Chapter 8

Bandpass Filters

In this chapter we consider the surface-wave bandpass filter, whose basic function is to pass signals with frequencies within a specified band, known as the pass-band, and to reject signals with frequencies outside this band. This is one of the commonest applications of surface waves, exemplified by the very large volume production of surface-wave filters for colour television receivers. A basic principle employed in most bandpass filters is that of apodisation, which enables an interdigital transducer to be designed such that its frequency reponse approximates a required response, as already discussed in Chapter 1. The earliest filters [231, 232] used this principle, and a variety of complementary techniques were developed later.

There are several important distinctions between the surface-wave filter and the more familiar *L–C* filter, comprising a network of inductors and capacitors. The response of an *L–C* filter is usually considered in terms of poles and zeros, in the complex frequency plane, and the design problem is expressed in terms of finding appropriate locations for these poles and zeros. For a surface-wave filter the response has no poles, and consequently quite different approaches must be used in design, similar to the methods used for digital finite-impulse-response filters. The number of zeros can be as large as the number of electrodes, typically several hundred; this is much larger than the number of poles or zeros for typical *L–C* filters. In addition, the contribution due to any one electrode is often accurate to 1% or better. These features enable surface-wave filters to achieve impressive performances with, for example, very flat pass-bands, sharp skirts and good stop-band attenuation. The phase of the frequency response may be either a linear or a non-linear function of frequency, and can be specified independently of the amplitude, a facility not available in the basic design method for *L–C* filters. Another distinction lies in the fact that surface-wave transducers give zero response at zero frequency. Thus, surface-wave devices cannot be used for low-pass filtering.

Suitable methods for analysis of interdigital bandpass filters are given in Chapter 4, and the relevant propagation effects are described in Chapter 6. This chapter is therefore concerned with the design and performance of the devices. The first two sections are concerned with apodised transducers. Section 8.1 considers the close analogy between an apodised transducer and a transversal filter, and Section 8.2 uses

this analogy in describing transducer design techniques. Section 8.3 is mainly concerned with withdrawal weighting, a technique alternative to apodisation. In Section 8.4 the design and performance of interdigital filters are described, concluding with some remarks on other types of surface-wave bandpass filter. Finally, Section 8.5 discusses filter banks, which are essentially arrays of bandpass filters.

In common with many other devices, including most surface-wave devices and L–C filters, surface-wave bandpass filters can be described as *linear filters*. The meaning of this term is explained in Appendix A, Section A.2, which also defines the important terms "frequency response" and "impulse response". In practice some non-linear effects do occur, as decribed in Section 6.3. However, if it is assumed that any signal applied to the device has a small enough power level, the device will behave as a linear filter to a very good approximation. This is nearly always true in practice, and is assumed to be the case throughout this chapter.

8.1. APODISED TRANSDUCER AS A TRANSVERSAL FILTER

We consider here a filter comprising two transducers, with one apodised and the other unapodised, as illustrated in Figure 8.1. The transducers are assumed to have regular electrodes and are taken to be of the two-terminal type, that is, multi-phase transducers are excluded. It is assumed that the only acoustic wave present is a piezoelectric Rayleigh wave, and that electrode interactions, propagation loss and surface-wave diffraction are negligible.

It was shown in Chapter 4 that the short-circuit response $H_{sc}(\omega)$ of this device is essentially the product of the two transducer responses $H_t^a(\omega)$ and $H_t^b(\omega)$. From equation (4.127),

$$H_{sc}(\omega) \equiv I_{sc}/V_t = H_t^a(\omega)\, H_t^b(\omega) \exp(-jk_0 d), \tag{8.1}$$

where d is the separation between the transducer acoustic ports and $k_0 = \omega/v_0$ is the free-surface wavenumber at frequency ω. V_t is the voltage applied to the input

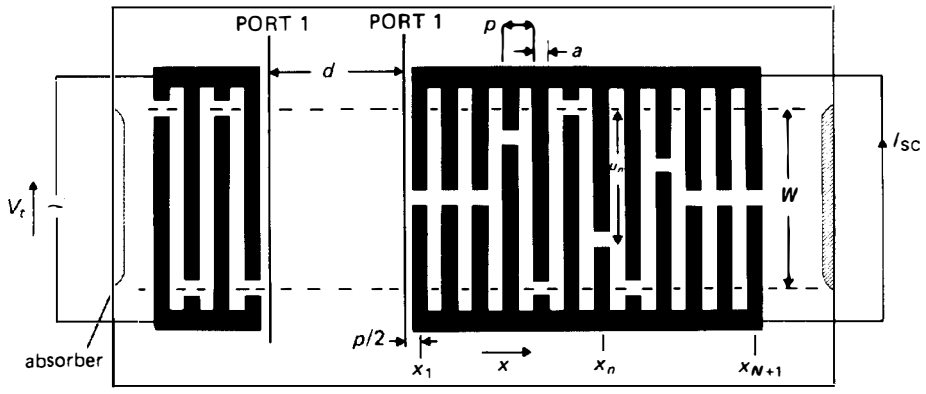

FIGURE 8.1. Bandpass filter using apodised transducer with regular electrodes.

transducer, and I_s the current produced by the shorted output transducer. Distortions due to the circuit effect, which arise when the transducers are connected to finite impedances, are ignored for the present but will be considered in Section 8.4.2.

Here we consider the design of the transducers such that the device response $H_s(\omega)$ meets some required specification. For convenience we assume initially that the unapodised transducer has already been designed, so that its response can be predicted. The problem is therefore reduced to that of designing the apodised transducer, whose required response is simply the required device response divided by the response of the unapodised transducer, as shown by equation (8.1).

In this section it is shown that the transducer geometry can be obtained by sampling a waveform $v(t)$, defined such that its Fourier transform $V(\omega)$ is closely related to the required frequency response of the transducer. The calculation of $v(t)$ itself is considered in Section 8.2.

8.1.1. Transversal Filter Analogy

The analysis of Section 4.7.3 shows that, for regular electrodes, the frequency response of an apodised transducer is given by equation (4.131). With minor changes of notation, this reads

$$H_t(\omega) = E(\omega) \sum_{n=1}^{N} v_n \exp\left[-jk_0(x_n - x_1 + p)\right], \tag{8.2}$$

where

$$v_n = (u_{n+1} - u_n)/W. \tag{8.3}$$

As shown in Figure 8.1, p is the electrode pitch, x_n is the centre location of electrode n, u_n is the location of the break in electrode n, N is the number of gaps and W is the aperture. The response can be regarded as a sum of contributions due to gap elements; the term $(x_n - x_1 + p)$ is the distance between element n, located at $x = x_n + p/2$, and the transducer acoustic port, which is here taken to be at $x = x_1 - p/2$. The term $E(\omega)$ is the gap element factor, given by

$$E(\omega) = (\omega W \Gamma_s)^{1/2} \bar{\varrho}_g(k_0), \tag{8.4}$$

where Γ_s is the piezoelectric coupling factor of equation (4.21). The function $\bar{\varrho}_g(k_0)$ is the Fourier transform of the elemental charge density for gap elements, given by equation (4.96), and is plotted in Figure 8.2 for two values of the metallisation ratio a/p. Note that $\bar{\varrho}_g(k_0)$ is zero when $k_0 p$ is a multiple of 2π, that is, at frequencies such that p is a multiple of the surface-wave wavelength; however, for frequencies not close to these points it varies quite slowly with frequency. Also, $\bar{\varrho}_g(k_0)$, and therefore $E(\omega)$, is imaginary. It is assumed that a few guard electrodes (Section 4.5.1) are included at each end of the transducer in order to minimise end effects.

The above form for the frequency response is also given quite simply by the delta-function model of Section 4.1, though this does not give the element factor. An alternative formulation in terms of elements associated with electrodes rather than gaps is given in Section 4.7.3, equation (4.129); for this case v_n in equation (8.2) is

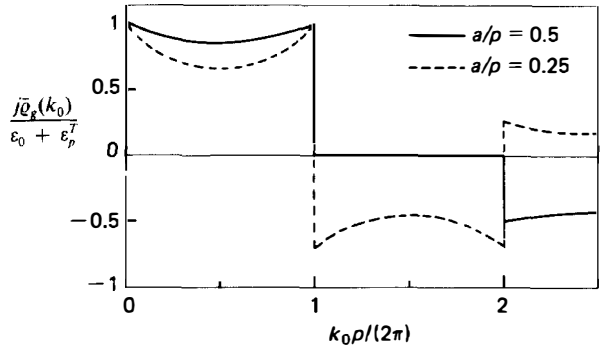

FIGURE 8.2. Elemental charge density for gaps, in the frequency domain.

taken as u_n/W and a different element factor is used. In this chapter the design is considered in terms of gap elements, though the use of electrode elements is equally valid and the design procedures are very similar.

It is convenient to define $\tau_s = p/v_0$ as the delay corresponding to the distance between successive gaps. Noting that $x_{n+1} = x_n + p$, equation (8.2) becomes

$$H_t(\omega) = E(\omega) \sum_{n=1}^{N} v_n \exp(-jn\omega\tau_s). \tag{8.5}$$

This shows that the transducer behaves essentially as a *transversal filter*, a conceptual device shown in Figure 8.3. Here an ideal tapped delay line, with regularly spaced taps, produces delayed replicas of an input waveform. These replicas, with delays $n\tau_s$, are weighted using real amplitude coefficients v_n and then summed to give the output waveform. Since the frequency response of an ideal delay line with delay τ is $\exp(-j\omega\tau)$, the frequency response of the transversal filter is

$$H_s(\omega) = \sum_{n=1}^{N} v_n \exp(-jn\omega\tau_s). \tag{8.6}$$

This is the same as the transducer response, equation (8.5), except for omission of the element factor $E(\omega)$. The significance of this comparison is that a variety of techniques have been developed for transversal filter design, determining the coefficients v_n required for the response $H_s(\omega)$ to meet a specification. These techniques can be

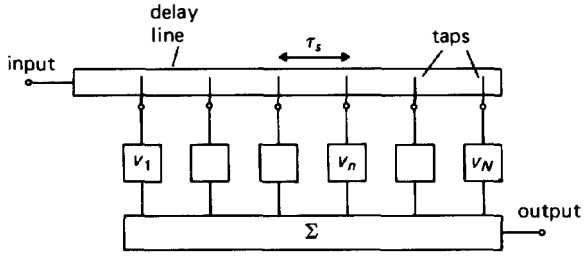

FIGURE 8.3. Transversal filter.

applied to transducer design if the required response is first divided by $E(\omega)$; the resulting values of v_n then give the transducer geometry, using equation (8.3).

The transversal filter concept was introduced by Kallman [233] as a versatile technique for obtaining accurate frequency responses. In addition to surface-wave devices, there are several other technologies using the same principle. For example, a transversal filter can be realised by means of a long cable with weakly-coupled taps, or by a charge-coupled device, in which charge packets representing the input signal are transferred sequentially along an array of electrodes in a metal-oxide-semiconductor structure [234]. Another realisation is a type of digital filter, which simply computes the output waveform in real time [235–237]. The digital realisation is called a "finite impulse response" (FIR) filter to distinguish it from the recursive digital filter, which employs feedback and gives an impulse response of infinite duration.

Although these devices are all based on the same concept, there are some important practical differences. Surface-wave devices give zero response at zero frequency, and so cannot be used for low-pass filtering. Also, the elements in a surface-wave device can have non-uniform spacing, though here uniform spacing is assumed except for some remarks at the end of Section 8.1.3. Another distinction is that a surface-wave device must have two transducers, and so is essentially *two* transversal filters, giving a useful degree of design flexibility. Charge-coupled devices and digital filters sample the input waveform before processing it, and so, unlike surface-wave devices, are not strictly linear filters.

8.1.2. Sampling and Surface-wave Transducers

Transforming equation (8.6) from the frequency domain to the time domain gives the impulse response $h_s(t)$ of the transversal filter, which is the delta-function sequence

$$h_s(t) = \sum_{n=1}^{N} v_n \delta(t - n\tau_s). \tag{8.7}$$

This equation also follows directly from Figure 8.3. To appreciate the design principle of a transversal filter, it is supposed that we can define a smooth real function $v(t)$, such that its values at times $n\tau_s$ are equal to the coefficients v_n. We also assume that $v(t) = 0$ for $t \leqslant 0$ and for $t > N\tau_s$. With these assumptions, equation (8.7) may be written in the form

$$\begin{aligned} h_s(t) &= \sum_{n=-\infty}^{\infty} v(n\tau_s)\delta(t - n\tau_s) \\ &= v(t) \sum_{n=-\infty}^{\infty} \delta(t - n\tau_s). \end{aligned} \tag{8.8}$$

This is a sampled form of the waveform $v(t)$. The quantity τ_s is called the *sampling interval*, and since we take $\tau_s = p/v_0$ this is equal to the delay corresponding to the electrode spacing in a transducer. The corresponding frequency $\omega_s = 2\pi/\tau_s$ is the

sampling frequency, and at this frequency the electrode pitch p is equal to the surface-wave wavelength.

The frequency response $H_s(\omega)$ of the transversal filter is the Fourier transform of $h_s(t)$. If $V(\omega)$ is the Fourier transform of $v(t)$, $H_s(\omega)$ can be expressed as

$$H_s(\omega) = \frac{\omega_s}{2\pi} \sum_{m=-\infty}^{\infty} V(\omega - m\omega_s). \tag{8.9}$$

This follows from standard theorems of Fourier analysis given in Appendix A, equations (A.20), (A.23) and (A.42). Figure 8.4 illustrates the relationship, showing the magnitudes of $V(\omega)$ and $H_s(\omega)$. The terms in the above equation with $m \neq 0$ are shown by broken lines. The magnitude of $V(\omega)$ is symmetrical about $\omega = 0$ because $v(t)$ is real. The figure assumes that $V(\omega)$ is a *bandpass* function, that is, for positive frequencies its magnitude is negligible except in the frequency range between two points ω_1 and ω_2. This condition is always valid for surface-wave transducer design.

In Figure 8.4 it has been assumed that the sampling frequency ω_s exceeds $2\omega_2$. For this case, the individual terms in equation (8.9) do not overlap, because $V(\omega)$ is negligible for $|\omega| > \omega_2$. In particular the "fundamental" component of $H_s(\omega)$, the term with $m = 0$, is essentially the same as $V(\omega)$. Thus,

$$H_s(\omega) = \omega_s V(\omega)/(2\pi), \quad \text{for } |\omega| \leqslant \tfrac{1}{2}\omega_s, \tag{8.10}$$

provided ω_s exceeds a minimum value $2\omega_2$, which is known as the *Nyquist frequency*. This result gives an important part of the design procedure for a transversal filter. It is assumed that the required response is specified for frequencies up to ω_2, at which point its magnitude has fallen to zero. We first generate a finite-length waveform $v(t)$ such that its transform $V(\omega)$ is a good approximation to the required response for $|\omega| \leqslant \omega_2$, and is negligible for $\omega > \omega_2$. We then sample $v(t)$ with some sampling frequency $\omega_s \geqslant 2\omega_2$, giving real weighting coefficients $v_n = v(n\tau_s)$. The frequency response $H_s(\omega)$ of the transversal filter is then a good approximation to the required response, for $|\omega| \leqslant \frac{1}{2}\omega_s$, apart from a multiplying constant. Since this procedure is valid for *any*

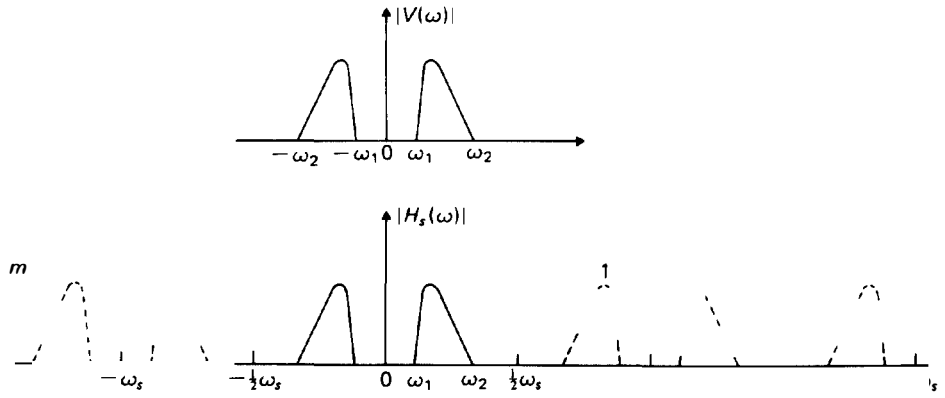

FIGURE 8.4. Frequency response of a transversal filter.

specified frequency response, the transversal filter is very flexible. In practice, some complications arise in the calculation of $v(t)$, and the methods used for this are considered in Section 8.2 below.

In addition to the required response, the transversal filter also gives "image" responses at frequencies above $\frac{1}{2}\omega_s$, as shown by the broken lines in Figure 8.4. It is usually necessary to suppress these by low-pass filtering. The sampling frequency ω_s is usually chosen to be somewhat larger than the Nyquist frequency $2\omega_2$, so that the low-pass filter does not need to have a sharp cut-off.

If the sampling frequency ω_s is less than the Nyquist frequency $2\omega_2$, the individual terms in equation (8.9) will in general overlap. This phenomenon is known as "aliasing". In this case, the original spectrum $V(\omega)$ cannot be recovered from $H_s(\omega)$ by low-pass filtering.

For a surface-wave transducer we have, from equations (8.5) and (8.6), $H_t(\omega) = E(\omega)H_s(\omega)$. Thus if the sampling frequency exceeds the Nyquist frequency, the transducer frequency response is given by

$$H_t(\omega) = \omega_s E(\omega)V(\omega)/(2\pi), \quad \text{for } |\omega| \leqslant \tfrac{1}{2}\omega_s. \tag{8.11}$$

The design procedure is therefore the same as for a transversal filter, except that the required frequency response is first divided by the element factor $E(\omega)$ in order to obtain $V(\omega)$. Although $E(\omega)$ has a sequence of zeros, these do not affect the method because they occur at zero frequency and at multiples of ω_s, where the transversal filter response $H_s(\omega)$ is zero.

It should be noted that, although the transversal filter has an impulse response with the simple form of equation (8.7), there is no corresponding expression for the transducer because of the frequency-domain distortion caused by the element factor, as shown by equation (8.5). However, the transducer response in the *fundamental* pass-band is given by equation (8.11), and if the bandwidth is not too large the slowly-varying element factor $E(\omega)$ has little effect, so that the response is approximately proportional to $V(\omega)$. Thus $v(t)$ can be regarded approximately as the transducer impulse response, provided the image responses are of no interest. This concept was mentioned earlier in Section 4.1, and was used in Hartmann's impulse model [198].

Note also that $V(\omega)$ was taken to be a bandpass function. This is not in fact valid, because the finite length of $v(t)$ implies, according to Fourier analysis, that $V(\omega)$ must have infinite extent in the frequency domain. However, it will be shown in Section 8.2 that $V(\omega)$ can be designed such that its amplitude outside a specified band is very small. Assuming that this is done, the above analysis can be taken to be valid to a good approximation, and can then provide the basis for a valid design technique.

8.1.3. Examples of Particular Cases

Here we give some illustrations of the relationship between the "impulse response" $v(t)$, the transducer geometry, and the transducer frequency response. It should be noted that $v(t)$ may be delayed without significantly affecting the transducer response – the delay of $v(t)$ causes a corresponding group delay, independent of frequency, in the transducer frequency response. For practical applications this

change of response is usually insignificant, though the transducer geometry may be changed substantially.

Since $V(\omega)$ is a bandpass function, the waveform $v(t)$ will be oscillatory and is conveniently written in the form

$$v(t) = \hat{a}(t) \cos [\omega_r t + \hat{\theta}(t)], \tag{8.12}$$

where ω_r is some reference frequency between ω_1 and ω_2, and $\hat{a}(t)$ is the envelope. If the phase $\hat{\theta}(t)$ is a non-linear function of t, the waveform is phase-modulated; it will also be amplitude-modulated if $\hat{a}(t)$ varies with t. The term "amplitude-modulated waveform" is used when there is no phase modulation, so that $\hat{\theta}(t)$ is a linear function of t. For this case ω_r may be chosen such that $\hat{\theta}(t)$ is a constant, and ω_r is then the carrier frequency, ω_c.

A simple example of an amplitude-modulated waveform is a pulse of carrier with a rectangular envelope, as shown in Figure 8.5(a). This figure shows samples taken at a sampling frequency $\omega_s = 4\omega_c$, so that there are four samples per period. With the waveform positioned such that the samples occur where the phase is a multiple of $\pi/2$, the weights v_n give the sequence 0, 1, 0, -1, 0, 1 . . . . When these are interpreted as interelectrode gaps $(u_{n+1} - u_n)/W$, as in equation (8.3), the transducer geometry is seen to be the conventional uniform double-electrode type, which is unapodised. A few zero-valued samples can be added beyond the ends of the waveform $v(t)$ to provide guard electrodes, minimizing end effects. Figure 8.5(b) shows the same case, but with the waveform $v(t)$ displaced slightly so that the samples occur where the phase is $\pi/4 + n\pi/2$. In this case the transducer is apodised, even though its frequency response is essentially the same. Similar observations apply if three samples are taken per wavelength, so that $\omega_s = 3\omega_c$. In this case a uniform $S_e = 3$ transducer is produced if the waveform is sampled at points where the phase is a multiple of $\pi/3$.

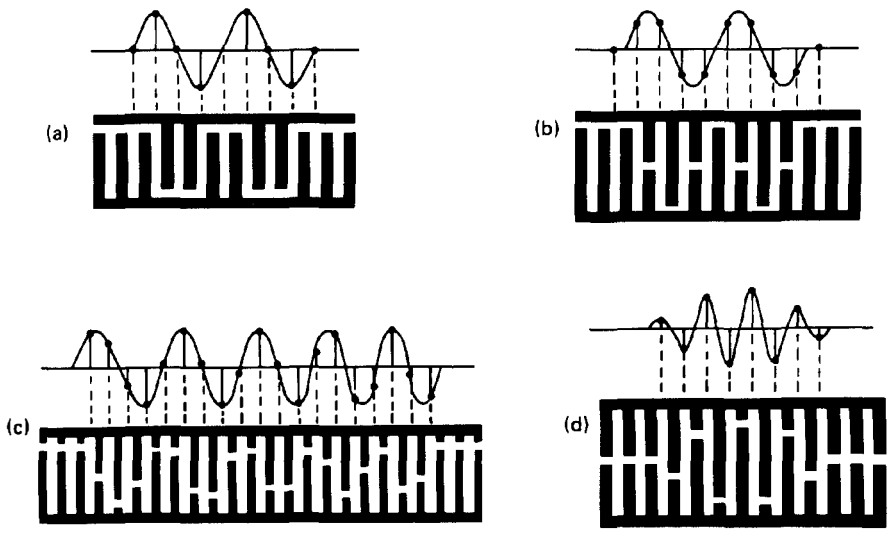

FIGURE 8.5. Examples of transducer design using regular sampling.

The transducer response remains essentially the same whatever value of the sampling frequency ω_s is used, provided it exceeds the Nyquist frequency $2\omega_2$. However, it is usually preferable to avoid apodised designs where possible, since they are more prone to errors due to diffraction. Unapodised designs can be obtained only when ω_s is a multiple of the carrier frequency ω_c. The image responses, shown by broken lines in Figure 8.4, are centred at $m\omega_s \pm \omega_c$, and become harmonics of the fundamental when ω_s is a multiple of ω_c. Similar considerations apply if the envelope of $v(t)$ is not flat; in this case rapid changes of apodisation are generally minimised by choosing ω_s to be a multiple of ω_c.

Figure 8.5(c) shows a transducer design obtained by sampling a phase-modulated waveform, taking a chirp waveform as an example. For a phase-modulated waveform an apodised transducer design is obtained if uniformly-spaced samples are used. However, chirp waveforms are usually sampled non-uniformly, and chirp transducers using this principle are discussed in Chapter 9.

Figure 8.5(d) shows an amplitude-modulated waveform, with carrier frequency ω_c, sampled with two samples per period, so that $\omega_s = 2\omega_c$. This gives a single-electrode apodised transducer, discussed using the delta-function model in Section 4.1. If the envelope of the waveform is flat, a uniform single-electrode transducer is produced. The sampling theory given above is *invalid* for this case because, since $\omega_2 > \omega_c$, the sampling frequency $\omega_s = 2\omega_c$ is less than the Nyquist frequency $2\omega_2$. However, it is shown in Section A.5 that, if $v(t)$ is an *amplitude-modulated* waveform, the sampling procedure remains valid if the particular sampling frequency $\omega_s = 2\omega_c$ is chosen. The transducer response $H_t(\omega)$ is then proportional to $E(\omega)V(\omega)$ for $|\omega| < \omega_s$. It is of course also necessary to ensure that the samples are not taken at the zeros of the waveform. This case is of practical importance because the low sampling frequency implies that the electrodes are relatively wide, making the fabrication easier.

Waveform Characteristics. Some relationships between a bandpass waveform $v(t)$ and is spectrum $V(\omega)$ are given in Appendix A, Section A.5. Bandpass filters are often required to have a frequency response whose phase is a linear function of frequency, since this phase variation causes a delay of an applied signal but does not distort it. This condition is satisfied if $V(\omega)$ has linear phase, since $H_t(\omega)$ is proportional to $E(\omega)V(\omega)$, by equation (8.11). Section A.5 shows that this implies that the waveform $v(t)$ of equation (8.12) must have an envelope $\hat{a}(t)$ which is symmetric or anti-symmetric about some time $t = t_0$. Also, the phase $\hat{\theta}(t)$ should be a constant plus a term anti-symmetric about t_0; it may for example be linear with t. The phase of $V(\omega)$ is then $-\omega t_0$ plus a constant, and the group delay is therefore t_0.

It is also shown in Section A.5 that, if $v(t)$ is an amplitude-modulated waveform, so that there is no phase modulation, its spectrum $V(\omega)$ has a magnitude symmetric about the carrier frequency ω_c, and the phase of $V(\omega)$ is a constant plus a function anti-symmetric about ω_c. For the particular cases when the envelope $\hat{a}(t)$ is symmetric or anti-symmetric about $t = t_0$, the phase of $V(\omega)$ varies linearly with frequency, and the group delay is t_0.

For many surface-wave filters the frequency response is required to have its magnitude symmetrical about a centre frequency ω_c, and its phase linear with ω. If

$V(\omega)$ is taken to have these characteristics, the waveform $v(t)$ will be an amplitude-modulated waveform, and can therefore be sampled with a sampling frequency $\omega_s = 2\omega_c$, giving a single-electrode transducer. However, the transducer response $H_t(\omega)$ is in fact proportional to $E(\omega)V(\omega)$ and, for accurate results, the distortion due to the element factor $E(\omega)$ must be compensated for, particularly if the bandwidth is large. If this is done, $V(\omega)$ will not have a symmetrical magnitude, and therefore $v(t)$ must have some phase modulation. It is then necessary to use a higher sampling frequency, above the Nyquist rate. Alternatively, the waveform $v(t)$ may be sampled non-uniformly, as discussed below.

Non-uniform Sampling. It has been assumed above that the waveform $v(t)$ is sampled uniformly, with sample spacing τ_s corresponding to the electrode spacing. However, for surface-wave transducers this is not a necessary constraint, as is clearly shown by considering the chirp waveform of Figure 8.5(c), for example. For this case it is usual to sample at the peaks and troughs of the waveform, so that the sample spacing varies and the electrode pitch p varies along the length; the result is a chirp transducer, considered further in Chapter 9.

This principle can also be applied to bandpass filter design, for cases where $v(t)$ has phase modulation. If this is done the analysis above is invalid, though the analysis in Chapter 9 can be applied provided the sample spacing does not vary rapidly. The non-uniform sampling causes some distortion in the response of the transducer. However, for bandpass filters it is often the case that the sample spacing is almost uniform and it is then found that the results of this section are approximately valid in the fundamental pass-band.

Non-uniform sampling has been used quite extensively in bandpass filters [238]. The main advantage is that for phase-modulated waveforms it is not necessary to sample above the Nyquist rate, so that a single-electrode transducer, with relatively wide electrodes, can be used. However, the method becomes invalid if the sample spacing varies rapidly, since the analysis breaks down in this case.

8.2 DESIGN OF APODISED TRANSDUCERS

It was shown above that an apodised transducer can be designed by sampling a continuous waveform $v(t)$, giving a transducer response proportional to $E(\omega)V(\omega)$, where $E(\omega)$ is the element factor. The sample values $v_n = v(n\tau_s)$ give the transducer geometry, as shown by equation (8.3). Methods of obtaining $v(t)$ are considered in Section 8.2.1 below. The main consideration is that $v(t)$ must have finite length. It is shown that the use of window functions enables $v(t)$ to be designed such that the transducer frequency response is a good approximation to some required response.

Several other design techniques are considered in Section 8.2.2. Most of these consider the problem in terms of designing the weights v_n, without considering the continuous waveform $v(t)$ as an intermediate step. Thus the sampling theory of Section 8.1 above need not be considered explicitly. Section 8.2.3 is concerned with

the design of minimum-phase filters, where only the amplitude of the frequency response is specified, and the design is required to minimise the group delay.

Although this section is primarily concerned with apodised transducers, the basic methods are applicable to a wide variety of finite-impulse-response filters; in particular they can be applied to surface-wave transducers with other types of weighting, and to digital filters. Digital design techniques are reviewed in, for example references [235–237], while the methods for surface-wave filters are reviewed in references [238–241].

8.2.1. Use of Window Functions

Suppose that the response required of the transducer is some function $H_0(\omega)$. Since the actual response $H_t(\omega)$ is proportional to $E(\omega)V(\omega)$, it appears that we can divide $H_0(\omega)$ by $E(\omega)$ to obtain $V(\omega)$, and then take the inverse Fourier transform to obtain $v(t)$. However, this is not acceptable in practice because it gives a waveform $v(t)$ of infinite length, and hence the transducer response cannot be exactly equal to $H_0(\omega)$.

To allow for this, we define a function $V_0(\omega) = H_0(\omega)/E(\omega)$ with inverse Fourier transform $v_0(t)$, which will have infinite duration. The design problem is then to find a function $v(t)$, of finite length, such that its transform $V(\omega)$ is a good approximation to $V_0(\omega)$; the transducer response $H_t(\omega)$ will then be a good approximation to $H_0(\omega)$. The function $v(t)$ is sampled to give the transducer geometry, as explained in the previous section. The accuracy of the approximation will depend on the method used to evaluate $v(t)$, and it will be shown that the accuracy obtainable generally increases with the duration of $v(t)$. Note that $V_0(\omega)$ must be a bandpass function, as required for the sampling technique, but is otherwise arbitrary.

An obvious method of obtaining $v(t)$ is simply to truncate $v_0(t)$ to a finite length. For present purposes, it is convenient to represent this as a multiplication by a *window* $W(t)$, so that

$$v(t) = W(t)v_0(t). \tag{8.13}$$

This equation truncates $v_0(t)$ if $W(t)$ is unity for some finite time interval, and is zero for other t; however other forms of $W(t)$ will be considered later. In the frequency domain, the multiplication in equation (8.13) causes $V_0(\omega)$ to be convolved with $\bar{W}(\omega)$, the Fourier transform of $W(t)$, so that

$$V(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} V_0(\omega')\bar{W}(\omega - \omega')\,\mathrm{d}\omega'. \tag{8.14}$$

Equation (8.13) truncates the waveform $v_0(t)$ to a length T if the window is taken as

$$W(t) = \mathrm{rect}(t/T) \tag{8.15}$$

where $\mathrm{rect}(x) = 1$ for $|x| \leqslant \frac{1}{2}$ and is zero for other x. Strictly speaking this form for $W(t)$ is not valid, because causality requires that $v(t)$ should be zero for $t < 0$. However causality can easily be satisfied at a later stage simply by delaying $v(t)$; in the frequency domain, this adds a phase term proportional to ω, as whown by the shifting theorem. The Fourier transform of equation (8.15) is

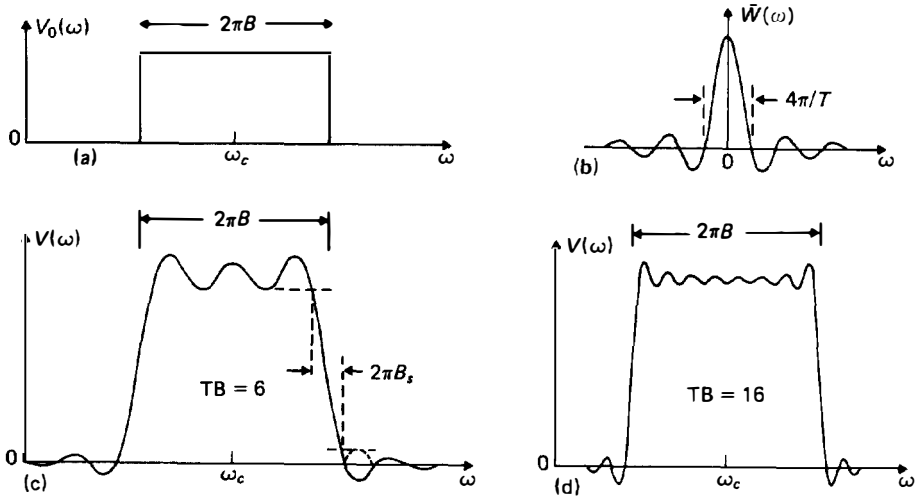

FIGURE 8.6. Effect of time-domain truncation, for a filter with an ideally rectangular frequency response.

$$\bar{W}(\omega) = T \operatorname{sinc}(\tfrac{1}{2}\omega T) \tag{8.16}$$

where sinc $x = (\sin x)/x$.

To illustrate the effect of truncation, suppose that the required frequency response $V_0(\omega)$ is flat within a band of width $\Delta\omega = 2\pi B$ and zero outside this band, so that

$$\begin{aligned} V_0(\omega) &= 1, \quad \text{for } |\omega - \omega_c| \leqslant \pi B, \\ &= 0, \quad \text{for } |\omega - \omega_c| > \pi B. \end{aligned} \tag{8.17}$$

The inverse transform, $v_0(t)$, of this function is an amplitude-modulated waveform with carrier frequency ω_c and envelope proportional to sinc (πBt), and is therefore infinite in length. Figure 8.6(a) shows $V_0(\omega)$, and Figure 8.6(b) shows the spectrum $\bar{W}(\omega)$ of the window function, equation (8.16). These are convolved in accordance with equation (8.14) to give $V(\omega)$, the spectrum of the finite-length waveform $v(t)$. This is shown in Figure 8.6(c) for $TB = 6$ and in Figure 8.6(d) for $TB = 16$. In comparison with the ideal response $V_0(\omega)$, the actual response $V(\omega)$ exhibits ripples in the pass-band and sidelobes in the stop bands. These are due to the sidelobes of the function $\bar{W}(\omega - \omega')$ in equation (8.14). At any ω, some of the sidelobes of $\bar{W}(\omega - \omega')$ are in the band occupied by $V_0(\omega')$ and therefore contribute to the integral; as ω changes, sidelobes enter the band at one side and leave at the other side, thus giving an oscillatory contribution. In addition the transitions at the band edges are no longer sharp, so that $V(\omega)$ has finite skirt widths; these are associated with the width of the main peak of the function $\bar{W}(\omega)$.

Although the above refers to a specific case, the pass-band ripple, stop-band sidelobes and skirt broadening are in fact quite general phenomena in finite-impulse-response filters, present to some extent irrespective of the design

method used. The distortions arise basically because the finite length of $v(t)$ implies that $V(\omega)$ cannot be strictly confined to a band of finite width, hence the presence of the stop-band sidelobes.

The distortion due to truncation can be reduced by increasing the window length T, as can be seen by comparing the results for $TB = 6$ and $TB = 16$ shown in Figures 8.6(c) and (d). A larger value of T reduces the skirt width and the ripple at the band centre. However, the ripple near the band edge, and the sidelobes near the band edge, have magnitudes almost independent of T; the largest sidelobe is 21 dB below the pass-band level then TB is large. For practical purposes this is usually unacceptable.

The solution to this problem is to modify the weighting function $W(t)$. The basic requirement is for a function of finite length, whose Fourier transform has sidelobes substantially less than those of the sinc function. This requirement arises in several fields; in addition to surface-wave and digital filters, it arises in spectral estimation from finite-length records [242] and in the design of antennas whose polar diagrams need to have low sidelobes. Consequently, a variety of suitable window functions are available. An example is the *Kaiser* window [235], given by

$$W_K(t) = \frac{I_0[\alpha\sqrt{1 - 4t^2/T^2}]}{I_0(\alpha)}, \quad \text{for } |t| \leqslant T/2,$$

$$= 0, \qquad \text{for } |t| > T/2, \tag{8.18}$$

where $I_0(x)$ is the modified Bessel function of the first kind and zero order. The parameter α is chosen to suit the application. Figure 8.7 shows $W_K(t)$ for $\alpha = 6$ and its transform $\bar{W}_K(\omega)$, which is given by a simple formula [241]. The figure includes for comparison the function sinc $(\frac{1}{2}\omega T)$, which is the transform of the rectangular window function, as in equation (8.16). These frequency-domain functions are plotted logarithmically. It can be seen that the Kaiser function $\bar{W}_K(\omega)$ gives much smaller sidelobes, though its main peak is about twice as wide as that of the sinc function. Thus, when used for filter design, the Kaiser window gives smaller ripples and sidelobes but, for a given value of T, the skirt width is larger. The skirt width can

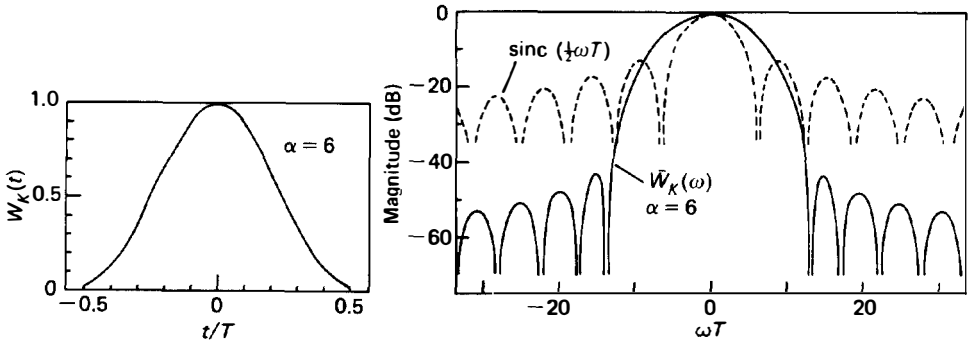

FIGURE 8.7. Kaiser window function $W_K(t)$ and its Fourier transform.

however be reduced by increasing T, since this simply reduces the width of $\bar{W}_K(\omega)$ in proportion.

The parameter α in equation (8.18) affects the sidelobe levels of $\bar{W}_K(\omega)$ and the width of the main peak. Larger values of α give smaller sidelobes and wider peaks, enabling these features to be traded against each other. When the Kaiser window is used to design a filter with an ideally rectangular frequency response [as in Figure 8.6(a)], α determines the pass-band ripple, stop-band attenuation and skirt width. Tancrell [239] gives these parameters as functions of α. For example, if $\alpha = 6$, the largest stop-band sidelobe is at -62 dB, the ripple at the band edge is 0.009 dB peak-to-peak, and the skirt width is $B_s = 4.0/T$, where B_s is in Hz and is defined as shown in Figure 8.6(c). To design the filter, the value of α is chosen first, such that the pass-band ripple and stop-band sidelobes are acceptable, and the value of T needed to give the required skirt width can then be deduced straightforwardly. Design data equivalent to Tancrell's are also given by Rabiner and Gold [237, p. 101].

A number of other window functions have been considered for surface-wave filters [239, 241], and Harris [242] has compared the performance of a large variety of windows for spectral estimation. In all cases, the window function in the time domain is real and symmetric about $t = 0$, so its transform in the frequency domain is real and symmetric about $\omega = 0$. None of these functions gives a performance appreciably better than the Kaiser window, in the context of filter design. A particular example is the *Dolph–Chebyshev* window [241, 243]. For a filter design using this window the stop-band sidelobes are all the same size, and the window is an optimum in the sense that the narrowest possible skirt width is obtained for a given sidelobe level. However, the Dolph–Chebyshev window is a complicated function, rather inconvenient for practical usage, and gives a performance little better than that of the Kaiser window [241]. Some other window functions are described in Section 9.2.3, in connection with chirp filters.

Having designed $v(t)$ with the aid of a window function, the final stage of the design procedure is to sample to obtain the weights $v_n = v(n\tau_s)$, as explained in Section 8.1. A complication arises here because the sampling theory of Section 8.1 assumes $V(\omega)$ to be a band-pass function. This is not in fact true here because $v(t)$ has finite length, hence the presence of the sidelobes of $V(\omega)$, extending indefinitely on either side of the pass-band. Thus, some aliasing occurs when $v(t)$ is sampled, so that the image responses generated by sampling contribute sidelobes in the required pass-band. However, the distortion due to this aliasing is usually small and acceptable, provided $V(\omega)$ is designed to have well-suppressed sidelobes. It is also necessary that the sidelobes should decrease in amplitude for frequencies more remote from the pass-band. This is the case for most window functions, including the Kaiser window. The Dolph–Chebyshev window is an exception, giving sidelobes of equal amplitude, but for this case a specially adapted form of the window function can be used to avoid the aliasing problem for sampled waveforms [243].

Note that, although the above discussion has illustrated the method for the case when the ideal response $V_0(\omega)$ is a rectangular function, the method is in fact quite general. Thus there are no constraints on either the amplitude or phase of $V_0(\omega)$, except that it must be a band-pass function.

8.2.2. Optimised Design Methods

Although the use of window functions described above is an effective and straightforward design technique, it does not give optimised designs. For example, the relative magnitude of the pass-band ripple and the stop-band sidelobes are determined by the weighting function, and the pass-band ripple is maximised at the band edges. For some cases a more sophisticated method is desirable, taking account of specified tolerances, so that an optimised design can be produced.

A quite simple approach is based on an iterative use of the Fourier transform. The required response $V_0(\omega)$ is first transformed to the time-domain, truncated to a finite length T, and then transformed back to the frequency domain. Due to the truncation, this new frequency response will generally fall outside the specified tolerances in some regions. The frequency response is modified to bring it within the tolerances and the procedure is the repeated, transforming to the time domain, truncating, and returning to the frequency domain. Provided T is large enough, iteration of this process generally gives progressive improvement, yielding finally a frequency response meeting the specification and a finite-length impulse response which gives the transducer geometry. This principle was used, with some modifications, by Moulding and Parker [244], and by De Vries [245]. Although the final design is not strictly optimal, the method has the advantages of simplicity and versatility; it can accommodate a non-linear phase in the frequency domain, and can allow for the tolerances being different in different frequency regions.

A variety of other optmisation techniques have been developed for design of digital finite-impulse-response filters [236, 237, 239], and these may be applied directly to the design of surface-wave transducers if the required response is first divided by $E(\omega)$ to give the required response $V_0(\omega)$ of the corresponding transversal filter. These methods give directly the weights v_n of the transversal filter, related to the transducer geometry by equation (8.3). The continuous waveform $v(t)$ is not considered explicitly, so the aliasing problem mentioned in Section 8.2.1 does not occur. A particular example, which has been used extensively for surface-wave filters, is that of McClellan and Parks [237, 246, 247]. Filters designed by this method have linear phase, so that the response $H_s(\omega)$ of the transversal filter has the form

$$H_s(\omega) = A(\omega) \exp [j(c - \omega t_0)], \tag{8.19}$$

for $\omega > 0$, where $A(\omega)$ is the amplitude and t_0 is a constant determined by the length of the response in the time domain. The constant c is either 0 or $\pi/2$. The method makes use of the Remez exchange algorithm to design the filter such that the amplitude response $A(\omega)$ is a good approximation to a required amplitude response $A_0(\omega)$. The design obtained minimises an error function E_A, defined by

$$E_A = \text{Max} \{e(\omega)|A(\omega) - A_0(\omega)|\}, \quad 0 < \omega < \tfrac{1}{2}\omega_s, \tag{8.20}$$

where $e(\omega)$ is an error weighting function chosen by the designer. The method thus allows the designer to specify that the response is required to be more accurate in some regions of the band than in others. For example, the tolerances on pass-band ripple and stop-band sidelobes can be specified independently. The accuracy obtained

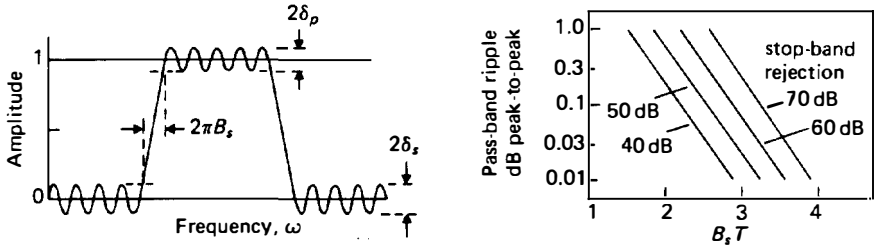

FIGURE 8.8. Response of equi-ripple filter, and approximate performance theoretically obtainable.

depends on the length of the impulse response, which is therefore increased if a better accuracy is needed.

If the required amplitude response $A_0(\omega)$ is rectangular, as in Figure 8.6(a), this method can give a design with an "equi-ripple" response, illustrated in Figure 8.8. This type of result is obtained if the tolerances in the pass-band and in the stop band are independent of frequency. The normalised amplitude in the pass-band oscillates between extrema at $1 \pm \delta_p$, and the sidelobes all have magnitude δ_s. Such a response is found to give the smallest possible skirt width consistent with a give maximum error in the pass-band and a given maximum sidelobe level. From studies of large numbers of designs of this type, it can be concluded that the skirt width B_s is approximately given by [248]

$$\log (\delta_p \delta_s) \approx -1.05 - 1.45 B_s T, \tag{8.21}$$

where T is the duration of the impulse response. This equation is useful for estimating the duration T needed to meet a given specification. Figure 8.8 shows the relation between pass-band ripple, stop-band rejection and skirt width B_s. Typical values of $B_s T$ are in the range 2 to 4.

Although the basic method assumes the required frequency response to have linear phase, it can be adapted quite readily to design filters with non-linear phase [249]. The adaptation exploits the fact that the basic method gives a response with the form of equation (8.19), where the phase constant c is either 0 or $\pi/2$. The constant is zero if the design is specified to be symmetric in the time domain, so that $v_n = v_{N+1-n}$, and is $\pi/2$ if the design is anti-symmetric, so that $v_n = -v_{N+1-n}$. The response of a filter with non-linear phase can be written

$$H(\omega) = A(\omega) \exp [j\phi(\omega)] \exp (-j\omega t_0) \tag{8.22}$$

This can be regarded as a sum of two linear-phase responses, an "in-phase" response with $c = 0$:

$$H^i(\omega) = [A(\omega) \cos \phi(\omega)] \exp (-j\omega t_0)$$

and a "quadrature" response with $c = \pi/2$:

$$H^q(\omega) = j[A(\omega) \sin \phi(\omega)] \exp (-j\omega t_0).$$

The two linear-phase responses can be designed by the basic method, taking the

required amplitude responses to be $A(\omega)\cos\phi(\omega)$ and $A(\omega)\sin\phi(\omega)$, with $c = 0$ and $\pi/2$ respectively. The number of time-domain samples, N, is taken to be the same for both responses, so that the value of t_0 is the same for both. The time-domain samples are at the same points, and are simply added to obtain the weights for the non-linear phase response. This method has been shown to be effective for design of a chirp type of response, where the phase $\phi(\omega)$ is a quadratic function of frequency [249].

A quite different design technique makes use of non-linear programming to optimise the response, and has recently been applied to surface-wave filter design [250].

8.2.3. Minimum-phase Filters

As already noted in Section 8.1.3, if the frequency response of a filter has a phase linear with ω, the group delay is equal to t_0, where t_0 is the centre point of the impulse response. Thus, if the impulse response has duration T, the group delay must be at least $T/2$. For some applications it is important to minimise the group delay, and we therefore consider whether the delay can be reduced without appreciably changing the magnitude of the frequency response. This implies that the frequency-domain phase will become a non-linear function of ω, causing some distortion of a signal applied to the device, but this will be acceptable if the phase non-linearity is small enough.

A particular case is the *minimum-phase* filter [251, 252]. If the response is $H(\omega) = A(\omega)\exp[j\phi(\omega)]$, the phase $\phi(\omega)$ is, for a minimum-phase filter, related to the amplitude by

$$\phi(\omega) = -\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\ln|A(\omega')|}{\omega-\omega'}\,d\omega' \tag{8.23}$$

Thus $\phi(\omega)$ is the Hilbert transform of $\ln|A(\omega)|$, that is, the convolution with $-1/(\pi\omega)$. The significance of this result is that, for any specified amplitude response $A(\omega)$, the magnitude of the phase $\phi(\omega)$ is smaller for a minimum-phase filter than for any other feasible filter response. It is assumed that the responses under consideration are causal, that is, the corresponding impulse responses are zero for $t < 0$. Equation (8.23) ensures this for the minimum-phase case. In comparison with a linear-phase filter with the same amplitude response, the minimum-phase filter is generally found to give a smaller group delay. Moreover, the phase non-linearity is often found to be acceptable. This is true particularly when the amplitude $A(\omega)$ varies relatively slowly with frequency, because functions of this type have Hilbert transforms that are relatively small and smoothly varying.

Given some required amplitude response, the phase required for a minimum-phase filter can be obtained from equation (8.23), and the filter may then be designed by the methods described earlier. For a transversal filter, an alternative method can be used [253, 254]. If we define $z = \exp(-j\omega\tau_s)$, the response $H_s(\omega)$ of a transversal filter, equation (8.6), becomes a polynomial in z, and can be specified by its zeros. These are the values of z, generally complex, at which the polynomial is zero. A linear-phase transversal filter is first designed, and its zeros evaluated. Some of the

zeros can then be deleted in a systematic way, such that the resulting new polynomial gives a minimum-phase filter with the same amplitude response.

For a surface-wave filter, the method is of course applied to the design of one of the two transducers [254]. In addition to giving a smaller delay, a minimum-phase design can also give a shorter transducer, making more economical use of the substrate area. It has also been found that the minimum-phase design is less affected by second-order effects [254].

8.3. THINNING AND WITHDRAWAL WEIGHTING

Withdrawal weighting is a technique in which selected sources are omitted from a transducer. This enables the transducer to be weighted without using apodisation, so that the apertures of all the remaining sources are the same. In narrow-band filters, which require transducers many wavelengths long, this is advantageous in reducing the perturbation due to diffraction. In addition, a withdrawal-weighted transducer can be used in conjunction with an apodised transducer, so that the flexibility obtained by using two weighted transducers becomes available for weakly piezoelectric substrates, such as quartz, on which a multi-strip coupler cannot be used.

It is convenient first to introduce a simple modification known as "thinning", which is often applied to narrow-band transducers. Figure 8.9(a) shows thinning as applied to a uniform unapodised transducer. The design is obtained simply by omitting some of the electrodes in a conventional transducer, such that the result obtained consists of a regular sequence of identical groups of electrodes. The nature of the response of this transducer is readily appreciated from sampling theory. Since the bandwidth of any one group of electrodes is much larger than the overall transducer bandwidth, the thinning process is approximately equivalent to sampling the impulse response with a sampling interval, τ_s', corresponding to the spacing of the groups. The sampling process causes image responses to appear in the frequency response, with frequency spacing $2\pi/\tau_s'$, as shown by equation (8.9). These image responses may be attenuated by filtering elsewhere, for example in a second transducer or in a matching network. Thus, for practical purposes the response of a thinned transducer can often be regarded as essentially the same as that of a

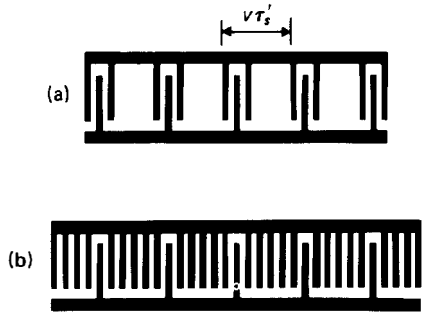

FIGURE 8.9. Two types of thinning, as applied to a uniform transducer.

conventional transducer, apart from a reduction of the strength of coupling to surface waves. An accurate analysis would of course need to account for the response of the individual groups of electrodes, and these are affected somewhat by end effects (Section 4.5.2).

The commonest reason for using thinning is to reduce electrode interaction effects in single-electrode transducers. For a shorted single-electrode transducer the acoustic reflection coefficient, which is due to electrode interactions, is approximately proportional to the number of electrodes and is therefore substantially reduced by thinning. Alternatively, interactions can be reduced by increasing the number of electrodes per period, as discussed in Section 4.2, but this requires narrower electrodes and so makes the fabrication more difficult. The scheme of Figure 8.9(a) is frequently used in delay-line oscillators, and also occurs in a modified form in PSK filters.

Thinning also has the effect of reducing the transducer admittance. The capacitance C_t and the centre-frequency conductance $G_a(\omega_c)$ are both reduced, while the transducer Q-factor $\omega_c C_t / G_a(\omega_c)$ is increased. These changes are sometimes beneficial in reducing the severity of circuit effects.

An alternative method of thinning, shown in Figure 8.9(b), is to change some of the electrode polarities so that selected gap elements are eliminated. This reduces the capacitance and conductance, though it does not substantially reduce electrode interaction effects. For this transducer the electrodes are regular and the individual groups are not affected by end effects. Either of the two types of thinning may also be applied to an apodised transducer.

Withdrawal weighting is illustrated in Figure 8.10. The technique is very similar to thinning, but here the thinning process is applied non-uniformly so that the remaining groups of sources have differing lengths. This enables the transducer to be designed such that its frequency response approximates a required response, without using apodisation. The technique was first demonstrated by Hartmann [255], who weighted

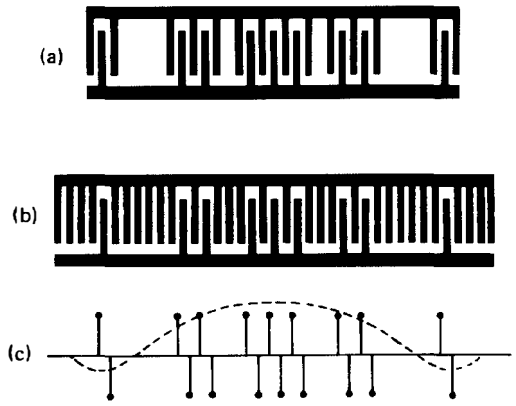

FIGURE 8.10. (a), (b): Two types of withdrawal weighting. (c): Approximate impulse response.

the transducer by omitting electrodes as in Figure 8.10(a). This is known as "electrode-withdrawal weighting". The alternative method of Figure 8.10(b), where selected gap sources are eliminated by changing the electrode polarities [256], is called "source-withdrawal weighting". Withdrawal weighting is suitable only for transducers with relatively narrow bandwidths. As in the case of a thinned transducer, additional responses arise at frequencies outside the main pass-band, so that the stop-band attenuation is poor. In practical devices, a withdrawal-weighted transducer is often used in conjunction with an apodised transducer, with the latter designed such that adequate stop-band attenuation is obtained.

Both types of withdrawal-weighted transducer may be analysed by the methods of Chapter 4, giving the frequency response, admittance and scattering parameters. For a source-withdrawal transducer the electrodes are regular and the response can be written in terms of an array factor and an element factor. This considerably simplifies the analysis, as explained in Section 4.5. Figure 8.10(c) shows the approximate impulse response, where each of the gap sources is represented by a delta function. The broken line indicates an approximate form for the envelope of the impulse response, obtained by smoothing out the actual envelope. The smoothed impulse response has a Fourier transform approximating the transducer frequency response, for frequencies in the main pass-band.

As for apodised transducers, withdrawal-weighted transducers can be designed by first generating an impulse response $v(t)$ of finite duration, such that its Fourier transform is a good approximation to the required frequency response. The envelope of $v(t)$ is exemplified by the broken line in Figure 8.10(c). Since the actual impulse response has sharp discontinuities and so is very different from $v(t)$, the design procedure seeks to synthesise the required frequency response only in the main pass-band, neglecting any stop-band requirements.

To design an electrode-withdrawal transducer, the responses of the possible types of electrode group are first evaluated, at the centre frequency. The ideal impulse response $v(t)$ is assumed to be an amplitude-modulated waveform, and its envelope is integrated to give a new function of time. The types and locations of the electrode groups are then chosen such that the integral of the designed impulse response envelope is as close as possible to the integral of the ideal impulse response envelope [255, 257]. The transducer design then gives a frequency response approximating the spectrum of $v(t)$, for frequencies close to the centre frequency. In practice, the design must allow for the fact that the responses of individual electrode groups are affected by end effects, as discussed in Section 4.5.2. In addition, the electrodes perturb the surface-wave velocity non-uniformly, owing to mechanical and electrical loading. This causes phase errors which must be compensated for by making small adjustments to the locations of individual electrode groups [257]. A sophisticated design procedure taking account of these complications is described by Laker *et al.* [258].

For a source-withdrawal transducer, Figure 8.10(b), the design procedure is less complicated. Because the electrodes are regular, end effects are virtually eliminated and the frequency response has the simple form of equation (8.2). Also, velocity perturbations due to the electrodes are uniform throughout the transducer, and so do not complicate the design.

8.4. FILTER DESIGN AND PERFORMANCE

8.4.1. Basic Types of Bandpass Filter

The simplest form of bandpass filter comprises one apodised transducer and one uniform transducer, as illustrated in Figure 8.1. The response of this device is essentially the product of the responses of the individual transducers, as shown by equation (8.1). To design the device, the length of the uniform transducer is chosen first, such that the variation of its frequency response over the band of interest is acceptable; if the specification calls for "traps", where the filter response is to have a particularly small amplitude, the uniform transducer is usually designed to give zeros at one or more of these locations. The response required for the filter is divided by the calculated response of the uniform transducer, giving the required response of the apodised transducer, and the latter is then designed by methods discussed in Section 8.2.

While this type of design is often acceptable, there are some significant limitations and hence a variety of other filter types are often used, the choice depending on the specific requirements. Nearly always, the design is chosen such that the filter response is essentially the product of two transducer responses, as this is very convenient for design purposes. For a strongly piezoelectric substrate material the two transducers can be located in different tracks and coupled via a multi-strip coupler, as illustrated in Figure 5.6. This reduces unwanted output signals due to bulk wave excitation, as discussed in Section 5.3, and is beneficial particularly for Y, Z lithium niobate substrates.

An additional advantage introduced by using a coupler is that the device response is still essentially the product of the two transducer responses, even if both transducers are apodised. This is not true for two apodised transducers located in the same track. The coupler thus enables two weighted transducers to be combined to synthesise the required response. This is often of considerable value when an exacting specification is to be met, because second-order effects, such as transverse end effects and diffraction, are often less deleterious if two weighted transducers are used. For example, if a rectangular frequency response is required, a combination of two similar apodised transducers enables smaller skirt widths and better stop-band rejection to be obtained [240]. The two transducers may have identical designs, in which case each is designed to approximate the square root of the required filter response. Strictly, the coupler response S_{14} should be allowed for, as shown by equation (5.38), though in fact S_{14} varies only slowly with frequency.

For narrow-band filters, long transducers are required. In this case withdrawal-weighting is often attractive since it gives transducers less susceptible to diffraction errors, as discussed in Section 8.3. In addition, since a withdrawal-weighted transducer is unapodised, it may be combined with an apodised transducer in the same track. The filter response is then the product of two weighted transducer responses, without requiring a multi-strip coupler. The advantages of using two weighted transducers are thus obtainable for weakly-piezoelectric substrates, such as quartz, on which a multi-strip coupler cannot be used. In view of the somewhat limited design methods for withdrawal weighting, it is usual to design the

withdrawal-weighted transducer first, and then to design the apodised transducer taking account of the predicted response of the withdrawal-weighted transducer.

8.4.2. Circuit Effect

The design procedures described earlier refer to the short-circuit response $H_{sc}(\omega)$ of the device. This is defined as the short-circuit output current I_{sc} produced when a given voltage V_t is applied to the input transducer, as in equation (8.1). In practice the transducers must be connected to finite impedances and this causes a distortion, due to the circuit effect, which may be calculated by methods described in Section 4.8. If the transducers are connected directly to a resistive source and a resistive load, the distortion is generally small. However, it is often necessary to tune the transducers in order to reduce the insertion loss, and we show here that this generally degrades the stop-band attenuation.

For the present discussion it is sufficient to consider only the input transducer, and it is assumed that this is tuned by a series inductor which connects it to a source with open-circuit voltage V_G and impedance R_G, as in Figure 8.11. The transducer is represented by a series equivalent circuit with radiation resistance $R_a(\omega)$, as discussed in Section 7.1.1. The transducer voltage, V_t, varies with frequency because of the inductor and the frequency variation of the transducer impedance. Since the short-circuit response $H_{sc}(\omega)$ gives the device output for a given value of V_t, the response allowing for the circuit effect can be obtained by multiplying by V_t/V_G, which is the circuit factor $F_c(\omega)$ defined in Section 4.8. From the circuit of Figure 8.11, this is approximately given by

$$\frac{V_t}{V_G} \approx \frac{-j/(\omega C_t)}{R_G + R_a(\omega) + j\omega L - j/(\omega C_t)}, \tag{8.24}$$

where C_t is the transducer capacitance. Here a term $R_a(\omega)$ has been omitted from the numerator on the assumption that the transducer Q-factor is large, which is often the case. The acoustic reactance $X_a(\omega)$ has also been neglected.

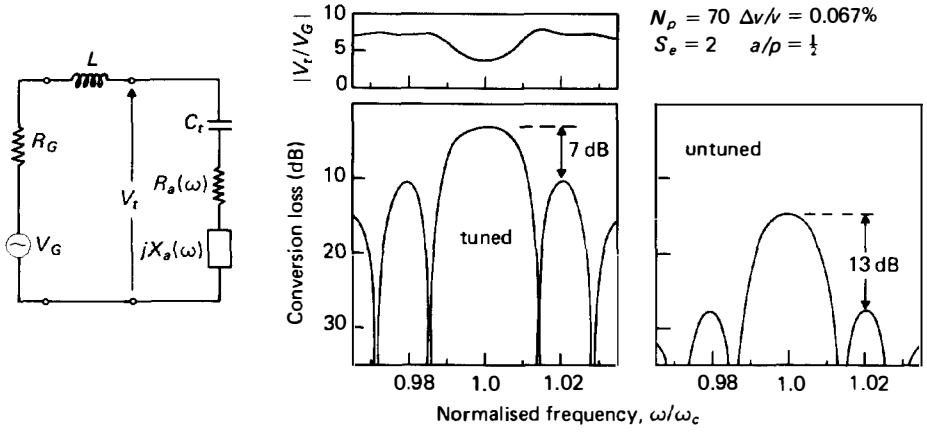

FIGURE 8.11. Circuit effect for a uniform single-electrode transducer.

Generally, $R_a(\omega)$ has a maximum at or near the centre frequency ω_c, and is small for frequencies outside the pass-band. It is assumed here that the transducer is matched at frequency ω_c, so that $R_a(\omega_c) = R_G$ and $\omega_c L - 1/(\omega_c C_t) = 0$. Thus, at frequency ω_c we have $V_t/V_G \approx -j/(2\omega_c C_t R_G)$. At other frequencies in the pass-band, the term $\omega L - 1/(\omega C_t)$ in equation (8.24) is still small – if this were not so, the response would be dominated by the electrical resonance of the circuit, and this condition is avoided in practical filters. Thus, for frequencies on either side of ω_c, where $R_a(\omega)$ is decreasing, the value of V_t/V_G increases because $R_a(\omega)$ is present in the denominator of equation (8.24). The circuit effect therefore tends to flatten the pass-band response. For frequencies in the stop-bands, but still close to the pass-band, $R_a(\omega)$ is small and hence $V_t/V_G \approx -j/(\omega C_t R_G)$. This is about twice the centre-frequency value, and hence the circuit effect increases the relative sidelobe levels by about 6 dB. For frequencies more remote from the centre frequency the electrical resonance is significant, so that $\omega L - 1/(\omega C_t)$ is no longer small, and the sidelobes are suppressed. Although a specific circuit has been considered here, distortions of this nature are generally observed when transducers are tuned; however they are generally less severe because accurate matching is avoided in order to ensure adequate triple-transit suppression, and because of the presence of parasitics.

The conversion loss curves in Figure 8.11 illustrate the circuit effect for a uniform transducer with $N_p = 70$ periods, calculated using the transducer admittance formulae in Section 4.6. For the tuned case the transducer is matched at the centre frequency, and the magnitude of V_t/V_G is also shown. The untuned conversion loss refers to the same case except that the inductor is omitted. The pass-band flattening and the increase of sidelobe level, due to the circuit effect, are clearly seen.

Because of the pass-band flattening, the circuit effect increases the bandwidth somewhat when N_p is large. From equation (8.24), V_t/V_G is approximately proportional to $1/(R_a + R_G)$, and hence the surface-wave power generated is proportional to $R_a/(R_a + R_G)^2$. For a uniform transducer, the radiation resistance $R_a(\omega)$ is approximately equal to $\hat{R}_a[(\sin X)/X]^2$, where $X = \pi N_p(\omega/\omega_c - 1)$, and here $\hat{R}_a = R_G$ because the transducer is matched. It follows that the 1.5 dB bandwidth $\Delta\omega \approx 1.14\omega_c/N_p$, when N_p is large. This formula is used in the discussion in Section 7.1.1.

8.4.3. Second-order Effects and Design

Many of the second-order effects that can affect the response of a bandpass filter are usually rendered negligible by an appropriate initial choice of the types of transducer and of the substrate material. For example, in a single-electrode transducer, electrode interactions are generally negligible if the number of electrodes, N, is such that $N\Delta v/v \ll 1$. If this is not the case, a multi-electrode transducer may be used, as described in Section 4.2, or, if the bandwidth is small, the transducer may be thinned as in Section 8.3. The triple-transit spurious signal, due to acoustic reflections, is normally suppressed adequately by ensuring that the transducers are not closely matched to the electrical source and load. As discussed for uniform transducers in Section 7.1.3, this implies a trade-off with the device insertion loss. Bulk-wave effects

can be minimised by an appropriate choice of substrate material or, if a strongly piezoelectric material is used, by incorporating a multi-strip coupler. Temperature changes cause the amplitude and phase of the frequency response to scale with frequency, as discussed in Section 6.4. This is sometimes significant for filters with narrow bandwidths, or with narrow skirts, requiring the choice of a temperature-stable substrate material such as *ST*, *X* quartz. Electromagnetic breakthrough between the input and output of the device must be adequately suppressed, and this is achieved by careful package design and lay-out; this is particularly important at high frequencies.

Generally, the only remaining second-order effects of significance are diffraction, the circuit effect and the transverse end effect, though diffraction is usually insignificant if the substrate orientation is a minimal-diffraction orientation. These effects are, when necessary, compensated for by modifying the transducer design. The analysis of a device allowing for diffraction is discussed in Section 6.2.5, while the analysis including the circuit effect is described in Section 4.8, which allows for arbitrary terminating circuits. For accurate results, it is often important to include stray components, as discussed in Section 7.1.2. To design a device allowing for these effects, the simplest approach is an iterative process. The device is first designed ignoring the second-order effects. The frequency response of the design is then calculated allowing for the second-order effects, so that the distortion due to these effects can be evaluated. The original specification is then modified to compensate for the distortion, and the design stage is repeated. This process may be repeated several times, and is usually effective if the distortion due to second-order effects is small. It is usual to add at least one iteration involving an experimental device, repeating the design procedure to compensate for the distortion observed experimentally.

In some cases diffraction effects are too severe for the above approach to be effective, so a more basic method of compensation, based on the diffraction analysis of Section 6.2, is necessary. This is a problem of considerable complexity and hence a variety of techniques have been considered, as discussed for example in Refs. [259–261]. The difficulties arise because of the complexity of the basic analysis for diffraction, and because the effect is frequency-dependent. For a narrow-band filter it is sufficient to consider diffraction at the centre frequency, ignoring the frequency dependence [259]. Assuming that at least one of the two transducers is apodised, the effect of diffraction is calculated for each element of this transducer in turn, allowing for the geometry of the other transducer. The aperture of each element is then adjusted such that it gives the required contribution when diffraction is present; this requires an iterative procedure, since the equations involved are transcendental. For wider bandwidth devices this method is of limited value because of the frequency variation of the diffraction effect, but some sophisticated and effective methods have been developed for this case [260, 261].

8.4.4. Performance

Generally, surface-wave bandpass filters are suitable for centre frequencies in the range 10 MHz to 2 GHz, with bandwidths ranging from a maximum of 50% of the centre frequency to a minimum of about 100 kHz. The minimum bandwidth is related

to the physical size, since narrower skirt widths generally require longer impulse responses. A stop-band rejection of 60 dB is often achievable. For a filter with a nominally rectangular frequency response a common figure of merit is the *shape factor*, defined as the ratio of the 40 dB bandwidth to the 3 dB bandwidth. Shape factors down to about 1.1 can be obtained, giving very sharp roll-off in the skirts of the response. Within the pass-band a relative amplitude accuracy of 0.1 dB is obtainable and, for a nominally linear-phase filter, the phase can be linear to within 1°. An important factor related to this fidelity is the triple-transit signal, which causes ripples in the amplitude and phase. For example, a triple-transit suppression of 40 dB corresponds to an amplitude ripple of 0.2 dB peak-to-peak, and a phase ripple of 1.1° peak-to-peak. It is thus important that the triple-transit signal should be adequately suppressed, and hence the insertion loss is not usually minimised. Typical insertion losses are in the range 15 to 30 dB. A more extensive discussion of the performance obtainable is given, with practical examples, by Hays and Hartmann [262].

Figure 8.12 shows the experimental response of a bandpass filter for a domestic television receiver [249]. This device had one uniform and one apodised transducer on a *Y*, *Z* lithium niobate substrate, with the transducers coupled by a multi-strip coupler. The hatched lines on the figure indicate the specification. The complexity of the specification illustrates well the versatility of surface-wave devices. Note that the phase $\phi(\omega)$ of the frequency response is required to be a non-linear function of frequency, so that the group delay $\tau_g(\omega) = -\mathrm{d}\phi(\omega)/\mathrm{d}\omega$ varies with frequency. The reason for this is that the surface-wave filter replaces an earlier *L*–*C* filter giving a non-linear phase, and consequently the broadcast signal is pre-distorted to compensate for this. The device is enclosed in a circular TO-8 package of 14 mm

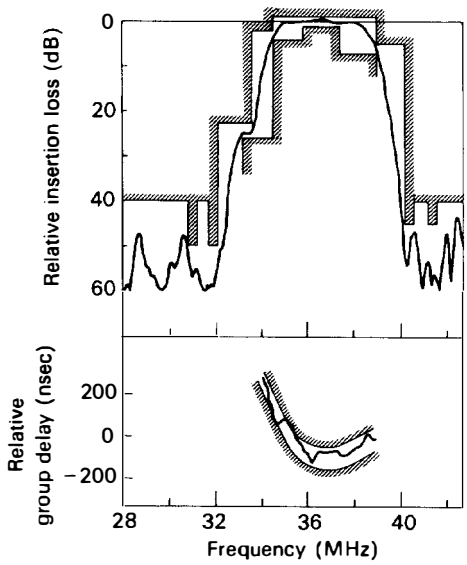

FIGURE 8.12. Response of a surface-wave bandpass filter for domestic television receivers. (Courtesy, Plessey Research)

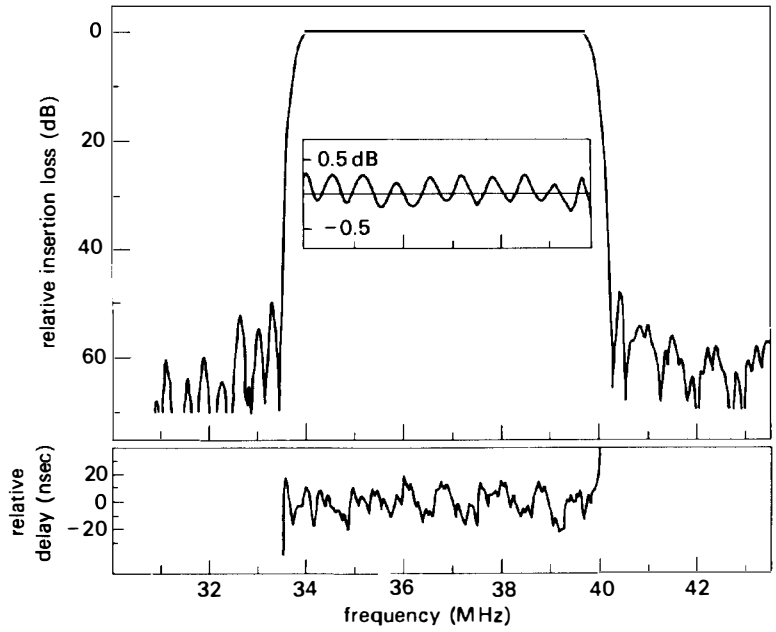

FIGURE 8.13. Response of a bandpass filter for television broadcast equipment (courtesy J. M. Deacon, Signal Technology Ltd.)

diameter. Such devices are now used almost universally in colour television receivers since, in comparison with the earlier L–C filters, they are smaller, give better performance, and do not require trimming after fabrication. In the interest of economy, the lithium niobate substrate may be replaced by a zinc oxide film on a glass substrate [263], a material mentioned previously in Section 3.5. The use of surface-wave filters in television receivers is discussed by Ash [264], who proposes an improved system configuration involving in addition a surface-wave oscillator.

An example of a filter response meeting very stringent requirements is shown in Figure 8.13. In this case the application is in television broadcasting equipment. The frequency response is required to have linear phase and a rectangular amplitude characteristic, with very narrow skirts. The upper part of the figure shows the insertion loss, with the pass-band detail shown in the inset, and the lower part shows the group delay. The device gave a shape factor (40 dB bandwidth divided by 3 dB bandwidth) of 1.2 and a pass-band ripple of 0.4 dB peak-to-peak. In this case two apodised transducers were used, coupled via a multi-strip coupler, on a Y, Z lithium niobate substrate. The mid-band insertion loss was 29 dB and the triple-transit suppression 63 dB. Kodama *et al.* [250] illustrate a similar result giving a pass-band ripple of 0.24 dB peak-to-peak, though with smaller out-of-band rejection.

Some other experimental results illustrating the performance of bandpass filters are given in, for example, references [262, 265–268], and in Sections 7.2.1 and 8.5.

8.4.5. Other Types of Bandpass Filter

Up to this point, this chapter has considered filters using transducers that are uniform,

withdrawal-weighted, or apodised with regular (or almost regular) electrodes. Most filters rely on one or more of these types of transducer. However, a variety of other techniques have been investigated, and these are considered briefly here.

Chirp transducers, in which the electrode pitch varies with position, can be used effectively in filters requiring narrow skirts [249]. For this case, a transducer with regular electrodes has several time-domain sidelobes, and hence there are many small-aperture sources. A chirp transducer gives fewer small-aperture sources and is found to be less affected by diffraction and electrode interactions. This type of transducer is considered in Chapter 9. Its frequency response is dispersive, but for applications requiring linear phase two similar transducers can be used, arranged such that the dispersion cancels.

For applications requiring low insertion loss, some special techniques can be used to maintain good triple-transit suppression, as mentioned in Section 7.1.3. In particular, multi-phase unidirectional transducers have been found to be very effective, and the performance of bandpass filters using these is discussed in Section 7.2.1.

In addition to apodisation and withdrawal weighting, several other weighting techniques have been demonstrated. These include a capacitive weighting technique [269] and other methods reviewed by Engan [270]. There are several methods in which the transducer generates a uniform surface-wave beam, minimising diffraction effects and also avoiding the design limitations of withdrawal weighting. However, none of these has come into common usage.

Bandpass filters have also been made using reflective arrays of either grooves [271, 272] or metal dots [273]. The reflective array is used to reflect surface waves through 90°, in a manner similar to the operation of a reflective array compressor, and the frequency-dependent behaviour of the array provides the required frequency selectivity. Reflective arrays are also used in the surface-wave resonator, which has two arrays each reflecting the waves through 180°, forming a resonant cavity. This principle has been considered extensively as a method of obtaining narrow-band filters [274], the advantage being that the high Q-factor enables the device to be much shorter than an interdigital device. Resonators are considered further in Chapter 10.

Several methods for introducing frequency selectivity in a multi-strip coupler have been discussed in Section 5.6. These are somewhat inflexible, but have the advantages that the coupler can have low loss and can readily be combined with weighted transducers.

Filters in which interdigital transducers are used to generate surface-skimming bulk waves are described in Appendix F.

8.5. FILTER BANKS

The term "filter bank" refers to a device with one input port and several output ports. The response measured between the input and any one output is a relatively narrow band-pass function. The responses for the various outputs are arranged to be contiguous in frequency so that, taken together, they cover a relatively wide band. The

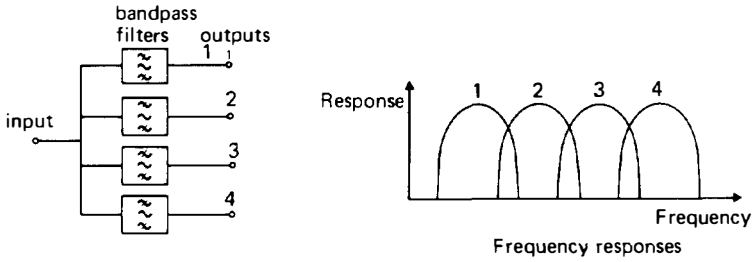

FIGURE 8.14. Principle of filter bank.

device functions as an array of bandpass filters with their inputs connected together, as shown in Figure 8.14. A narrow-band signal applied to the input will appear predominantly at one or two of the outputs, the choice of output depending on the centre frequency of the signal.

This device has two main applications. Firstly, it may be used for frequency measurement by detecting the output signals and comparing their amplitudes. Receivers which perform this function are known as channelised receivers. The second application is in frequency synthesis, where a comb spectrum is first generated and the filter bank is used to select the individual spectral components. An electronic switch is connected to each output of the filter bank, so that the required component can be selected.

A surface-wave filter bank may be made quite straightforwardly by connecting the inputs of several bandpass filters. For example, Slobodnik *et al.* [275] describe a filter bank which had 9 channels with 3 MHz spacing, covering an overall band of 321 to 345 MHz. Each channel had a 3 dB bandwidth of about 1 MHz, with an insertion loss of 20 dB and a relative stop band rejection of 61 dB or more. The individual filters were made on *ST*-cut quartz substrate, each filter using one withdrawal-weighted and one apodised transducer. A compact arrangement was obtained by fabricating three filters on each substrate, so that only three substrates were needed for the 9 channels.

Another type of filter bank, introduced by Solie, makes use of 3 dB multistrip couplers [276]. In this device a surface wave generated by an input transducer is routed to one of several output transducers, depending on its frequency. The principle is shown in Figure 8.15(a). A broad surface-wave beam is directed toward a 3 dB coupler in which the electrodes have an offset of $v\tau$, corresponding to a delay τ, where v is the velocity. If a beam is incident in *one* track of a 3 dB coupler, output beams are produced in both tracks with equal amplitude, as explained in Chapter 5. If there were no offset, the output beams would differ in phase by $\pi/2$. In Figure 8.15(a) the coupler outputs, at A and B, are the sums of signals due to inputs on both tracks, and the offset introduces an additional phase change of $\omega\tau$. If $v\tau = (n - \frac{1}{4})\lambda$, where λ is the surface-wave wavelength, all of the incident power emerges at A; if $v\tau = (n + \frac{1}{4})\lambda$ all of the power emerges at B. This behaviour is essentially the same as that of the beam compressor using 3 dB couplers, described in Section 5.7, though here a larger offset is used. If the coupler frequency response is ignored, a good approximation here, and

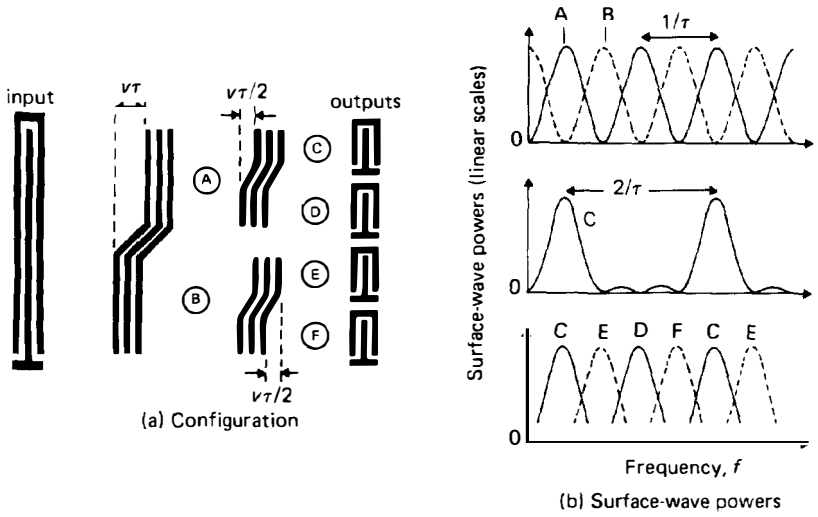

FIGURE 8.15. Filter bank using 3 dB multi-strip couplers.

if the input wave power is independent of frequency, the wave amplitudes at A and B are proportional to $\sin(\omega\tau/2 - \pi/4)$ and $\sin(\omega\tau/2 + \pi/4)$ respectively. The powers of the waves, which repeat with a frequency interval $\Delta f = 1/\tau$, are shown in the upper part of Figure 8.15(b).

The waves at A and B are applied to two further 3 dB couplers which act in a similar manner, except that the offset is approximately $v\tau/2$ instead of $v\tau$. Thus the responses give peaks with spacing $\Delta f = 2/\tau$ instead of $1/\tau$, and half of the peaks observed at A or B are eliminated. The response observed at C is shown in the centre of Figure 8.15(b), and the peaks of the responses for the four output channels C, D, E, F are shown below. The four output transducers, one in each channel, can be weighted in order to shape the response and to improve the stop-band rejection. Additional couplers, with smaller offsets, can be added if a larger number of channels is required.

An experimental device of this type [276] had 16 channels with 0.5 MHz channel spacing, covering the band 81.5 to 89 MHz. Each channel had a 3 dB bandwidth of 1 MHz, an insertion loss of 16 dB and a stop-band rejection of typically 30 dB. The relatively low loss is obtained because the signal is routed to an output dependent on its frequency, instead of being divided initially between a set of parallel filters. The authors describe an application in an FMCW radar system, where the filter bank is used for frequency measurement, determining the range to the target.

Another type of filter bank, using a fanned multi-strip coupler, is described in Section 5.6.