SURFACE ACOUSTIC WAVE DEVICES

INTRODUCTION

Surface Acoustic Wave (SAW) devices were first considered for electronics in the 1960s, and since then a vast array of devices has emerged, with wide-ranging applications. Like other acoustic waves, SAW's have the advantage of a low velocity, around 3500 m/s, so that large delays are obtainable in a small space, and in suitable materials they can give very low propagation loss. A familiar device exploiting these properties is the quartz crystal oscillator, found almost everywhere in clocks and wristwatches as well as in many professional systems. The oscillator relies on propagation of a bulk acoustic wave in a quartz crystal, with reflections from the two major parallel faces so that a resonant cavity is formed. Surface wave devices exploit the same two advantages, low velocity and low loss, but with an additional feature. Because the wave travels along a surface it is accessible throughout its propagation path to components for generation, reflection, waveguiding and so on, so that substantial versatility is obtained. The versatility is well illustrated by the enormous range of devices developed since the inception of the subject.

This subject developed initially in response to a demand for dispersive delay lines (chirp filters) for pulse compression radar. In this application the SAW filters perform a signal processing function, giving an improvement in signal-to-noise ratio and increasing the range capability of the radar. Other devices perform a similar function for spread-spectrum signals in communication systems. Another type of device is the bandpass filter, initially developed for I.F. filtering in domestic TV receivers. This has been an outstanding success for SAW, with practically all TV sets now having a SAW filter. The bandpass filter has also been widely used in professional systems such as radar, and more recently the demand for filters has increased greatly with the growth of wireless telecommunications, particularly mobile telephones. Many different types of SAW bandpass filter, suitable for different applications, have been developed, and a typical present-day mobile phone handset will contain 6 to 10 SAW filters. Another type of SAW device is the oscillator, in which a SAW delay line or resonator is used as a feedback element to control the frequency of an oscillating circuit. Compared with a bulk-wave oscillator, the SAW device can operate at higher frequencies, up to 500 MHz, though its stability is not quite as good. The popularity of SAW devices, particularly for wireless communications, is demonstrated by the world-wide production quantities of several billion devices annually.

Nature of Surface Waves. In a solid material, acoustic waves (or 'elastic' waves) are associated with distortions of the material, such that distances between particular atoms are changed. These distortions, formally expressed as 'strains', are accompanied by internal forces known as 'stresses'. In an acoustic wave, energy is continuously exchanged between elastic and kinetic forms. An infinite material can support propagation of a bulk wave, that is, a freely propagating wave unaffected by boundaries. Bulk acoustic waves are of two basic types: (a) longitudinal waves, in which the particle motion is along the wave vector, and (b) transverse (or shear) waves, with particle motion normal to the wave vector. Typical velocities are 6000 m/s for longitudinal waves and 3000 m/s for transverse waves.

The surface wave solution in an isotropic material was found by Lord Rayleigh in 1885 [1], and is often called a Rayleigh wave. Here, the material is assumed to have a planar surface, free of mechanical forces. Rayleigh found a propagating wave solution, essentially a sum of longitudinal and transverse components, with amplitude decaying exponentially away from the surface. Typically, most of the energy is within a depth equal to one wavelength. The wave is of seismological interest because it explains a component of the signal detected by a seismograph following a ground shock. It travels somewhat more slowly than transverse bulk waves and is non-dispersive, with velocity independent of frequency. The motion of the material is in the sagittal plane, that is, the plane containing the surface normal and the propagation direction.

Piezoelectric Materials. As will become clear later, SAW devices always use piezoelectric materials. The meaning of this term is that stresses and strains within the material are accompanied by electric fields, which arise from fundamental interactions at the atomic level. The electric fields are central to the operation of the majority of SAW components, particularly transducers for generation and detection of the waves. The presence of piezoelectricity complicates the nature of the wave, which is now (in most cases) accompanied by an electric field. But there is another complication in that to be piezoelectric a material has to be anisotropic, so that its macroscopic properties depend on the orientation of the internal structure. Usually crystalline materials are chosen because they can give low losses. The wave properties depend on the orientation of the crystal lattice relative to the macroscopic object. In contrast to the isotropic case considered by Rayleigh, anisotropy and piezoelectricity introduce immense complications when investigating properties of SAW's. As a consequence, the search for materials and orientations suitable for SAW devices has been an ongoing activity throughout the development of SAW.

In general, a numerical investigation of a piezoelectric crystal usually yields surface-wave solutions similar to the isotropic case except for the associated electric field. This is often called a piezoelectric Rayleigh wave. Various properties relevant to SAW usage are considered, in order to find a suitable orientation. As a result, a variety of 'standard' materials and orientations have been established. In particular, crystals of quartz, lithium niobate and lithium tantalate are used. Owing to anisotropy, it is essential to quote the orientation when specifying a SAW material.

In some cases other types of wave are found, in particular various types of 'leaky wave'. Some of these can be used for SAW-type devices, though their behaviour is somewhat different. In this article, the wave is assumed to be a piezo-

electric Rayleigh wave unless otherwise stated; other types of wave are considered later.

Basic SAW Device. Piezoelectricity enables SAW's to be generated conveniently by interdigital transducers (IDT's) placed on the surface of the material, which is often called the substrate (Fig. 1). The IDT consists of a set of interleaved metal electrodes, alternately connected to two bus bars. When a voltage is applied, this sets up a spatially alternating electric field at the surface, coupled to acoustic stresses by the piezoelectric effect. A surface wave is generated if the frequency is such that the SAW wavelength is similar to the transducer periodicity. The IDT is most effective for a frequency such that its periodicity equals the SAW wavelength. At other frequencies the waves generated in different regions of the transducer are subject to progressive phase differences, so that the total wave amplitude is smaller. The IDT is therefore frequency-selective. A similar IDT can be used to detect an incident SAW, producing an electrical output signal. Thus, the device has electrical input and output signals, with electro-acoustic conversion at both transducers.

The wave propagation inside the device is almost ideal in that propagation loss is low, diffraction effects are small (because the beam of surface waves is typically more than 20 wavelengths wide), and dispersion is negligible. For typical power levels the material behaves linearly, so that harmonic generation and intermodulation products are negligible. Hence for many purposes second-order effects such as these can be ignored. However, they do need to be considered, and sometimes compensated for, when an exacting specification is to be satisfied.

Because of the selectivity of the IDT's, this device can be regarded as a bandpass filter. It can also be regarded as a delay line, with delay determined by the velocity and the distance between transducer centres, typically 3 μs/cm. As a delay line, it is suitable for signals whose spectrum is within the frequency band of the transducers. The input transducer generates SAW's travelling in both directions, and it is often necessary to add absorbing material to suppress reflections of waves from the ends of the substrate. The basic design is capable of a huge range of variations for special purposes. In particular, bandpass filters employ modifications to ensure good suppression in the stop band, a flat response in the passband, narrow skirts separating these regions, and minimal group delay variation within the passband.

Fabrication and Performance. The device basically consists of a piezoelectric substrate with a specially-shaped metal film (usually of aluminium) on the surface. This structure is well suited to fabrication by established photolithographic techniques, similar to those used in semiconductor device processing. Many patterns can be made simultaneously by exposing a layer of photoresist on a large circular wafer (or 'slice'), followed by development and chemical etching. The patterns are separated by cutting them into the individual devices, then known as 'chips', which are sealed into hermetic packages to protect them from moisture and contamination. This process is well established in the semiconductor industry and well suited to production of large quantities of devices, thus giving economy of scale. For large quantities, costs can be less than US$ 1.00 per device. Moreover, the method is very versatile since almost arbitrary geometries can be produced.

The fabrication process sets some limits on device performance. The linewidth of the basic IDT, Fig. 1, is a quarter of the centre-frequency wavelength, and the minimum linewidth usable in commercial production is around 0.3 μm. For a typical SAW velocity of 3500 m/s this sets the maximum centre frequency at about 3 GHz, though special techniques can increase this number a little. The minimum frequency is in the region of 10 MHz, simply because in this region the devices become rather clumsy and expensive. The velocity determines the delay for a given length. A device 2 cm. long gives about 6 μs delay, and lengths of 15 cm. are quite feasible though they become expensive because they are not suited for large-quantity production. Bandwidths range substantially, from 0.1 % to 50 % of the centre frequency.

Generally, internal losses in the device are small, so one might expect a small insertion loss to be obtainable. However, in practice this is limited by distortion arising from multiple SAW transits. To give low loss, the device of Fig. 1 would need good electrical matching between the transducers and the source and load, normally involving inductors to tune out the transducer capacitances. For this case, the acousto-electric conversion loss is ideally 3 dB (because half of the energy applied to a transducer emerges as a wave generated in the unwanted direction). Hence a device insertion loss of 6 dB is expected. This is indeed close to reality, but unfortunately there is also an unwanted signal because the transducers reflect the waves. This signal, called the *triple-transit* signal, occurs after 3 transits of the device, involving a reflection from each transducer. When electrically matched, the transducer reflection coefficient is -6 dB, and hence the triple-transit signal is 12 dB below the main signal. This level is quite unacceptable for most applications, and often it is reduced by deliberately mismatching the transducers (for example, omitting the inductors). For this reason, many SAW devices have relatively high insertion losses, 15 to 30 dB. However, a wide range of techniques have been applied to this problem, yielding a variety of low-loss types of bandpass filter.

Some general accounts of the subject will be found in references [2–6].

INTERDIGITAL TRANSDUCERS

The interdigital transducer (IDT) is a basic component in all SAW devices. As explained in earlier, a simple IDT consists of a set of electrodes alternately connected to two busbars, as shown in Fig. 2a. The electrodes are regular, that is, they have constant width a and pitch p.

The operation of the IDT can be described approximately by a simple delta-function model [7]. Suppose initially that a voltage V is applied to only one electrode, with the others grounded. Surface waves will be generated with amplitude proportional to an element factor denoted by $E(\omega)$. The SAW amplitude is expressed in terms of the associated surface potential $\phi_s(x)$, which is thus given by

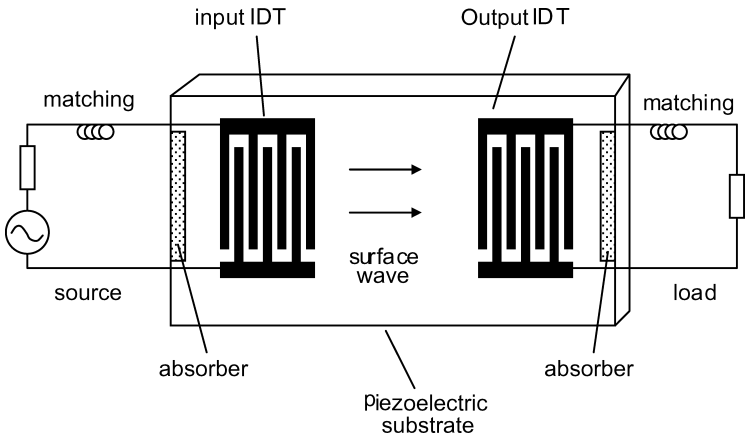


Figure 1. Basic SAW delay line

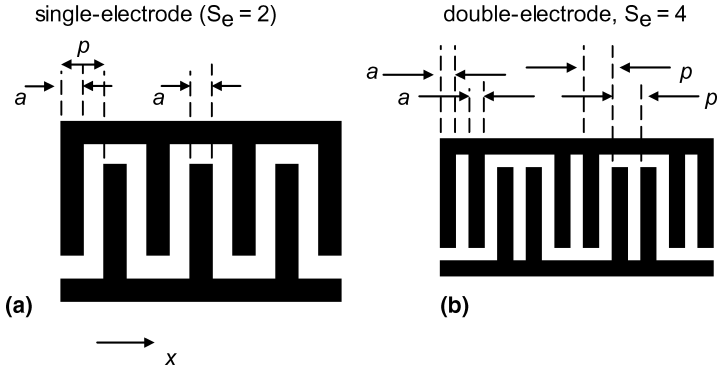


Figure 2. Transducer types. (a) single-electrode; (b) double-electrode

$V \cdot E(\omega) \cdot \exp[\pm j\,k(x - x_n)]$, where k is SAW wavenumber and x_n is the location of the electrode centre. Here k is related to the velocity v by $k = \omega/v$. In Fig. 2a we take the lower bus bar to be grounded and apply the above formula to each live electrode, adding the waves generated. We define a polarity $\hat{P}_n$ for electrode n, such that $\hat{P}_n = 1$ or zero for a live or grounded electrode, respectively. The potential of the wave emerging at the left is thus

$$\phi_s(\omega) = V.E(\omega)e^{jkx}\sum_{n=1}^{N}\hat{P}_n\exp(-jkx_n)$$
$$= V.E(\omega)e^{jkx}.A(\omega) \qquad (1)$$

where N is the number of electrodes, which are centred at x_n. The summation is denoted by $A(\omega)$, which is known as the array factor. Usually the array factor varies with ω much faster than the element factor $E(\omega)$, so the array factor is essentially the frequency response. Fourier transformation of $A(\omega)$ into the time domain gives the function $a(t) = \sum_{n=1}^{N}\hat{P}_n\delta(t - x_n/v)$, which is a series of delta-functions at times corresponding to the centres of the live electrodes.

For the transducer of Fig. 2a the electrode positions are regular and we can take x_n = np, where p is the electrode pitch. Also, the polarity sequence is $\hat{P}_n = 1, 0, 1, 0, \ldots$, since the electrode voltages alternate between V and 0. Substituting these, $A(\omega)$ becomes a simple geometrical progression which is easily summed. Its magnitude is

$$|A(\omega)| = \left|\frac{\sin(N_p\theta/2)}{\sin(\theta/2)}\right| \qquad (2)$$

where $\theta = 2$kp and N_p is the number of live electrodes (equal to the number of transducer periods). This equation predicts a series of peaks. The fundamental response, at which the transducer pitch equals the SAW wavelength, occurs for $\theta = 2\pi$. The formula remains unchanged if we replace θ by $\phi = \theta - 2\pi$. In the region of the fundamental passband, ϕ is small and the function becomes $|A(\omega)| \approx N_p.|\sin(X)/X|$, with $X = N_p\phi/2$. And X can be written as $X = \pi N_p.(f/f_0 - 1)$, where $f_0 = v/(2p)$ is the centre frequency. From this it follows that the 4 dB points are approximately at $\phi = \pm\pi/N_p$, giving a fractional bandwidth of $\Delta f/f_0 = 1/N_p$. The bandwidth is $\Delta f = 1/T$, where $T = 2\,pN_p/v$ is the transducer length in time units.

The above delta-function model is applicable because in practice the wave propagation is almost ideal. Provided suitable materials are chosen for the substrate and the electrodes (usually aluminium), the waves propagate with very little loss or dispersion. Also, diffraction effects are usually small, though this often needs consideration and we return to it in later.

In addition, we have assumed that the electrodes don't reflect the waves. For the transducer of Fig. 2a reflections can be significant, though the reflectivity can be minimised by recessing the electrodes into the substrate surface. Individual electrodes have a small reflection coefficient, typ-

ically 1 or 2%. The reflections add coherently at the centre frequency, because the electrode pitch is $\lambda_0/2$ (where λ_0 is the centre-frequency wavelength). This problem can be avoided by using a 'double-electrode' transducer, as in Fig. 2b, where each electrode is split into two. The electrode spacing is now $\lambda_0/4$, so reflections from adjacent electrodes are cancelled at f_0. The main penalty is that the electrode width is $\lambda_0/8$ instead of $\lambda_0/4$, so that fabrication becomes more difficult. The original transducer of Fig. 2a is in contrast called a 'single-electrode' transducer. The above analysis, eq. (1), applies for both types of transducer if reflections are ignored; it is only necessary to change the polarity sequence $\hat{P}_n$.

The transducer admittance Y_t, defined assuming that no waves are incident, is important when electrical matching is considered. Often the admittance is dominated by a capacitance C_t, so this is usually written explicitly. Y_t can be written as $Y_t = G_a(\omega) + jB_a(\omega) + j\omega C_t$, where G_a and B_a are the acoustic conductance and susceptance, due to the excitation of surface waves. For a uniform transducer such as those in Fig. 2, these terms have the approximate forms

$$\begin{aligned} G_a(\omega) &\approx G_a(\omega_0)[\sin X/X]^2; \\ B_a(\omega) &\approx G_a(\omega_0)[\sin(2X) - 2X]/(2X^2) \end{aligned} \tag{3}$$

These formulae assume that electrode reflections are not significant. The conductance is essentially $A^2(\omega)$. The formula for B_a is deduced from G_a by using causality. This implies that B_a must be the Hilbert transform of G_a for any transducer.

For a two-transducer device, such as that in Fig. 1, the terminal admittances Y_{11} and Y_{22} are just the transducer admittances Y_t. This follows because Y_{11} is the admittance seen at one transducer when the other is shorted, and we have assumed that a shorted transducer does not reflect SAW's. Also, the transadmittance Y_{12} can be shown to have magnitude given by $|Y_{12}| = [G_{a1}.G_{a2}]^{1/2}$, where G_{a1} and G_{a2} are the two transducer conductances.

As noted above, electrode reflections can be eliminated by using a double-electrode transducer. However, these transducers still have a finite SAW reflection coefficient when connected to finite electrical loads, thus giving rise to the unwanted triple-transit signal. Special methods can be applied to deal with this, notably the use of specially-designed internal reflections in SPUDT's, described later.

Figure 3 shows the theoretical conductance of a 30-period single-electrode transducer. The solid line assumes that electrode reflections are negligible. The conductance has the form of eq. (3). For the broken line electrode reflections have been included, using the COM theory described later.

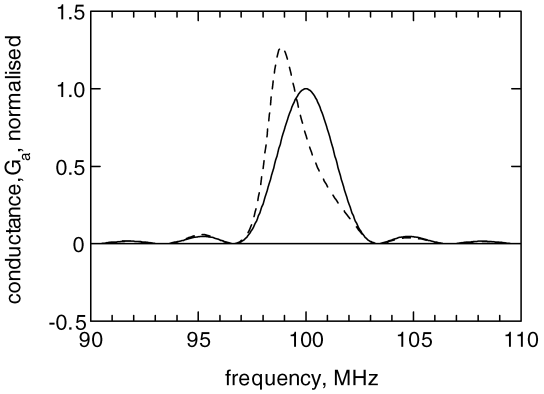


Figure 3. Conductance G_a (ω) of a 30-period uniform single-electrode transducer. Solid line: without electrode reflections. Broken line: with reflections, $r_e = -0.017$ j.

PROPAGATION EFFECTS

The choice of suitable materials for SAW devices is intimately concerned with a variety of propagation effects, particularly piezoelectric coupling, diffraction and temperature effects. The choice of material is a vital part of the design procedure because the material properties have a strong bearing on the performance obtainable.

SAW Solutions and Velocities

Many of the required SAW properties of a material can be deduced from the SAW velocity, and the velocity itself is important, of course. The SAW solution is found starting from the bulk material constants: the stiffness c_{ijkl}, piezoelectric d_{ijk} and dielectric ε_{ij} constants, together with the mass density ρ. The values originate from bulk measurements on the crystals, and large amounts of data are given by Slobodnik [8] and Gualtieri *et al* [9]. Kovacs *et al* [10] give values for lithium niobate and tantalate.

The calculation of SAW solutions in an anisotropic piezoelectric material is a complex numerical process [11] which will not be detailed here. Because of the anisotropy, the assessment of a material for SAW applications involves calculation of its SAW properties for all possible orientations, and since the orientation is governed by three angles this task is very extensive. The substantial literature on these topics includes Slobodnik's Handbook [8] and a review [12].

The SAW solution for a free surface has velocity denoted here by v_f. It is also common to examine a 'metallised' surface, meaning that an idealised metal film is assumed to be on the surface, shorting out the parallel electric field but thin enough to have no mechanical effect. In this case the velocity is v_m. The fractional difference between the two velocities is defined as

$$\Delta v/v \equiv (v_f - v_m)v_f \tag{4}$$

and this indicates the strength of electrical coupling to the SAW. The element factor $E(\omega)$ mentioned earlier is proportional to $\Delta v/v$, so this governs the conversion efficiency of an IDT. Figure 4 shows velocities for Y-cut lithium niobate, as functions of propagation direction. The coupling is strongest for propagation in the Z-direction and this orientation, known as Y-Z lithium niobate, is a popular choice. For this case, a 100 MHz wave on a free surface, with power density 1 mW/mm, gives a surface potential of about 0.8 V. The particle displacements at the surface are about 0.08 nm (0.8 Å), which is comparable with interatomic distances and far less than the SAW wavelength of 35 μm. It is also common to define a coupling constant as $k^2 = 2\,\Delta v/v$.

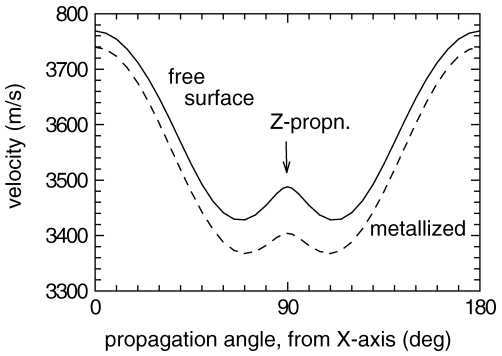


Figure 4. SAW velocities on Y-cut lithium niobate, as functions of propagation direction.

Diffraction

SAW diffraction can be analysed by a method similar to that often used in optics, using a scalar to represent the wave amplitude [12]. Perturbations due to components such as transducers and gratings are ignored in this method. Except for the most sophisticated devices, these approximations are usually adequate. The analysis may then be done based on data for the SAW velocity. As in conventional optics, SAW diffraction gives a near-field (Fresnel) region, in which the amplitude distribution is similar to that of the source, and a far-field (Fraunhofer) region where the amplitude corresponds to the Fourier transform of the source. Several complications occur, however, because of the anisotropy of the substrate material. In summary:

(a) *Beam steering* can occur, that is, the energy flow direction **E** in a wide beam of SAW's is not parallel to the wave vector **k**. The angle γ between **E** and **k** is the beam-steering angle (or 'power flow angle'), given by $\tan\gamma = (1/\mathrm{v})\,\mathrm{d}\mathrm{v}/\mathrm{d}\theta$, where $\mathrm{v}(\theta)$ is the SAW velocity at an angle θ. Usually the substrate is chosen such that $\gamma = 0$, but the effect can still occur due to misalignment.

(b) *Minimal-diffraction Orientations*. For some substrates, notably Y-Z lithium niobate, the anisotropy is such that diffraction spreading is substantially reduced. This occurs when $\mathrm{d}\gamma/\mathrm{d}\theta \approx -1$. This condition can be very helpful when designing high-performance devices, because it eliminates the need to compensate for diffraction.

(c) *Parabolic Approximation and Scaling*. In many cases the SAW velocity variation $\mathrm{v}(\theta)$ can be approximated by writing k_x as a quadratic function of k_y (where x is the principal propagation direction). Assuming no beam steering, we can write $k_x(k_y) = k_x(0) - 0.5\,bk_y^2/k_x(0)$. The constant b is unity for an isotropic material, and zero for a minimal-diffraction orientation, and it is given by $b = 1 + \mathrm{d}\gamma/\mathrm{d}\theta$. If the approximation is valid, the wave amplitude distribution is simply scaled in the x-direction. Specifically, if $\psi_\mathrm{i}(\mathrm{x}, \mathrm{y})$ is the distribution for the isotropic case, the anisotropic case gives a distribution $\psi(x, y) = \psi_i(bx, y)\exp[\,j(b-1)x\,k_x(0)]$.

As in optics, the amplitude distribution for the isotropic case is given by formulae involving Fresnel integrals. When the parabolic approximation is invalid it is necessary to use a more fundamental approach, commonly the Angular Spectrum of Plane Waves (ASPW) method. For this method, the source is represented as a sum of plane wave elements, with distribution determined by a Fourier transform. Each plane wave propagates according to the known velocity $\mathrm{v}(\theta)$. The above statements can all be deduced from the ASPW formulation.

In SAW devices, diffraction is usually a second-order effect, that is, the receiving transducer is in the near-field region of the transmitter. However, in an apodised bandpass filter, described below, diffraction can cause some unacceptable signals in the stopband above the passband. These can be reduced by sophisticated design techniques which compensate for the diffraction.

Temperature Effects

Temperature stability is usually of vital interest when designing a SAW device. Temperature changes cause SAW velocity changes and also thermal expansion, and both these factors affect the delay given by a device. The velocity variation can be calculated using the temperature coefficients of the bulk piezo-elastic constants, and with the expansion coefficient this gives the temperature coefficient of delay (TCD). In most cases the delay varies approximately linearly with temperature. However, for some cases the velocity and expansion effects cancel, at one temperature. The delay $\mathrm{T}(\theta)$ then has the quadratic form $\mathrm{T}(\theta) = \mathrm{T}(\theta_0)[1 + c(\theta - \theta_0)^2]$, where θ is the temperature and θ_0 is the 'turn-over temperature', at which the TCD is zero. In particular, for ST-X quartz, θ_0 = 21°C and $c = 32 \times 10^{-9}$ $(^\circ\mathrm{C})^{-2}$. This orientation is a rotated Y-cut, with propagation along the crystal X-axis and a rotation angle of 42.75°. It is commonly used for temperature-stable devices, though the rotation angle is usually reduced a little, increasing θ_0. This is done because the centre of the operational temperature range is usually higher than 21°C. It also compensates for a reduction of θ_0 associated with the crystal mounting and the presence of transducers.

The effect of a temperature change is to scale the device impulse response $h(t)$, which changes to $h'(t) = h[t/(1+\varepsilon)]$, where ε is a small quantity. Correspondingly, the frequency response changes from $H(\omega)$ to $H'(\omega) = H[(1+\varepsilon)\omega]$. Hence the centre frequency is scaled by a factor $\Delta f_0/f_0 = -\Delta\mathrm{T}/\mathrm{T} = -\varepsilon$, and the bandwidth is also scaled by the same factor.

Other Propagation Effects

For standard SAW materials, propagation loss is usually small provided the crystals are well polished. For example, at 1 GHz the loss for a free surface is about 1 dB/μs for Y-Z lithium niobate, and 3 dB/μs for ST-X quartz [12]. These figures are approximately proportional to f_2, and they are increased a little when transducers are present. Non-linear effects are usually negligible in practical devices. Theoretically a free surface gives no dispersion, but in practice there

is a little dispersion due to transducers and gratings. However, these effects are not usually of practical significance.

A more significant effect is waveguiding, which can occur because the SAW velocity in components such as transducers is slightly lower than the free-surface velocity outside. In relatively long devices the waves propagate as a series of modes, with some dispersion. For moderate apertures there is often a dominant fundamental mode plus some unwanted higher modes. The unwanted modes cause small perturbations, usually at frequencies somewhat higher than the centre frequency. These distortions are not usually significant in devices with wide apertures, greater than typically 20 wavelengths.

MATERIALS

A considerable variety of suitable materials has been established, bearing in mind the above propagation effects and the strength of piezoelectric coupling ($\Delta v/v$). The orientation is often specified by two axes. For example, Y-Z lithium niobate indicates a Y-cut crystal with SAW propagation along the crystal X-axis. An example of a rotated Y-cut is ST-X quartz, which can be written as 42.75°Y-X quartz. This has propagation along the crystal X-axis (in the surface). Looking down the X-axis, toward the origin, the crystal Y-axis is rotated by 42.74° clockwise from the surface normal.

For more general cases, it is common to use Euler angles, λ, μ and θ. We define X, Y and Z as crystal axes and x, y and z as device axes, with x as the SAW propagation direction and z as the outward-directed normal to the surface. Initially the angles are zero and x, y, z are the same as X, Y, Z. There are three rotations, as follows:

1. Rotate the x, y and z axes anticlockwise about z through and angle λ (as seen by an observer looking down the z-axis towards the origin). Call the new axes x_1, y_1, z_1.
2. Rotate the x_1, y_1, z_1 axes anticlockwise about x_1 through an angle μ. Call the new axes x_2, y_2, z_2.
3. Rotate the x_2, y_2, z_2 axes anticlockwise about z_2 through an angle θ. Call the new axes x_3, y_3, z_3.

In the final orientation, z_3 is the surface normal and x_3 is the propagation direction. For a rotated Y-cut, with cut angle ψ, we have $\lambda = \theta = 0$ and $\mu = \psi - 90°$.

Table 1 gives data for some common SAW materials. Y-Z lithium niobate ($LiNbO_3$) has strong piezoelectric coupling and minimal-diffraction properties, advantageous for wide band and low loss devices, but its temperature stability is poor. This material generates unwanted bulk waves rather strongly. The 128° rotated cut minimises bulk wave generation, but this is not a minimal-diffraction orientation. ST-X quartz (SiO_2) has weak piezoelectric coupling and is generally limited to narrow-band devices, but its good temperature stability is attractive for narrow-band filters and resonators. X-112°Y lithium tantalate ($LiTaO_3$) is an intermediate case, as is lithium tetraborate ($Li_2B_4O_7$). The rotated quartz cut shown gives both minimal diffraction and zero TCD [13].

Zinc oxide (ZnO) films on glass substrates have been used for TV IF filters, for reasons of economy. This film has also been used on diamond or sapphire substrates, giving a high velocity attractive for high-frequency devices. Using sapphire substrates, filters at 1.5 GHz have been produced [14]. For many substrate materials a SiO_2 film can be added to improve the temperature stability, and the Table shows data for 128°Y-X lithium niobate. Langasite ($La_3Ga_5SiO_{14}$) has recently become available for SAW applications, giving properties quite similar to quartz but with stronger piezoelectric coupling [15]. The case shown gives good temperature stability and minimal diffraction. A variety of other new crystals with similar properties, including langatate and langatite, are also becoming available.

Some cases in the Table show the N-SPUDT (Natural Single-Phase Unidirectional Transducer) effect. This means that a geometrically symmetrical transducer, with an applied voltage, will generate waves with different amplitudes in the two directions. The effect arises from asymmetry in the substrate properties. This effect is of little practical value in itself, but it can occur in materials which are attractive because of other SAW properties. Generally, the effect needs to be allowed for when devices are designed using these materials. Some unidirectional transducers using geometrical asymmetry are described later.

The leaky SAW and STW cases will be described later.

BASIC SAW FILTERS

Bandpass Filtering

This is one of the commonest applications of SAW devices, and apodisation is a fundamental method of achieving this.

Apodised Transducers. Apodisation, that is, variation of the electrode lengths, is a common method for weighting SAW IDT's. The principle is illustrated in Fig. 5, which shows a device with two transducers, one apodised and the other unapodised. In the apodised transducer, at each x_n there are electrodes extending from both bus bars, with a small break at a position which determines the apodisation. The SAW sources can be taken to be the live electrodes, or alternatively the gaps between electrodes with different voltages. The figure shows single-electrode transducers for clarity, though in practice they are more commonly the double-electrode type to avoid the electrode reflection problem described above.

To appreciate the operation, it is first necessary to consider reception of waves by the unapodised transducer at the left. For the moment, we will simply state that the frequency response, for waves incident from the right, is essentially the same as the response when the transducer generates waves to the right. This is a result of reciprocity, considered later. For a transducer receiving SAW's, the response is defined by assuming the transducer to be shorted, and considering the current I_{sc} in the short-circuit when a wave with specified potential ϕ_s is incident. The ratio I_{sc}/ϕ_s is almost the same as eq. (1), except that a slightly different element factor is used.

Table 1. Data for Common Materials

Material	v_f, m/s	$\Delta v/v$	$\varepsilon_\infty/\varepsilon_0$	TCD ppm/°C
Y-Z $LiNbO_3$ (MD)	3488	2.4%	46	94
128°Y-X $LiNbO_3$	3979	2.7%	56	75
ST-X quartz, SiO_2	3159	0.06%	5.6	0
Quartz[a] (MD, NS)	3162	0.07%	5.5	0
X-112°Y $LiTaO_3$	3300	0.35%	48	18
45°X-Z $Li_2B_4O_7$	3350	0.45%	11	0
$La_3Ga_5SiO_{14}$[b] (MD, NS)	2730	0.16%	-	0
ZnO on sapphire* (higher mode, fh = 870 m/s)	5800	6%	-	-
36°Y-X $LiTaO_3$ (LSAW)	4212	2.4%	50	32
64°Y-X $LiNbO_3$ (LSAW)	4742	5.5%	52	80
Y-Z $LiNbO_3$ (LLSAW)	7285	1.5 %	-	-
Quartz, 36 °Y-X+90° (STW)	5100	-	5.6	0
$SiO_2/LiNbO_3$ 128 °Y-X*	3990	3 %	-	0

LSAW = leaky surface acoustic wave.
LLSAW = longitudinal LSAW.
TCD = temperature coefficient of delay
MD = minimal-diffraction orientation.
NS = shows N-SPUDT effect
[a]Euler angles 0, 43, 23.7°.
[b]Euler angles 0, 138.5, 26.6°
*data depend on film thickness h and frequency f; typical data are shown.

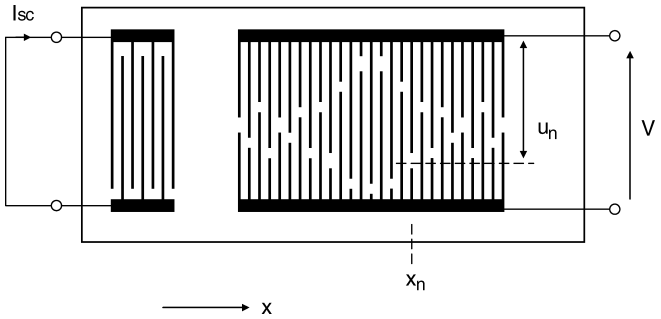


Figure 5. Transversal bandpass filter using apodised transducer.

The above argument applies if the beam of surface waves has its width the same as the transducer aperture. When dealing with an apodised launching transducer, this is no longer true. To simplify, consider a voltage applied to only one electrode of the apodised transducer. The electrode extends a distance u_n, say, from the live bus bar. Now, the short-circuit current induced in the left transducer will be proportional to u_n. To see this, note that apodisation has not affected the amplitude of the wave, or the charge density which the wave induces on the receiving transducer. However, the charges only exist within the SAW beamwidth, and this width is u_n. Hence, when integrated to obtain the short-circuit current I_{sc}, the result will be proportional to u_n. The response of the apodised transducer is thus similar to an unapodised one, eq. (1), except that the source strength is proportional to u_n. To maintain the units, we use u_n/W, where W is the total aperture, so the response has the form

$$H_{ap}(\omega) = E(\omega)\sum_{n=1}^{N}\frac{u_n}{W}\exp(-jkx_n) \qquad (5)$$

The device response is basically $H_{ap}(\omega)\cdot H_{un}(\omega)$, where $H_{un}(\omega)$ is the response of the unapodised transducer, eq. (1). To be more specific, this gives the transadmittance Y_{12}, which is the current obtained when the output transducer is shorted and unit voltage is applied to the input transducer. This formulation is valid only if one of the two transducers is unapodised. If we ignore the slow frequency variation of $E(\omega)$, the Fourier transform of eq. (5) gives a series of impulses with amplitude proportional to u_n. Hence the *time-domain* response is directly related to the *spatial* variation of the electrode lengths, i.e. by the apodisation.

In practice the observed response is affected by the electrical source and load impedances, and to allow for these we also need the remaining Y-matrix components Y_{11} and Y_{22}. Since electrode reflections are here assumed to be negligible, these Y components are just the transducer admittances.

A general technique for analysing an apodised device is to divide it into imaginary parallel 'channels', extending in the x-direction [7]. Each channel can be regarded as an unapodised device, connected electrically in parallel with the others. However, provided there are no electrode reflections the analysis can be done without chanelling.

Transversal filtering. If we ignore the slowly-varying function E(ω), eq. (5) has the form

$$H(\omega) = \sum_{n=1}^{N} A_n \exp(-j\omega n\tau) \qquad (6)$$

where A_n are real constants and τ = p/v is the electrode spacing in time units. This is the response of a *transversal filter*, which can be thought of as a tapped delay line with taps spaced in time by τ and connected to an output via circuits with amplitude response A_n. The response could have been arrived at by considering a continuous waveform and then sampling it with a sampling frequency $f_s = 1/\tau$. In this context we can apply the sampling theorem, which states that any bandlimited waveform can be reproduced by sampling at a high enough rate, and then filtering to eliminate some high-frequency components introduced by the sampling process. Hence, *any* required frequency response can be synthesised by appropriate choice of the real constants A_n, provided (a) f_s is high enough, (b) the original time-domain waveform is of finite length, and (c) technological constraints are satisfied (maximum frequency, device length etc.). It is a property of transversal filters that the amplitude response is symmetrical about frequency $f_s/2$, and this is the centre frequency of a single-electrode transducer. Hence, for the general case we need more than two electrodes per centre-frequency wavelength, four being quite common. This actually brings a benefit in that electrode reflections are minimised, as discussed earlier.

Another realisation of the transversal filter is the digital Finite Impulse Response (FIR) filter. This comparison is of great value because the design of FIR filters has been considered in much detail. A simple design method is to Fourier transform the required frequency response to the time domain and then, in order to obtain finite length, multiply by a finite-length window function such as the Hamming function or the Kaiser function [3]. A very effective design method is provided by the Remez algorithm, which can be used to design a filter with an amplitude response of arbitrary shape, subject to a tolerance expressed as an arbitrary function of frequency [16]. Alternatively, linear programming can be used [17].

In practice, the design procedure needs to compensate for a variety of second-order effects, particularly distortion associated with the terminating impedances (the 'circuit effect') and diffraction. These can be compensated by recursive methods; it is even possible to compensate for distortions of unknown physical origin, provided experimental results are sufficiently repeatable.

Sometimes a transversal filter will incorporate a *multistrip coupler* (MSC). This consists of a sequence of disconnected metal strips perpendicular to the SAW wavevector and spanning two SAW tracks. The MSC can couple SAW's from one track to the other with high efficiency and with wide bandwidth. In the transversal bandpass filter, it has been widely used in order to minimise interference from unwanted bulk waves. The MSC is often used when the substrate is Y-Z lithium niobate; this material gives significant bulk wave generation but is otherwise attractive because it minimises diffraction effects. The coupler also enables two apodised transducers to be used, giving improved flexibility. The number of strips needed for a full transfer of the SAW's is in the region of 2.5/(Δv/v). This prohibits use of the coupler on weakly-piezoelectric substrates such as quartz, since the number of strips would be unrealistic.

Withdrawal Weighting. In this technique, selected SAW sources are withdrawn from a uniform transducer, in such a way that the density of remaining sources mimics the amplitude variation of an apodised transducer [18]. This is shown in Fig. 6. Withdrawal-weighted transducers are suitable for filters with small bandwidths, less than say 5%, and they have the merit of being insensitive to diffraction effects. The electrodes can still be regular, so the response is still governed by the simple delta-function analysis described earlier.

The two weighted transducers in Figs. 5 and 6 have rather different impulse responses. However, the withdrawal-weighted transducer can easily be designed such that the *integral* of its impulse response is similar to the *integral* of the impulse response of the apodised transducer. When this is done, the frequency responses of the two transducers will be similar. The main difference is that the stepped nature of the withdrawal-weighted transducer introduces unwanted sidelobes at some distance from the passband of the filter. For this reason, the technique is only applicable for narrow-band devices.

Performance of Transversal Bandpass Filters. Filters of the types described here can give excellent performance because their behaviour is well understood and the devices are very reproducible, so that small errors can be corrected by redesigning. The main limitation is that insertion losses must be high in order to minimise the triple-transit signal. For the best results, single-electrode transducers are avoided and great care is taken in the design – errors due to the circuit effect, diffraction and other second-order effects are compensated for [19]. Filters for vestigial sideband (VSB) applications in TV are made with 6 MHz bandwidth centred at 40 MHz, in-band variations of only $\pm$ 0.2 dB in amplitude and $\pm$ 20 nsec in delay, and skirt width 500 kHz. Ganss-Puchstein [20] developed filters for digital radio with 64-QAM (quadrature amplitude modulation), with 20 MHz bandwidth and an amplitude ripple of only $\pm$ 0.05 dB. Peach [17] shows filters at 160 MHz, with a shape factor (ratio of 40 and 3 dB bandwidths) of 1.04, and with an in-band ripple of $\pm$ 0.15 dB.

Correlation

In the context of SAW devices, correlation refers to the process of applying a complex waveform to a linear device whose impulse response is the time-reverse of the input waveform. The device is known as a 'matched filter'. The output waveform has a prominent peak called the 'correlation peak'. If the input waveform is accompanied by noise, as is always the case for a radio receiver, the correlation process increases the signal-to-noise ratio (SNR), giving the maximum SNR obtainable by linear signal processing. Several types of SAW device have been developed for this purpose.

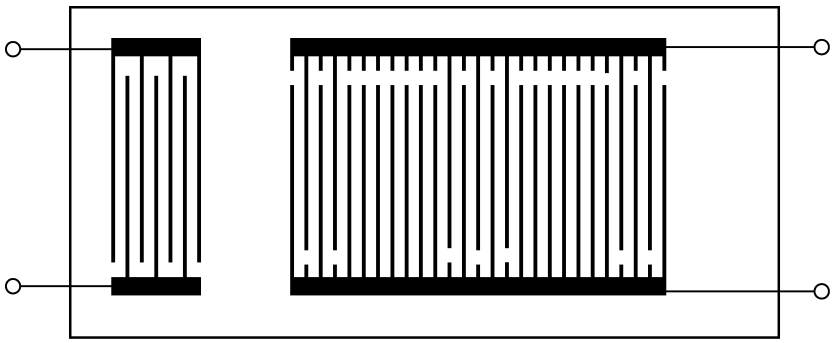


Figure 6. Transversal filter using withdrawal-weighted transducer.

Pulse Compression. This technique, used in radar systems, increases the maximum range at which a target can be detected without needing an increase in the peak transmitted power. In principle any type of complex waveform is applicable, provided the time-bandwidth product (TB) is large, where T is the waveform duration and B is its bandwidth. The matched filter increases the signal-to-noise power ratio by a factor TB, equivalent to increasing the transmitted peak power by this factor. For radar systems, TB is typically 50 to 500.

In practice the transmitted waveform is usually a chirp, that is, a waveform whose frequency varies monotonically with time. The echo due to a point target has the same form. The matched filter needs to have its impulse response as the time-reverse of this, another chirp, and this can be regarded as a dispersive delay line. In essence, the filter applies a delay dependent on frequency, such that all frequency components arrive at the output at the same time. Figure 7 shows a typical arrangement, where the right transducer has graded periodicity. For a single-frequency SAW, the transducer responds most strongly where its pitch corresponds to the SAW wavelength, so that different delays are obtained for different frequencies. The left transducer can be uniform, as shown, or it could be another graded-periodicity transducer.

In the radar receiver, the device usually has one transducer apodised in order to give the impulse response a humped envelope. This reduces the sidelobes of the output signal, which need to be small so that they are not falsely interpreted as extra targets. The sidelobes can be 35 dB below the main peak. Typically, bandwidths are in the region 10 to 100 MHz and durations (T) are 5 to 20 μsec. For the longer devices quartz is the preferred material because temperature stability is needed, while shorter devices use strong piezoelectrics (lithium niobate or tantalate) to obtain acceptable losses.

Interdigital devices generally give larger insertion losses as the TB product is increased. This limitation is overcome in the Reflective Array Compressor (RAC), which has two arrays of inclined grooves, each reflecting the waves through 90°. The groove arrays reflect strongly when the pitch corresponds to the wavelength, and the RAC has graded pitch so that different frequencies are subject to different path lengths and hence delays. With this technique, devices with TB products up to 16000 have been produced [21].

PSK Filters. While chirp waveforms are common in radar systems, communication systems often use phase-shift-keyed (PSK) waveforms, or other related waveforms. A PSK waveform consists of a sequence of contiguous pulses of carrier, with relative phase 0 or 180°. In spread-spectrum communications, each bit of data can be represented by a sequence of this type, and correlation is needed to optimise the SNR. This can be achieved using a SAW tapped delay line, with taps corresponding to the pulses; the taps are connected in a polarity sequence corresponding to the coding.

Convolvers. Another type of correlator is the non-linear convolver, shown in Fig. 8. In this device, RF signals are introduced at both ends of a lithium niobate substrate, so that they propagate towards each other. Where they overlap, a weak non-linear mechanism causes mixing and a sum-frequency signal is produced, with amplitude proportional to the product of the input wave amplitudes. The non-linearity is an intrinsic property of the material. The product signal appears as an electric field which can be sensed by a uniform metal electrode (the parametric electrode) extending over most of the area between the input transducers.

If the input waveforms are $f_1(t)$ and $f_2(t)$, the wave amplitudes within the device have the forms $f_1(t–x/v)$ and $f_2(t + x/v)$ (ignoring some delays). The output waveform g(t) is proportional to the integrated product of these, so that

$$g(t) \propto \int f_1(t - x/v) \cdot f_2(t + x/v)dx$$

$$= (1/v) \int f_1(2t - \tau) \cdot f_2(\tau)d\tau \qquad (7)$$

where $\tau = t + x/v$. It is assumed here that the travelling SAW signals have finite length and that they overlap only under the parametric electrode. With this assumption, the integrals in eq. (7) can have limits $\pm\infty$. The second form in eq. (7) is a standard convolution integral, except that t has been replaced by $2t$ because of the relative motion of the two waveforms. As is well known from systems theory, convolution describes mathematically the process performed by any linear device. Hence the convolver acts as a *linear* filter, but with one of the input waveforms acting as the impulse response. In particular, it can correlate any type of coded waveform, including the chirp and PSK waveforms

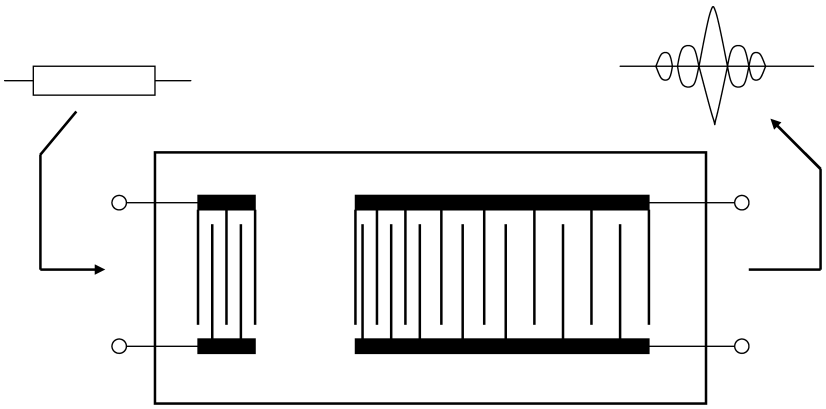


Figure 7. Interdigital pulse compression filter

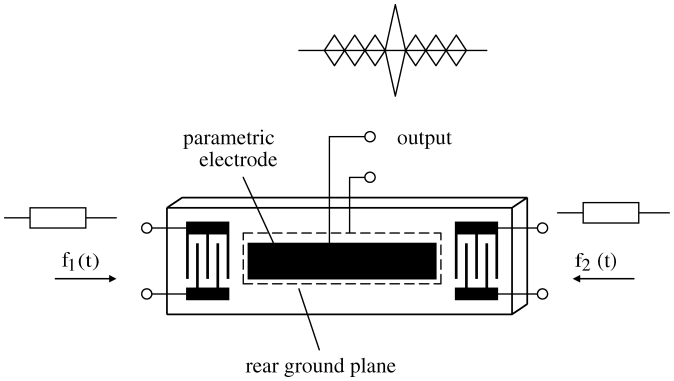


Figure 8. SAW convolver

considered above, and the coding can be changed instantaneously.

For the configuration of Fig. 8 the device produces a very weak output signal. To improve the efficiency, convolvers have been developed with the SAW beam compressed to a narrow width, typically 3 wavelengths, thus increasing the power density. Typically, the output power for CW signals is $P_O = P_1 + P_2 - 70$ dBm, where P_1 and P_2 are input powers and all powers are measured in dBm. Clearly the interaction is still weak. However, input power levels of 25 dBm are quite feasible, giving output levels strong enough for practical applications. Practical devices can have parametric lengths of 20 μsec and bandwidths of 100 MHz. Interactions in semiconductors have also been used to improve the efficiency.

The convolver shows exceptional versatility. However, it is limited by the need to provide an input 'reference' waveform to act as the impulse response, and it has not found widespread application.

Resonators

As for microwave devices, SAW resonators can be used either for stabilisation of oscillators or as the basis for low-loss bandpass filters. A basic difference is that for Rayleigh waves there is no effective localised reflection mechanism – a single strip, for example, gives mode conversion as well as reflection and is therefore unsuitable for a resonator. However, strong reflectivity can be obtained from a grating consisting of an array of strips, with period equal to $\lambda/2$ at the centre frequency. Typically, each strip has a reflection coefficient in the region of 1%, and there may be say 200 strips. The strips are commonly shorted metal electrodes, though grooves are often used for oscillators because they give better Q values. The grating may be analysed using the COM equations given later.

As shown in Fig. 9, a resonant cavity can be formed by two reflecting gratings [22]. The waves penetrate some distance into the gratings, so that the cavity length is effectively larger than the separation of the gratings. Consequently, the frequency spacing of the resonances can be quite small, for example 1% of the centre frequency. The gratings reflect strongly only over a narrow bandwidth, and the device can be designed such that only one resonance appears.

The addition of one or two transducers between the gratings provides coupling to one electrical port (Fig. 9) or two electrical ports (Fig. 11). The transducers are often designed to have negligible electrode reflections. In high-Q devices for oscillator applications, this condition can be obtained by recessing the electrodes into grooves on the surface. Alternatively, it is possible to use double-electrode transducers. Such transducers have little effect on the cavity resonances. The substrate is usually ST-X quartz, because the high Q-value implies that temperature stability is important. Experimental Q-factors can be up to 10^5 at 100 MHz, or 10^4 at 1000 MHz [22].

Properly-designed resonators behave approximately as the equivalent circuits shown in Figs. 9 and 11, with a series resonant branch connected to capacitances equal to the transducer capacitances C_t. The one-port device has a reso-

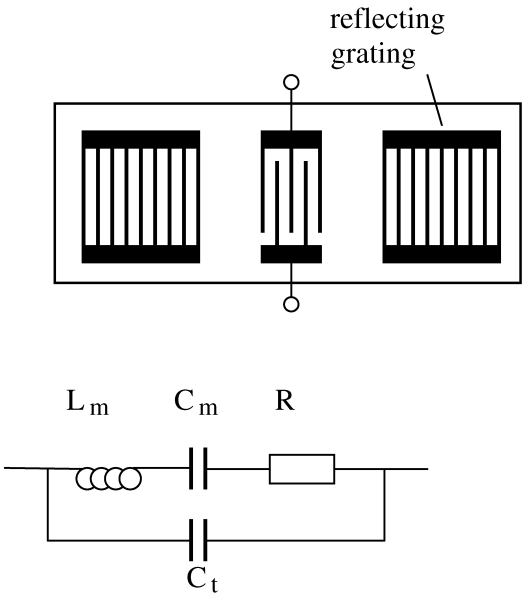


Figure 9. One-port resonator and its equivalent circuit

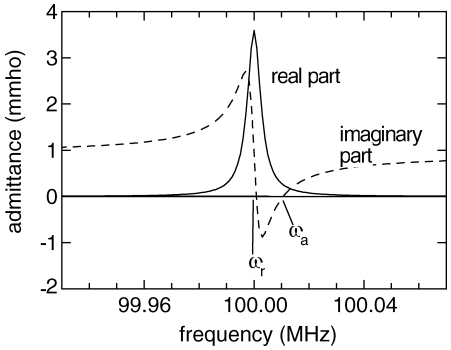


Figure 10. Admittance of a single-port resonator on ST-X quartz, using a non-reflective transducer, calculated using COM. The grating separation is 110.75 wavelengths.

nance frequency $\omega_r = [L_m C_m]^{-1/2}$, at which the admittance is maximised. There is also an anti-resonance frequency ω_a at which the admittance is zero, related to the capacitance ratio by $(\omega_a - \omega_r)/\omega_r \approx 0.5\, C_m/C_t$. These frequencies are shown on the admittance curves of Fig. 10.

Figure 12 shows the insertion loss of a two-port resonator. The main peak is due to the resonance. At frequencies away from the peak the response has shoulders due to direct coupling between the transducers, and these are not predicted by the equivalent circuit.

An alternative type of two-port resonator uses single-electrode transducers with the electrode pattern made as an extension of the adjacent grating pattern. In this case the device can be made entirely from a single metal film, and it is not necessary to recess the electrodes. This 'synchronous' resonator is much easier to make, especially at high frequencies, and Q's of 2250 have been obtained at 2.6 GHz [23]. The response appears as a distorted version of Fig. 12, but it still has a sharp peak usable for oscillator applications.

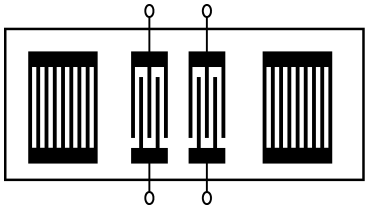


Figure 11. Two-port resonator and its equivalent circuit.

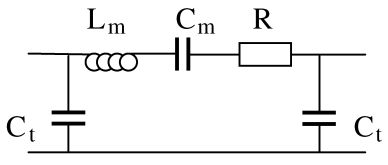
Oscillators

A two-port resonator can be used to realise an oscillator by connecting one port to the other via an amplifying feedback circuit, with the amplifier gain exceeding the resonator loss [22]. Alternatively, a SAW delay line can be used [24]. The delay line is designed with one or both transducers long enough so that it can oscillate at only one frequency. It is also possible to use a one-port resonator, with a special type of circuit [22].

The frequency stability of the oscillator is a crucial issue. Short-term stability relates to fluctuations associated with noise in the loop, causing the output signal to have a spectrum of finite width. Lewis's analysis [24] showed that for a delay line the 3 dB width of the spectrum is

$$\Delta\omega \approx 4\,k\,\theta(NF)G^2\omega 0^2/[Q^2P_0]$$

where k = Boltzmann's constant, θ = absolute temperature and $Q = \omega_0\tau$ is the Q-factor of the delay line, with delay τ. G, P_0 and NF are respectively the amplitude gain, output



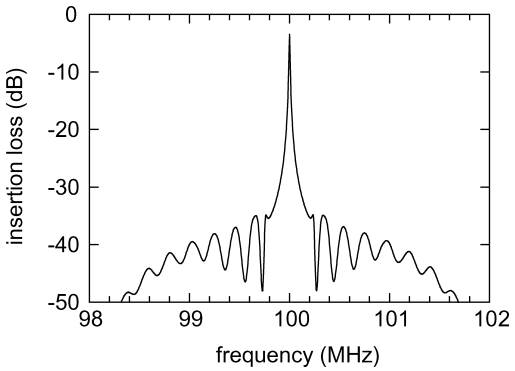


Figure 12. Theoretical insertion loss of a two-port resonator on ST-X quartz, with non-reflective transducers, calculated using COM.

power and noise figure of the amplifier, when saturated as in the oscillation condition. This formula also applies for a resonator, though it does not include flicker noise. A delay line needs to be physically much larger than a resonator to obtain the same Q. On the other hand, the maximum input power for a resonator is less than that of a delay line because of the magnification associated with the resonance.

Medium-term stability is associated with the temperature stability of the quartz, discussed above. Long-term stability refers to frequency drift over periods of months or years, and the mechanisms determining this are not clear. However, Parker [22] has shown that use of a package made of quartz can enable the ageing rate to be 0.1 ppm per year or less.

Improved results are obtainable using the surface transverse wave (STW) on quartz. This orientation gives high velocity, attractive for high-frequency devices, and can tolerate higher power levels than ST-X quartz, leading to better short-term stability [25].

Wireless Interrogation-Sensors and Tags

The temperature sensitivity of SAW devices is usually seen as a limitation. However, it can be exploited to measure temperature. Many other quantities can also be measured, and related devices are used as identification tags. A typical device is shown in Fig. 13 [26]. An interdigital transducer is connected directly to an antenna, generating SAW's when a RF pulse signal is received. A series of reflectors, usually short gratings, generates a characteristic pattern of reflected waves, and these are received by the transducer and radiated by the antenna. Typically, the 2.4 GHz ISM band is used. The delay in the SAW device helps to distinguish the return signal from the transmitted signal.

The return signal generated by this device is dependent on environmental factors such as temperature or stress, so these can be measured. It is also characteristic of the device design, and can therefore be used for identification. Identity tags can be made by using a sequence of reflectors coded in some way. For example, they might be in a regular sequence except that selected reflectors are omitted. Such devices potentially have wide-ranging applications. The number of possible codes is envisaged to be up to 10^{19} [27].

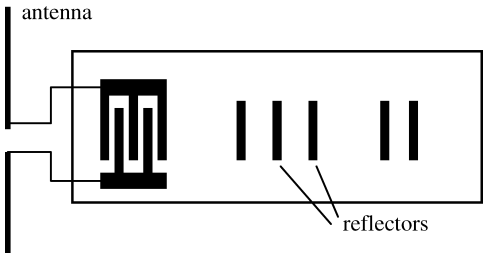


Figure 13. Reflective delay line for wireless tag or sensor.

Temperature can be measured by comparing the phases of waves from individual reflectors. Strain can also be measured, and with the device mounted on a sealed cavity this can also indicate pressure. Chemicals can be detected by coating the device with a film which absorbs the target chemical, causing a change of density and hence SAW velocity. As an alternative to the delay line of Fig. 13, one-port resonators can be used.

LOW-LOSS BANDPASS FILTERS

In the last two decades, the rapid expansion of wireless systems, particularly mobile telephony, has led to a strong demand for bandpass filters with low insertion losses. Losses often need to be less than 10 dB for IF filters, and 2 dB for RF filters. As discussed above, this requirement cannot be met using traditional transversal filters because of the triple-transit problem, and consequently a variety of new devices have been developed. There are many approaches for this problem [28], but here we consider only the main ones. They can be classified as follows:

1. Use a unidirectional transducer, which generates waves more strongly in one direction than the other. With this feature, triple-transit signals can be reduced while still maintaining low losses. It can be shown that the transducer must have internal reflectivity to satisfy this condition.
2. Develop filters based on SAW resonators, analogous to microwave resonator filters. In effect, an infinity of multiple-transit signals are used here.

Single-Phase Unidirectional Transducers (SPUDTs)

This term describes a class of unidirectional transducers; the phrase 'single-phase' distinguishes them from some earlier types which are of little interest now. The dominant type of SPUDT is known as the Distributed Array Reflective Transducer (DART), illustrated in Fig. 14. Each cell of this transducer has two narrow electrodes of width $\lambda_0/8$ and a wide electrode of width $3\lambda_0/8$, and the interelectrode gaps are all $\lambda_0/8$. Here λ_0 is the cell length, equal to the centre-frequency wavelength. Reflections from the narrow electrodes are negligible because they occur in pairs with spacing $\lambda_0/4$, so that reflections cancel at the centre frequency. Hence, in effect, only the wide electrodes act as SAW reflectors. The waves are effectively generated at the centres of the live electrodes. In general, it can be shown that for best directivity the distance between a generation

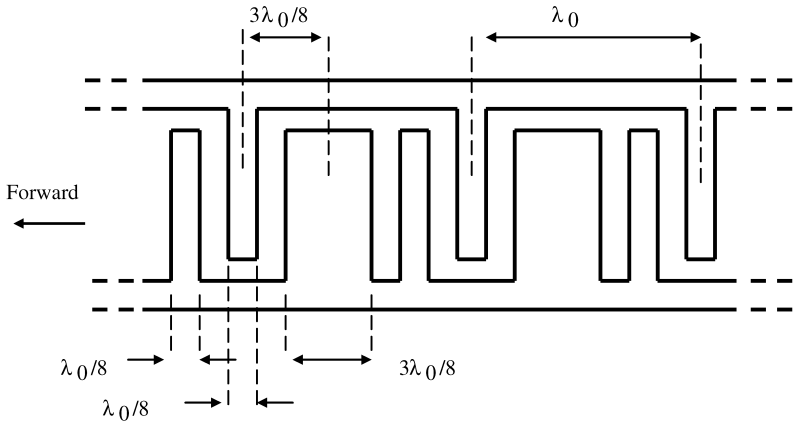


Figure 14. Structure of DART transducer.

centre and a reflection centre needs to be $(n/2 \pm 1/8)\lambda_0$, where the factor 1/8 appears because the reflection coefficient is known to be imaginary. In the DART, the distance between the two centres is $3\lambda_0/8$ and therefore satisfies this requirement [29].

Like any other transducer, a SPUDT gives optimum acousto-electric conversion if it is electrically matched. If it has enough directivity, it can be shown that this is also close to the condition that minimises the acoustic reflection. Thus, low loss and good triple-transit suppression are obtained simultaneously. In practical DART's it is common to have a directivity of 6 dB or more, and this is sufficient for these purposes. However, this is limited by the electrode reflectivity obtainable, and for this reason the DART needs to be fairly long, $100\lambda_0$ or more. Hence the bandwidth is quite small. The substrate is a temperature-stable material such as quartz.

Weighting can be applied to modify the response. The transduction in one cell can be removed by connecting the second electrode to the ground bus instead of the live one. Reflection can be removed by replacing the wide electrode by two narrow ones (each $\lambda_0/8$ wide, with a $\lambda_0/8$ space). Moreover, the transduction and reflection can be regarded as smooth functions of position, using withdrawal weighting to obtain the actual designs for individual cells. In this way, the design can satisfy performance criteria such as bandwidth, ripple, skirt width and stopband rejection. A considerable advance was made when it was realised that the reflection and transduction functions can also change sign, these changes being realised by displacing the corresponding electrodes. The use of different-signed reflectors is equivalent to incorporation of resonators into the structure, so the transducer became known as a resonant SPUDT, or R-SPUDT [29]. The resonances enable the device length to be substantially reduced, for a given specification. A further reduction was obtained by using a two-track arrangement, in which each track consists of two R-SPUDT's connected in parallel with the R-SPUDT's in the other track [30]. These size reductions have been very significant in the main application area, which is for IF filtering in CDMA mobile phone handsets, with bandwidth around 1 MHz. The design process is an extremely sophisticated 'global' optimisation procedure, taking account of the amplitude and phase ripple requirements as well as the amplitude response and tolerances. These devices can actually exploit the triple-transit effect, using it to reduce the skirt widths; in the passband, the triple-transit is usually minimised. This represents a considerable extension of the initial objective of the SPUDT, which was simply to reduce the triple transit level.

Another modification of the basic SPUDT is to vary the pitch transversely, so that the pitch is smaller at one side of the surface-wave track and larger at the other side. This gives a 'fanned' or 'tapered' transducer. The modification helps to obtain a flat passband with good sidelobe rejection. Such devices have been developed for IF filtering in mobile-phone basestations [13], using the minimal-diffraction orientation of quartz shown on Table 1.

Resonator Filters

As in microwave technology, SAW resonators can be coupled in order to realize a bandpass function. However, a SAW device only behaves as a resonator within the reflection band of the gratings. The width of this band is approximately $\Delta f/f \approx 2|r_s|/\pi$. Here $|r_s|$ is the strip reflection coefficient, typically a few percent, so the filter bandwidth obtainable is quite small.

An example is the longitudinally coupled resonator (LCR) filter, also called the dual-mode SAW (DMS) filter [31]. As shown in Fig. 15, this has three transducers between two reflecting gratings, with the outer two transducers connected together. In effect, each outer transducer plus its adjacent grating serves to act as a SPUDT, launching SAW's towards the centre of the device when a voltage is applied. The device as a whole behaves as a cavity with two resonances. This arrangement is often used for RF filters, with centre frequency 900 MHz or more, in which case the transducers are usually the single-electrode type. The response is typical of a two-pole filter, giving a flat passband with low loss when electrically matched. It is common to cascade two devices electrically in order to improve the stopband rejection. The device can give low loss and good stopband rejection, but the skirt width is relatively wide because of complications arising from reflections within the device.

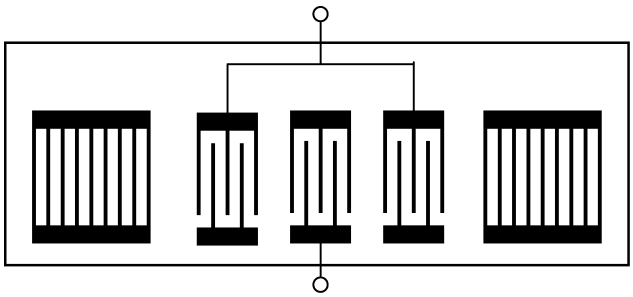


Figure 15. Longitudinally-coupled resonator (LCR) filter. Busbars shown without connections are grounded.

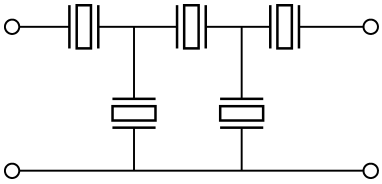


Figure 16. Typical configuration for impedance element filter (IEF). Usually, each resonator is a single-electrode transducer with a grating on each side.

Another resonator filter is known as the impedance element filter (IEF). This consists of a sequence of SAW resonators connected as an electrical circuit, without any acoustic coupling between them. Resonant behaviour can be obtained simply by using long single-electrode transducers. It is known that, for a large enough number of electrodes, the reflections can cause the transducer to behave as a resonator, with equivalent circuit approximately as in Fig. 9 [32]. Usually, reflecting gratings are also added at the ends to increase the Q-factor. The IEF uses a sequence of series and parallel resonators (Fig. 16), where the former have higher resonant frequencies. The filter centre frequency is at the antiresonance of the parallel resonator (where $Y \rightarrow 0$) and at the resonance of the series resonator (where $Z \rightarrow 0$), so that the signal is transmitted with little attenuation. Outside the passband there are deep nulls due to the resonance of the parallel resonator and the antiresonance of the series one. Beyond that are the stopbands where there is little acoustic activity, and here the response is determined mainly by the static capacitances and the electrical loading.

Like the LCR filter, the IEF can provide very low loss (1 dB) at RF frequencies (900 MHz and above) [33]. Its stopband rejection is not so good, but the skirts can be narrower. There is also a key advantage that quite high power levels can be tolerated, around 2 Watts. This ability is essential for some mobile phone transmitter filters. Many systems use a duplexer consisting of two IEF's, both connected to the antenna. One of these IEF's is at the transmitter output, and the other is at the receiver input. These devices normally use leaky waves on lithium tantalate, a subject described later.

The transverse-coupled resonator (TCR) filter is another resonant device, though it operates on entirely different principles [34]. As shown in Fig. 17 it has two identical tracks, each containing a single-electrode transducer

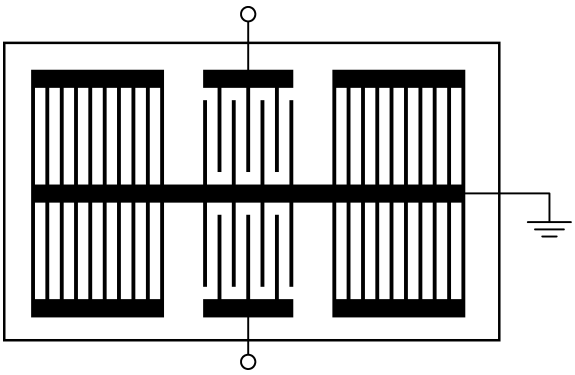


Figure 17. Transverse-coupled resonator (TCR) filter.

with a grating on either side. Coupling between the tracks occurs because the device behaves as a waveguide which supports two waveguide modes, one symmetric and one antisymmetric. This behaviour is associated with the narrow width, typically 6λ. Excitation of one transducer causes generation of both modes, and because of the different mode velocities a signal appears in the other track. For each mode the device behaves as a conventional resonator. The presence of two modes gives the device a 2-pole response. The pole frequencies correspond to the wave velocities of the modes, and because these are very close together the fractional bandwidth of the filter is very small, typically 0.1%. Using a ST-X quartz substrate, this filter has been widely used in the IF circuit of GSM mobile phone handsets, with typically 200 kHz bandwidth and 200 MHz centre frequency.

Waveguiding is also present in other SAW devices because the transducers reduce the SAW velocity, so that it is less than the free-surface velocity outside the structure. However, this is not usually significant unless an exacting performance is required. The distortion due to waveguiding can be reduced by designing a device to use a 'piston mode' of the waveguide, such that the amplitude is constant over the region occupied by the electrodes. In this case, the electrodes couple preferentially to the piston mode, reducing the distortion due to other modes. SPUDT filters have been designed using this principle [35].

Capabilities of Bandpass Filters

Table 2 gives a summary of the capabilities of the various types of filter. When considering a particular requirement, the different capabilities of the various devices need to be considered, including the temperature stability. A variety of packages are used, including metal and surface-mount ceramic types. Flip-chip mounting is often used to minimise the size, which may be as small as 2×2 mm^2 for GHz devices. The frequency response depends on the external terminating impedances, often specified as 50 Ω, and often the transducer capacitances need to be tuned out using inductors. It is also common to use two-component circuits to transform the impedances. Other factors, especially important at high frequencies, are stray capacitance or inductance, and electromagnetic feedthrough.

Table 2. Typical Performance for Main Types of Bandpass Filter

Type	Transversal	DART	LCR	IEF	TCR
Substrate	Various	ST-X quartz	42°Y-X $LiTaO_3$	42°Y-X $LiTaO_3$	ST-X quartz
Centre frequency (MHz)	50–500	80–500	500–2000	600–3000	50–200
Bandwidth	1–50 %	1–3 %	1–3%	1–4%	0.05–0.15%
Insertion loss	15–40 dB	6–10 dB	1–3 dB	1–3 dB	5–7 dB
Max. rejection	60 dB	45 dB	55 dB	45 dB	50 dB
Minimum shape factor*	1.1	1.5	3.0	1.5	2.0

*shape factor = ratio of 40 dB and 3 dB bandwidths

Many filters can be arranged in a balanced form, that is, at each port there are two live terminals designed to accept voltages with opposite polarities. This is needed for operation with some modern communication circuitry. It is also possible to have one port balanced and the other unbalanced, so that the device acts as a balun transformer in addition to its filtering function.

LEAKY SURFACE WAVES

In addition to the Rayleigh-type surface wave described earlier, some devices use alternative wave types which behave in a somewhat similar manner. A particular case occurs on 36°Y-X lithium tantalate. The wave is rather like a Rayleigh wave, but the particle displacement is approximately in the transverse direction (parallel to the surface) instead of being in or near the sagittal plane. This wave is not a true surface wave, as shown by the fact that it can radiate some energy into the bulk of the material. Hence the wave is called a 'leaky surface acoustic wave', or LSAW. It is related to a bulk shear wave. A numerical search for wave solutions gives a wavenumber with a small imaginary part, and hence there is some attenuation. The attenuation depends on rotation angle, and becomes small, negligible for practical purposes, at an angle of 36°. For this angle, the LSAW interacts with interdigital transducers and gratings in a manner similar to a true surface wave, so devices can be made using the same principles. However there are significant differences, notably that the LSAW gives rise to substantial bulk wave generation at frequencies above the passband.

This leaky wave has the advantage that, compared with SAW's on lithium tantalate, it gives stronger piezoelectric coupling without an increased TCD (in contrast, the SAW on Y-Z lithium niobate gives stronger coupling but the TCD is larger). In addition, the leaky wave can tolerate higher power levels, because the wave penetrates more deeply in to the substrate. It also has a higher velocity, increasing the electrode widths needed for a specified frequency. All these features are attractive for low-loss RF filters, and hence IEF and LCR filters for these applications often use this wave. Recently a modified orientation of 42° has been preferred, because this is found to give less loss when in the presence of transducers and gratings. For the specific application to IEF's, a further rotation to 48° has been shown to be beneficial [36].

A similar case occurs in 36°Y-cut quartz, with propagation normal to X. Again, this is related to a bulk shear wave, giving a high velocity with some attenuation. This was originally known as a surface-skimming bulk wave (SSBW). In practical applications the wave is trapped at the surface by transducers and gratings, and it is then known as a surface transverse wave (STW). The high velocity (5000 m/s) is attractive for high frequency devices and losses can be small, as shown by high-Q resonators at 1 GHz [25]. It has also been shown that waves of this type can be reflected efficiently by a small number of electrodes made of a heavy metal such as gold. In this way a resonator can use gratings having only 10 or so electrodes, so that it is much shorter that a Rayleigh-wave device [37]. It has also been shown that such waves can be efficiently reflected at an abrupt end of the substrate, so that the substrate ends form a resonant cavity [37].

Leaky waves have been found in a substantial variety of materials, often giving elevated piezoelectric coupling [38]. In addition to the above types, with shear polarisation, there are also 'longitudinal leaky surface acoustic waves' (LLSAW) with particle motion parallel, or almost parallel, to the propagation direction. These give even higher velocities. In 47°Y-X+90° lithium tetraborate there is a LLSAW with velocity 7000 m/s and $\Delta v/v = 0.7\%$. The Y-Z orientation of lithium niobate has an LLSAW with velocity 7285 m/s. Some IEF filters using this substrate had centre frequency 5.2 GHz, bandwidth 300 MHz and insertion loss 4 dB [39].

ANALYSIS METHODS

Analysis techniques for SAW devices vary greatly in sophistication. The simple delta-function model is very effective for illustrating some basic features, and it can provide the basis for optimised design methods such as the Remez algorithm. At the other extreme, numerical techniques can take account of complicated wave behaviour, together with effects due to the mechanical vibration of the electrodes. Such methods can simulate very complex second-order effects and they can be beneficial when demanding requirements are to be met, even though the analysis make take many hours on a modern computer. The following presents some notes on methods for unapodised devices, omitting derivations.

SAW Analysis for Thin Films-Effective Permittivity

The electrodes in a SAW device perturb the waves due to electrical and mechanical effects. For practical purposes, electrical effects are independent of the film thickness. Mechanical effects do depend on thickness, causing reflections of the waves. However, these are not important if the film

is thin enough, or if electrode reflections are cancelled as in the double-electrode transducer.

Assuming initially that mechanical effects can be ignored, an excellent basis for analysis is provided by the effective permittivity $\varepsilon_s(\beta)$. We define this by assuming a surface potential of the form $\bar{\phi}(\beta)\exp(j\,\beta x)$ and a charge density of the form $\bar{\sigma}(\beta)\exp(j\,\beta x)$. A factor $\exp(j\omega t)$ is omitted from the equations. We then write

$$\bar{\sigma}(\beta) = \varepsilon_s(\beta)|\beta|\,\bar{\phi}(\beta) \tag{8}$$

It can be shown that $\varepsilon_s(\beta)$ is actually a function of the slowness β/ω. It can be calculated numerically using the electromechanical constants of the material, and its orientation. Applying appropriate boundary conditions, this leads to an analysis allowing for electrode reflections and different wave modes (including bulk waves, for example) [40]. However, for many cases it is adequate to ignore bulk wave generation. Noting that $\varepsilon_s(\beta)$ must be zero for a free surface and infinite for a metallized surface, we can approximate it by using first-order Taylor expansions, leading to Ingebrigtsen's form

$$\varepsilon_s(\beta) \approx \varepsilon_\infty(\beta^2 - k_f^2)/(\beta^2 - k_m^2) \tag{9}$$

where $k_f = \omega/v_f$ and $k_m = \omega/v_m$ are respectively the wavenumbers for a free surface and a metallized surface. Also, ε_∞ is an abbreviation for $\varepsilon_s(\infty)$. This formulation can also be written in terms of a Green's function G(x), relating potential $\phi(x)$ to charge density $\sigma(x)$, such that

$$\phi(x) = G(x) * \sigma(x) = \int_{-\infty}^{\infty} G(x - x')\sigma(x')dx' \tag{10}$$

where the asterisk indicates convolution. It can be shown that the Green's function is the sum of an electrostatic term $G_e(x) = -\ln|x|/[\pi\varepsilon_\infty]$ and a surface-wave term $G_s(x) = j\Gamma_s\exp(-j\,k_f|x|)$, so that $G(x) \approx G_e(x) + G_s(x)$. Here Γ_s is given by

$$\Gamma_s \approx [(v_f - v_m)/v_f]/\varepsilon_\infty = (\Delta v/v)/\varepsilon_\infty \tag{11}$$

When a voltage is applied to a transducer, the surface wave generated in the $-x$ direction has surface potential $\phi_s(x) = j\,\Gamma_s\bar{\sigma}(k_f)\exp(j\,k_f x)$, where $\bar{\sigma}(\beta)$ is the Fourier transform of the charge density $\sigma(x)$. This arises from the SAW Green's function $G_s(x)$. For the same transducer receiving SAW's, with potential ϕ_i, the short-circuit current is given by the reciprocity relation

$$[I_{sc}/\phi_i]_{receive} = -(\omega W/\Gamma_s)[\phi_s/V]_{launch} \tag{12}$$

provided the two potentials are evaluated at the same location. Here W is the beam width.

Quasi-static Solution. The above formulation leads to complex solutions because it includes the effects of electrode reflections. However, it can be shown that these reflections can be eliminated from the analysis if the charge density $\sigma(x)$ is replaced by the electrostatic charge density $\sigma_e(x)$, that is, the charge density for the non-piezoelectric case $\Gamma_s = 0$. The SAW potential generated is then $\phi_s(x) = j\Gamma_s\bar{\sigma}_e(k_f)\exp(jk_f x)$. This is a much simpler formulation because $\sigma_e(x)$ is independent of frequency. Integration of $\sigma_e(x)$ over individual electrodes also gives the transducer capacitance, C_t. Established numerical techniques [41] can be used to calculate this function.

In many transducers (including withdrawal weighted transducers) the electrodes are regular, that is, they have constant width and spacing. In this case there is a simple algebraic method. First, imagine an infinite array of electrodes with width a and pitch p, and suppose that unit voltage is applied to one electrode, centred at $x = 0$, with all other electrodes grounded. The electrostatic charge density for this case is denoted by $\rho_f(x)$ (which is finite on all electrodes). The total charge density on a transducer can be obtained simply by superposition, using $\rho_f(x)$ for each live electrode. The potential of the SAW generated is simply $\phi_s = j\Gamma_s\sum_n \bar{\rho}_f(k_f)\exp[jk_f(x - x_n)]$, where the sum is over the live electrodes, which are centred at x_n. Here $\bar{\rho}_f(\beta)$ is the Fourier transform of $\rho_f(x)$, given by the formula

$$\bar{\rho}_f(\beta) = 2\varepsilon_\infty\sin(\pi s)P_m(\cos\Delta)/P_{-s}(-\cos\Delta) \quad \text{for } 0 < s < 1 \tag{13}$$

where $\Delta = \pi a/p$, $s = \beta p/(2\pi) - m$, and the integer m is chosen such that $0 < s < 1$. $P_\mu(x)$ is a Legendre function given by $P_\mu(x) = \sum_{n=0}^{\infty} a_n$, with $a_0 = 1$ and

$$a_n/a_{n-1} = (n - 1 - \mu)(n + \mu)(1 - x)/(2n^2) \tag{14}$$

When μ is an integer, the Legendre function becomes a polynomial. This formulation provides a simple way of analysing electrostatic phenomena, including the effect of the electrode width a on the fundamental and harmonic responses. For example, it shows that the 3rd harmonic of a single-electrode transducer vanishes when a/p = 0.5.

The capacitance of the transducer is obtained by integrating the electrostatic charges on the live electrodes. To facilitate this, we define Q_n as the integral of $\rho_f(x)$ over electrode n. Generally, this requires a numerical integration, but for the special case a/p = 0.5 it is found that $Q_n = 4\varepsilon_\infty/[\pi(1 - 4n^2)]$. With this formula, the transducer capacitance is easily calculated using $C_t = W\sum_n\sum_m \hat{P}_n\hat{P}_m Q_{m-n}$, where $\hat{P}_n$ are electrode polarities, equal to 1 for a live electrode and 0 for a grounded electrode. We find $C_t = \alpha\varepsilon_\infty W N_p$, where W is the aperture, N_p is the number of periods, and α = 1 for a single-electrode transducer and 1.414 for a double-electrode transducer.

Electrode Reflections in Regular Arrays. Reflections are particularly relevant in reflective gratings and single-electrode transducers. They arise from electrical and mechanical loading of the surface, and to first order these effects can be analysed separately and then added. Electrical loading is usually negligible in weakly-piezoelectric materials such as quartz.

Electrical reflections can be deduced from eq. (10) [2], giving a formula involving Legendre functions. For a/p = 0.5 and a frequency such that $\lambda = 2p$, the reflection coefficient of one electrode is $r_e \approx \pm 0.72\,j(\Delta v/v)$. Here the + (−) sign is used for open-circuit (shorted) electrodes. This takes account of the influence of neighbouring electrodes. The phase of r_e is defined by referring it to the centre of the electrode. The fact that r_e is imaginary is a consequence of power conservation.

For mechanical loading, a first-order analysis based on perturbation theory [2] gives $r_e = C \cdot h/(2p)$ for a/p = 0.5, where h = film thickness. The constant C is typically in the range -0.5 to -0.3 for common SAW materials, though for 128°Y-X lithium niobate $C = +0.8$. For grooves on Y-Z lithium niobate or ST-X quartz, $C \approx +0.6$.

For transducer analysis including electrode reflections, the quasi-static method above can be adapted, giving the Reflective Array Model [42]. Alternatively, the COM equations below can be used.

P-matrix Description. For general purposes it is common to specify the behaviour of a transducer or grating in terms of a matrix P_{ij}, defined by

$$[A_{t1}, A_{t2}, I]^T = [P][A_{i1}, A_{i2}, V]^T \tag{15}$$

where superscript T indicates a transpose. Here I and V are the transducer current and voltage. The A's are SAW amplitudes, with subscript i (t) to indicate waves incident on (leaving) the transducer. These have additional subscripts 1 or 2 to indicate the acoustic port. The ports are imaginary lines near the edges of the transducer, and their precise location is unimportant. The amplitudes are defined such that $|A|^2/2$ equals the power of the surface wave, and the phase of A is the same as that of the surface potential ϕ_s. This requires the relation $A = \phi_s[\omega W/(2\Gamma_s)]^{1/2}$. Reciprocity leads to the relations $P_{21} = P_{12}$, $P_{31} = -2P_{13}$ and $P_{32} = -2P_{23}$. For a grating, the P_{ij} with i or j = 3 are not used.

The P-matrix can be applied quite generally, though it is not valid if more than one type of wave is present. For example, it cannot be applied to a SAW device which is also generating bulk waves. However, it is very useful for analysing devices with several components, such as two-port resonators. Simple formulae [43] can be applied to cascade components together, leading to a device Y-matrix.

Coupled-Mode Equations. These equations, also called coupling-of-modes (COM) equations, are often used for unapodised components with internal electrode reflections. They are written in the form [43, 44]

$$\begin{aligned} dC(x)/dx &= -j\,\delta C(x) + c_{12}B(x) + \alpha_1 V \\ dB(x)/dx &= j\,\delta B(x) + c_{12}^* C(x) + \alpha_1^* V \\ dI(x)/dx &= 2\,\alpha_1^* C(x) - 2\,\alpha_1 B(x) + j\,\omega C_1 \end{aligned} \tag{16}$$

where $\delta = k - k_0$ and k_0 is the wavenumber at the Bragg frequency, determined by the structure. For example, at the first stop band of a grating with electrode pitch p we have $k_0 = \pi/p$. C(x) and B(x) are slowly-varying wave amplitudes defined such that the 'real' amplitudes are $c(x) = C(x)\exp(-j\,k_0 x)$ and $b(x) = B(x)\exp(j\,k_0 x)$. The parameters c_{12} and α_1 give the reflection and transduction per unit length, and their presence in different places is needed to satisfy reciprocity. I(x) is the bus-bar current, which needs to be evaluated at one end to give the transducer current. C_1 is the capacitance per unit length.

For a single-electrode transducer, the parameters are given by

$$c_{12} = -r_s^*/p; \quad \alpha_1 = j\,\bar{\rho}_f(k)[2\,\omega W \Gamma_s]^{1/2}(4p) \tag{17}$$

assuming that the acoustic ports are taken at the centre of the end electrodes. Here c_{12} and α_1 are independent of x, and the COM equations have an algebraic solution. We first consider the case V = 0, which applies for a grating or a shorted transducer. We look for a 'grating mode', in which C(x) and B(x) are both proportional to $\exp(-j\,sx)$. Equations (16) give

$$s^2 = \delta^2 - |c_{12}|^2 \tag{18}$$

When there are no losses, δ is real. Then s is also real, except for a stop band with edges at $\delta = \pm|c_{12}|$. The stop band has a fractional width $\Delta f/f = 2|r_s|/\pi$. Equations (16) can also be solved to find the P-matrix of a uniform transducer, assumed to extend from x = 0 to L. Defining $D = s\cos(sL) + j\delta\sin(sL)$, this is found to be

$$\begin{aligned} &P_{11} = -c_{12}^*\sin(sL)/D \\ &P_{12} = P_{21} = s\exp(-j\,k_0 L)/D \\ &P_{22} = c_{12}\sin(sL)\exp(-2j\,k_0 L)/D \\ &D \cdot P_{31} = 2\,\alpha_1^*\sin(sL) - 2\,s\,K_2[\cos(sL) - 1] \\ &D \cdot P_{32} = \{-2\,\alpha_1\sin(sL) + 2\,s\,K_1[1 - \cos(sL) - 1]\}\exp(-j\,k_0 L) \\ &P_{33} = -K_1 P_{31} - K_2 P_{32}\exp(j\,k_0 L) + 2(\alpha_1^* K_1 - \alpha_1 K_2)L + j\,\omega C_t \end{aligned} \tag{19}$$

where $C_t = LC_1$ is the transducer capacitance, $K_1 = (\alpha_1^* c_{12} - j\,\delta\alpha_1)/s^2$ and $K_2 = (\alpha_1^* c_{12} + j\,\delta\alpha_1^*)/s^2$. In eqs. (19), the first three equations apply for a grating as well as a transducer. We also find $P_{13} = -P_{31}/2$ and $P_{23} = -P_{32}/2$, as required by reciprocity.

To use the COM equations, c_{12} and α_1 must be known. Equation (17) gives approximate formulae for a single-electrode transducer, but it is also common to use values derived from more accurate analysis, or from experimental measurements. The equations are valid for a wide variety of devices. For example, they can also be used for SPUDT's, with appropriate values for c_{12} and α_1. For DART's, approximate formulae are available [42].

The COM equations can predict directivity (i.e. $|P_{23}| \neq |P_{23}|$), depending on the phases θ of c_{12} and ϕ of α_1. Maximum directivity occurs when $2\theta - \phi = n\pi$. The transducer is bidirectional ($|P_{23}| = |P_{13}|$) if $2\theta - \phi = (2n + 1)\pi/2$. Usually the transducer geometry is designed to give directivity if needed, as in the DART. However, a symmetrical transducer, which may be the single-electrode type, can give directivity because of asymmetry in the substrate properties. This is known as the Natural SPUDT effect. It occurs in many materials to some extent (Table 1), but for common SAW materials and orientations it is absent or negligible.

Accurate Analysis Methods. Although the above methods are often adequate, many situations call for more detailed techniques. This is the case when better accuracy is needed, for example when the electrodes are relatively thick or have non-uniform geometry as in the DART. For such cases, numerical methods based on finite-element (FEM) or boundary-element (BEM) methods have been developed [45–48]. For long transducers, such methods can be very time-consuming. In this case, a possible approach is to use an accurate analysis for a short transducer and use it to deduce the COM parameters; the latter can then be used to analyze a long transducer. Another approach is to use FEM to analyze infinite gratings corresponding to the transducer under short-circuit and open-circuit conditions.

From the stop-band edge frequencies, the COM parameters c_{12} and α_1 can be deduced [49, 50].

Accurate analysis of leaky wave and STW devices call for different approaches, because these waves are fundamentally different from Rayleigh SAW's. The waves are substantially affected by surface gratings and transducers. For example, a grating can serve to trap the wave at the surface, thus reducing the attenuation. However, at frequencies above the stop band it can cause bulk wave radiation and thus have the opposite effect. In a transducer, the coupling parameter α_1 is known to increase with the film thickness h of the electrodes, in contrast to Rayeligh waves which give α_1 independent of h. For these reasons, analysis of these waves is best done with the grating included a priori. Because of these complications, analysis of these waves has been mostly confined to uniform transducers and gratings. For leaky waves, centre-frequency COM parameters (dependent on film thickness) have been deduced from FEM analysis of gratings [49]. Hashimoto and Yamaguchi [51] analyzed a metal grating starting from the effective permittivity and adding mechanical loading using FEM analysis. The numerical results were fitted to an empirical formula due to Plessky [52]. This enabled frequency-dependent COM parameters to be deduced, including the simulation of loss due to bulk wave generation. This approach was used for leaky waves on 36°Y-X lithium tantalate [53] and STW's on quartz [54].

REFERENCES

1. Lord Rayleigh (John Strutt). 'On waves propagated along the plane surface of an elastic solid'. *Proc. London Math. Soc.*, **17**, 4–11 (1885)
2. S. Datta. *Surface Acoustic Wave Devices*. Prentice-Hall, 1986.
3. D. P. Morgan. *Surface-wave Devices for Signal Processing*. Elsevier, 1991.
4. C. K. Campbell. *Surface Acoustic Wave Devices for Mobile and Wireless Communications*. Academic Press, 1998.
5. K.-Y. Hashimoto. *Surface Acoustic Wave Devices in Telecommunications*. Springer, 2000.
6. C. C. W. Ruppel and T. Fjeldy (eds.). *Advances in Surface Acoustic Wave Technology, Systems and Applications*. World Scientific Publishing, 2000 (vol. 1), 2001 (vol. 2).
7. R. H. Tancrell and M. G. Holland. 'Acoustic surface wave filters'. *Proc. IEEE*, **59**, 393–409 (1971) .
8. A. J. Slobodnik, E. D. Conway and R. T. Demonico. *Microwave Acoustics Handbook*, vol. 1A, *Surface Wave Velocities*. US Air Force Cambridge Research Labs., Report No. AFCRL-TR-73-0-597 (1973). Accession No. AD 780172.
9. J. G. Gualtieri, J. A. Kosinski and A. Ballato. 'Piezoelectric materials for acoustic wave applications. *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, **41**, 53–59 (1994).
10. G. Kovacs et al. 'Improved material constants for $LiNbO_3$ and $LiTaO_3$'. *IEEE Ultrasonics Symp.*, 1990, pp. 435–438.
11. G. W. Farnell. 'Elastic surface waves'. InH. Matthews (ed.), *Surface Wave Filters*, Wiley, 1977, pp. 1–53.
12. A. J. Slobodnik. 'Materials and their influence on performance'. InA. A. Oliner (ed.), *Acoustic Surface Waves*, Springer, 1978, pp. 225–303.
13. B. P. Abbott and L. Solie. 'A minimal diffraction cut of quartz for high performance SAW filters'. *IEEE Ultrasonics Symp.*, 2000, pp. 235–240.
14. J. Koike, K. Shimoe & H. Ieki. '1.5 GHz low-loss SAW filter using ZnO/sapphire substrate'. *Japan. J. Appl. Phys.*, **32**, 2337–2340 (1993).
15. N. Naumenko and L. Solie. 'Optimal cuts of langasite, $La_3Ga_5SiO_{14}$ for SAW devices'. *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, **48**, 530–537 (2001).
16. L. R. Rabiner and B. Gold. *Theory and Application of Digital Signal Processing*. Prentice-Hall, 1975.
17. R. Peach. 'The use of linear programming for the design of SAW filters and filterbanks'. *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, **41**, 532–541 (1994).
18. C. S. Hartmann. 'Weighting interdigital surface wave transducers by selective withdrawal of electrodes'. *IEEE Ultrasonics Symp.*, 1973, pp. 423–426.
19. G. Visintini, C. Kappacher and C. C. W. Ruppel. 'Modular two-dimensional analysis of surface-acoustic-wave filters - Part II'. *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, **39**, 73–81 (1992).
20. R. Ganss-Puchstein, C. Ruppel and H. R. Stocker. 'Spectrum shaping SAW filters for high bit rate digital radio'. *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, **35**, 673–684 (1988).
21. H. M. Gerard, P. S. Yao and O. W. Otto. 'Performance of a programmable radar pulse compression filter based on a chirp transformation with RAC filters'. *IEEE Ultrasonics Symp.*, 1977, pp. 947–951.
22. T. E. Parker and G. K. Montress. 'Precision surface-acoustic-wave (SAW) oscillators'. *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, **35**, 342–364 (1988).
23. L. L. Pendergrass and L. G. Studebaker. 'SAW resonator design and fabrication for 2.0, 2.6 and 3.3 GHz'. *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, **35**, 372–379 (1988).
24. M. Lewis. 'The surface acoustic wave oscillator – a natural and timely development of the quartz crystal oscillator'. *Annual Freq. Control Symp. (NTIS)*, 1974, pp. 304–314.
25. I. D. Avramov et al. 'Extremely low thermal noise floor high power oscillators using surface transverse wave devices'. *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, **43**, 20–29 (1996).
26. A. Pohl. 'A review of wireless SAW sensors'. *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, **47**, 317–332 (2000).
27. C. S. Hartmann. 'A global SAW ID tag with large data capacity'. *IEEE Ultrasonics Symp.*, 2002, pp. 65–69.
28. C. C. W. Ruppel et al. 'SAW devices for consumer applications'. *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, **40**, 438–452 (1993).
29. J. M. Hodé et al. 'SPUDT-based filters: design principles and optimization'. *IEEE Ultrasonics Symp.*, 1995, pp. 39–50.
30. M. Solal et al. 'Advanced design techniques for high-performance IF and RF SAW filters'. *Intl. Symp. Acoustic Devices for Future Mobile Comm. Systems.*, Chiba University, Chiba, Japan, 2001, paper 3A-1.
31. T. Morita et al. 'Wideband low loss double mode SAW filters'. *IEEE Ultrasonics Symp.*, 1992, pp. 95–104.
32. V. P. Plessky. 'SAW impedance elements'. *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, **42**, 870–875 (1995).
33. X. Perois et al. 'An accurate design and modeling tool for the design of RF SAW filters'. *IEEE Ultrasonics Symp.*, 2001, pp. 75–80.

34. S. V. Biryukov et al. 'Consistent generalisation of COM equations to three-dimensional structures and the theory of the SAW transversely coupled waveguide resonator filter'. *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, **42**, 612–618 (1995)
35. M. Solal, O. Holmgren and K. Kokkonen. 'Design modeling and visualization of low transverse modes R-SPUDT devices'. *IEEE Ultrasonics Symp.*, 2006, pp. 82–87
36. N. Naumenko and B. Abbott. 'Optimum cut of LiTaO3 for resonator filters with improved performance'. *IEEE Ultrasonics Symp.*, 2003, pp. 385–390.
37. M. Kadota, T. Yoneda, K. Fujimoto, T. Nakao and E. Takata. 'Resonator filters using shear horizontal-type leaky surface acoustic wave consisting of heavy-metal electrode and quartz substrate'. *IEEE Trans. Ultrason., Ferroelec. Freq. Contr.*, **51**, 202–210 (2004)
38. M. P. daCunha. 'Pseudo and high-velocity pseudo SAW's'. In C. C. W. Ruppel and T. Fjeldy (eds.). *Advances in Surface Acoustic Wave Technology, Systems and Applications*. World Scientific Publishing, vol. 2, 2001, pp. 203–243.
39. T. Makkonen, V. P. Plessky, W. Steichen, S. Chamaly, C. Poirel, M. Solal and M. M. Salomaa. 'Properties of LLSAW on YZ-cut LiNbO3; modelling and experiment'. *IEEE Ultrasonics Symp.*, 2003, pp. 613–616.
40. R. F. Milsom, N. H. C. Reilly and M. Redwood. 'Analysis of generation and detection of surface and bulk acoustic waves by interdigital transducers'. *IEEE Trans. Sonics and Ultrasonics*, **SU-24**, 147–166 (1977).
41. S. V. Biryukov and V. G. Polevoi. 'The electrostatic problem for the SAW interdigital transducer in an external electric field'. *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, **43**, pp. 1150–1159 (part I); pp. 1160–1170 (part II), (1996).
42. D. P. Morgan. 'Reflective array modelling for reflective and directional SAW transducers'. *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, **45**, 152–157 (1998).
43. V. Plessky and J. Koskela. 'Coupling-of-modes analysis of SAW devices'. In C. C. W. Ruppel and T. Fjeldy (eds.). *Advances in Surface Acoustic Wave Technology, Systems and Applications*. World Scientific Publishing, vol. 2, 2001, pp. 1–81.
44. D. P. Chen and H. A. Haus. 'Analysis of metal-strip SAW gratings and transducers'. *IEEE Trans. Sonics and Ultrasonics*, **SU-32**, 395–408 (1985).
45. P. Ventura. 'Full strip reflectivity study on quartz'. *IEEE Ultrasonics Symp.*, 1994, p. 245–248.
46. K. Hashimoto and M. Yamaguchi. 'Derivation of coupling-of-modes parameters for SAW device analysis by means of boundary element method'. *IEEE Ultrasonics Symp.*, 1991, p. 21–26.
47. P. Ventura, J. M. Hodé and B. Lopes. 'Rigorous analysis of finite SAW devices with arbitrary electrode geometries'. *IEEE Ultrasonics Symp.*, 1995, p. 257–262.
48. P. Ventura, J. M. Hodé and M. Solal. 'A new efficient combined FEM and periodic Green's function formalism for the analysis of periodic SAW structures'. *IEEE Ultrasonics Symp.*, 1995, p. 263–268.
49. Y. Suzuki et al. 'Coupled-mode theory of SAW periodic structures'. *Electronics and Commun. Jap.*, Part 3, vol.**76**, p. 87–98 (1993).
50. Z. H. Chen, M. Takeuchi and K. Yamanouchi. 'Analysis of the film thickness dependence of a single-phase unidirectional transducer using the coupling-of-modes theory and the finite-element method'. *IEEE Trans. Ultrason., Ferroelec. Freq. Contr.*, **39**, p. 82–94 (1992).
51. K. Hashimoto and M. Yamaguchi. 'Precise simulation of surface transverse wave devices by discrete Green function theory'. *IEEE Ultrasonics Symp.*, 1994, p. 253–258.
52. V. P. Plessky. 'A two parameter coupling-of-modes model for shear horizontal type SAW propagation in periodic gratings'. *IEEE Ultrasonics Symp.*, 1993, p. 195–200.
53. K. Hashimoto and M. Yamaguchi. 'General-purpose simulator for leaky surface acoustic wave devices based on coupling-of-modes theory'. *IEEE Ultrasonics Symp.*, 1996, p. 117–122.
54. B. I. Boyanov et al. 'COM-theory analysis of STW resonator structures'. *IEEE Ultrasonics Symp.*, 1995, p. 317–320.

DAVID P. MORGAN
Impulse Consulting,
Northampton,, England