good temperature stability when ST, X quartz is used. For Y, Z lithium niobate the two temperature coefficients of delay are more similar, enabling RAC's to operate successfully over temperature ranges of

typically several tens of degrees. The response obtained is of course quite temperature-sensitive because the delays involved vary with temperature; quantitatively, temperature changes cause effects similar to those found in interdigital devices, as discussed in Section 9.5.3. Thus, for individual lithium niobate RAC's temperature regulation is usually mandatory, though for a matched pair comprising an expander and a compressor regulation is often unnecessary.

Comparison with Interdigital Devices. RAC's enable time-bandwidth products up to about 10,000 to be obtained, in contrast to interdigital devices which are limited to *TB*-products below about 1,000. One reason for this is that, for large *TB*-products, the insertion loss of an interdigital device becomes unacceptable because the inactive regions of the transducer act as a capacitive shunt, reducing the impedance. In a RAC, the inactive regions of the groove arrays have little effect and hence the length of the arrays can be increased without appreciably changing the loss. In addition, RAC's are found to give responses less affected by second-order effects, when compared with interdigital devices with similar *TB*-products. This is partly because the RAC structure discriminates against bulk waves generated by the input transducer, since bulk waves reflected by the groove arrays are generally found to have wavefronts not parallel to the electrodes of the output transducer. Further, second-order effects in groove arrays can be made less significant quite simply by reducing the groove depth. The phase plate is another useful feature, enabling phase errors to be reduced as noted above. The RAC also enables a larger time dispersion to be obtained for a given length of substrate, since the surface waves traverse the length of the substrate twice. Thus, a time dispersion up to about 100 μsec can be obtained, while interdigital devices are limited to about 50 μsec. At high frequencies a RAC can be somewhat easier to fabricate than a corresponding interdigital device, because the groove width, approximately $\lambda/(2\sqrt{2})$, is rather larger than the $\lambda/4$ width of the electrodes in a single-electrode transducer; in addition, defects that can cause the electrodes of a transducer to be shorted have little effect on the RAC.

The main disadvantage of the RAC is that a more complex fabrication procedure is needed and, since this is time-consuming, RAC's are relatively costly. Also, since a temperature-stable substrate cannot normally be used, RAC's are generally quite temperature-sensitive.

9.6.2. Analysis and Performance

In analysing RAC's a geometrical complication arises because the wave reflected by a groove in one array illuminates several grooves in the other array, and generally the grooves in the second array are only partially illuminated. In addition multiple reflections can be significant, and there are also some second-order effects associated with the physics of the reflection process. It is convenient to assume initially that all of these effects are negligible, which is the case if the grooves are shallow enough and if the width of the arrays, in the y-direction, is only a few surface-wave wavelengths. This gives a formulation convenient for design purposes. The complications mentioned above generally give quite small perturbations and are discussed later.

(a) First-order Analysis. We first consider reflection of surface waves through 90° by a single groove with depth h, assuming that h is much less than the wavelength λ of the incident waves. The amplitude reflection coefficient for this case is known to have the form [311]

$$r_g(\omega) = -2jC\frac{h}{\lambda}\sin(\tfrac{1}{2}ak_x), \tag{9.91}$$

where $k_x = 2\pi/\lambda$ is the wavenumber, a is the groove width measured in the x-direction (the propagation direction of the incident waves), and C is a constant depending only on the substrate material and orientation. This formula has been established experimentally for several substrate materials [311]. For Y, Z lithium niobate, the constant C is found experimentally to be 0.51. The formula is obtained by adding waves reflected at the two edges of the groove, with reflection coefficients $\pm Ch/\lambda$, and the phase of $r_g(\omega)$ is referred to the centre of the groove. The reflection coefficient is maximized when $a = \lambda/2$, which gives $r_g(\omega) = -2jCh/\lambda$. In a RAC, a is usually proportional to the pitch p in Figure 9.14, and since the reflections occur primarily in the region where p equals the wavelength it is a good approximation to take ak_x to be a constant in equation (9.91). Thus $r_g(\omega)$ is proportional to ω.

For convenience, it is assumed here that the two groove arrays in the RAC are identical and have constant width W, as in Figure 9.14. The groove depths h are the same in both arrays, though h will vary along the length of the device. Diffraction and propagation loss are neglected here. The response of the two arrays is described by a function $R(\omega)$, defined as the ratio of the surface-wave amplitude incident on the output transducer to the amplitude of the wave launched by the input transducer. Both amplitudes are measured at the transducer ports, taken to be at $x = 0$. Consider first the wave reflected by groove n of one array, and then by groove n of the other array, where both grooves are centred at $x = x_n$. For this process the contribution to $R(\omega)$ is

$$[r_{gn}(\omega)]^2 \exp(-2jk_x x_n)\exp(-jk_y d),$$

where $r_{gn}(\omega)$ is the reflection coefficient of groove n, $k_y = \omega/v_y$ is the wavenumber in the y-direction and d is the centre-to-centre spacing of the two arrays. If the aperture W were very small, $R(\omega)$ would be given simply by summing the above expression with respect to n. In practice however, the wave reflected by one groove in one array illuminates several grooves in the other array, and the number of grooves illuminated is inversely proportional to the pitch, p. To allow for this we multiply by a factor K/p_n, where K is a constant and p_n is the groove pitch in the vicinity of groove n. Thus

$$R(\omega) \approx \exp(-jk_y d)\sum_{n=1}^{N}(K/p_n)[r_{gn}(\omega)]^2\exp(-2jk_x x_n), \tag{9.92}$$

where N is the number of grooves in each array. It will be seen later that K can be taken as a constant provided the aperture W is not too large. The overall device has

a short-circuit response $H_{sc}(\omega) = H_t^a(\omega)\, H_t^b(\omega)\, R(\omega)$, where $H_t^a(\omega)$ and $H_t^b(\omega)$ are the transducer responses.

To simplify further we make use of the stationary-phase approximation. From equation (9.91), the groove reflection coefficient $r_{gn}(\omega)$ is proportional to ωh_n, where h_n is the depth of groove n. At any ω, the reflections occur mainly in the region where the pitch p is approximately equal to the wavelength. Thus, for groove n, ω can be replaced by the instantaneous frequency Ω_n in the region of this groove, where $\Omega_n = 2\pi v_x/p_n$. For convenience we ignore the initial complex exponential in equation (9.92), and also a constant multiplier, giving

$$R(\omega) \propto \sum_{n=1}^{N} \frac{h_n^2 \Omega_n^2}{p_n} e^{-j\omega t_n}, \tag{9.93}$$

where we define times $t_n = 2x_n/v_x$ corresponding to the groove positions. It is assumed that these times are obtained from a time-domain phase function $\theta(t)$, such that

$$\theta(t_n) = 4n\pi. \tag{9.94}$$

This definition is chosen such that the instantaneous frequency $\Omega_i(t) = \dot{\theta}(t)$, evaluated at $t = t_n$, is equal to Ω_n defined above.

Equation (9.93) shows that the response is that of a transversal filter, as in equation (9.67) of Section 9.3.3. Comparing with equations (9.62) to (9.64), it is seen that the fundamental component of the impulse response has the form $a(t)\,\dot{\theta}(t) \cos[\theta(t)]$, where $a(t)$ is defined such that $a(t_n) = h_n^2 \Omega_n^2/p_n$. It follows that if $\theta(t)$ is the required phase in the time domain, the groove positions are given by $x_n = v_x t_n/2$, where t_n are the solutions of equation (9.94). The frequency response $R(\omega)$ is readily found, approximately, from the stationary-phase approximation. Comparing with equations (9.3) and (9.12a), the magnitude of $R(\omega)$ in the fundamental region is, ignoring a constant multiplier,

$$|R(\omega)| \propto \frac{\omega^4 [h(T_s)]^2}{\sqrt{|\ddot{\theta}(T_s)|}}, \tag{9.95}$$

where $T_s(\omega)$ is the stationary-phase point and the groove depths are expressed by the function $h(t)$, such that $h(t_n) = h_n$. This equation gives the required variation of groove depth. For a linear chirp, $\ddot{\theta}(T_s)$ is a constant. Thus, if $|R(\omega)|$ is required to be flat in the band, as in an expander, h_n should be made proportional to Ω_n^{-2}. In practice it is usually necessary to modify this to allow for the responses of the transducers, and for several second-order effects.

(b) Geometrical Complications. The above analysis does not properly allow for the fact that waves reflected by a groove in one array illuminate several grooves in the other array. A more accurate analysis can be obtained with the aid of Figure 9.15(a), which illustrates some of the reflections involved. For convenience the

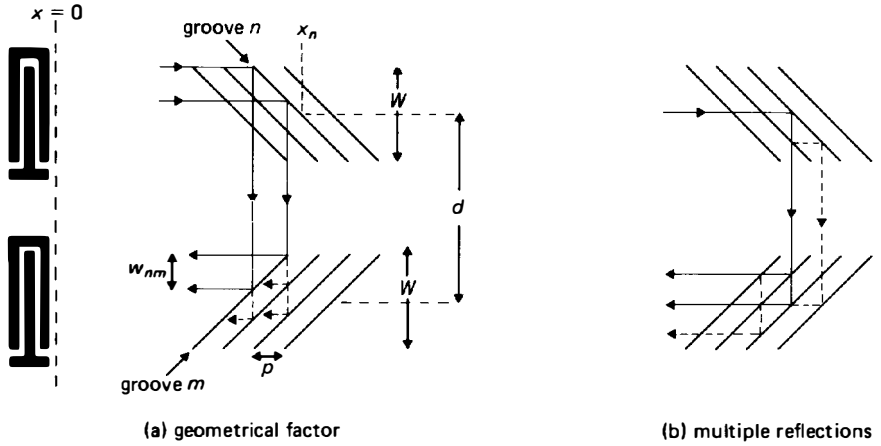


FIGURE 9.15. Factors in RAC analysis.

grooves are here represented as straight lines. It is assumed that the reflection coefficient is small, so that multiple reflections can be neglected and the transmission coefficient of each groove can be taken to be unity. Some surface-wave rays reflected by groove n of the upper array and groove m of the lower array are shown as continuous lines. All such rays have the same phase shift, and are therefore added coherently by the output transducer. Some rays reflected by subsequent grooves in the lower array are indicated by broken lines. Relative to waves reflected by groove m, the reflections from subsequent grooves have phase shifts $k_x p$, $2k_x p$, and so on. Consequently, the lowest frequency at which strong reflections are obtained is the frequency at which the pitch p is approximately equal to the surface-wave wavelength. This condition is generally used in practical devices, though there are also strong reflections at the odd harmonics.

In general, groove n of one array and groove m of the other array will overlap only partially, and in consequence the output beam due to reflection from these grooves has a width less than the width of the groove arrays. This can be seen in Figure 9.15(a), where the two rays shown as continuous lines illustrate the limits imposed by the overlap; thus the output rays, directed toward the output transducer, define the edges of the output beam. Now, the output current produced by a uniform transducer is proportional to the average of the surface-wave amplitude at its input port, as shown in Section 4.7.2, equation (4.126). The output current is therefore proportional to the width w_{nm} of the output beam. To allow for this we define an effective reflection contribution $R_{nm}(\omega)$ proportional to w_{nm}. Taking the two groove arrays to be identical and to have width W, $R_{nm}(\omega)$ is given by

$$R_{nm}(\omega) = \frac{w_{nm}}{W} r_{gn}(\omega) r_{gm}(\omega) \exp\left[-jk_x(x_n + x_m)\right] \exp\left(-jk_y d\right), \tag{9.96}$$

where $r_{gn}(\omega)$ is the reflection coefficient of groove n in either array. The total response $R(\omega)$ of the groove arrays is obtained by summing $R_{nm}(\omega)$ over both n and m, though

the sum over m needs to include only those grooves in the lower array that overlap groove n in the upper array. The output beam width w_{nm} is equal to $W - |x_n - x_m|$ for $|x_n - x_m| < W$, and is zero otherwise; this follows by taking $\alpha = 45°$, a good approximation here.

The above formulation is quite time-consuming to compute, owing to the double summation required. A convenient approximate form for $R(\omega)$ has been derived by Martin [313], making use of the stationary-phase approximation. The result is expressed in terms of a parameter

$$N_{\text{eff}}(\omega) = \frac{\omega}{2\pi\sqrt{2|\ddot{\theta}(T_s)|}}. \tag{9.97}$$

We can regard $N_{\text{eff}}(\omega)$ as the effective number of grooves in either array that are active at frequency ω. Martin's derivation shows that the magnitude of the response is approximately [310, 313]

$$|R(\omega)| \approx |N_{\text{eff}}(\omega) r_g(\omega)|^2 g\left(\frac{W}{\lambda N_{\text{eff}}}\right), \tag{9.98}$$

where $r_g(\omega)$ is the reflection coefficient for a groove in the region active at frequency ω. The function $g(z)$ is a geometrical factor allowing for the fact that reflections from many grooves are involved, and for $z \leqslant 1$ is well approximated by $g(z) \approx z/\sqrt{2}$. When this is valid, that is, when W is less than the effective length λN_{eff}, the form of equation (9.98) is the same as that of the result derived earlier, equation (9.95). This confirms the validity of the assumption that K can be taken to be a constant in equation (9.92). However, for $W > \lambda N_{\text{eff}}$, that is, for $z > 1$, $g(z)$ increases more slowly with z and the results given earlier are invalid. The slower increase of $g(z)$ with z is due to the fact that a ray propagating in the y-direction encounters grooves whose pitch is varying, so that the phases of the reflected waves are not quite the same. For $z > 1$ the phase differences involved are sufficient to cause a significant reduction in the amplitude of the output beam.

(c) Multiple Reflections. The term "multiple reflections" refers to waves that are successively reflected two or more times in one or both of the groove arrays. This is illustrated in Figure 9.15(b), where rays due to first-order reflections are shown as continuous lines and those due to multiple reflections are shown as broken lines. The multiple reflections, ignored in the analysis described above, are negligible if the groove reflectivity is low enough.

Multiple reflections in one-dimensional structures, such as unapodised transducers or multi-strip couplers, can be analysed by using a transmission matrix for each element and cascading the matrices, as shown in Appendix E. For RAC's, the two-dimensional nature of the problem introduces considerable extra complexity. Otto [314] has developed a two-dimensional analysis for a periodic array of grooves, representing it as a two-dimensional array of cells with each cell small enough for multiple reflections within it to be neglected. A two-dimensional transmission matrix

was used for each cell. The application of this method to non-periodic grooves is described by Bloch *et al.* [315]. Alternatively, Wright and Haus [316] have developed an analysis based on coupled mode theory, and this enables the response to be computed more easily.

In a RAC, it is found that the main effect of multiple reflections is to reduce the amplitude of the frequency response $R(\omega)$ by an amount that varies slowly with frequency [317]. The groove depths can be modified in order to compensate this. However, it is also found [317] that multiple-reflection effects are generally insignificant if the total reflection loss exceeds about 15 dB, that is, if $20 \log |R(\omega)| < -15$. In practical devices this condition is nearly always acceptable, and if so multiple reflections need not be considered.

(d) Other Second-order Effects. The interaction between surface waves and grooves is found to be accompanied by some additional second-order effects, associated with the physical processes involved. The most important effect is termed *stored energy*. It is found that a periodic array of grooves causes a slight reduction of the surface-wave velocity, by an amount dependent on the groove depth h. This phenomenon is attributed to evanescent acoustic fields which store energy in the regions near the groove edges, but do not cause any loss of power. The perturbation can be expressed in terms of a surface-wave phase shift $\Delta\phi = -\hat{B}/2$ at each groove edge. This gives a perturbed wavenumber $k = k_0 + \hat{B}/p$, where k_0 is the free-surface wavenumber and p is the pitch, which in a RAC varies with distance. From experimental measurements [311] the parameter $\hat{B}$ is known to have the form

$$\hat{B} = 2C'(h/\lambda)^2, \tag{9.99}$$

where C' is a material-dependent constant, equal to 4.5 for a RAC on Y, Z lithium niobate. Thus, $\hat{B}$ is proportional to ω^2.

For a RAC, the main consequence of stored energy is that a phase error arises in the device frequency response. This error can be estimated using the stationary-phase approximation [306, 311], though some care is necessary since the perturbed velocity is frequency-dependent and varies with position. For a linear chirp device the error is found to include a cubic term of the form $\Phi_3 P_3(x)$, causing a spurious sidelobe to appear when the device is used for pulse compression, as explained in Section 9.5.1. In large RAC's Φ_3 is often 50° or more. The error can of course of corrected by means of a phase plate, or alternatively the groove positions can be adjusted. Another consequence of stored energy is that a groove array is found to give responses at the even harmonics, not predicted by the first-order analysis.

Propagation loss is often significant in RAC's, and can be allowed for by adjusting the groove depths. The loss for free-surface propagation is described in Section 6.3. It is also found that the grooves cause some additional loss, attributed to excitation of non-coherent surface or bulk waves, called the "non-synchronous scattering loss" [318]. The attenuation coefficient due to this effect is dependent on position. In large RAC's, both types of loss can cause the amplitude of the frequency response to vary by 10 dB or more, so it is important to design the device to compensate for them.

For most RAC's, using Y, Z lithium niobate substrates, diffraction is not

significant. However, other substrate materials are sometimes used, and it is then usually necessary to design the device compensating for diffraction.

(e) Performance. Figure 9.16 shows the performance of a linear-chirp RAC with 200 MHz centre frequency, using a *Y*, *Z* lithium niobate substrate. The bandwidth was 80 MHz and the time dispersion 80 μsec, giving a time-bandwidth product of 6400. The device was designed to have a flat amplitude characteristic in the frequency demain, and to obtain this the groove depths were varied from about 500 to 2000 Å. The overall insertion loss was 38 dB. A phase plate was used to minimise the phase error. Ideally, the phase of the frequency response is a quadratic function of frequency, and the phase error curve in Figure 9.16 shows the result obtained when a quadratic was subtracted from the measured phase. The error is less than 10° at most frequencies. It should be remembered that the ideal phase varies very rapidly with frequency; in the present case the variation over the band is some 6×10^5 degrees, so the phase error corresponds to an accuracy of a few parts in 10^5. In laboratory tests this device was used as an expander in conjunction with a second RAC used as a compressor, with the latter having amplitude weighting, and a time-sidelobe suppression of 33 dB was obtained. The compressor had a bandwidth of 40 MHz, half that of the expander, and therefore had a time-dispersion of 40 μsec and a time-bandwidth product of 1600. The difference in the bandwidths is due to the intended application of the devices in a compressive receiver, as explained in Section 9.7 below.

Some examples of RAC and IMCON performance are summarised in Table 9.2, where the fifth column refers to the device described above. For the remaining columns, the time-sidelobe rejection refers to a pair of devices with the same bandwidth. It can be seen that RAC's can give very large time-bandwidth products,

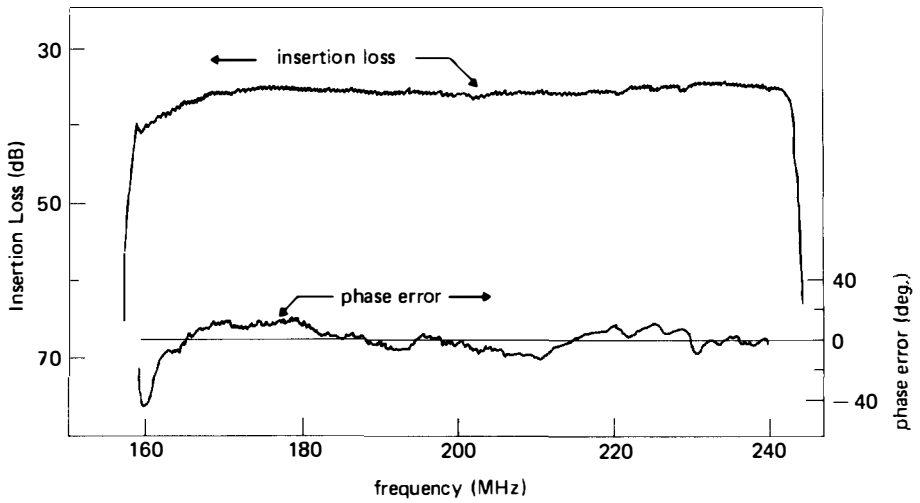


FIGURE 9.16. Frequency response of a RAC with 80 MHz bandwidth and 80 μsec dispersion. (Courtesy Plessey Research)

TABLE 9.2

Performance examples for reflective array compressors. All entries refer to linear-chirp devices

Type†	IM	IM	R	R	R	R	R	L	IL	RDA	RDA, IL
Centre frequency (MHz)	20	15	60	1000	200	400	60	1000	150	60	150
Bandwidth, B (MHz)	10	6	6	512	80	180	2.5	500	50	20	50
Duration, T (μsec)	250	100	100	10	80	90	125	1	20	10	20
TB	2500	600	600	5120	6400	16,200	312	500	1000	200	1000
Time-sidelobe rejection, dB	–	48	37	–	33*	–	33	–	30	26	–
C.W. insertion loss (dB)	–	30	34	52	38	56	33	45	47	56	–
Substrate‡	Steel	Steel	LN	LN	LN	LN	BGO	LN	LN	LN	LN
Reference	[313]	[319]	[320]	[310]	*	[335]	[321]	[322]	[323]	[324]	[325]

* See text.

† R = conventional RAC; IM = IMCON; L = length weighted; IL = in-line RAC; RDA = reflective dot array.

‡ LN = lithium niobate, BGO = bismuth germanium oxide.

16,200 for one device, and the dispersion T is typically up to 100 μsec. One column refers to devices using bismuth germanium oxide substrates, which give an unusually low surface-wave velocity and thus enable a larger time dispersion to be obtained. The last four columns refer to some novel types of RAC, discussed in the next Section.

9.6.3. Other Types of RAC

While the results discussed above show RAC's to be capable of impressive performance, there are some practical disadvantages. Notably, the time required to etch an array of grooves with varying depth renders the devices relatively expensive, and the devices are quite temperature sensitive. Some modified forms of RAC have been developed to overcome these disadvantages, though with some loss of performance.

The fabrication time required can be reduced substantially if all the grooves have the same depth, since this avoids the need to constrain the ion beam by a narrow aperture. However, this implies that some novel method of weighting is necessary. Figure 9.17 shows two possibilities, for which experimental data are included in Table 9.2. In the length-weighted RAC [322] the groove lengths are varied, and "dummy" grooves at a different angle are included so that the velocity perturbation is uniform. The in-line RAC [323] uses a 3 dB multi-strip coupler to partition the input wave between two arrays of grooves, which reflect it through 180°. Amplitude weighting is obtained by slightly displacing the grooves in one track relative to those in the other track, though length weighting would alternatively be applicable.

Further simplification of the fabrication can be obtained by using metal structures to reflect the waves instead of grooves, so that the etching procedure can be omitted altogether. Metal strips can sometimes be used, but for lithium niobate substrates this is not feasible because the electrical loading causes too strong a perturbation. To overcome this, Solie [324] introduced the "reflective dot array" in which each reflector is divided into a large number of metal segments, known as dots. This also enables weighting to be obtained by varying the density of the dots. Table 9.2 includes some results, and also some data for an in-line version of the dot array device [325].

Continuous metal strips can be used as reflectors on some substrates, as shown by a matched pair of devices on quartz [326] with a TB-product of 500. These devices used 90° reflections and were length-weighted. A time-sidelobe suppression of 28 dB was demonstrated.

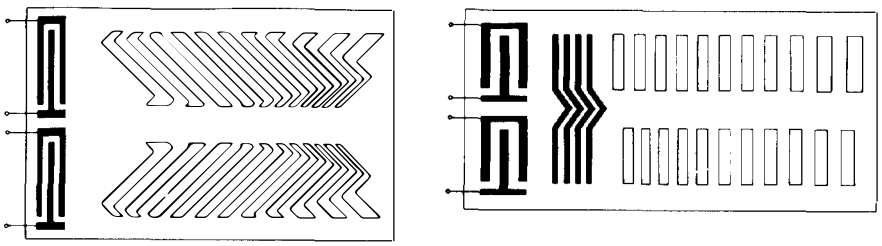


FIGURE 9.17. RAC's using grooves of uniform depth.

9.7. SPECTRUM ANALYSIS AND OTHER TYPES OF SIGNAL PROCESSING

In Section 9.1 it was seen that chirp filters play a central role in the operation of pulse compression radar systems. In this section we consider a number of other applications, mostly for filters whose impulse responses are linear chirps. The most prominent of these is a system for frequency measurement, known as the *compressive receiver*. It was shown in Section 9.5.3 that, for linear chirps, the main effect of a doppler shift on a pulse-compression system is to delay the output pulse. The compressive receiver uses a similar principle to measure the frequency of a C.W. input signal; the delay of the output pulse is measured in order to determine the input frequency. The delay is a linear function of frequency, and the system is also linear in the sense that a signal comprising several C.W. components can be analysed, giving output pulses corresponding to the individual input components. From this it can be concluded that, for a more general form of input signal, the output waveform must be closely related to the Fourier transform of the input, that is, to its spectrum. In fact, Fourier transformation can be achieved using a system similar to the compressive receiver, though three chirp filters are required rather than two. This enables a complex function to be transformed, presenting the complex output as an amplitude and phase modulated waveform, or as a pair of waveforms. In both the compressive receiver and the Fourier transform system, the time taken to perform the transform is typically a few times $1/\Delta f$, where Δf is the frequency resolution. Thus the transformation is achieved much more rapidly than in a conventional spectrum analyser, which mixes the signal with a slowly-swept local oscillator and applies the output to a narrow-band filter.

The development of such systems has been given considerable impetus by the high performance standards achievable using surface-wave devices, particularly RAC's. However, the principles do not depend on the technology. Thus any of the technologies suitable for chirp filters, mentioned in Section 9.1, would be applicable subject to performance constraints. In fact, the compressive receiver and Fourier transform systems have been studied extensively using digital techniques and using charge-coupled devices, with modifications owing to the fact that sampled data are used.

Section 9.7.1 below gives a descriptive account of the compressive receiver. The basic mathematical relations involved are given in Section 9.7.2, which includes systems for Fourier transformation, and experimental results are discussed in Section 9.7.3. Section 9.7.4 considers a variety of other signal processing functions that can be accomplished using chirp filters, including a variable delay line and a chirp filter system with variable chirp rate.

9.7.1. Compressive Receiver Principles

The principle involved here is closely related to the observation by Klauder *et al.* [277, p. 761] that a pulse compression system can be used to perform Fourier transformation. Subsequently, Darlington [327] showed in detail how the method could be applied to frequency measurement, and Darlington's system is now called

a compressive receiver, or sometimes microscan receiver. Recently, reviews of the topic have been given by Jack *et al.* [328, 329] and by Roberts *et al.* [330].

The system is illustrated in Figure 9.18. The input signal, taken to be C.W. with angular frequency Ω, is mixed with a linear-chirp waveform obtained by impulsing an expander, and the resulting output is applied to a compressor with a chirp slope matching that of the expander. For the moment, both chirp filters are taken to have impulse responses with flat envelopes. The lower part of the figure shows the operation of the system in terms of the frequency-time curves of the devices; the curves for the expanded pulse are shown as they appear at the compressor input, and thus shift in frequency according to the value of Ω. It is assumed that the frequency band of the signal at the compressor input overlaps completely the band occupied by the compressor response. To ensure this, the centre frequencies of the two filters must be different and the expander bandwidth must be larger than that of the compressor. In addition, Ω must be constrained to a finite band, equal to the difference between the two device bandwidths. The curves in Figure 9.18 refer to the two extreme values of Ω, which are denoted Ω_1 and Ω_2.

For a particular input frequency, the compressor output waveform has an envelope given approximately by sinc $(\pi B_c t)$, where B_c is the compressor bandwidth. This is similar to the output of an unweighted linear-chirp pulse compression system. However, it can be seen in Figure 9.18 that the expanded pulse sweeps through the band of the compressor at times which vary according to the value of Ω. Thus the time of the output pulse varies with the input frequency, and the relationship is linear because linear chirps are used. The input frequency can therefore be found by detecting the output pulse and measuring its delay relative to the pulse used to excite the expander.

Important parameters are the frequency resolution of the system and the number of resolvable points. If μ is the chirp slope of the expander, the time τ of the output

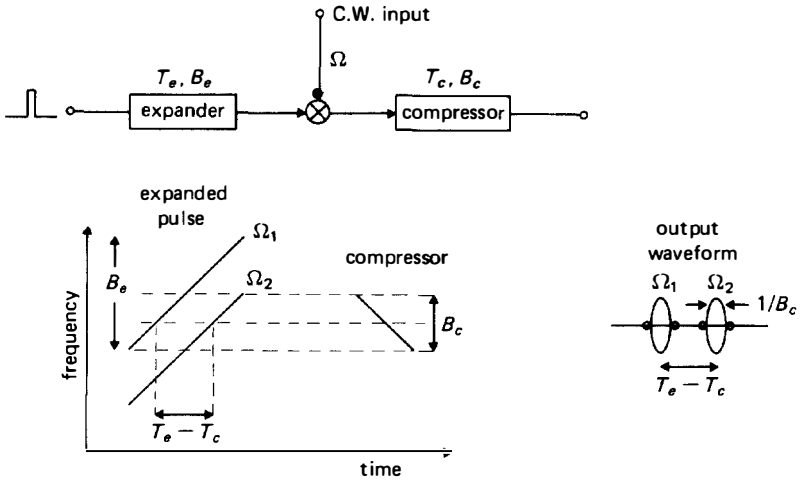


FIGURE 9.18. Compressive receiver.

pulse is given by $\Omega = -2\pi\mu\tau + \text{constant}$. The width of the output pulse is approximately $1/B_c$, and hence the resolution is $\Delta f = |\mu|/B_c = 1/T_c$, where T_c is the length of the compressor impulse response. The range of acceptable input frequencies is $(B_e - B_c)$, where B_e is the expander bandwidth, and hence the number of resolvable points is $(B_e - B_c)T_c$. For technological reasons, it is usually advisable to use the minimum possible value of the expander time-bandwidth product, T_eB_e. This value can be deduced by writing $B_c = kB_e$, from which it follows that the number of points is $(k - k^2)T_eB_e$. Differentiating this shows that, for a given number of points, T_eB_e is minimised by taking $k = \frac{1}{2}$, so that the expander bandwidth is twice that of the compressor. The number of points is then $T_eB_e/4$.

Provided the multiplier used is linear with respect to the signal input the system behaves linearly, so that for an input signal comprising several C.W. components the output gives pulses at the corresponding times. However, the output does not in this case give the Fourier transform of the input; it corresponds instead to a function known as the *sliding* Fourier transform, as will be shown in Section 9.7.2 below. In practice, this distinction is not usually very consequential. For the multiplier a balanced mixer is generally used, and for optimum dynamic range one of the inputs, the "local oscillator" input, is driven into saturation. For the compressive receiver the expanded pulse is applied to the local oscillator input, and saturation is acceptable here because this pulse has a flat envelope.

In practical systems it is usual to apply amplitude weighting in the compressor so as to reduce the time-sidelobes of the output pulse, as described for pulse-compression systems in Section 9.2.3. This enables the system to resolve two C.W. components of a signal, with closely similar frequencies but substantially different power levels. In comparison with an unweighted system, this incurs some loss of resolution and a corresponding reduction in the number of points, but the device bandwidths can be increased to compensate for this.

It has been assumed so far that the expander has a bandwidth larger than that of the compressor. An alternative arrangement, with the compressor having the larger bandwidth, operates in a very similar manner. In this case the expander can be weighted to reduce the time-sidelobes. However, this implies that the multiplier used must be linear with respect to both input ports, and this imposes technological constraints which limit the performance obtainable [329]. For this reason, the expander usually has the wider bandwidth in practical compressive receivers.

9.7.2. Analysis of Compressive Receivers and Fourier Transform Systems

Here the compressive receiver is considered in more detail, and in view of the close similarity we also consider systems for Fourier transformation. Both cases are related to a fundamental concept which expresses the Fourier transform in terms of chirp waveforms. Consider a waveform $f(t)$, with Fourier transform $F(\omega)$. Writing τ instead of t, and $2\pi\mu t$ instead of ω, we have

$$F(2\pi\mu t) = \int_{-\infty}^{\infty} f(\tau) \exp(-j2\pi\mu t\tau)\,d\tau \tag{9.100}$$

Using the identity $2t\tau = t^2 + \tau^2 - (t - \tau)^2$, this becomes [331]

$$F(2\pi\mu t) = \exp(-j\pi\mu t^2) \int_{-\infty}^{\infty} [f(\tau) \exp(-j\pi\mu\tau^2)] \exp[j\pi\mu(t-\tau)^2] d\tau \quad (9.101)$$

This form of the Fourier transform is called *chirp transform*. An equivalent time-discrete form, suitable for digital or charge-coupled devices, is known as the chirp z transform [332, 237]. The chirp transform, equation (9.101), can be seen to be equivalent to multiplication by a complex chirp, convolution with another chirp, and finally multiplication by a third chirp. Consequently, this is called the M–C–M scheme. If we define

$$\begin{aligned} h_e(t) &= h_m(t) = \exp(-j\pi\mu t^2), \\ h_c(t) &= \exp(j\pi\mu t^2), \end{aligned} \quad (9.102)$$

then equation (9.101) becomes

$$F(2\pi\mu t) = h_m(t)[\{f(t)h_e(t)\} * h_c(t)]. \quad (9.103)$$

Figure 9.19 shows a representation of this process in terms of chirp filters. The waveforms $h_e(t)$ and $h_m(t)$ are obtained by impulsing chirp filters having $h_e(t)$ and $h_m(t)$ as their impulse responses. The convolution with $h_c(t)$ is obtained by applying the signal to a filter with impulse response $h_c(t)$. The final multiplication by $h_m(t)$ changes the phase of the output waveform, but not its amplitude. Thus, if only the amplitude of the spectrum is required, the final multiplication can be omitted. In this case, the system becomes essentially the same as the compressive receiver of Figure 9.18. The representation in Figure 9.19 is actually a rather naive interpretation of equation (9.103), because in practice the device impulse responses must be real and must have finite lengths. A more detailed analysis allowing for these factors is given below.

A dual form of the system, shown in Figure 9.20, involves convolution followed by multiplication and then another convolution, and is called the C–M–C scheme. The

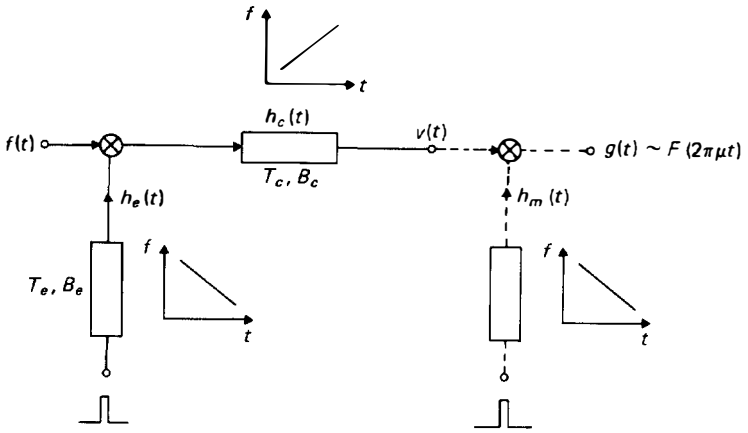


FIGURE 9.19. M–C–M Fourier transform system.

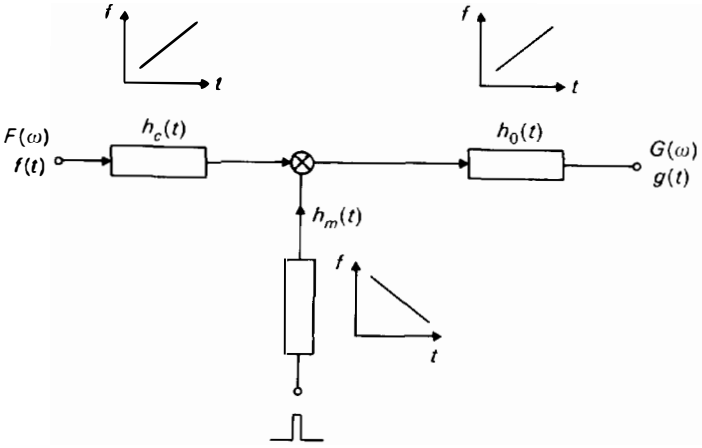


FIGURE 9.20. C–M–C Fourier transform system.

input $f(t)$ is convolved with $h_c(t)$, multiplied by $h_m(t)$, and then convolved with $h_0(t)$. To show that this gives the Fourier transform, we consider the waveforms in the frequency domain. Define $H_c(\omega)$, $H_m(\omega)$ and $H_0(\omega)$ as the filter frequency responses, that is, the transforms of $h_c(t)$, $h_m(t)$ and $h_0(t)$ respectively. The multiplication by $h_m(t)$ becomes, in the frequency domain, a convolution with $H_m(\omega)$, as shown by the convolution theorem. Thus if $F(\omega)$ is the spectrum of the input waveform, the spectrum $G(\omega)$ of the output waveform is

$$G(\omega) = \frac{1}{2\pi} H_0(\omega)[\{F(\omega)H_c(\omega)\} * H_m(\omega)], \tag{9.104}$$

which is analogous to equation (9.103). To show that this gives the Fourier transform of the input, note that the transform of $\exp(j\pi\mu t^2)$ is proportional to $\exp(-j\alpha\omega^2)$, with $\alpha = 1/(4\pi\mu)$. We can thus take $H_c(\omega) = H_0(\omega) = \exp(-j\alpha\omega^2)$ and $H_m(\omega) = \exp(j\alpha\omega^2)$, ignoring multiplying constants. Substituting these into equation (9.104) gives

$$G(\omega) = f\left(-\frac{\omega}{2\pi\mu}\right), \tag{9.105}$$

where $f(t)$ is the original input waveform. From the definition of the Fourier integral, it then follows that the output waveform is $g(t) = \mu F(2\pi\mu t)$, which is essentially the same as the output given by the M–C–M scheme.

Analysis for Finite Length Devices (M–C–M scheme). To appreciate the system operation in more detail, we must take account of the fact that the device impulse responses have finite lengths. For brevity, only the M–C–M scheme is considered here. The following account is a summary of that given by Jack *et al.*

[328, 329], simplified by taking the filter impulse responses to be centred at $t = 0$. In practice, causality requires each impulse response to be centred at some positive t, so that it is zero for $t < 0$, but this makes no difference to the output waveform except for a constant delay.

In the M–C–M system of Figure 9.19, the compressor output waveform $v(t)$ is

$$v(t) = [f(t)h_e(t)] * h_c(t) \tag{9.106}$$

and this also gives the output of a compressive receiver. The expander impulse response is taken to have a center frequency ω_e, chirp slope $-\mu$ and duration T_e, so that

$$h_e(t) = \text{rect}\,(t/T_e)\cos(\omega_e t - \pi\mu t^2). \tag{9.107}$$

For the compressor, ω_c and T_c are defined similarly, so that

$$h_c(t) = \text{rect}\,(t/T_c)\cos(\omega_c t + \pi\mu t^2), \tag{9.108}$$

where the chirp slope is μ. Substituting into equation (9.106) gives

$$v(t) = \int_{-\infty}^{\infty} f(\tau)\,\text{rect}\,(\tau/T_e)\cos(\omega_e\tau - \pi\mu\tau^2)\,\text{rect}\left(\frac{t-\tau}{T_c}\right) \times \cos[\omega_c(t-\tau) + \pi\mu(t-\tau)^2]\,d\tau. \tag{9.109}$$

Here the product of cosines gives cosines of the sum and difference of the arguments. However, the difference term can be neglected because the waveforms involved can be taken to be band-pass functions; the difference term arises from the negative-frequency part of $h_e(t)$ and the positive-frequency part of $h_c(t)$ and vice versa. Neglecting this term, we have

$$v(t) = \frac{1}{2}\int_{-\infty}^{\infty} f(\tau)\,\text{rect}\,(\tau/T_e)\,\text{rect}\left(\frac{t-\tau}{T_c}\right)\cos(\omega_c t + \pi\mu t^2 - \Omega\tau)\,d\tau, \tag{9.110}$$

where Ω is a linear function of t, defined by

$$\Omega(t) = \omega_c - \omega_e + 2\pi\mu t. \tag{9.111}$$

We now consider separately the case when the expander has the shorter impulse response ($T_e < T_c$), and the case when the expander is longer ($T_e > T_c$). As explained earlier, the latter case is usually chosen for a compressive receiver.

Case 1: Short Expander. It is first necessary to consider the role of the rect functions in equation (9.110). If all the other functions in the integrand are replaced by unity, it is found that, for $T_e < T_c$, the waveform $v(t)$ has a flat region for $|t| < \frac{1}{2}(T_c - T_e)$ and sloping regions of length T_e at each end, so that the total duration of the output waveform is $T_c + T_e$. The flat region corresponds to the time interval when the expanded pulse is entirely "inside" the compressor. Clearly, the system can only be expected to give a Fourier transform for times corresponding to the flat region, so only this region is considered here.

For this region, that is, for $|t| < \frac{1}{2}(T_c - T_e)$, it is found that the presence of the first rect function in equation (9.110) implies that the second rect function can be omitted. We thus find

$$v(t) = \tfrac{1}{4} \exp(j\omega_c t + j\pi\mu t^2) \int_{-\infty}^{\infty} f_1(\tau) \exp(-j\Omega\tau)\, d\tau + \text{c.c.}, \quad (9.112)$$

where "c.c." indicates complex conjugate and we define

$$f_1(t) = f(t) \text{ rect } (t/T_e), \quad (9.113)$$

which is a segment of the input waveform $f(t)$, with length T_e. The integral in equation (9.112) is clearly equal to $F_1(\Omega)$, where $F_1(\omega)$ is the Fourier transform of $f_1(t)$. If we define $\Phi_1(\omega)$ as the phase of $F_1(\omega)$, so that $F_1(\omega) = |F_1(\omega)| \exp[j\Phi_1(\omega)]$, then the compressor output is

$$v(t) = \tfrac{1}{2}|F_1(\Omega)| \cos[\omega_c t + \pi\mu t^2 + \Phi_1(\Omega)], \quad \text{for } |t| < \tfrac{1}{2}(T_c - T_e). \quad (9.114)$$

This shows that the envelope of the output waveform is proportional to the magnitude of the spectrum of $f_1(t)$, given by equation (9.113). The spectral frequency Ω is linearly related to the time-axis of the output waveform, as shown by equation (9.111). Note that, since the system only transforms a segment of input waveform of length T_c, the frequency resolution, in Hz, is approximately $1/T_c$. In addition, the output gives the spectrum only over a finite band of frequencies. This band, of width $\Delta\Omega$, can be deduced from the time interval for which equation (9.114) is valid, substituting into equation (9.111) to give $\Delta\Omega = 2\pi(B_c - B_e)$. The chirp slope μ in the above equations can be positive or negative; in the latter case, the output gives the inverse Fourier transform instead of the forward transform.

Although the input waveform $f(t)$ must be real, the system can in fact transform complex data. This is easily seen by writing $f(t) = f_0(t) \exp(j\omega_0 t) + \text{conjugate}$, where $f_0(t)$ is a complex function and ω_0 is a carrier frequency. If $F_0(\omega)$ is the transform of $f_0(t)$, the transform of $f(t)$ includes a term $F_0(\omega - \omega_0)$. The system can therefore be used to determine $|F_0(\omega - \omega_0)|$, from which $|F_0(\omega)|$ follows.

If the phase of the spectrum is required in addition to its amplitude, an additional multiplication by a chirp waveform $h_m(t)$ is included, as shown in Figure 9.19. This removes the quadratic phase term $\pi\mu t^2$ in the compressor output of equation (9.114). The chirp slope of $h_m(t)$ is $-\mu$, and its duration must be at least $(T_c - T_e)$. If ω_m is the centre frequency of $h_m(t)$, the output waveform has the form

$$g(t) = \tfrac{1}{4}|F_1(\Omega)| \cos[(\omega_c + \omega_m)t + \Phi_1(\Omega)], \quad \text{for } |t| < \tfrac{1}{2}(T_c - T_e), \quad (9.115)$$

where it is assumed that a difference-frequency term has been eliminated by appropriate filtering. It is usual to demodulate this waveform using two synchronous detectors in phase quadrature, with reference frequency $\omega_c + \omega_m$. With the phases properly adjusted this gives outputs of the form $|F_1(\Omega)| \cos[\Phi_1(\Omega)]$ and $|F_1(\Omega)| \sin[\Phi_1(\Omega)]$, the real and imaginary parts of $F_1(\Omega)$, as noted by Darlington [327].

Case 2: Long Expander. For $T_e > T_c$, which is usually the case for a compressive receiver, we again deduce the compressor output $v(t)$ from equation (9.110). For this case the rect functions alone give a flat region for $|t| < \frac{1}{2}(T_e - T_c)$. The total length of the output waveform is $T_e + T_c$, as before. For the interval $|t| < \frac{1}{2}(T_e - T_c)$, the first rect function in equation (9.110) can be omitted, giving

$$v(t) = \tfrac{1}{4} \exp(j\omega_c t + j\pi\mu t^2) \int_{-\infty}^{\infty} f_2(\tau, t) \exp(-j\Omega\tau)\,d\tau + \text{c.c.}, \qquad (9.116)$$

where

$$f_2(\tau, t) = f(\tau) \operatorname{rect}\left(\frac{t - \tau}{T_c}\right), \qquad (9.117)$$

and $\Omega(t)$ is given by equation (9.111) as before. The integral in equation (9.116) is denoted $F_2(\Omega, t)$, the transform of $f_2(\tau, t)$ from the τ-domain to the Ω-domain. If $\Phi_2(\Omega, t)$ is the phase of $F_2(\Omega, t)$, we have

$$v(t) = \tfrac{1}{2}|F_2(\Omega, t)| \cos[\omega_c t + \pi\mu t^2 + \Phi_2(\Omega, t)], \quad \text{for } |t| < \tfrac{1}{2}(T_e - T_c). \qquad (9.118)$$

This is similar to the short-expander case, but here the transform obtained does not quite correspond to the Fourier transform of the input. Transforming equation (9.117) shows that $F_2(\Omega, t)$ can be written

$$F_2(\Omega, t) = \int_{t - T_c/2}^{t + T_c/2} f(\tau) \exp(-j\Omega\tau)\,d\tau \qquad (9.119)$$

This shows that, for any t, the output gives one point in the spectrum of a segment of the input $f(t)$. However, the location of the segment is varying with t, and so equation (9.119) is called the *sliding* Fourier transform. Generally, the Fourier components of an input signal cannot be deduced from the sliding transform. However, if the signal comprises a finite sum of C.W. waveforms, as is often the case for compressive receiver applications, the magnitude of the output given by the sliding transform is the same as that given by the conventional transform; the distinction is then of no significance, provided the spectral phase is not required.

9.7.3. Experimental Compressive Receivers and Fourier Transform Systems

We consider here a few of the many experimental examples described in the literature. RAC's are often used since, for a large number of resolvable points, a large TB-product is needed. With TB-products up to about 10,000, the number of points can be up to about 2500. For such RAC's the time dispersion is typically 20 to 100 μsec, and hence the frequency resolution of the system is typically 20 to 100 kHz.

Figure 9.21 shows the output of a compressive receiver for several C.W. input signals with different frequencies, where the traces have been displaced vertically for clarity. The system used two RAC's, an expander with 80 MHz bandwidth and 34 μsec dispersion, and a compressor with 40 MHz bandwidth and 17 μsec dispersion. The compressor incorporated amplitude weighting, giving a time-sidelobe

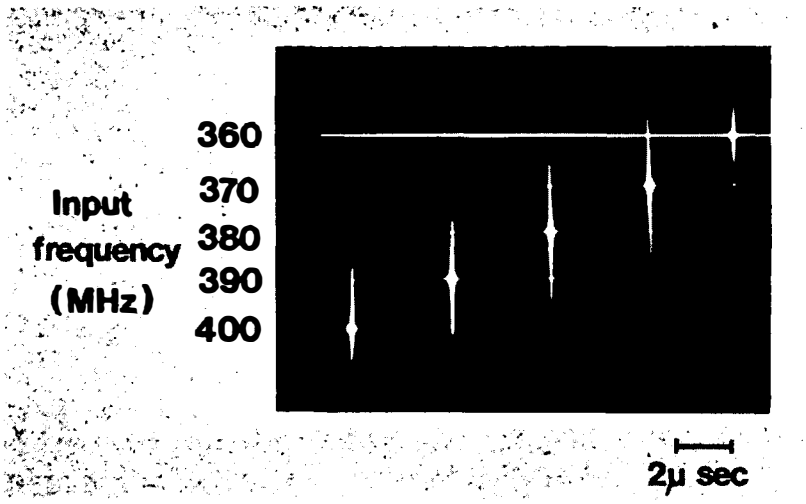


FIGURE 9.21. Performance of a compressive receiver using RAC's (Courtesy, Plessey Research).

suppression of 34 dB. This system enabled signals over a 40 MHz band to be analysed, with a resolution of 75 kHz.

Some other examples are given in References [333–340]. The performance achievable using RAC's is well illustrated by the system of Gerard *et al.* [335], which transformed signals with up to 60 MHz bandwidth and 60 μsec duration, giving a frequency resolution of 17 kHz. For this system the largest RAC had an impressive time-bandwidth product of 16,200, with other parameters as shown in Table 9.2. A Fourier transform system with digital input and output, incorporating D-to-A and A-to-D convertors, was demonstrated by Dolat *et al.* [336], and analysed a 10 MHz band with 40 kHz resolution. This was seen as an alternative to an all-digital approach using the Fast Fourier Transform. In comparison, the surface-wave approach offers rapid operation with reduced weight and power consumption, though the accuracy obtainable is not as great. In a later development of this system [337] an automatic temperature compensation technique, using recurrent test signals, eliminated the need for bulky ovens to control the temperature.

Moule *et al.* [338] developed a wide-band compressive receiver, analysing 250 MHz bandwidth with 4 MHz resolution. This used interdigital devices with 500 MHz bandwidth, shown on Table 9.1. A direct consequence of the wide bandwidth is that the output waveform had a very fine time resolution of 4 nsec. In order to process this, the waveform was detected and entered into a digital shift register clocked at 256 MHz. The data was subsequently read out at a slower rate for further processing.

Williamson *et al.* [339] describe a satellite-borne processor for demodulation of frequency-shift-keyed communication signals. This system simultaneously handled 100 users, each using a 4-ary code with 4 tones, so that there were 400 tones in all. The tones had a separation of 38 kHz, and were 50 μsec long. In this application the system must process the entire length of each input tone in order to preserve the

information content of the signal, so synchronisation is necessary and the system must have a 100% duty factor. A system using four RAC's was used to meet these requirements.

Another application for such systems is in measurement of doppler frequency shifts of radar signals, thus determining target velocities. Generally, the resolution given by surface-wave systems is insufficient for doppler measurements. However for a radar using optical rather than radio-wave propagation the doppler shifts are relatively large, so that surface-wave systems become applicable. For example, a compressive receiver for an infra-red radar [340] had 84 kHz resolution, corresponding to a velocity resolution of 1.6 km/h. Alternatively, for radar systems using radio-wave propagation there is the possibility of contracting the time-scale of the input waveform, thus increasing the doppler shift sufficiently for measurement by a surface-wave system. Roberts *et al.* [341] have demonstrated this by entering the received waveform into a charge-coupled device and then reading it out at a faster rate, suitable for a subsequent surface-wave Fourier transform system. Using a time-compression factor of 1000 in the charge-coupled device, a doppler resolution of 40 Hz was obtained.

Time compression can also be achieved digitally, by entering the waveform into a shift register and then reading out using a faster clock rate. This requires the surface-wave system to have a digital input, as in Dolat's system [336] mentioned above. The additional flexibility introduced by combining surface-wave and digital techniques leads to a variety of novel applications, which have been discussed by Gautier and Tournois [342]. For example, in a sonar system with a phased-array receiver, beam forming can be accomplished by sampling the receiving elements in turn and Fourier transforming the resulting waveform. The rapid processing performed by the surface-wave system enables the beam forming to be done during the time interval corresponding to the range resolution of the sonar system. Several other applications were considered [342], including two-dimensional Fourier transformation.

9.7.4. Other Types of Signal Processing

In addition to Fourier transformation, considered above, surface-wave chirp filters have been used to perform a variety of other signal processing functions. Some of these make use of the Fourier transform system as a building block, giving systems with a higher level of complexity. For example, suppose an input waveform $s(t)$ is Fourier transformed to give $S(\Omega)$, which is multiplied by some waveform $u(t)$ and then inverse transformed to give an output $g(t)$. This configuration, involving two Fourier transform units, can perform a variety of signal processing functions, depending on the waveform $u(t)$. For example, if $u(t)$ is sinusoidal the output is a delayed version of the input, with delay determined by the frequency of $u(t)$, while if $u(t)$ is a linear ramp the output is the time-differential of the input. Atzeni *et al.* [334] demonstrated these functions experimentally. If $u(t)$ is a rectangular pulse with a gating action, so that only a short segment of $S(\Omega)$ is allowed to pass, the system behaves as a variable bandpass filter whose bandwidth and centre frequency are determined by the pulse timing [333, 334, 343]. Alternatively, if the gate is inverted so that a segment of $S(\Omega)$

is suppressed, the system becomes a variable band-stop filter. This principle has been used to demonstrate suppression of narrow-band interference corrupting a wide-band signal [333].

Some signal-processing functions can be accomplished quite simply using the configuration shown in Figure 9.22(a). Here the input signal is passed through two chirp filters in sequence. The first filter has a higher centre frequency than the second, and the signal is up-converted before entering it and down-converted afterwards, using mixers with a C.W. reference at frequency Ω. Using linear chirp devices with matching chirp slopes this system can act as a *variable delay line*, with delay determined by the value of Ω. Figure 9.22(b) illustrates the delays involved for this case. The delay for filter 1, a down-chirp, is shown as it appears between points A and B, before the up-conversion and after the down-conversion. This delay curve therefore moves along the frequency axis as Ω is varied. The bandwidth of filter 1 is made larger than that of filter 2, and Ω is constrained so that the delay curves of the two filters overlap in frequency. It can be seen that the total delay is independent of signal frequency, but its value varies linearly with Ω. A system of this type [344] enabled the delay to be varied by 30 μsec, for signals with bandwidths up to 10 MHz. A similar configuration can be used as a *variable bandpass filter*. In this case the two devices have the same bandwidth, and the bands overlap by an amount dependent on the value of Ω. This principle was demonstrated, in a modified form, by Melngailis *et al.* [345].

The configuration of Figure 9.22(a) can also be used to give a *variable chirp filter*, that is, a chirp filter whose chirp slope can be varied electronically. In this case the two

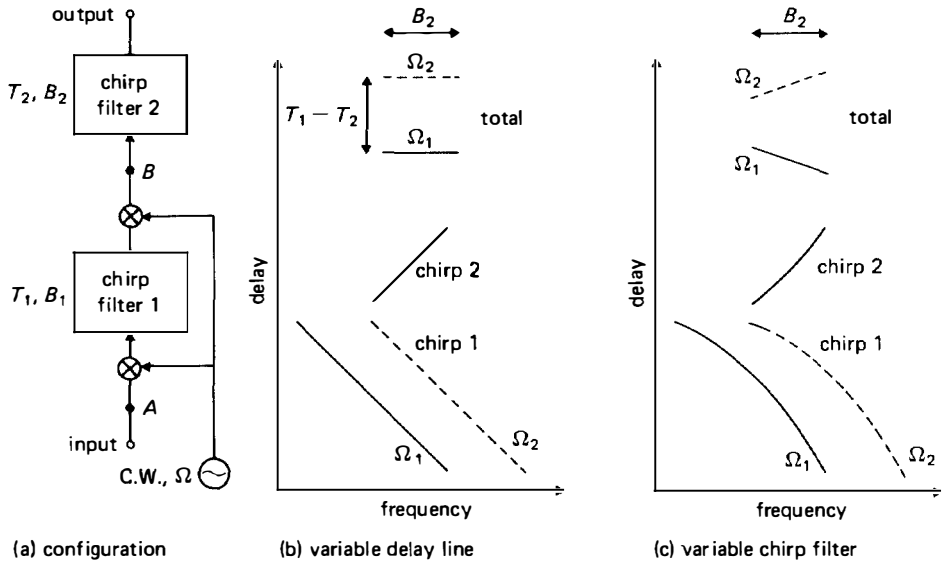


FIGURE 9.22. Signal processing using two chirp filters.

devices have non-linear chirps, with delays varying quadratically with frequency. The delay curves are shown in Figure 9.22(c). The quadratic terms are arranged to cancel so that the total delay is a linear function of frequency. Thus the overall system behaves as a linear chirp filter with bandwidth equal to B_2, the bandwidth of filter 2, and with dispersion T dependent on Ω. The system can even be designed such that the chirp slope can change sign, as shown in the figure. An experimental system [346] had a 100 MHz bandwidth, and enabled the dispersion to be varied from 0 to 1 μsec. It should be noted that systems using the configuration of Figure 9.22 are asynchronous, that is, there is no internal timing so that the signal processing is carried out irrespective of the signal time of arrival.

Chirp filters can also be used to perform *time scaling* [347, 348], using the system of Figure 9.23. Here there are two convolutions with chirp waveforms and two multiplications with chirps. The chirp slopes of the first three chirps are made to satisfy the relation $\alpha + \beta + \gamma = 0$. If $f(t)$ is the input waveform, it can be shown [348, p. 203] that the output $g(t)$ is given by

$$g(t) = f(-\gamma t/\alpha)$$

apart from a constant multiplier. Thus, if α and γ have different signs the output is an expanded or compressed version of the input waveform, with a scaling factor of $|\gamma/\alpha|$. If α and γ have the same sign, the output waveform is also reversed in time. The final multiplication may of course be omitted if the phase of the output waveform is not required.

Many of the above examples have counterparts in optical systems, as discussed by Papoulis [348]. Multiplication by a linear chirp waveform is mathematically equivalent to the action of a lens, while convolution with a chirp is equivalent to optical propagation in the space between two parallel planes. Thus the C–M–C Fourier transform system corresponds to the optical system in which the light distribution arriving at one focal plane of a lens is the Fourier transform of the distribution incident at the other focal plane. The time-scaling system described above corresponds to image formation by a lens; the time-scaling corresponds to the change of image size, related to the distances of the object and image planes from the lens, while time-reversal corresponds to the formation of an inverted image. Papoulis gives several other examples.

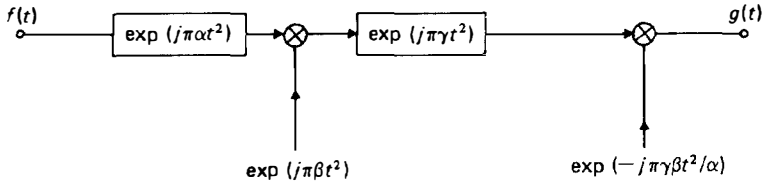


FIGURE 9.23. Time scaling using chirp filters.