Electromagnetic, elastic and electro-elastic waves

C. A. A. J. Greebe

Introduction

Electronics is very much concerned with electromagnetic waves — their generation, modulation, propagation, reception and processing. Electromagnetic waves may be transmitted through space or they may be guided by wires or other types of transmission line. Sometimes, especially in the microwave region, electromagnetic waves appear in the generating or processing equipment itself. Two important examples are the resonant cavity — where energy can be stored in the form of standing waves — and the delay line — where information can be stored in the form of modulated travelling waves.

Electronics also makes use of elastic waves: the quartz-crystal resonator is a very early and well known example. The use of elastic waves offers in many cases two notable advantages: the velocity of propagation is some $10^5$ smaller than that of electromagnetic waves — so that, in 1 cm of a solid, elastic waves are delayed by the same amount as electromagnetic waves in 1 km of a cable; also, in certain carefully prepared materials, the attenuation of elastic waves can be relatively very small.

Elastic waves in solids are almost always generated and detected electrically. The conversion of electric signals into mechanical signals and vice versa is usually done by means of piezoelectric materials such as quartz; sometimes magnetostrictive materials are used.

Attempts to generate high-frequency elastic waves were for a long time limited to frequencies below 100 MHz because the electromechanical conversion was always done with mechanically resonant transducers. Such transducers must be only one or a few half wavelengths in thickness and above 100 MHz they became so thin as to be difficult to make or too fragile for practical use. This difficulty was surmounted during the fifties [1] and progress was such that some years later (1966) it was possible to generate and detect coherent waves of no less than 114 GHz [2]. One of the features of this breakthrough was the integration of transducer and medium: for example, elastic waves in a quartz crystal were generated and detected by virtue of the piezoelectric property of the crystal itself [3].

The use of elastic waves in electronics only really got under way after another development: the application of elastic surface waves [4]. As the name implies, these waves propagate only on the surface, leaving the bulk of the solid undisturbed. Like elastic waves in the bulk material they have a low velocity and, for well-prepared surfaces, a low attenuation. They have however a great extra advantage: they are accessible over the whole length of their trajectory. This unique property opens up a whole range of possibilities which are easy to put into practice when the substrate is piezoelectric.

The waves can then be generated, processed and detected by means of simple comb-shaped surface electrodes (interdigital transducers, see fig. 1); for example, filters with a wide range of characteristics can be made simply by choosing the shape, spacing and number of the ‘teeth’ of the electrodes [4]. Layer structures on the medium can be used to guide the waves or to give local changes in their dispersion. Delay lines based on surface waves can be provided with a large number of points where the signal may be tapped off [5]. The waves can be amplified by drift electrons in an adjacent semiconductor [6]. Finally, surface waves are particularly well adapted to systems of planar integrated circuits.

![Fig. 1. Interdigital electrodes on a slice of a piezoelectric material (interdigital transducer) for the generation of elastic surface waves. The temporal frequency of the applied a.c. voltage, and the spatial frequency of the ‘fingers’ must correspond to the frequency and the wave number of the wave to be excited.](image-url)

---

In recent years there has therefore been a growing interest in all sorts of wave phenomena — bulk waves and surface waves, electromagnetic waves and elastic waves and combinations of these in piezoelectric materials. It seemed to be useful to attempt a systematic review of these various forms against a background of conventional, well known forms of wave propagation. This article, therefore, is meant as a sort of introduction to waves, and is of a tutorial nature; it gives no scientific ‘news’ but presents known material and points out relationships. The opportunity will also be taken of discussing certain perhaps in practice less important but nevertheless remarkable wave phenomena such as helicon waves.

In the first part of the article we shall consider wave propagation in unbounded media — for example in free space, in optically anisotropic media and in conductors with and without a magnetic field — starting from the differential equations for the appropriate variables of the medium. The travelling waves that we find are characterized by an angular frequency $\omega$ ($2\pi \times$ the frequency) and a wave vector $\mathbf{k}$ (whose components $k_x$, $k_y$ and $k_z$ are respectively $2\pi$ divided by the wavelengths in the $x$-, $y$- and $z$-directions). The waves may grow or diminish in both space and time (see fig. 3, p....), which is indicated by $k$ or $\omega$ having an imaginary part. A very important aspect of a wave phenomenon is the dispersion relation which is the relation between $\omega$ and $\mathbf{k}$. Among the subjects dealt with in this first part of the article are the familiar waves of light and sound; the strongly attenuated propagation in metals resulting in the skin effect; a variant of this in a strong magnetic field, the ‘helicon’ waves, and some longitudinal electric waves. We shall also consider a situation where the wave does not propagate in the direction of the wave vector $\mathbf{k}$, a matter to be borne in mind when considering anisotropic materials, such as crystals, whether carrying bulk or surface waves.

The second part of the article deals with the coupling of waves in unbounded media. Wave propagation in piezoelectric materials can be very complicated because the electric and elastic variables are not independent of each other. If, however, the coupling is weak, the problem can be considerably simplified by regarding the waves as coupled electric and elastic waves, each of which would propagate independently if the coupling were in fact zero. This method of attack can also be useful in other cases where there are many coupled variables. Among the examples discussed here is the amplification of acoustic waves (‘acoustic amplifier’).

In the third part of the article combinations of waves that can exist in two adjacent media are discussed. These include combinations of incident, refracted and reflected waves and also — our particular concern here — surface waves. A surface wave occurs in the well known phenomenon of total internal reflection, but in this case it occurs only in combination with the incident wave and the reflected wave. Modern developments in electronics, however, are concerned with true surface waves that are independent of any bulk waves. A simple example — the Bleustein-Gulyaev wave — will be discussed at length.

In concluding this introduction attention should be directed to a problem that will not be dealt with in this article but is of the greatest importance to investigations of wave behaviour in unusual, novel media. In general a travelling wave transports energy of which, usually, a fraction is dissipated in the medium. For a given real frequency the wave amplitude then diminishes in the direction of propagation ($k$ is partly imaginary); the medium is passive. There are, however, media which can be activated in one way or another; in such media waves are possible that become larger in the direction of propagation. In the acoustic amplifier, for example, the medium is a piezoelectric semiconductor which is fed with energy by means of a d.c. current; this energy is partly taken up by the acoustic wave. In a well designed device, the input signal re-appears, after traversing the medium, amplified at the output. It is, however, not at all certain that a medium in which such ‘amplifying’ waves are theoretically possible will necessarily be potentially useful as an amplifier. It is possible, for example, that the medium will exhibit ‘absolute instabilities’ and reacts to an input signal with an explosive increase of the variables. In this case the output signal is no longer related in any way to the input signal. A. Bers and R. J. Briggs have given a theoretical approach to the problem of how to decide, on the basis of the dispersion relation, whether a new medium will have absolute instabilities or whether it can be used for amplification [7]. The investigation of how the medium reacts to an excitation (input signal, source) is inherent to this analysis. We shall leave this question completely aside and consider only freely propagating waves, without enquiring how they are generated.

Waves in unbounded homogeneous media

Main features of the analysis

The method of analysis of wave phenomena in unbounded media will be illustrated by means of a simple one-dimensional example: an infinitely long uniform transmission line with a capacitance of \( \frac{C}{F/m} \) per unit length and an inductance of \( \frac{L}{H/m} \) per unit length; see fig. 2. Changes in current in this transmission line give rise to voltage differences across the inductances. The capacitances are charged by the difference in the currents before and after them, so that the voltages across the capacitances also change.

These qualitative relations between the two wave variables of this problem, the voltage \( V \) and the current \( I \) can be quantified in two homogeneous linear differential equations in the spatial coordinate \( z \) and the time \( t \):

\[
\begin{align*}
\frac{\partial V}{\partial z} + L \frac{\partial I}{\partial t} &= 0, \\
C \frac{\partial V}{\partial t} + \frac{\partial I}{\partial z} &= 0.
\end{align*}
\]

The solutions of these equations are exponential functions of \( z \) and \( t \):

\[
\begin{align*}
V &= V_0 \exp j(\omega t - k z), \\
I &= I_0 \exp j(\omega t - k z).
\end{align*}
\]

Substituting (2) in (1) yields two homogeneous linear equations for the complex amplitudes:

\[
\begin{align*}
kV_0 - \omega LI_0 &= 0, \\
\omega CV_0 - kI_0 &= 0.
\end{align*}
\]

There are solutions to (1) only when the determinant of the coefficients of (3) is zero and this condition gives the dispersion relation

\[
k^2 - \omega^2 LC = 0.
\]

From this we derive the phase velocity

\[ v = \omega/k = \pm 1/\sqrt{LC}, \]

which, combined with (3), gives the following ratio of the complex amplitudes:

\[
V_0/I_0 = \pm \sqrt{L/C}.
\]

The positive root is called the characteristic impedance of the transmission line.

The steps outlined above are typical of many problems of wave motion. We shall always express the properties of the medium or the physical system in terms of differential equations in the wave variables. We shall restrict ourselves, as above, to homogeneous linear differential equations, whose coefficients are independent of time and place: this expresses the fact that the properties of the medium remain constant and are spatially homogeneous. Substitution of harmonic waves leads to homogeneous linear algebraic equations for the complex amplitudes. In a well formulated problem, the number of these equations is equal to the number of variables. Putting the determinant of these equations equal to zero yields the dispersion relation. Subsequently, we can in general calculate all the complex amplitudes in terms of one of them and so find the ratios of all the real amplitudes as well as all the phase differences — that is to say, the 'structure' of the wave.

If there are several harmonic solutions these can be quite freely superposed. Superposition implies, by its nature, that the behaviour of each component wave is entirely independent of the presence of the others: there is no interaction between the components. The situation is quite different when the differential equations contain nonlinear terms. If such terms are sufficiently small, it is often possible to consider a solution as the sum of several approximately harmonic components, but the behaviour of each component will now depend on the presence of the other components: the components interact, some becoming stronger, others weaker.
The dispersion equation; dispersion

The left-hand side of the dispersion relation (4), i.e. the determinant of (3), can be factorized into two factors. If one of these is set equal to zero we get the dispersion relation for one type of wave, e.g. a wave travelling to the left ($v < 0$). The other factor set equal to zero gives a wave travelling to the right ($v > 0$).

This is a trivial example of what one always tries to do: to resolve the determinant of the wave problem into factors — setting each factor equal to zero gives a dispersion relation for one type of wave. A less trivial example is found in the problem of sound waves in an isotropic solid: when the equations are set up sufficiently generally, two factors are found in the determinant, one corresponding to longitudinal waves and the other to transverse waves. It is also possible to reverse this whole approach. For example, in this article we shall assume — to stay with sound waves in solids — a longitudinal wave in the $z$-direction in the given medium, and the structure of these waves. In the case of an isotropic substance, the characteristics of longitudinal waves in any direction would then also be known. However, whether other waves could exist in the medium then remains an open question.

In the transmission line all harmonic waves travelling to the right have the same velocity $v$. If a disturbance consists only of waves travelling in this direction, therefore, these all continue to proceed together along the line, i.e. they do not disperse from one another, so that the disturbance or signal retains its form while propagating at a velocity $v$ to the right. The transmission line is then called a dispersionless system. We shall encounter many other dispersionless systems but also systems with dispersion in which $v$ is a function of $k$ and where the shape of a disturbance in general changes as it is propagated.

Complex wave number and complex frequency

If the transmission line of fig. 2 has not only series inductance but also series resistance ($R$ per unit length, in $\Omega$/m), a term $IR$ must be added to the left-hand side of the first equation (1). Repeating the procedure outlined above, we arrive at the dispersion relation

$$k^2 + j\omega RC - \omega^2 LC = 0.$$  

Expressions (2) are thus no longer solutions for real $\omega$ and $k$. This is obvious physically: the line is no longer lossless so that the waves are attenuated as they are propagated. Our whole scheme can however still be retained and the attenuation included if $\omega$ and $k$ are allowed to be complex.

When $\omega$ and $k$ are written as the sums of real and imaginary parts:

$$\omega = \omega_r + j\omega_i,$$
$$k = k_r + jk_i,$$

the waveform (2) can be expressed as the product of an exponential and a harmonic factor:

$$\exp(j\omega t - kz) = \exp(-\omega_r t + k_z) \exp(j\omega_r t - k_z). \tag{6}$$

This represents a sinusoidal wave (the second factor) whose amplitude diminishes (or grows) both with time and from place to place. The general case is illustrated in the central curve of fig. 3. The other curves show the nature and behaviour of the excitation if $\omega$ or $k$ is purely real or purely imaginary. All these and the intermediate cases can be regarded as kinds of wave. Among them are phenomena which in ordinary experience would not be called waves, for example the alternating field in a waveguide when this is excited at a (real) frequency below the cut-off frequency; $k$ is then purely imaginary. Such a cut-off wave (or evanescent mode) is shown in fig. 3c.

Waves in three dimensions

For wave phenomena in three dimensions, the term $kz$ in equation (2) must be replaced by $k \cdot r$, where $r$ is the radius vector of a point in space defined by coordinates $x$, $y$, $z$, and $k$ is the wave vector having the components $k_x$, $k_y$, $k_z$ along these coordinates:

$$k = k_x i_x + k_y j_y + k_z k_z.$$  

If $k$ is complex it can be represented by the two real vectors $k_1$ and $k_2$:

$$k(k_x, k_y, k_z) = k_1(k_{x1}, k_{y1}, k_{z1}) + jk_2(k_{x2}, k_{y2}, k_{z2}),$$

$$k_x = k_{x1} + jk_{x2},$$
$$k_y = k_{y1} + jk_{y2},$$
$$k_z = k_{z1} + jk_{z2}.$$  

If $k_1$ and $k_2$ are parallel to one another, in other words, if the ratios $k_{x2}/k_{x1}$, $k_{y2}/k_{y1}$, $k_{z2}/k_{z1}$ are equal, then the problem can be reduced once more to a one-dimensional one. We only have to rotate the coordinate system until the new $z$-axis coincides with the common direction of $k_1$ and $k_2$; then $k \cdot r = k_z$. The waves are in this case essentially one-dimensional and plane waves: the wave variables are independent of the (new) $x$- and $y$-coordinates. The (new) $x$, $y$-planes (perpendicular to $k_1$ and $k_2$) are wavefronts.

When $k_1$ and $k_2$ are not parallel, the wave is essentially not one-dimensional. This is the case, for example, for surface waves, which are propagated parallel to the surface ($k_2$ // surface) but usually fall off exponentially in the perpendicular direction ($k_1 \perp$ surface).
With more than one dimension there is still only one dispersion relation. This implies that a great variety of waves is possible since, of the four complex quantities \( k_x, k_y, k_z \) and \( \omega \), three are in general independent.

**Notation**

In order to avoid more indices than are really necessary, we shall usually make no distinction between a complex variable, its real part (i.e. the actual physical quantity) and the complex amplitude. In equations of the type (3) and (5) we shall therefore omit the indices 0. This should give no difficulties: where the distinction is important it is usually clear from the context what is meant. Some care may be necessary with nonlinear combinations and relations; an expression such as \( IV \) for power, for example, is correct only if \( I \) and \( V \) are the actual instantaneous current and voltage.

The differential operators \( \partial / \partial t, \partial / \partial x, \ldots \) will be abbreviated to \( \partial_t, \partial_x, \ldots \). The algebraic equations of the type (3) are obtained from the differential equations of the type (1) by replacing the operator \( \partial_t \) by the factor \( j\omega \), \( \partial_x \) by the factor \( -jk_x \), etc.

We shall also be concerned below with curls and divergences of the vectorial wave variables. In terms of Cartesian coordinates the curl and divergence of an arbitrary vector \( a \) are defined as follows:

\[
\text{curl } a = \frac{\partial a_z}{\partial y} - \frac{\partial a_y}{\partial z},
\]

\[
\frac{\partial a_x}{\partial y} - \frac{\partial a_y}{\partial x},
\]

\[
\frac{\partial a_z}{\partial x} - \frac{\partial a_x}{\partial z}.
\]

\[
\text{div } a = \frac{\partial a_x}{\partial x} + \frac{\partial a_y}{\partial y} + \frac{\partial a_z}{\partial z}.
\]

Fig. 3. The character of the various waves represented by the expression \( \exp(j(\omega t - k_z)) \), classified according to whether \( \omega \) and \( k \) are real, imaginary or complex (see eq. 6). Solid curves: the waveform at a given instant. Dashed lines: a fraction of a period later. Dotted lines (e and f only): half a period later. Only cases (a), (b), (d) and (e) represent travelling waves in the conventional sense. It is assumed that \( \omega_r \) and \( k_r \) have the same sign: the waves travel from left to right (+z-direction). It is assumed that \( \omega_r \) is positive and \( k_r \) is negative, where they arise: the amplitudes decrease with time and from left to right. The opposite sign for \( \omega_r \) or \( k_r \) would imply waves of increasing amplitude.

\[\text{(8)}\]

The power is thus \( (\Re I)(\Re V) \) in terms of complex variables \( I \) and \( V \). Usually only the time average of such a product, \( (\Re I)(\Re V) \), is of interest; this is given by \( \Re(IV^*) \), where the asterisk denotes a complex conjugate.
Electromagnetic waves

Maxwell’s equations

Many natural phenomena are wholly or partly electromagnetic. The set of differential equations describing a wave phenomenon often, therefore, involves Maxwell’s equations in one way or another. In their most general form these are, in SI units:

\[ \text{curl } \mathbf{H} = \dot{\mathbf{D}} + \mathbf{J}, \quad \text{div } \mathbf{B} = 0, \quad \text{(9)} \]
\[ \text{curl } \mathbf{E} = -\dot{\mathbf{B}}, \quad \text{div } \mathbf{D} = \varepsilon_0 \mathbf{J}, \quad \text{(10)} \]

where \( \mathbf{H} \) is the magnetic field, \( \mathbf{E} \) the electric field, \( \mathbf{B} \) the magnetic flux density, \( \mathbf{D} \) the dielectric displacement, \( \mathbf{J} \) the current density and \( \varepsilon_0 \) the charge density.

‘Divergence’ can be interpreted as ‘strength of source’. Thus (10) states that charge is the source of the \( \mathbf{D} \) field and (9) states that the \( \mathbf{B} \) field has no sources. Similarly we can say that ‘curl’ is equivalent to ‘vortex strength’ [10].

If we take the divergence of (7), remembering that the div curl of any vector is zero, and combine the result with (10), we find the continuity equation for the charge:

\[ \text{div } \mathbf{J} = -\varepsilon \mathbf{J}. \quad \text{(11)} \]

This equation states that the charge decreases at locations where there is a source of current.

Finally, there is an important energy equation:

\[ -\text{div} [\mathbf{E} \times \mathbf{H}] = \mathbf{E} \cdot \dot{\mathbf{J}} + \mathbf{E} \cdot \dot{\mathbf{D}} + \mathbf{H} \cdot \dot{\mathbf{B}}, \quad \text{(12)} \]

which is found by combining (7) and (8) with the vector identity \(-\text{div} [a \times \mathbf{b}] = a \cdot \text{curl } b - b \cdot \text{curl } a\). Equation (12) may be interpreted as follows: energy is transported by the electromagnetic field with an energy-flow density given by the Poynting vector [10]

\[ \mathbf{S} = \mathbf{E} \times \mathbf{H}. \quad \text{(13)} \]

The three terms on the right-hand side of (12) thus represent sinks (negative sources) for the energy flow. The first term represents the development of ohmic heat, the second the storage of electrical energy and dielectric losses, and the third the storage of magnetic energy and magnetic losses. We shall encounter the second term \( \mathbf{E} \cdot \mathbf{D} \) again in our considerations below.

Maxwell’s equations are ‘laws of nature’ in the sense that they are always and everywhere valid. However, they leave a considerable freedom in the behaviour of the electromagnetic variables: they give only 8 scalar relations as against 16 scalar variables. The properties of the wave are further determined by the properties of the medium. Therefore one can expect new and unusual electromagnetic phenomena if new and unusual media become available. An example is furnished by the remarkable helicon waves, first discovered on paper, which can be generated in very pure sodium at very low temperatures (4 K) in a strong magnetic field (10⁴ Oe). These electromagnetic waves, which will be discussed in more detail below, propagate with almost no attenuation at the unusually low velocity (for electromagnetic waves) of, say, 10 cm per second.

In simple cases the properties of the medium can be specified by three constants of the material: the permittivity \( \varepsilon \), the magnetic permeability \( \mu \) and the conductivity \( \sigma \). The following three equations then describe the medium:

\[ \mathbf{D} = \varepsilon \mathbf{E}, \quad \mathbf{B} = \mu \mathbf{H}, \quad \mathbf{J} = \sigma \mathbf{E}. \quad \text{(14)} \]

In the special case where the medium is free space,

\[ \mathbf{D} = \varepsilon_0 \mathbf{E}, \quad \mathbf{B} = \mu_0 \mathbf{H}, \quad \mathbf{J} = 0, \quad \sigma_0 = 0. \quad \text{(15)} \]

When \( \mathbf{D} \) and \( \varepsilon_0 \mathbf{E} \) differ, as they do for a physical medium, this is a consequence of the electric polarization of the medium. Equally, any difference between \( \mathbf{B} \) and \( \mu_0 \mathbf{H} \) is a consequence of the magnetization of the medium.

If we introduce the electric polarization \( \mathbf{P} \) via the definition \( \mathbf{D} = \varepsilon \mathbf{E} + \mathbf{P} \), the expression \( \mathbf{E} \mathbf{D} \) for the electrical energy delivered by the electromagnetic field in time \( \text{dt} \) (see the text referring to eq. (12) and (13)) becomes clearer physically: 

\[ \mathbf{E} \mathbf{D} = \mathbf{E} \mathbf{(\varepsilon \mathbf{E} + \mathbf{P})} = \mathbf{(\varepsilon \mathbf{E})}^2 + \mathbf{E} \mathbf{P}. \]

The first term is the increase in the free-space field energy and the second term is the work done on the medium by the field (force \( \times \) displacement).

Taking the equations (14) together with (7), (8), (9) and (10), the medium seems to be ‘overdetermined’: we have five vectors \( \mathbf{D}, \mathbf{J}, \mathbf{E}, \mathbf{B} \) and \( \mathbf{H} \) and one scalar \( \rho_0 \) and also five vector equations in (7), (8) and (14) but two scalar equations (9) and (10). However, the derivative of (9) with respect to time, \( \text{div } \dot{\mathbf{B}} = 0 \), is a direct consequence of (8) (because \( \text{div curl } \mathbf{E} = 0 \)). For our time-dependent waves, with \( \mathbf{B} = \omega \mathbf{B} \) and \( \omega \neq 0 \), this means that (8) implies (9). In a more general situation, the independent information given by (9) is concerned only with the constant (time-independent) part of \( \mathbf{B} \).

In what follows we shall first derive the velocity and the structure of electromagnetic waves in free space. Then we shall consider other non-conducting media (\( \sigma = 0 \)). If in such media \( \varepsilon \) and \( \mu \) are truly constants of the material, i.e. wholly determined by the medium and not at all by the wave, then we should find waves that are qualitatively the same as in free space. An interesting phenomenon that does not occur in free space, double refraction, can be related to anisotropy of the medium; to describe the medium in such a case, instead of the constant \( \varepsilon \), six constants are necessary (in the worst case) combined in the permittivity tensor \( \varepsilon \).

Other phenomena that do not occur in free space, dispersion and absorption, can be described by a formal extension of the concept of permittivity to a
frequency-dependent complex permittivity (which thus also depends on the wave). Of course, this does not explain dispersion and absorption; to do this the required \( \varepsilon(\omega) \) has to be related to the structure of the medium. Faraday rotation, as we shall see, can also be described formally in the same way, using a particular complex permittivity tensor.

The same methods can be used to describe wave propagation in conducting media, because the conductivity can be represented by an imaginary part of the permittivity. In this case we shall proceed less formally and derive for example an effective \( \varepsilon \)-tensor that describes the propagation of helicon waves on the basis of the behaviour of the conduction electrons in a strong magnetic field. We shall also encounter longitudinal electric waves which are not possible in free space because \( \varepsilon_0 \) is not zero, but which may occur in conductors under certain conditions when the effective permittivity is zero.

**Electromagnetic waves in non-conducting media**

**Free space**

For the analysis of electromagnetic waves in free space we start with Maxwell's equations, combined with the equations (15). (We omit here the suffix 0 from \( s \) and \( u \); some of the results can then be used later.)

When we substitute \( j\omega \) for \( \partial_t \) and assume non-zero \( \omega \), the equations become considerably simpler. In view of the identity \( \text{div} \, \text{curl} = 0 \), not only does (9) follow from (8) but also (10) follows from (7) because \( J \) and \( q_0 \) are both zero. Two vector equations are thus left over for the two vectors \( E \) and \( H \):

\[
\begin{align*}
\text{curl } H &= j\omega \varepsilon E, \\
\text{curl } E &= -j\omega \mu H. 
\end{align*}
\]

For plane waves propagating in the \( z \)-direction we have therefore (with \( \partial_z = -jk, \partial_x = \partial_y = 0 \)):

\[
\begin{align*}
\omega \varepsilon E_z + kH_y &= 0 \quad (a) \\
kE_z - \omega \mu H_y &= 0 \quad (b) \\
\omega \varepsilon E_y + kH_z &= 0 \quad (c) \\
kE_y + \omega \mu H_z &= 0 \quad (d)
\end{align*}
\]

The determinant of these equations can be seen to factorize into four factors, and by putting each factor separately equal to zero we find in principle (see p. 314) four dispersion relations, each representing a wave. In each of the four waves, given by (a), (b), (c) and (d) the wave variables are different.

Now the pair of equations (a), with \( \varepsilon = \varepsilon_0 \) and \( \mu = \mu_0 \), already describe all the properties of electromagnetic waves in free space (see fig. 4a). Their velocity — the velocity of light in free space — follows from the dispersion relation for (a):

\[
k^2 - \varepsilon_0 \mu_0 \omega^2 = 0, 
\]

and is therefore \( 1/\sqrt{\varepsilon_0 \mu_0} \). The waves are transverse (\( E \) and \( H \) both perpendicular to \( k \)) and \( E \) and \( H \) are also perpendicular to each other. The ratio of the complex amplitudes \( E_x/H_y \) is equal to \( k/\varepsilon_0 \omega = \sqrt{\mu_0/\varepsilon_0} \), the intrinsic impedance of free space; and since this ratio is real, \( E_x \) and \( H_y \) are in phase. (The concept of the intrinsic impedance of a medium is directly analogous to that of the characteristic impedance of a line.)

**Fig. 4. Structure of electromagnetic waves in free space, a) corresponding to (17a), b) corresponding to (17b). The wave (b) is simply the wave (a) rotated through 90° about the \( z \)-axis. These waves are plane polarized. A circularly polarized wave (c) is obtained by superposition of waves (a) and (b) of equal amplitude but with 90° phase difference.**


[10] In (13) \( E \) and \( H \) are the actual electric and magnetic fields. Using complex wave variables, the time-averaged Poynting vector is \( \mathbf{S} = 4 \text{Re}(|E \times H|^2) \); cf. note [8].
The practical significance of the expression $1/V_{\varepsilon_0}$ for the velocity of light in free space is that, in the construction of a system of units such as SI, although there is some freedom of choice with respect to $\varepsilon_0$ and $\mu_0$, the combination $1/V_{\varepsilon_0}$ must always be equal to the velocity of light.

Other solutions are obtained by applying a rotation to that of fig. 4a; $k$ (or $\omega$) and $E$ (or $H$) can be freely chosen. The solution of the equations (17b) is simply the wave of (17a) rotated through 90° about the $z$-axis (fig. 4b). Any wave in free space can be described by a superposition of such waves.

One rather special case of superposition is the superposition of (a) and (b) where $E_x$ in (a) and $E_y$ in (b) are of the same amplitude but differ by 90° in phase:

$$E_x = \pm jE_y. \quad (19)$$

This is a circularly polarized wave (fig. 4c). The upper sign ($+$) represents a vector rotating clockwise, and the lower sign ($-$) represents a vector rotating anti-clockwise, as seen by an observer looking in the $+z$-direction, and assuming $\omega$ to be positive.

The end points of the vectors lie on a helix. The relation between the sense of this helix, the sense of rotation of the vectors and the direction of wave propagation can best be formulated by adopting the convention used in optics. In this convention a sense of rotation is defined as that seen by an observer receiving the wavefronts. Then the sense of the helix is the same as that of the rotation of the vectors (in whatever direction the wave propagates), and this is by definition the sense of the circular polarization. According to this definition the upper sign in (19) represents left-handed circular polarization for a wave travelling in the $-z$-direction.

The equations (17c) and (17d) would represent longitudinal electric waves and longitudinal magnetic waves respectively, but their dispersion relations, $\omega_\varepsilon = 0$, $\omega_\mu = 0$, are not satisfied in free space (for $\omega \neq 0$). For this reason longitudinal electromagnetic waves cannot exist in free space.

For a medium whose electric and magnetic properties are described by (14), where $\varepsilon$ and $\mu$ are true constants of the material and where $\sigma$ is zero, electromagnetic waves entirely analogous to those in free-space waves are possible; their velocity is $1/V_{\varepsilon\mu}$ and the intrinsic impedance $E/H = V_{\varepsilon\mu}/\sigma$. Such media do not really exist but in certain cases — for example that of low-frequency waves in an isotropic lossless insulator — the wave propagation is well described in this way.

We shall now examine what happens when the medium is not isotropic.

**Anisotropy; double refraction**

Let us consider a crystal that is not equally polarizable in all directions. The relation between $D$ and $E$ can now no longer be characterized by a single scalar quantity. We shall assume that an orthogonal coordinate system $\xi$, $\eta$, $\zeta$ exists in which

$$D_\xi = \varepsilon_1 E_\xi, \quad D_\eta = \varepsilon_1 E_\eta, \quad D_\zeta = \varepsilon_2 E_\zeta, \quad (20)$$

where $\varepsilon_2 \geq \varepsilon_1$; the polarizability is thus larger in the $\zeta$-direction than in the $\xi,\eta$-plane. We are then concerned with a uniaxial crystal in which the $\zeta$-axis is the optical axis. From (20) we can immediately conclude that $D$ and $E$ are no longer parallel to each other, unless they happen to be parallel or perpendicular to the $\zeta$-axis.

For light propagated along the optical axis, the calculation of p. 317 can again be used, with $\varepsilon = \varepsilon_1$. For propagation perpendicular to the optical axis, the calculation is also entirely analogous, except that in (17a), $\varepsilon = \varepsilon_2$ and in (17b), $\varepsilon = \varepsilon_1$; we therefore get two waves, one polarized along the optical axis and the other perpendicular to it, with different velocities.

The situation becomes really interesting when we consider plane waves whose $k$ vector makes an angle other than 90° with the $\zeta$-axis. Going back to Maxwell’s equations (7) and (8) we find, taking a coordinate system $x,y,z$ in which $k$ is parallel to the $z$-axis, and taking $B = \mu H$, $J = 0$, $\mu \neq 0$ and $\omega \neq 0$:

$$kH_y = -\omega D_\eta, \quad \omega_\mu H_x = -kE_\eta, \quad kH_x = -\omega D_\eta, \quad \omega_\mu H_y = kE_\eta, \quad (21)$$

$$0 = D_\eta, \quad H_z = 0.$$

It follows that $D$, $H$ and $k$ are perpendicular to one another. $D$ and $H$ are still transverse: the wavefronts are $D,H$-planes. If $D$ is taken perpendicular to the $\zeta$-axis (fig. 6a) the situation is still quite unremarkable: $E$ is again parallel to $D$ and the wave has the same nature as we have already encountered. If, however, $D$...
is taken to be in the \( k, \zeta \)-plane (fig. 6b) then \( D \) is neither parallel nor perpendicular to the \( \zeta \)-axis, so that \( E \) is no longer parallel to \( D \) (see fig. 5). The Poynting vector \( S = E \times H \) is therefore no longer parallel to the wave vector \( k \). Since \( S \) gives the direction of the energy flow and therefore, in the case of a parallel beam, the direction of the beam, the wavefronts lie obliquely to this direction (fig. 7). An unpolarized beam falling perpendicularly on the \( x,y \)-plane in fig. 6a and b (this being the surface of the crystal) is therefore split into a 'straight-through' beam (fig. 6a) and an 'oblique' beam (fig. 6b): this is double refraction.

The permittivity tensor

We have seen that in an anisotropic material \( D \) and \( E \) are in general not parallel to one another (see fig. 5). It follows that, in a coordinate system \( x,y,z \) not parallel to the coordinate system \( \xi, \eta, \zeta \), used above, each component of \( D \) may depend on all the components of \( E \). In general we must write:

\[
D_k = \sum \epsilon_{kl} E_l. \quad (k,l = x,y,z) \tag{22}
\]

This applies also to a biaxial crystal for which three different \( \epsilon \)'s occur in (20). The two or three \( \epsilon \)'s in (20) and the nine \( \epsilon_{kl} \)'s in (22) give the same linear relationship between the physical vector fields \( D \) and \( E \); if yet another coordinate system is chosen, the nine quantities describing this relationship will have other values. Such a tensor relation between \( D \) and \( E \) is written:

\[
D = \epsilon E, \tag{23}
\]

where \( \epsilon \), the permittivity tensor, is thus a property of the crystal which, for each coordinate system \( x,y,z \), is defined by another array

\[
\begin{pmatrix}
\epsilon_{xx} & \epsilon_{xy} & \epsilon_{xz} \\
\epsilon_{yx} & \epsilon_{yy} & \epsilon_{yz} \\
\epsilon_{zx} & \epsilon_{zy} & \epsilon_{zz}
\end{pmatrix}
\tag{24}
\]

of nine scalar quantities.

In such a crystal the electric state and therefore the electric energy per unit volume \( U_E \) are determined by the values of \( E_x, E_y \) and \( E_z \). In a change of state in which \( D \) increases by \( dD \), the crystal takes up an energy of \( E \cdot dD \) per unit volume from the field (see p. 316) and \( U_E \) thus changes by this amount so that

\[
dU_E = E \cdot dD = \sum \epsilon_{kl} E_k dE_l. \tag{25}
\]

It follows that

\[
dU_E/dE_l = \sum \epsilon_{kl} E_k,
\]

and

\[
d^2U_E/dE_k dE_l = \epsilon_{kl}. \tag{26}
\]

---

[11] A right-handed coordinate system is assumed here as in fig. 4. It is further assumed that all the wave variables are proportional to \( \exp (+jwt) \) and not to \( \exp (-jwt) \), as is sometimes done. We shall continue to use these conventions.
From this it can be seen that $\varepsilon$ is symmetrical:

$$\varepsilon_{kl} = \varepsilon_{lk}. \quad (25)$$

Hence there are at most six different quantities in (24). The proofs that symmetries of the same nature must exist for elasticity and piezoelectricity run along the same lines.

The description of a crystal by means of six constants of the material $\varepsilon_{kl}$ is satisfactory for static and for slowly varying fields. Losses and dispersion which become important at higher frequencies are not encompassed by this description. They can however be included in the formal framework of the permittivity, if the latter concept is extended in the following manner.

**Complex and frequency-dependent permittivity**

If, for an isotropic material, $\varepsilon$ is regarded strictly as a constant of the material, then $D$ is everywhere and at every instant (independently of the situation elsewhere and of previous events or states) given by the value of $E$ at the same place and the same instant. In other words, $D = \varepsilon E$ is a *local, instantaneous* relation. In reality such a rigorous relation between $D$ and $E$ exists only in free space. For example, when $E$ varies very rapidly, the polarization and therefore $D$ in most materials also usually varies at the same frequency but the ratio of the amplitudes of $D$ and $E$ may well depend on the frequency and $D$ often lags behind $E$. At a given moment $D$ may thus depend not only on the value of $E$ at that moment but also on previous values. Such a non-instantaneous relationship will be accounted for, as usual, by regarding the expression $D = \varepsilon E$ as a relationship between the complex quantities $D$ and $E$, where $\varepsilon$ is then a quantity that may be complex and frequency-dependent:

$$\varepsilon = e'(\omega) - je''(\omega).$$

Since a minus sign is conventionally used here, a positive value of $e''$ indicates a lag of $D$ behind $E$ and this implies losses, as can be seen by calculating the mean energy dissipated by the dielectric per second and per unit volume, $(ReE)(ReD)$, which is found to be $1/2\omega e''E^*E$. At zero frequency we must recover the original relation between $D$ and $E$: this means that $e''(0)$ must be zero and $e'(0)$ must be the permittivity for static fields.

The complex representation and method of calculation thus allows us to take account of non-instantaneous relations between $D$ and $E$ and so to describe losses ($e'' \neq 0$) and dispersion ($e'$ is a function of $\omega$). Non-local relations between $D$ and $E$ will not be considered here. Later on, the conductivity $\sigma$ in (14) will sometimes also be taken as complex and frequency-depend-ent. The permeability $\mu$, on the other hand, will be considered here always as a real constant (and not complex as for example in problems related to electron and nuclear spin resonance).

Analogously we shall hereafter consider that $\varepsilon$ in (23) may be a complex and frequency-dependent tensor. Instead of (25) we must then have the symmetry relation

$$\varepsilon_{kl} = \varepsilon_{lk}^{*} \quad (26)$$

if the material is lossless. This follows since the time-average of $(ReE) \cdot (ReD)$ can be shown to be

$$1/2 \omega \sum_{kl} (\varepsilon_{kl} - \varepsilon_{kl}^{*})E_{k}E_{l}^{*},$$

and in a lossless crystal this must be zero for all $E$. At zero frequency the real part of $\varepsilon$ must again reduce to the original tensor for static fields and the imaginary part must again vanish. The relation (26) then reduces again to (25).

**The Onsager relations**

The conclusion that the permittivity tensor $\varepsilon$ must satisfy the relation (26) is based on the assumption that the medium is lossless. It can be shown in quite a different way that certain relations must in any case exist between the elements of $\varepsilon$, whether there are losses or not. However, other factors then have to be taken into account, e.g. whether or not the medium is subjected to a magnetic field $H_{0}$. If this is the case (and if this is the only other factor involved), then

$$\varepsilon_{kl}(-H_{0}) = \varepsilon_{lk}(H_{0}). \quad (27)$$

These are the well known *Onsager relations* \[12\] applied to $\varepsilon$. Thus, if there is no magnetic field, $\varepsilon$ is symmetric, whether there are losses or not. Only if $\varepsilon_{kl}$ is real does (27) reduce to (26) when $H_{0} = 0$.

The Onsager relations are applicable to the coefficients of many kinds of linear relationships in physics and engineering and are of a fundamental nature. They are based on the reversibility (in time) of micro-processes. They are valid only for coefficients relating variables that are conjugated in a prescribed manner. The derivation of the Onsager relations cannot be dealt with here.

Reversibility in a system of particles implies that all the particles would retrace their paths exactly if at a given moment all the velocities were reversed. All external influences that are antisymmetric in time must then also be reversed, for example electric currents and magnetic fields (which can always be considered as deriving from currents). The only influence of this kind mentioned in the foregoing was an applied magnetic field.

**Faraday rotation**

On the basis of the symmetry relations we shall now set up a very simple \( \varepsilon \)-tensor with which the rotation of the plane of polarization in a constant magnetic field (Faraday rotation) can be formally described. The question of how the form of the tensor depends on the structure of the medium will not be discussed.

Restricting ourselves to lossless media, we resolve \( \varepsilon \), element for element, into real and imaginary parts. From (26) the real part is symmetric, the imaginary part antisymmetric. We can therefore write:

\[
\varepsilon = \varepsilon_a + j\varepsilon_a,
\]

where \( \varepsilon_a \) is symmetric, \( \varepsilon_a \) is antisymmetric, and both are real.

Next, by a suitable rotation of coordinates, we reduce the \( \varepsilon \)-tensor to diagonal form. It can be shown that this is always possible for a real and symmetric matrix. \( \varepsilon_a \) then remains antisymmetric and real so that \( \varepsilon \) takes the form:

\[
\varepsilon = \begin{pmatrix}
\varepsilon_{a1} & 0 & 0 \\
0 & \varepsilon_{a2} & 0 \\
0 & 0 & \varepsilon_{a3}
\end{pmatrix} + j \begin{pmatrix}
0 & \varepsilon_{a3} & \varepsilon_{a1} \\
-\varepsilon_{a0} & 0 & \varepsilon_{a2} \\
-\varepsilon_{a2} & -\varepsilon_{a1} & 0
\end{pmatrix}.
\]

In our previously considered isotropic case (free space) all the \( \varepsilon_a \)'s would be equal and all the \( \varepsilon_a \)'s would be zero. One of the simplest deviations from isotropy — all the \( \varepsilon_a \)'s zero but one of the \( \varepsilon_a \)'s different from the other two — has also been discussed; this was the case of the uniaxial crystal (see 20) and it leads, as we have seen, to double refraction. If we now take all the \( \varepsilon_a \)'s equal but make one of the \( \varepsilon_a \)'s not zero:

\[
\varepsilon = \begin{pmatrix}
\varepsilon_a & j\varepsilon_a & 0 \\
-j\varepsilon_a & \varepsilon_a & 0 \\
0 & 0 & \varepsilon_a
\end{pmatrix},
\]

we have the tensor with which the Faraday rotation can be described. We note, first of all, that according to the Onsager relations, \( \varepsilon_a \) in (28) can be non-zero only if a constant magnetic field \( H_0 \) is present: for we must have \( \varepsilon_a(H_0) = -\varepsilon_a(-H_0) \). In the coordinate system \( x, y, z \) in which (28) is valid, the \( z \)-axis differs from the \( x \)- and \( y \)-axes: this must therefore be the direction of the magnetic field. The simplest case in which (27) is satisfied is that with \( \varepsilon_a \) proportional to \( H_0 \), so that its sign reverses if the field \( H_0 \) is reversed. If the frequency goes to zero (25) must again be satisfied and so \( \varepsilon_a \) must become zero.

Next we show that (28) leads to a rotation of the plane of polarization. For a wave propagating along the \( z \)-axis, Maxwell's equations, with \( B = \mu H, J = 0, \mu \neq 0 \) and \( \omega \neq 0 \), lead again to the equations (21). The waves are thus purely transverse; from (21), \( D_z \) and \( H_x \) are zero and (28) then shows that this is also the case for \( E_z \). Combining (21) and (28) we find for the transverse components of \( E \) and \( H \):

\[
\begin{align*}
\omega \varepsilon_a E_x - \mu H_y &= 0, \\
\mu E_x - \omega H_y &= 0,
\end{align*}
\]

\[
\begin{align*}
-j\omega \varepsilon_a E_x + \omega E_x + kH_x &= 0, \\
\mu E_x + \omega H_y &= 0.
\end{align*}
\]

The terms are arranged in the same way as in (17a,b). The \( \varepsilon_a \) term now, however, couples the pair of equations (a) and (b), so that independent linearly polarized \( E_x, H_y \)-waves and \( E_y, H_x \)-waves are no longer possible. From (29) we find as dispersion relation:

\[
\frac{1}{v^2} = \frac{k^2}{\omega^2} = \mu(\varepsilon_a \pm \varepsilon_a).
\]

We see here again what we already knew: propagation of undamped waves is possible only for real \( \varepsilon_a \) and real \( \varepsilon_a \). Eliminating \( H_y \) from (29) and using (30) leads to:

\[
E_x = \pm jE_y.
\]

We thus find a left-handed and a right-handed circularly polarized wave with different velocities. For a small difference in velocity (|\( \varepsilon_a \)| \( \ll \varepsilon_a \)), these waves, if of equal amplitude, can be combined to give a plane-polarized wave with a slowly rotating plane of polarization.

The plane of polarization forms a helix whose sense is the same as that of the circularly polarized wave with the smallest pitch, i.e. the slower of the two waves, that having the shorter wavelength. If the magnetic field has a polarity such that \( \varepsilon_a \) is positive, then the slower wave corresponds to the upper sign in (30) and in (31) and is thus left-handed if it travels in the +z-direction and right-handed if it travels in the −z-direction (optical convention, see (19) and the accompanying text); for negative \( \varepsilon_a \) (reversed polarity of magnetic field) the reverse is true. For a given polarity of the magnetic field, the sense of polarization of the slower wave, and consequently the sense of rotation of the plane of polarization (optical convention) is thus opposite for waves propagated in opposite directions; also, for both cases, the rotation reverses its sense when the magnetic field is reversed.

The situation is thus essentially different from that with natural optical activity, e.g. in quartz or in sugar solutions, where the sense of rotation (optical convention) is the same in both directions. The above description is therefore not applicable to natural optical activity. This follows, too, from the Onsager relation for the field-free case: \( \varepsilon_{kl} = \varepsilon_{lk} \), which is inconsistent with (28).

Returning to the magnetic rotation, the situation is
Electromagnetic waves in conducting media

Conductors obviously cannot support undamped waves; the field \( E \) and the consequent currents \( J \) give rise to losses. However, the losses may be small if \( J \) differs in phase from \( E \) by about 90°. Some examples of both strongly attenuated waves and almost unattenuated waves in conductors will now be discussed.

We assume that the current is carried by free electrons in a crystal lattice, and also that the material has no pronounced magnetic or dielectric properties; for convenience we put \( B = \mu_0 H \) and \( D = \varepsilon_1 E \), and assume that the \( \mu_0 \) and the \( \varepsilon_1 \) are little different from the permeability and permittivity of free space. This implies a wave of the type shown in fig. 8. For real \( \omega \), (36) represents strongly attenuated travelling waves of the type shown in fig. 3b; in particular the real and imaginary parts of \( k \) are equal in magnitude. This implies a wave of the type shown in fig. 8. Such waves can exist only in the neighbourhood of the surface of a metal. They propagate inwards from the surface and die out within a small distance, the penetration depth or skin depth. At high frequencies the skin depth is very small.

The above is a description of the skin effect at high frequencies (or in very thick wires). An a.c. current through a conducting wire is not distributed uniformly over the whole cross-section of the wire as is a direct current: the amplitude and phase of the current density are functions of distance from the surface and at high frequencies the current is confined to a thin layer under the surface. To discover the distribution of the current and its magnetic field (see inset, fig. 8) it is only necessary to consider a layer of thickness equal to a few times the penetration depth. If the penetration depth is small compared to the wire diameter, the surface can be considered as flat and in this case the current and field distribution can be calculated from (36), and the result is that shown in fig. 8.

For the values used above, \( \sigma_0 = 10^8 \, \text{m}^{-1}, \omega = 10^{10} \, \text{s}^{-1} \), and \( 1/\mu_0 = 4\pi \times 10^{-2} \, \text{H/m} \), the classical skin depth \( \delta_b = k_0 = 10^{-3} \). For a metal whose intrinsic impedance \( E_{\text{in}}/H_{\text{in}} \) for a transverse wave is found from (17a) and (36):

\[ E_{\text{in}}/H_{\text{in}} = \omega \mu_0 \mu_1 |k| = \pm (1 + j) \sqrt{\omega \mu_1 \sigma_0}. \]

Because the value of \( \sigma_0/\omega \) is so very much larger than \( \varepsilon_0 \), the modulus of \( E_{\text{in}}/H_{\text{in}} \) is many orders of magnitude less than \( \sqrt{\mu_0/\varepsilon_0} \), the intrinsic impedance of free space. This implies a virtually complete mismatch between free space and metal. For this reason an electromagnetic wave in space incident on a metal surface is almost completely reflected (see p. 341/42).
It follows from this that \( J \) and \( E \) are no longer linked by a scalar relation such as (32), but by a tensor relation. Table I shows how in the normal situation, in the absence of a magnetic field, the usual scalar relation 

\[
J = \sigma_0 E 
\]

is obtained. The equation (I.1) in the table expresses the fact that the conduction electrons (charge \(-q\), mass \( m \), concentration \( n \)) are accelerated by the field but also — because of collisions — are subject to an averaged frictional force; \( \nu_d \) is the resulting mean drift velocity of the electrons. The 'coefficient of friction' is the reciprocal of the relaxation time \( \tau \) approximately equal to the mean time between collisions. When the left-hand side is neglected — justifiable if the frequency is not too high — then making use of (I.3), (I.4) and (I.6) we find the usual relation (I.5); \( \mu_e \) is the 'mobility' of the electrons.

In order to include the effect of a static magnetic field, a term representing the Lorentz force has to be added to the right-hand side of (I.1):

\[
m \ddot{\nu}_d = -qE - q\nu_d \times B_0 - mn_0/\tau
\]

(37)

\( B_0 \) is the static magnetic flux density. We shall presently study waves that propagate in the direction of \( B_0 \), or in the opposite direction, and we therefore choose a coordinate system with the \( z \)-axis in this direction \((B_{0z} = B_{0y} = 0, B_{0x} = \pm B_0)\). If the vector equation (37) is written out as three equations for the components of \( E \) and \( \nu_d \), we find (because \( B_{0z} = B_{0y} = 0 \))

Table I. Summary of the theory of conduction in metals (Drude) at low frequencies (\( \omega = 0 \)) and in the absence of a magnetic field. \( m \) mass of charge carriers, \( n \) their concentration, \( \nu_d \) drift velocity, \( \tau \) relaxation time.

<table>
<thead>
<tr>
<th>( \omega )</th>
<th>( \nu_d )</th>
<th>( \mu_e )</th>
<th>( \sigma_0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>( -qE - q\nu_d/\tau )</td>
<td>( + q\tau/m )</td>
<td>( + nq\mu_e = nq^2\tau/m )</td>
</tr>
<tr>
<td>( \mu_e )</td>
<td>( \sigma_0 )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Helicon waves

In a conductor with a high concentration of high-mobility electrons (e.g. a pure metal at low temperature) situated in a strong magnetic field, circularly polarized waves can propagate in the direction of the magnetic field. Under certain circumstances, these waves are practically unattenuated and propagate at an extremely low velocity. These waves are the helicon waves noted earlier. Their existence was predicted theoretically \([13]\) in 1960 and in 1961 they were demonstrated experimentally \([14]\). There is a certain kinship with the Hall effect: as in that case, the current and electric field are not parallel — indeed, if the magnetic field and the mobility of the charge carriers are large enough, current and field may be almost perpendicular to one another.


two equations in the transverse components \( v_{dx}, v_{dy}, E_x \) and \( E_y \) in which the magnetic field appears, and one equation in \( v_{dx} \) and \( E_x \) which is independent of the other two and in which the magnetic field does not occur. The equation in \( v_{dx} \) and \( E_x \) is of no interest to us and will not be treated further. Neglecting the left-hand side again and making use of (I.3) then for the transverse components, instead of (I.2) we find:

\[
v_{dx} = -\mu_e E_x - \beta v_{dy},
\]

\[
v_{dy} = -\mu_e E_y + \beta v_{dx},
\]

Equations (38) and (I.4) and Maxwell’s equations lead to the wave phenomena called helicon waves. Before going further we should note that the quantity \( \beta \) — which for electrons is the opposite of the Hall ratio (see fig. 9) — is a kind of quality factor: only for \( |\beta| > 1 \) are the electric field and the drift velocity (i.e. current) nearly perpendicular to one another, the condition for virtually unattenuated helicon waves. It can be seen that this condition is only satisfied in quite extreme circumstances: for example, in a field \( B_0 = 1 \text{T} = 1 \text{Vs/m}^2 = 10000 \text{ gauss} \), we must have \( \mu_e > 1 \text{ m}^2/\text{Vs} \), whereas in copper at room temperature \( \mu_e \) is only about 6\times10^{-8} \text{ m}^2/\text{Vs}. Indeed, the first helicon experiment \[14\] was done with exceptionally pure sodium at 4 K: \( \beta \) was \( \approx 40 \) for \( B_0 = 1 \text{T} \), so that \( \mu_e \approx 40 \text{ m}^2/\text{Vs} \).

Solving (38) for \( v_{dx} \) and \( v_{dy} \) and using (I.4) yields a tensor relation between \( J \) and \( E \) instead of (1.5):

\[
J = \sigma E,
\]

where the tensor \( \sigma \) is given by:

\[
\sigma = \left( \begin{array}{cc} \frac{\sigma_0}{1 + \beta^2} & -\beta \\ \beta & \frac{1}{1 + \beta^2} \end{array} \right); \quad (40)
\]

\( \sigma_0 \) is the conductivity (I.6) when there is no static magnetic field.

The dispersion relation for helicon waves propagating along the \( z \)-axis can now be easily derived because of the following.

1) We consider good conductors for which we may write \( \varepsilon_{\text{eff}} = \sigma/j\omega \).
2) The \( \varepsilon \)-tensor that then follows from (40):

\[
\varepsilon_{\text{eff}} = \left( \begin{array}{cc} \varepsilon_{\text{eff} \parallel} & -j\varepsilon_{\text{eff} \parallel} \\ j\varepsilon_{\text{eff} \parallel} & \varepsilon_{\text{eff} \parallel} \end{array} \right) = \frac{\sigma_0}{j\omega(1 + \beta^2)} \left( \begin{array}{cc} 1 & -\beta \\ \beta & 1 \end{array} \right),
\]

has the same form as the transverse part of (28), although \( \varepsilon_{\text{eff} \parallel} = \sigma_0/j\omega(1 + \beta^2) \) is no longer real for real \( \omega \) (whereas \( \varepsilon_{\text{eff} \parallel} = \beta_0\sigma_0/j\omega(1 + \beta^2) \) is still real).

3) The calculation based on (28) and using (29) which leads to (30) and (31) is straightforward and is therefore also applicable for \( \varepsilon_h \) and \( \varepsilon_a \) not real.

Application of (30), with \( \mu = \mu_1 \), therefore yields the required dispersion relation. The result is:

\[
\omega = \pm \frac{\beta + j}{\mu_1\sigma_0} k^2.
\]

The tensor (41) satisfies the Onsager relations (\( \beta \) changes sign with \( B_0 \)) but no longer represents lossless propagation, as was to be expected, because \( \varepsilon_{\text{eff} \parallel} \) is no longer real. The medium is however virtually lossless when \( |\beta| > 1 \) because the real quantity \( \varepsilon_{\text{eff} \parallel} \) in (41) then dominates the imaginary quantity \( \varepsilon_{\text{eff} \parallel} \). In this situation \( |\beta| > 1 \) the properties of helicon waves are most clearly manifested. The term \( j \) in (42) can then be neglected and we find, using (I.6) and (39):

\[
\omega = \pm \frac{B_0 k^2}{\mu_1\sigma_0}.
\]

For real \( \omega \) we find a real \( k \), which means travelling waves, for the upper sign (\( + \)) if \( B_0 \) is positive. This corresponds to the upper sign in (31), i.e. to waves in which the vectors rotate clockwise as seen by an observer looking in the \( +z \)-direction. This is true for waves propagating in both directions along the \( z \)-axis (\( k > 0 \) and \( k < 0 \); see fig. 10). The sense of rotation using the optical convention (see p. 318) is thus, as in Faraday rotation, opposite for waves in the two directions, and reverses if the magnetic field is reversed. For positive \( B_0 \) the waves in which the vectors rotate anticlockwise (for an observer looking in the \( +z \) direction) have imaginary \( k \) and are thus cut off. The whole of the discussion above has been based on the assumption that the charge carriers are electrons; if the conduction were to take place via holes, the senses of rotation would all be reversed.
Helicon waves exhibit a strong dispersion: the phase velocity \( v = \omega/k \), which from (43) is proportional to \( k \) or \( \sqrt{\omega} \), can have widely differing values, depending on the frequency. In particular the velocity can be exceedingly low. For example, in a metal with a low intrinsic resistance \( \sigma_0/\alpha \) is so many orders of magnitude larger than \( \epsilon_0 \) implies that the intrinsic impedance of a metal for helicon waves — as for the classical skin effect ‘waves’ — is many orders of magnitude less than that of free space for conventional electromagnetic waves. From (29a) and (42), neglecting losses:

\[
\frac{E_x}{H_y} = \frac{\alpha_0 \mu_1 / \kappa}{\sqrt{\mu_1 / (\sigma_0 / \alpha)}} \ll \sqrt{\mu_0 / \epsilon_0}.
\]

(Although \( |\beta| \gg 1 \), it is negligible compared to the very high value of the ratio of \( \sigma_0 / \alpha \) to \( \epsilon_0 \).) We therefore again have a complete mismatch between the medium and free space, so that both normal electromagnetic waves in free space and helicon waves in the medium are almost completely reflected at the interface.

As a result, in a configuration like that shown in fig. 11, standing waves can be set up whose attenuation is determined entirely by \( \beta \). For example under the same conditions as above (\( n = 6 \times 10^{28} \) m\(^{-3} \), \( B_0 = 1 \) T) in a plate of thickness 3 mm (= \( \lambda/2 \)), standing waves of 17 Hz can be expected.

In helicon experiments the sample is usually arranged with a primary coil and a secondary coil as in fig. 11. If a d.c. current is switched on or off in the primary, a series of standing helicon waves are excited and the corresponding damped oscillations induced in the secondary can be observed (fig. 12). Crossed coils are particularly well adapted for the experiment: the only coupling between them is via the (circularly polarized) helicon waves.

By means of such experiments, the elements of the \( \sigma \)-tensor and hence the Hall constant and the magnetoresistance [15] can be determined relatively easily and very accurately as functions of \( B_0 \) [16]. The determinations of such experiments, the elements of the \( \sigma \)-tensor and hence the Hall constant and the magnetoresistance [15] can be determined relatively easily and very accurately as functions of \( B_0 \) [16]. The determinations of such experiments, the elements of the \( \sigma \)-tensor and hence the Hall constant and the magnetoresistance [15] can be determined relatively easily and very accurately as functions of \( B_0 \) [16].

[15] These are \( g_{xx}(B_0)/B_0 \) and \( g_{xy}(B_0) \) respectively, where \( g_{xx} \) and \( g_{xy} \) are the elements of \( g \) (the inverse of the tensor \( \sigma \)) which expresses \( E \) in terms of \( J \) through the relation \( E = gJ \).


tion of these quantities by conventional methods usually requires difficult precision measurements of very small resistances and voltages between accurately located contacts. Helicon measurements are made without contacts on the sample.

Secondly, there is the question of 'open cyclotron orbits'. In a metal with a simple Fermi surface (e.g., an alkali metal) in a magnetic field, the momentum vector of an electron describes a closed orbit on the Fermi surface. In metals such as copper, silver, and gold, however, the Fermi surface is so anisotropic that in certain directions the cyclotron orbits are 'open'. Because of this the helicon waves may be plane-polarized and strongly attenuated. This effect is also used for the study of the Fermi surface \(^{16}\).

Reflection and transmission of optical waves in metals

We shall now leave situations involving magnetic fields to enquire what happens when the frequency of an electromagnetic wave is raised to the optical region. Important changes occur in the skin effect, primarily because the term \(m^2 \omega_0^2\) in (I.1) can no longer be neglected. Instead of (I.3), with \(\tilde{v}_d = j\omega_0 q\) we find:

\[
J = \sigma_0 E / (1 + j\omega t).
\]

Neglecting \(m^2 \omega_0^2\) in (I.1) is clearly justified only when \(\omega t \ll 1\). Let us now assume that the frequency is so high that \(\omega t \gg 1\); there is then an effective conductivity

\[
\sigma_{\text{eff}} = \sigma_0 j\omega t = -juq^2 / m \omega_0,
\]

which is purely imaginary so that there are no losses (\(J\) and \(E\) differ in phase by 90°). Substituting (45) for \(\sigma_0\) in (35) gives the dispersion relation:

\[
k^2 = -\mu \omega_0 q^2 / m.
\]

Since \(k\) is imaginary, the waves are evanescent (fig. 3c, for real \(\omega\)). As with the classical skin effect, these 'waves' are restricted to a thin layer at the surface of the metal. When electromagnetic waves are incident on such a surface, it follows that no power can be transmitted through the metal and there are also no losses; the waves are reflected completely. The shiny appearance of most metals is explained in this manner.

At still higher frequencies the term \(\varepsilon_1\) in (33) can no longer be neglected. This means that we have an additional term \(\varepsilon_1 \mu_1 \omega^2\) in (46):

\[
k^2 = \varepsilon_1 \mu_1 \omega^2 - \mu_1 q^2 / m = \varepsilon_1 \mu_1 (\omega^2 - \omega_p^2),
\]

where

\[
\omega_p = \sqrt{q^2 / \varepsilon_1 m}
\]

is called the plasma frequency. This is a critical frequency: for \(\omega < \omega_p\) \(k\) is imaginary so that the waves are evanescent; for \(\omega > \omega_p\) \(k\) is real so that the waves are propagated through the metal. In the first case there is complete reflection at the surface; in the second case there is partial reflection and partial transmission, dependent on the ratio of the intrinsic impedances of metal and free space (see Part III of this article). This impedance ratio passes through the value unity in the transition region, around the plasma frequency, and this implies zero reflection and 100% transmission.

Fig. 12. Voltages induced by helicon waves, after R. Bowers, C. Legendy and F. Rose \(^{14}\). The diagrams show the voltages across the secondary coil (see fig. 11) as a function of time, after interruption of the primary current, with the sample in a magnetic field of strength (from top to bottom) 0 Oe, 3600 Oe, 7200 Oe and 10 800 Oe. (In this first helicon experiment, the sample was not of plate form as in fig. 11 but a cylinder of diameter 4 mm, and the coils were not crossed.)

Two complications that can arise with helicon waves should be mentioned. These do not come within the framework of purely local relations, characterized by effective \(\varepsilon\)'s and \(q\)'s to which we have previously confined ourselves (see p. 320).

Firstly, there is the absorption arising from Doppler-shifted cyclotron resonance. The electrons responsible for conduction move in all directions through the metal at a high velocity, the Fermi velocity \(v_F\) (not to be confused with the drift velocity \(v_d\)). An electron with a Fermi velocity in the direction of the helicon wave runs through the wavefronts and is thus subject to an alternating field of frequency \(\omega_F\) (the velocity of the slow helicon wave is neglected here). If \(\omega_F\) equals to \(\omega_0\) the electron undergoes cyclotron resonance and so absorbs energy from the wave and attenuates it. If \(\omega_F\) is greater than \(\omega_0\) then there are some electrons moving obliquely to the wave which come into resonance. For a given \(B_0\) there is thus an absorption edge at \(k = \omega_0 / v_F\). Measurement of this absorption edge in single crystals for various direction of \(B_0\) with respect to the crystal axes yields data on the anisotropy of the Fermi velocity and hence information about the shape of the Fermi surface \(^{17}\).
For most metals $\omega_p$ lies in the ultraviolet. With $n = 6 \times 10^{28} \text{m}^{-3}$, $\varepsilon_1 = \varepsilon_0$, and the usual values for the electronic charge $q$ and mass $m$, we find $\omega_p \approx 1.4 \times 10^{16} \text{s}^{-1}$ which corresponds to a free-space wavelength of 140 nm. In this way the transparency of alkali metals in the ultraviolet region can be understood [10].

**Longitudinal electric waves in conducting media**

Introduction of an effective dielectric constant $\varepsilon_{\text{eff}} = \varepsilon_1 + \sigma j/\omega$ allowed us to make use of (16) for the problem of electromagnetic waves in conductors. For plane waves propagated in the $z$-direction we arrived at (17) and it was noted that longitudinal electric waves would be possible if $\varepsilon_{\text{eff}}$ were to be zero. It can in fact be seen from (17) that if

$$\varepsilon_{\text{eff}} \equiv \varepsilon_1 + \sigma j/\omega = 0,$$  

(48)

then all components of $E$ and $H$ in (17) must be zero except $E_z$. (It follows directly from (8) that all magnetic components must be absent in a wave with a longitudinal electric field, since longitudinal vectors all have zero curl.) Now $\varepsilon_1$ and $\sigma$ themselves are not zero in (48), so that not only $E$ but also $D$ and $J$ have longitudinal components. Since we also have $\partial_\sigma \neq 0$ it follows that the divergences of $D$ and $J$ ($\partial_\sigma D_\sigma$ and $\partial_\sigma J_\sigma$), and hence $\varrho_e$ and $\varrho_0$ (see (10) and (11)), are neither of them zero. These waves are thus characterized by fluctuations in charge density: the electrons bunch together and disperse again. This is different from the case of purely transverse waves, where the charge density is everywhere zero (local electroneutrality): the divergence of a transverse vector is zero.

In what circumstances is the dispersion relation (48) satisfied? For real conductivity, $\sigma = \sigma_0$, we have for the first time the situation that $\omega$ cannot be real; $\omega$ must be purely imaginary, $\omega = j \sigma_0/\varepsilon_1$, and $k$ is completely arbitrary. Any charge distribution $\varrho_0(z)$ with its corresponding field therefore dies away exponentially (see fig. 3g, h, i). The characteristic time for this process is $\tau_\varrho = \omega_1^{-1} = \varepsilon_1/\sigma_0$, the dielectric relaxation time. For metals $\tau_\varrho$ has no physical significance: $\varepsilon_1/\sigma_0 \approx 10^{-11}/10^8 = 10^{-19}$ s and, for processes taking place in such a short time, the assumption that $\sigma$ equals $\sigma_0$ is certainly incorrect. Certainly at radio frequencies we can conclude that there is always local electroneutrality in metals. In semiconductors, however, local space-charge variations do play a role and $\tau_\varrho$ is an important quantity as we shall see presently.

In media and under circumstances where $\omega \tau \gg 1$ (e.g. in metals at optical frequencies) the conductivity is purely imaginary, as we saw earlier (see 45). Longitudinal waves of real frequency are then possible, for substitution of (45) in (48) gives:

$$\varepsilon_1 - nq^2/m\omega^2 = 0,$$

so that

$$\omega = \omega_p = \sqrt{nq^2/\varepsilon_1 m}.$$

We see that the plasma frequency (47) is not only the critical frequency for the propagation of transverse waves but it is also the frequency at which longitudinal waves can exist, if $\omega \tau \gg 1$. As with dielectric relaxation, $k$ is arbitrary. Such waves do not transfer energy: the Poynting vector $S = E \times H$ is zero because there are no magnetic fields.

The plasma frequency is a quantity continually encountered in 'plasma physics', which is the basic discipline for a number of quite diverse subjects such as travelling-wave amplifiers, astrophysics and controlled nuclear fusion. A plasma is a medium whose behaviour depends primarily on the charge and mass of the charge carriers and in which collisions play only a minor role. (Helicon waves are thus waves in a plasma.) The term was introduced in the twenties by Irving Langmuir, in connection with his investigations into gas discharges, to describe a dilute, strongly ionized but electrically neutral gas [19]. In this work Langmuir discovered that, surprisingly, electrons injected into the plasma rapidly came into thermal equilibrium with the plasma in spite of the very long mean free path. High frequency oscillations of the plasma would explain this. Such plasma oscillations had in fact been observed earlier by F. M. Penning [20]. The frequency found by Penning was $10^8$ to $10^9$ Hz, corresponding to wavelengths of several decimetres. From (47) this would imply an electron density of the order of $10^{17}$ m$^{-3}$, which is indeed typical for low-pressure gas discharges such as those used by Penning.

Summarizing we can say that local space-charge fluctuations die away exponentially in a time $\tau_\varrho$ if electron collisions play the dominant role ($\sigma = \sigma_0$), or oscillate at the frequency $\omega_p$ if collisions can be neglected ($\omega \tau \gg 1$). For charge variations that are very steep (large $k$) it is necessary to take into account a phenomenon that has not yet been discussed in this article, and which is of a non-electromagnetic nature: the diffusion of the electrons from regions of high concentration to regions of low concentration. We shall now look into this, but only for the case of low frequencies. The extra electron current due to diffusion is $-D_\varrho$ and $n_0$ to collisions can be neglected ($\omega \tau \gg 1$). For charge variations that are very steep (large $k$) it is necessary to take into account a phenomenon that has not yet been discussed in this article, and which is of a non-electromagnetic nature: the diffusion of the electrons from regions of high concentration to regions of low concentration. We shall now look into this, but only for the case of low frequencies. The extra electron current due to diffusion is $-D_\varrho$ and $n_0$ to collisions can be neglected ($\omega \tau \gg 1$). For charge variations that are very steep (large $k$) it is necessary to take into account a phenomenon that has not yet been discussed in this article, and which is of a non-electromagnetic nature: the diffusion of the electrons from regions of high concentration to regions of low concentration.
\[ J = \sigma_0 E - D_n \text{grad} \varphi_0. \]

(49)

For plane longitudinal waves this can easily be reduced to the form (32). With (grad \( \varphi_0 \)) \( = -jke \) and \( \varphi_0 = \partial_z D_2 = -jke_1 E_z \), we find:

\[ J_z = (\sigma_0 + k^2 e_1 D_n) E_z. \]

The factor in the bracket is again an effective conductivity. Together with (48) it leads to the following improved dispersion relation for longitudinal waves:

\[ j\omega e_1 + \sigma_0 + k^2 e_1 D_n = 0. \]

(50)

For waves of infinite wavelength (\( k \to 0 \)), we find again the relaxation behaviour discussed above; from (50) we find — as was to be expected — that for shorter, steeper waves (steeper charge variations) the relaxation is more rapid.

We now consider infinitely slow waves (\( \omega \to 0 \)) instead of infinitely long wavelengths. From (50) we find that these are exponential charge distributions having the characteristic length \( k_1^{-1} = \sqrt{e_1 D_n / \sigma_0} = \sqrt{D_n c} \). This is the Debye-Hückel length \( \lambda_D \). A surface inside a conductor covered with a uniform charge is screened by a layer in which the charge density at the distance \( \lambda_D \) has fallen off by a factor \( e \).

In the Debye and Hückel theory \cite{22} of the conduction of electrolytes the quantities \( \tau_1 \) and \( \lambda_D \) both play a role. Generally speaking, a positive ion is surrounded by a cloud of negative ions of radius \( \lambda_D \); the positive ion experiences a frictional force, because relaxation causes the cloud to lag behind the positive ion when this moves.

In (50) we first neglected the third term and then we neglected the first term. Suppose we now neglect the second term (\( \sigma_0 \to 0 \)):

\[ e_1 \text{ then also disappears from the equation and all purely electric variables have vanished.} \]

With \( \omega e_1 = \tau_D \), and \( k_1^{-1} = \lambda_D \), (50) reduces to the familiar relation common to diffusion problems, \( L_D = \sqrt{D \tau_D} \).

**Elastic waves**

Elastic waves in solids can be of a very complex nature. We shall introduce only a few elementary elastic waves here, but in passing we shall see how complications can easily arise. Later on we shall consider coupling between the waves introduced here and electromagnetic waves, and we shall then see that this can give rise to some remarkable effects in piezoelectrics.

We shall find in this section the well known result that the wave velocity (the velocity of sound) is highest in rigid and light substances. More specifically, in substances with a high resistance to pressure and tension but not to shear, longitudinal waves are fast but transverse waves are slow; in those with a high resistance to shear as well, transverse waves are also fast. In most substances the velocities of longitudinal and transverse waves do not differ greatly. Gelatine is an example in which transverse waves are much slower than longitudinal waves.

In elastic waves we are concerned with non-uniform displacements of volume elements, i.e. deformation of the material; this implies internal mechanical stresses in the material which, in turn, react on the displacements. The linear equations (algebraic and differential) between these quantities again define the wave problem.

**Displacements, strains and stresses**

Starting with the displacement \( u \) of each point of the material from its equilibrium position \( x,y,z \), in which \( u \) is thus a function of \( x, y \) and \( z \), the six strain components \( S_1, S_2, \ldots S_6 \) are defined as follows:

\[ S_1 = S_{xx} = \partial_x u_x, \quad S_4 = S_{yz} = \partial_y u_z, \]

\[ S_2 = S_{yy} = \partial_y u_y, \quad S_5 = S_{zx} = \partial_z u_x, \]

\[ S_3 = S_{zz} = \partial_z u_z, \quad S_6 = S_{xy} = \partial_x u_y. \]

(51)

There is deformation of the medium only if the displacement \( u \) is a function of the coordinates; if it is not, either there has been no displacement (\( u = 0 \)) or the medium has been displaced as a whole (\( \nabla u \neq 0 \)). The strain components are therefore derivatives of \( u \). \( S_1, S_2 \) and \( S_3 \) give the extensions in the \( x-, y-, z- \) directions (fig. 13a) and \( S_4, S_5 \) and \( S_6 \) give the shear (fig. 13b); the latter consist of combinations of the derivatives such as \( \partial_y u_x + \partial_x u_y \) because \( \partial_y u_x \neq 0 \) alone does not necessarily imply deformation as is explained in fig. 13.

The internal stress is the force per unit area exerted by material on one side of a given internal plane on material on the other side. In tension or compression the force is directed normally to the plane, in shear tangentially (fig. 14). We can therefore expect nine stress components \( T_{xx}, T_{yy}, \ldots T_{zz} \); the first suffix indicates the direction of the force and the second the normal to the plane considered. The net couple on each volume element must be zero, which reduces the number of components to six (because \( T_{yz} = T_{xy} \), \( T_{xx} = T_{zz} \), \( T_{xy} = T_{yz} \)); this is explained in fig. 15.

There remain the six stress components:

\[ T_1 = T_{xx}, \quad T_4 = T_{yz} = T_{xy}, \]

\[ T_2 = T_{yy}, \quad T_5 = T_{zx} = T_{zx}, \]

\[ T_3 = T_{zz}, \quad T_6 = T_{xy} = T_{yx}. \]

In equilibrium the net force on each volume element must also be zero which means that the stress field is

\cite{22} P. Debye and E. Hückel, Phys. Z. 24, 305, 1923.

L. Onsager, Phys. Z. 27, 388, 1926.
homogeneous (fig. 15). We shall return to this later (p. 330).

Provided the strains are small, they are linearly related to the stresses (Hooke's law):

\[ T_k = \sum_i c_{kl} S_l \quad (k,l = 1, 2, \ldots 6). \]  

(52)

The 36 coefficients \( c_{kl} \) are called the elastic moduli (or stiffness constants). As in the derivation of (25) (and assuming that any changes in \( S \) and \( T \) are so slow that the \( c_{kl} \) remain real) it can be shown that the elastic moduli are symmetric \( (c_{kl} = c_{lk}) \) if no mechanical energy is transformed into other forms of energy. (The work done on the medium per unit volume in the elastic case is \( \sum T_k \delta S_k \). This is explained in fig. 16.)

In the case of greatest anisotropy (triclinic crystal) we need however 6 independent constants; the array of constants still need however \( 6 + \frac{1}{2} \times (36 - 6) = 21 \) different constants to describe the elastic properties of the material. Clearly complications can arise all too easily. Crystal symmetries, however, restrict the number of independent constants and with a favourable choice of coordinate system this restriction becomes apparent through the appearance of many zeros and many equal-valued constants. In particular, isotropic material has only two independent constants; the array of constants then has the following form:

<table>
<thead>
<tr>
<th>( T )</th>
<th>( S_1 )</th>
<th>( S_2 )</th>
<th>( S_3 )</th>
<th>( S_4 )</th>
<th>( S_5 )</th>
<th>( S_6 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T_1 )</td>
<td>( c_{11} )</td>
<td>( c_{12} )</td>
<td>( c_{13} )</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( T_2 )</td>
<td>( c_{12} )</td>
<td>( c_{22} )</td>
<td>( c_{23} )</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( T_3 )</td>
<td>( c_{13} )</td>
<td>( c_{23} )</td>
<td>( c_{33} )</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( T_4 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( c_{44} )</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( T_5 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( c_{55} )</td>
<td>0</td>
</tr>
<tr>
<td>( T_6 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( c_{66} )</td>
</tr>
</tbody>
</table>

(53a)

Fig. 13. Various types of deformation: \( a \) extension, \( b \) shear. In \( a \) a volume element is extended in the \( x \)-direction; \( \partial u_x / \partial x \) is positive. A negative \( \partial u_x / \partial x \) means a compression in the \( x \)-direction. By definition \( \partial u_x / \partial x \) is the first deformation component \( S_{xx} = S_1 \). In \( b \) layers perpendicular to the \( z \)-axis are displaced with respect to each other in the \( y \)-direction; \( \partial u_y / \partial y \) is positive. A positive or a negative value of \( \partial u_y / \partial y \) does not always imply deformation, however: in \( c \), where a volume element has been rotated without deformation, \( \partial u_y / \partial y \) is also positive. But \( \partial u_y / \partial y \) is equally large and of opposite sign. If the sum \( \partial u_x / \partial x + \partial u_y / \partial y \) is not zero, the element does undergo deformation. This is the deformation component \( S_{yy} = S_2 \). Analogously, \( S_3 = S_{zy}, S_4 = S_{zz} \) and \( S_6 = S_{zz}, S_4 = S_{zy} \).

Fig. 14. Tensile stress and shear stress. The material on the + \( x \)-side of an elementary area perpendicular to the \( x \)-axis exercises a force on the material on the − \( x \)-side. The stress components \( T_{xx}, T_{yz} \) and \( T_{zx} \) are by definition the \( x \)-, \( y \)- and \( z \)-components respectively of that force per unit area of the elementary area. \( T_{xx} \) is a tensile stress \( (T_{xx} > 0) \) or a compression stress \( (T_{xx} < 0) \). \( T_{yz} \) and \( T_{zx} \) are shear stresses. The stress components \( T_{xy}, T_{yx}, T_{zx}, T_{yz}, T_{xz} \) and \( T_{zy} \) are defined analogously. See also fig. 15.

Fig. 15. \( a \) In equilibrium the net force on a volume element is zero. Since the force per unit area exerted on the volume element on its face that faces left (or down or backwards) is equal but opposite to the stress there, we have \( T_{xx}(x_2) \) is equal to \( T_{xx}(x_1) \), \( T_{xy}(x_2), T_{yx}(x_1), \) etc; in other words the stress field is uniform. \( b \) The uniform stress field \( T_{yy} \) exerts a couple on the volume element. In equilibrium, this must be balanced by the opposite couple resulting from \( T_{xy} \). It follows that \( T_{xy} = -T_{yx} \), and equally, \( T_{yz} = -T_{zy}, T_{zx} = -T_{xz} \). There are thus six independent stress components: \( T_1 = T_{xx}, T_3 = T_{yy}, T_5 = T_{zz}, T_4 = T_{xy}, T_6 = T_{yz}, T_0 = T_{xy} \).
where \( c_{11}, c_{12} \) and \( c_{44} \) are related as follows:
\[
c_{11} = c_{12} + 2c_{44}.
\]  
(53b)

One or two comments will serve to illustrate the significance of the constants \( c_{11}, c_{12} \) and \( c_{44} \) in the isotropic case. If \( c_{11} \) and \( c_{12} \) are non-zero and positive, a tensile stress \( T_1 \) implies not only an extension in the \( x \)-direction but also a lateral contraction: when \( T_2 = T_3 = 0 \), and \( S_1 \) is positive, it follows from (53a) that \( S_2 = S_3 < 0 \). That the shear modulus (or modulus of rigidity) \( c_{44} \) must be closely related to \( c_{11} \) and \( c_{12} \) is illustrated in fig. 17. It is shown there that the shear resulting from a shear stress \( \sigma_{46} \) in one coordinate system is equivalent to an extension and a lateral contraction resulting from the combination of a tension and a lateral pressure \( (c_{11}, c_{12}) \) in an other coordinate system; and because of the isotropy, the constants must be independent of the coordinate system chosen. Finally, if \( c_{44} \) is zero (zero rigidity), \( c_{11} = c_{12} \) so that \( T_1 = T_2 = T_3 = c_{11}(S_1 + S_2 + S_3) \). It is easily shown that for any deformation, the relative change in volume \( \Delta V/V \) is given by \( S_1 + S_2 + S_3 \). Thus \( c_{44} = 0 \) implies that when a deformation takes place involving no change of volume no stresses are set up. This is the situation with fluids (gases and liquids). In this sense gelatine and rubber are ‘near-liquids’ because \( c_{44} \) is small, and very much smaller than \( c_{11} \) and \( c_{12} \).

Finally we must consider the dynamic influence of the stresses \( \sigma \) on the displacements \( u \). If the stress field is non-uniform, the volume elements undergo a net force which accelerates them. The net force in the \( x \)-direction on a volume element \( dx dy dz \) is (see fig. 18): \( \sigma_{xx} \frac{d^2 u_x}{dx^2} dx dy dz + \sigma_{xy} \frac{d^2 u_y}{dy dx} dx dy dz + \sigma_{xz} \frac{d^2 u_z}{dz dx} dx dy dz \). This force is equal to the product of the mass \( \rho m \) and the acceleration \( \rho m \frac{d^2 u_x}{dx^2} \) in the \( x \)-direction \( (\rho m \) is the density). In this way we find the equations of motion:

\[
\rho m \frac{d^2 u_x}{dx^2} = \sum_{l} \sigma_{xl} = \sum_{l} \frac{\partial}{\partial x} T_{kl} (k,l = x, y, z).
\]  
(54)

The six defining equations for the strain (51), the six Hooke equations (52) and the three equations of motion (54) give altogether 15 linear homogeneous equations for the 15 variables \( u_x, u_y, u_z, S_1, \ldots S_6, T_3, \ldots T_6 \), thus describing fully the wave phenomena. Of the possible solutions of these equations, we shall consider only two simple cases.
Transverse waves in isotropic materials

For transverse waves in an isotropic medium (see 53), propagating in the +z-direction \((\delta_x = \delta_y = 0, \delta_z = -jk, \delta_t = j\omega)\) and with displacements only in the x-direction \((u_x = u_y = 0)\), many of the 15 variables are zero. There remain:

\[
\begin{align*}
(51) & \rightarrow S_5 = -jku_x, \\
(52) & \rightarrow T_5 = c_{44}S_5, \\
(54) & \rightarrow -\varrho_m\omega^2u_x = -jkT_5,
\end{align*}
\]

from which the dispersion relation follows immediately:

\[
\omega^2/k^2 = c_{44}/\varrho_m. \tag{56}
\]

The waves are thus dispersionless; at all frequencies the velocity of propagation is \(v_{\text{lat}} = \sqrt{c_{44}/\varrho_m}\).

Longitudinal waves in isotropic materials

For longitudinal waves in an isotropic medium propagating in the +z-direction \((u_x = u_y = 0, \delta_x = \delta_y = 0, \delta_z = -jk, \delta_t = j\omega)\), (51), (52) and (54) reduce to:

\[
\begin{align*}
(51) & \rightarrow S_3 = -jku_x, \\
(52) & \rightarrow T_1 = T_2 = c_{12}S_3, T_3 = c_{11}S_3, \tag{57} \\
(54) & \rightarrow -\varrho_m\omega^2u_x = -jkT_3,
\end{align*}
\]

Combination of the three equations in \(u_x, S_3\) and \(T_3\) yields the dispersion relation:

\[
\omega^2/k^2 = c_{11}/\varrho_m. \tag{58}
\]

(Coupling of waves in an unbounded homogeneous medium)

Up to now we have studied separately two classes of waves, electromagnetic and elastic waves. We shall now go on to consider coupled waves, in particular coupling between these two classes of waves, as found for example in piezoelectric materials. To give an illustration of what is involved in the coupling of waves, we shall take as example the longitudinal and transverse elastic waves in an isotropic medium which we have just treated separately. We shall now consider them together: we assume a wave propagated in the z-direction in which displacements are allowed both in the z- and the x-directions (but not in the y-direction). We then find again equations (55) and (57), now together. These are written symbolically as follows:

\[
\begin{array}{cccc|c}
\hline
u_x & S_5 & T_5 & u_x & S_3 & T_3 \\
\hline
x & x & & \bullet & \hline
x & x & & & & \hline
x & x & & \bullet & \hline
\bullet & x & x & & \\
\hline
\end{array}
\tag{60}
\]

Each row represents an equation and the crosses indicate which variables are involved. (The equations for \(T_1\) and \(T_2\), here omitted, are of no importance at the moment.) The 6 \times 6 determinant of the equations is the product of two 3 \times 3 determinants, \(f(\omega,k)\) and \(g(\omega,k)\). The dispersion relation is thus

\[
f(\omega,k)g(\omega,k) = 0. \tag{60}
\]

There are therefore two independent solutions, representing two types of wave:

- \(f = 0\), variables \(u_x, S_5, T_5\) \((u_x = S_3 = T_3 = 0)\): transverse waves;
- \(g = 0\), variables \(u_x, S_3, T_3\) \((u_x = S_3 = T_3 = 0)\): longitudinal waves.

Let us now suppose that the medium loses its isotropy in such a way that a tension \(T_3\) results not only in an extension \((S_3)\) but also in a shear \((S_3)\). The purely longitudinal wave can then no longer exist: the tensile stress associated with it would cause transverse displacements \(u_x\) via the term \(S_3\). This is expressed in (59) by terms at the black dots \((c_{35}\) and thus \(c_{35} = 0\) no longer zero). The 6 \times 6 determinant then no longer factorizes: longitudinal and transverse waves are coupled.

In this way the \(E_{xy}H_y\) waves and the \(E_{xy}H_x\) waves of \((1a,b)\) are mixed in (29) by the \(e_0\) term. Similarly, electromagnetic and elastic waves in piezoelectric materials are coupled because the electric fields give rise to mechanical stresses.
Weak coupling

In the above we found pure longitudinal waves and pure transverse waves for $C_{35} = 0$. If $C_{35}$ is not zero but still very small — weak coupling — we may expect 'nearly-pure' longitudinal and transverse waves. In such a case we can start from the dispersion relations $f = 0, g = 0$ as a zero-order approximation to find the relation between $\omega$ and $k$ for the new waves. If, for example, the terms at the dots in (59) are small, although not zero, the dispersion relation becomes, instead of (60):

$$f(\omega, k)g(\omega, k) = \delta(\omega, k),$$

where $\delta$ is a measure of the coupling. More complicated situations can arise giving, for example, a dispersion relation of the form $fg^2 = \delta$; however, the simpler form (61) is usually found and we shall restrict our discussion to this form.

For a given frequency $\omega_0$, the new waves will have wave numbers in the neighbourhood of those of the old $f$-wave and $g$-wave (fig. 19). Suppose that the wave number of the old $f$-wave is $k_0$, so that $f(\omega_0, k_0) = 0$, and let the new wave number closest to $k_0$ be $k_0 + \Delta k$. For $\omega = \omega_0$, it follows from (61) that

$$(f_0 + f_{k_0} \Delta k)(g_0 + g_{k_0} \Delta k) = \delta,$$  

where $f_0, g_0, f_{k_0}, g_{k_0}$ are the values of $f, g, df/dk, dg/dk$ at $\omega_0, k_0$. According to our assumption, $f_0 = 0$; on the other hand, $g_0$ will in general differ substantially from zero, so that $g_{k_0} \Delta k$ can be neglected. To a first approximation, therefore,

$$\Delta k = \delta f_{k_0}^{-1} g_0 (\omega = \omega_0).$$

Similarly, for a given wave number $k_0$, the difference in frequency $\Delta \omega$ between the new 'near-$f$-wave' and the old $f$-wave is

$$\Delta \omega = \delta f_{\omega_0}^{-1} g_0 (k = k_0).$$

In a similar way we can find the difference in wave number or the difference in frequency between the new 'near-$g$-wave' and the old $g$-wave. These expressions will be useful later on.

Resonant coupling

Suppose that the curves $f = 0, g = 0$ intersect at some point with real $k$ and $\omega$. Such an intersection can occur, apart from the trivial case of $k = 0, \omega = 0$, only if at least one of the waves is dispersive. An example is the combination helicon wave/sound wave; see fig. 20. At the point of intersection $k_0, \omega_0$ the two waves are 'resonant': their phase relation is constant in both space and time. Even a weak coupling can then give a strong effect. In particular, for a given $\delta, k$ and $\omega$ will exhibit larger changes than in the non-resonant case. This can be seen directly by applying (62) to the point of intersection. Because $f_0 = 0, g_0 = 0$ we find (see fig. 21):

$$\Delta k = \pm \sqrt{\delta f_{k_0}^{-1} g_0}$$

at $\omega = \omega_0$,  

and

$$\Delta \omega = \pm \sqrt{\delta f_{\omega_0}^{-1} g_0}$$

at $k = k_0$.

![Fig. 19. The dispersion relations $f = 0$ and $g = 0$ (solid curves) for two independent types of waves. The dashed curves represent the dispersion relations of waves that can result from the coupling of the $f$ and $g$ waves.](image)

![Fig. 20. The dispersion relation for helicon waves (43) is a square law in $k$, that for acoustic waves is linear. They therefore intersect not only at the origin but also at another point ($k_0, \omega_0$). At this intersection even a weak interaction gives rise to strong effects (resonance). For a given magnetic field, the intersection takes place at lower frequencies as the concentration of the electrons is lower (see eq. 43). For $B_0 = 1$ T, the resonant frequency in metals lies in the gigacycle region. In semiconductors the resonance lies in the megacycle region or lower.](image)
As $\delta$ approaches zero, $\sqrt{\delta}$ approaches zero less rapidly and, at sufficiently small $\delta$, the resonant effects (65) and (66) are much larger than the non-resonant effects (63) and (64). Owing to the coupling, the intersection disappears; in its neighbourhood the curves $f = 0$, $g = 0$ become two hyperbolae asymptotic to the original intersecting lines. If, in (65) and (66), $\Delta k$, $\Delta \omega$ or both happen to be imaginary, the curves shown respectively in the second, third, and fourth diagrams of fig. 22 are relevant. One of the main results of the analysis given in [7] is, that in the fourth situation we have travelling-wave amplification while the third gives rise to the ‘absolute instability’ which was briefly mentioned in the introduction. The second situation was already referred to as ‘cut-off’ on page 314.

For the combination helicon wave/sound wave there is a weak electric coupling between the electron motion of the helicon wave and the ion motion of the sound wave [23]. The combined dispersion relation of fig. 20 thus splits into two branches which nowhere intersect except at the origin (fig. 23); along each branch, the corresponding wave gradually changes from ‘near-sound’ to ‘near-helicon wave’ or vice versa.

The coupling of helicon waves with sound was first demonstrated experimentally in potassium [24]. However, the point of intersection of the uncoupled dispersion curves for metals in ‘convenient’ magnetic fields (up to $1 \text{T} = 10 \text{ kilogauss}$) lies in the none too easily accessible gigacycle region. In this respect, semiconductors are more convenient: owing to the lower electron concentration the helicon curve in fig. 20 is steeper so that the point of intersection is displaced to lower frequencies. W. Schilz [25] has demonstrated the coupling in $N$-PbTe with $n \approx 10^{24} \text{ m}^{-3}$. He used a plate of PbTe fitted with two quartz transducers (fig. 24). The three elements were all acoustically a half wavelength long at the same frequency ($\approx 800 \text{ kHz}$) and thus resonant. The acoustic transmission of the system peaked sharply at this frequency (curve a, fig. 24). The PbTe plate was also fitted with crossed coils (as in fig. 11) by means of which a $\lambda/2$ helicon resonance could be set up and detected at this same frequency when the magnetic field was about $0.4 \text{T}$ ($= 4 \text{ kG}$). This field and this frequency correspond to the point of intersection shown in fig. 20. If, at this field, a signal of varying frequency was applied to one of the helicon coils, then the transducers gave a sharply peaked output close to the resonant frequency (curve b, fig. 24). The height of this peak varied with the magnetic field in approximately the same way as the helicon resonance itself. The high electron mobility necessary for this experiment was obtained by using a single crystal of PbTe at a temperature of $4 \text{ K}$ ($\mu \approx 100 \text{ m}^2/\text{Vs}$).

---


Piezoelectric coupling

In piezoelectric materials mechanical variables such as stress and displacement are coupled with electrical variables such as field and polarization. Since the twenties piezoelectric materials have been used extensively for the conversion of acoustic or mechanical signals into electrical signals and vice versa: in microphones, loudspeakers, pick-ups, crystal oscillators, filters, transducers, etc. For a long time the well known materials tourmaline, Rochelle salt and quartz have been of practical importance and indeed quartz still is. Since the fifties, however, a large number of synthetic piezoelectric materials have entered the field.

Materials exhibiting piezoelectricity all have a crystal structure without a centre of symmetry. It can easily be seen that a centro-symmetric crystal cannot be piezoelectric: for instance, a uniform deformation cannot cause any separation of the centres of positive and negative charges, i.e. a polarization, in such a crystal.

Conduction can mask the piezoelectricity of a crystal, for the motion of free charges tends to destroy an electric field if there is sufficient time available. All the piezoelectric materials widely used in practice are consequently insulators. Nevertheless, it is in fact the interaction of a piezoelectric crystal lattice with free charge carriers — in piezoelectric semiconductors — that can offer so many interesting effects. Such possibilities form the main theme of the next section. Cadmium sulphide and tellurium are examples of such piezoelectric semiconductors.

The piezoelectric coupling is expressed through electric terms in the elastic equations (52) and mechanical terms in the electric equations (22):

\[ T_k = \sum_i c_{ik} S_i - \sum_n e_{nk} E_n, \quad (67a) \]
\[ D_m = \sum_i e_{mi} S_i + \sum_n e_{mn} E_n. \quad (67b) \]

\((k,l = 1, \ldots 6; \quad m,n = x,y,z)\)

The nine linear equations (67) with their 81 coefficients express the nine variables \(T_1, \ldots, T_6, D_x, D_y, D_z\) in terms of the nine variables \(S_1, \ldots, S_6, E_x, E_y, E_z\). Apart from the minus signs in (67a), the \(9 \times 9\) array of coefficients is again symmetric if the electrical and mechanical energy is not transformed into other forms such as heat. With regard to the coefficients \(e\), this symmetry has already been incorporated in the notation of (67), whereas for the \(c\)'s and \(e\)'s we must have, as before:

\[ c_{kl} = c_{lk}, \quad e_{mn} = e_{nm}. \quad (68) \]

This can be shown by reasoning analogous to that on p. 319/320. Again we assume that the coefficients are real constants; this means that the variables may not vary too rapidly.

If we assume that the electromechanical state of the material is completely determined by the six quantities \(T\) or \(S\) and the three quantities \(D\) or \(E\), the expression

\[ dU = \Sigma T dS_k + \Sigma E_m dD_m, \]

i.e. the mechanical and the electrical work done when the state of the material changes must again be a total differential; in other words the 'internal energy' \(U\) is again a variable of state. It follows in the same way as before, that the \(9 \times 9\) array of coefficients will be completely symmetric if \(T\) and \(E\) are expressed in terms of \(S\) and \(D\). In (67), however, as in the literature on acoustic amplification and electro-acoustic surface waves, we have expressed \(T\) and \(D\) in terms of \(S\) and \(E\). A difference in sign for the \(e\) then occurs between (67a) and (67b); this can be seen as follows. Under the above assumption, not only \(U\) but also the quantity \(H_S = U - \Sigma E_m D_m\) is a variable of state (i.e. it is fully determined by the nine quantities \(T\) or \(S\) and \(D\) or \(E\)), so that

\[ dH_S = \Sigma T dS_k - \Sigma D_m dE_m \]

is a total differential. From this follows, in the usual manner, the symmetry of the array of coefficients in (67), including the minus signs. \(H_S\) is called the electric enthalpy.

Fortunately, a drastically simplified form of (67) is sufficient for the discussion of a number of typical cases. This is because, in certain materials possessing a sufficient symmetry, waves in certain directions are...
possible in which only one component of each of $T$, $S$, $u$, $D$, and $E$ is of importance. We shall restrict the discussion to such situations. We then have:

$$T = cS - eE,$$  \hspace{1cm} (69a)

$$D = eS + eE.$$  \hspace{1cm} (69b)

Strictly speaking, in each actual case we should indicate which component of $T$, $S$, $D$ and $E$ is involved. We shall do this for $D$ and $E$ — we shall discuss waves for which $D$ and $E$ are either transverse or longitudinal — but we shall not specify which components of $T$, $S$ and $u$ are involved.

The positive dimensionless constant $e^2/\varepsilon c$ will be found to be a convenient measure of the relative strength of the piezoelectric coupling. In most piezoelectric materials, $e^2/\varepsilon c$ is much less than 1, i.e. the coupling is weak. In certain specially prepared ceramics, e.g. certain lead zirconium titanates, $e^2/\varepsilon c$ may approach the value unity.

As for the purely electromagnetic waves, we shall first discuss insulators here and then examine what complications and possibilities arise as a result of conduction. These considerations follow, somewhat freely, the analysis of A. R. Hutson and D. L. White [86].

Near-light and near-sound

We shall first consider electric transverse waves in weakly piezoelectric insulators. As might be expected, these are near-electromagnetic waves or near-acoustic waves. The deviations in velocity from ‘pure light’ and ‘pure sound’ respectively will be calculated from the weak-coupling relation (63).

We take the direction of propagation as the $z$-axis and the direction of $D$ and $E$ as the $x$-axis. Maxwell’s equations reduce to the first two, (7) and (8), with $J = 0$. Because $\partial_x$ and $\partial_y$ are zero and, according to our assumptions, only one component of each of $S$, $T$ and $u$ are involved, only one of the defining equations (51) remains ($S = \partial_y u$). Similarly, only one of the equations of motion (54) remains ($\rho_m \partial_y^2 u = \partial_z T$). Using (69) as well we thus obtain six equations with six variables. With $\partial_x = 0$, $\partial_y = 0$, $\partial_z = -jk$ and $\partial_t = j\omega$, these equations become:

$$k H_y = \omega D_s, \hspace{1cm} \omega \mu H_y = k E_z; \hspace{1cm} (\text{Maxwell}) \hspace{1cm} (a,b)$$

$$D_x = eS + eS, \hspace{1cm} T = cS - eE_x; \hspace{1cm} (\text{mixed}) \hspace{1cm} (c,d)$$

$$-jk u = S, \hspace{1cm} \rho_m \omega^2 u = jk T; \hspace{1cm} (\text{elastic}) \hspace{1cm} (e,f)$$

Equations (a,b) are purely electromagnetic, (c) is electric with some elastic admixture, (d) is elastic with some electrical admixture and (e,f) are purely elastic. We can find the dispersion equation by putting

the $6 \times 6$ determinant of (70a ... f) equal to zero or by successive elimination of the wave variables. The result can be written as:

$$(\omega^2 - v_1^2 k^2)(\omega^2 - v_2^2 k^2) = -\frac{e^2}{\varepsilon c} k^2 v_3^2.$$  \hspace{1cm} (71)

In this equation $v_1$ and $v_2$ are the velocities of light and sound respectively in the ‘unperturbed’ medium ($v_1 = 1/\sqrt{\varepsilon \mu}$, $v_2 = \sqrt{\varepsilon / \gamma_m}$). As anticipated, equation (71) is of the form (61), $f_{\beta} = \delta$. For zero coupling ($\delta = 0$) we return to the two independent waves: light waves of velocity $v_1$ and sound waves of velocity $v_2$. With (weak) coupling ($e^2/\varepsilon c \ll 1$) we find near-light and near-sound. To find the velocity of the near-light wave we put $\omega = v_k k_0$. Using (63), together with

$$f = \omega^2 - v_1^2 k^2 \hspace{1cm} g = \omega^2 - v_2^2 k^2 \hspace{1cm} \delta = \frac{e^2}{\varepsilon c} \omega^2 k^2 v_3^2,$$

we obtain:

$$\frac{\Delta k}{k_0} = -\frac{\frac{e^2}{\varepsilon c} v_3^2}{v_2^2}. \hspace{1cm} (72)$$

For the near-sound wave, putting $\omega = v_k k_0$, we find analogously:

$$\frac{\Delta k}{k_0} = +\frac{\frac{e^2}{\varepsilon c} v_3^2}{v_1^2}. \hspace{1cm} (73)$$

(In (72) and (73), $v_3^2$ in the denominator has been neglected in comparison with $v_1^2$.) Since $\Delta u/v_1$ is equal to $-\Delta k/k$, the velocity of the near-light wave is a fraction $\frac{1}{2}(e^2/\varepsilon c) v_3^2/v_1^2$ greater than that of light in a non-piezoelectric medium with the same $\varepsilon$ and $\mu$. Similarly, the velocity of the near-sound wave is the same fraction smaller than the velocity of sound in a non-piezoelectric medium of the same $\gamma_m$ and $c$. The effect is extremely small ($\ll 10^{-10}$), not primarily because the coupling factor $e^2/\varepsilon c$ is small but because the velocities of light and sound differ by so large a factor (10$^9$). This is a matter of some generality: between waves of greatly differing velocities there is little interaction.

Piezoelectric stiffening

We now consider electric longitudinal waves in piezoelectric insulators. Since in non-piezoelectric materials, as in free space, electromagnetic waves of this type do not exist, there is no question of near-light in this case; we can expect only an acoustic wave with some admixture of an electric-longitudinal character. Its velocity is easily found. Since $E$ is longitudinal, so that curl $E$ is zero, it follows from the Maxwell equation (8) (with $\mu \neq 0$, $\omega \neq 0$), that $H = 0$; also from (7) (with $J = 0$, $\omega \neq 0$), that $D = 0$. (For longitudinal $D$, this also follows from the fact that

div $D = 0$ for an insulator.) From the mixed equations:

$$0 = eS + eE,$$  \hspace{1cm} (74)

$$T = cS - eE,$$  \hspace{1cm} (75)

it follows by elimination of $E$ that

$$T = c(1 + e^2/|c|) S.$$  \hspace{1cm} (76)

We see that only elastic variable remains for which the conventional elastic equations are valid — with this difference that $c$ is replaced by $c(1 + e^2/|c|)$. The material has thus become effectively stiffer. This phenomenon [27] is known as piezoelectric stiffening. The velocity of the near-sound wave is thus equal to

$$v_s = \sqrt{c(1 + e^2/|c|)/\mu_m} \approx (1 + \frac{1}{3}e^2/|c|)\sqrt{c/\mu_m},$$

so that the relative effect on the velocity is a factor $v_s^2/v_e^2 (\approx 10^{16})$ larger than above.

For the elastic wave, the distinction between electric-longitudinal and electric-transverse admixture is essentially the distinction between electrostatic and electrodynamic phenomena. This explains the relatively small effect of the electric-transverse admixture: electrodynamic effects are small for slow processes and an elastic wave is slow compared with a light wave. Let us now elaborate on this a little further. In the neutral insulator with which we are concerned, the divergence of $D$ must be zero. For the longitudinal wave, this also implies that $D = 0$ and that therefore the polarization wave $eS$ gives rise to a field wave $E$ of significant amplitude: $eE = -eS$. According to (75) this results in a considerable change in the elastic stiffness of the medium. In the transverse case there is no electrostatic reason for $D$ to be zero: all transverse vectors have zero divergence. The (transverse) polarization $eS$ now excites a field $E$ only by induction. This however is very small. From (70a,b) we find (with $\omega/k = v_E$):

$$eE_x = e_0v_2^2D_z = (v_2^2/v_1^2)D_x.$$  \hspace{1cm} (77a)

The terms in $E_x$ in (70c,d) can therefore be neglected so that $D_x$ is virtually equal to $eS$ and, from (70d), the stiffness is hardly affected. Summarizing, the electric field is significant in the longitudinal case ($E = -eS/e$) but not in the transverse case. For this reason we may restrict ourselves to longitudinal fields and currents in our discussion of piezoelectric semiconductors, which now follows.

**Acoustic waves in piezoelectric semiconductors**

Mobile charge carriers in a piezoelectric material cause attenuation of an acoustic wave: the electric fields associated with the wave cause currents and thus ohmic losses. On the other hand, however, there is the interesting possibility of negative attenuation, i.e. amplification of an acoustic wave, through activation of the medium by a constant electric drift field [28]. A simple criterion decides whether amplification is possible or not: the drift velocity $v_d$ must be higher than the wave velocity $v_s$. As noted above, we are exclusively concerned here with longitudinal fields and currents.

We shall treat the attenuation and amplification of acoustic waves in two stages. First we shall establish formally the relation between attenuation or amplification and the complex conductivity $\sigma = \sigma_r + j\sigma_i$, which is the ratio of the wave variables $J$ and $E$. It will be found that a negative $\sigma_i$ implies amplification. After this we shall show that with a constant applied electric field, a negative $\sigma_r$ can be obtained.

For the first stage we write, as usual (eq. 32):

$$J = \sigma E.$$  \hspace{1cm} (78)

As in the case of piezoelectric stiffening, we shall derive a new effective stiffness of the material by elimination of the electric wave variables. From Maxwell’s equations (using $\operatorname{curl} E = 0$ for longitudinal $E$, whence $H = 0$ and therefore $J + D = 0$), and (76):

$$D = -\sigma E/j\omega.$$  \hspace{1cm} (79)

Substituting this in (69) now leads not to (74) and (75) but to:

$$T = cS - eE,$$  \hspace{1cm} (77a)

$$0 = eS + (e + \sigma_0/j\omega)E,$$  \hspace{1cm} (77b)

and hence to:

$$T = c \left(1 + \frac{e^2}{|c|} \frac{1}{1 + \sigma_0/j\omega} \right) S.$$  \hspace{1cm} (80)

We now have an effective elastic stiffness of

$$c' = c \left(1 + \frac{e^2}{|c|} \frac{1}{1 + \sigma_0/j\omega} \right).$$

There is thus stiffening of the medium again but it is now a complex stiffening. (Comparing this with the stiffening in insulators, we see that, as before, the conduction could have been taken into account by introducing an effective dielectric constant $\epsilon + \sigma_0/j\omega$.)

We assume that $\Delta c = c' - c$ is small compared with $c$. This is certainly the case for weak piezoelectric coupling, provided $|1 + \sigma_0/j\omega|$ is not too much less than unity. We then obtain directly from the dispersion relation $k^2/\omega^2 = \mu_m/c$ for sound waves, the relative change $\Delta k/k$ occurring at a given $\omega$ as a result of the relative complex stiffening $\Delta c/c$:

$$\frac{\Delta k}{k} = -\frac{\Delta c}{c} = -\frac{e^2}{|c|} \frac{1}{1 + \sigma_0/j\omega}.$$  \hspace{1cm} (81)

We are interested only in the imaginary part $\Delta k_1$ of $\Delta k$. Negative $\Delta k_1/k$ implies attenuation, positive $\Delta k_1/k$ amplification of the sound wave. (See (6) and the caption to fig. 3. If $\Delta k_1$ is sufficiently small, the relative
change in amplitude per unit length is equal to $\Delta k_1$ and the relative change per wavelength is $2\pi \Delta k_1 / k$.) Remembering that $\sigma$ can be complex, we find:

$$\frac{\Delta k_1}{k} = -\frac{\frac{e^2}{sc} \frac{\sigma_i/\omega \epsilon}{1 + (\sigma_i/\omega \epsilon)^2 + (\sigma_0/\omega \epsilon)^2}}{\sigma_0/\omega \epsilon^2}.$$  (78)

This shows that amplification occurs if $\sigma_i$ is negative.

The small change in the real part of $k$, i.e. in the wave velocity, can be neglected in the calculation of the attenuation or amplification. In the following, therefore, we put $\omega/k = v_0$.

In the simple cases where $\sigma$ merely represents the static conductivity ($\sigma_i = \sigma_0, \sigma_1 = 0$), the attenuation is given by

$$\frac{\Delta k_1}{k} = -\frac{\frac{e^2}{sc} \frac{\sigma_i/\omega \epsilon}{1 + (\sigma_i/\omega \epsilon)^2 + (\sigma_0/\omega \epsilon)^2}}{\sigma_0/\omega \epsilon^2},$$

where $\omega_i$ is called the relaxation frequency, $\omega_0 = \sigma_0/\epsilon$. This is the reciprocal of the dielectric relaxation time $\tau$, discussed on p. 327. The attenuation is a maximum for $\omega = \omega_0$, and has a significant value only when $\omega$ and $\omega_0$ are of the same order of magnitude. It can be shown from this that, even for frequencies in the GHz region, attenuation of this nature is found only in semiconductors with a very low concentration of charge carriers. In photoconductors $\sigma_0$, and hence $\omega_0$, is dependent on the intensity of the incident light. The light-sensitive acoustic attenuation found in CdS [29] can be explained in this way [30].

**Amplification of acoustic waves**

We now come to the second stage: to show that $\sigma_1$ can become negative by the application of an electric field $E_0$ that gives the electrons a constant drift velocity $v_{d0}$, thus enabling acoustic waves to be amplified.

The acoustic wave in a piezoelectric semiconductor produces variations in three directly related quantities via the lattice polarization. These are the longitudinal electric field, the charge density (electron concentration) and the current. There are now two factors in the expression for the current density that vary: the drift velocity and the electron concentration. We are thus concerned with the product of two varying quantities and we can therefore no longer use complex quantities. We now denote the actual field by $E_0 + E_1$, the actual drift velocity by $v_{d0} + v_{d1}$ and the actual electron concentration by $n_0 + n_1$, where $E_0$, $v_{d0}$ and $n_0$ are constant and $E_1$, $v_{d1}$ and $n_1$ are real quantities with frequency $\omega$ and wave number $k$. We then have

$$v_{d1} = -\mu_0 E_1.$$  (79)

The current density is

$$J(\omega_0 + n_1)(v_{d0} + v_{d1}) =
-qn_{d0}v_{d0} - qn_{d0}v_{d1} - qn_1v_{d0} - qn_1v_{d1}.$$  (80)

There is thus a d.c. component $-qn_{d0}v_{d0}$ and a wave component $-qn_{d1}v_{d1} - qn_1v_{d0}$. The last term $-qn_1v_{d1}$ is of second order and can be neglected. This term does contain a d.c. component $-q(n_1v_{d1})$ to which we shall return in the next section. At the moment, however, we are only concerned with the wave component. Using complex wave variables $J, n$ and $E$ again ($n_1 = Re n, E_1 = Re E$) we find, with $n_0\mu_0 = \sigma_0$ (see Table I, p. 323, eq. I.6) and $-qn = qe_0$:

$$J = \sigma_0 E + v_{d0}q_0e_0.$$  (81)

The current in the wave thus consists of a direct reaction $\sigma_0 E$ on the field and a current $v_{d0}q_0$ which is automatically present in each charge fluctuation $q_0$ because the electron gas as a whole has the drift velocity $v_{d0}$.

Equation (79) is not quite complete. Because there are charge fluctuations there is also diffusion, so that besides the current (79) there is also a diffusion current. We shall neglect this diffusion current for the moment, but will return to it presently.

Apart from (79) there is a further relation between the wave variables $J$ and $q_0$: the continuity equation (11). Because $J$ is longitudinal, this assumes the form

$$-jkJ = -joq_0e_0$$  or, with $\omega/k = v_0$:

$$J = v_0q_0e_0.$$  (80)

Elimination of $q_0$ from (79) and (80) gives the ratio between $J$ and $E$ and hence the complex conductivity that we require. The result is:

$$J = \sigma_0 E / (1 - v_{d0}/v_0),$$

and thus

$$\sigma_1 = \sigma_0 / (1 - v_{d0}/v_0), \quad \sigma_1 = 0.$$  (81)

This shows that by giving the electrons a drift velocity $v_{d0}$ greater than the wave velocity $v_0$, $\sigma_1$ can indeed be made negative and amplification of the wave can be achieved.

The operation of the acoustic amplifier is therefore as follows. Let us assume that the acoustic wave propagates at a velocity $v_0$ to the right. The associated longitudinal alternating electric field causes charge fluctuations that propagate with the wave. Now a positive net space charge only moves to the right because it is continually built up at the front and continually dispersed at the back by an a.c. current moving

in the direction of the wave. This current must thus be directed to the right in a positive space charge and to the left in a negative space charge (see fig. 25). This is expressed by the continuity equation (80) for longitudinal waves.

If the mean drift velocity is zero, this current is directly caused by the alternating field, which thus maintains the charge wave. Alternating field and current are in phase, so that there are losses. When, however, the electron gas as a whole has a velocity \( v_0 \) to the right, each net positive space charge now implies a current to the right (and each negative space charge a current to the left) so that a weaker alternating field is necessary to maintain the charge wave. Alternatively we see that, with the same alternating field, larger charge fluctuations will arise. If \( v_0 \) is equal to \( v_0 \), then no field at all is necessary to maintain the charge wave — it propagates "by itself" — and, for a given alternating field, infinitely large space charges would arise (in the absence of diffusion). Finally, if \( v_0 \) is larger than \( v_0 \), then a positive space charge must now be dispersed at the front and built up at the back to maintain the stationary state, and this has to be done by the alternating field (fig. 26). Field and current are therefore in antiphase, so that the field, and hence the wave, take up energy from the charge carriers.

By ignoring the diffusion we have been able to give a fairly simple description of the essentials of the acoustic amplifier. To discuss the practical possibilities, however, it is necessary to take the diffusion into account. This can be done by adding a diffusion term \(-D_n \text{grad} \rho_0 = jkD_n \rho_0 \) to the right-hand side of (79), see also (49). It is then easy to show that (81) becomes:

\[
\frac{\sigma_r}{\sigma_0} = \frac{\gamma}{\gamma^2 + \omega_D^2\omega_D^2}, \quad \frac{\sigma_1}{\sigma_0} = \frac{\omega_0}{\omega_D} \frac{\omega_0}{\omega_D^2 + \omega_D^2} \tag{82}
\]

where \( \gamma = 1 - v_0/v_0 \) and \( \omega_D = v_0^2/\nu_0 \).

To find the amplification given by the negative conductivity, we must combine (82) with (78). The result for the amplification per unit length is found, after some manipulation, to take the form:

\[
\Delta k_1 = -\frac{1}{2} \frac{\omega_0}{\gamma} \frac{\omega_0}{v_0} \frac{\gamma}{\gamma^2 + (\omega_0/\omega + \omega_0/\omega_D)^2} \tag{83}
\]

In (83) \( \omega/k \) is again written as \( v_0 \). From this equation it can be seen that two restrictions are imposed on the frequency: there is substantially no amplification when \( \omega \) is much smaller than \( \omega_0 \) or much greater than \( \omega_D \).

The physical reason is as follows. When \( \omega \) is much smaller than \( \omega_0 \), the charge fluctuations relax so rapidly that they completely screen off the piezoelectric field. If, on the other hand, \( \omega \) is much larger than \( \omega_D \), then the waves are so short and the wave flanks so steep that diffusion erases the charge fluctuations. Amplification is thus possible only in materials for which \( \omega_0 < \omega_D \). Now in most materials \( \omega_D \) is no larger than 10 GHz. With \( \epsilon \approx \epsilon_0 \approx 10^{-11} \text{F/m} \), it then follows that to make an acoustic amplifier, the piezoelectric semiconductor must have a conductivity \( \sigma_0 = \epsilon_0 \omega_0 \) smaller than \( 10^{-11} \times 10^{10} = 10^{-1} \Omega^{-1} \text{m}^{-1} \).

The acoustic amplifier is an example of the general phenomenon in which a stream of particles can interact in some way with a travelling wave, and give up energy to that wave when their velocity is larger than the wave velocity. Another well-known example is the travelling-wave tube, in which electrons are injected into an electromagnetic wave of relatively low velocity.

Acoustic amplification was first demonstrated experimentally by A. R. Hutson, J. H. McFee and D. L. White [28] in a photoconducting CdS crystal. The acoustic transmission they measured is plotted as a function of the applied field in fig. 27. Free charge carriers were produced by illumination of the crystal. In
the dark the crystal was a good insulator; the concentration of charge carriers, and hence \( \omega_{dr} \), was adjusted by the incident light intensity. The zero level in fig. 27 refers to acoustic transmission through the crystal in the dark (the attenuation at 45 MHz was less than 7 dB and at 15 MHz less than 2 dB); positive dBs mean attenuation, negative dBs amplification. The results are qualitatively in good agreement with equation (83). In particular, the charge mobility can be derived from the zero intersection at \( \approx 700 \text{ V/cm} \), by putting the drift velocity corresponding to this field equal to the velocity of the transverse acoustic wave used, \( v_0 = 2 \times 10^6 \text{ cm/s} \). The mobility of 285 cm\(^2\)/Vs found in this way agrees well with values found by other method (\( \approx 300 \text{ cm}^2/\text{Vs} \)).

The acoustic amplifier we have dealt with can be regarded as the precursor of the amplifier of acoustic surface waves (p. 347-348) which has attracted attention in recent years.

![Fig. 27. Acoustic amplification as shown by A. R. Hutson, J. H. McFfee and D. L. White](image)

Acoustic amplification considered as weak coupling of two types of wave

We have considered the acoustic attenuation and amplification as a result of the ‘complex stiffening’ of the material. In this connection we came across the phenomena of dielectric relaxation and diffusion; these are ‘longitudinal-electrical’ wave phenomena (p. 327). Acoustic attenuation and amplification can also be considered as the result of weak piezoelectric coupling between an acoustic wave and a longitudinal electrical ‘wave’ \[^{[31]}\]. To show this, we shall start with a complete set of equations in the variables \( J, D, E \) (all longitudinal), \( \theta_e, T, S \) and \( u \), in, for example, the following form:

\[
\begin{align*}
J + \dot{D} &= 0, \quad T = S + eE, \quad S = \partial_t u, \\
\text{div } J &= -\partial_t \theta_e, \quad D = eS + eE, \quad \partial_t T = \partial_t \partial_t u, \\
J &= \sigma_0 E + \sigma_0 \theta_e - D \text{grad } \theta_e.
\end{align*}
\]

The dispersion relation that follows from these equations can easily be written in the form \( f_g = \delta \) (eq. 61), remembering that \( e^2/\varepsilon_c \) is small. The result is:

\[
\begin{align*}
\frac{(\omega^2/v_s^2 - k^2)}{f} [\omega_e + j(\omega - \omega_D k) + k^2 D_2] &= \frac{\varepsilon}{f}, \\
&= (\varepsilon^2/\varepsilon_c) k^2 [(\omega - \omega_D k) + k^2 D_2].
\end{align*}
\]

By putting \( f = 0 \), we have once more the unperturbed sound wave, while \( \omega = 0 \) yields an unperturbed longitudinal electric wave — another version of (50) somewhat complicated by the occurrence of \( \omega_0 \). Applying (63), with \( \omega = \omega_0 \), to (84) we obtain after some manipulation a complex expression for \( \Delta k/k \), the imaginary part of which is indeed given by (83).

**The acousto-electric effect; Weinreich’s relation**

We have seen that a d.c. current in a piezoelectric semiconductor can affect an acoustic wave in that material. Conversely, an acoustic wave in the material can give rise to a d.c. current or to a static electric field (depending on whether the semiconductor is short-circuited or open-circuited). The charge carriers are ‘carried along’ by the wave. This is known as the acousto-electric effect. It also occurs, to a lesser extent, in other materials. When R.H. Parmenter predicted the effect \[^{[33]}\] in 1952, he remarked on the correspondence with two other quite different phenomena — the linear accelerometer, in which electrons are accelerated by oscillating fields that are synchronized to form an (accelerating) travelling wave, and the thermoelectric effect where charge carriers are borne along by the stream of phonons (acoustic quanta) caused by a temperature gradient.

The effect can be attributed to the d.c. component that is present in the current density because both the electron concentration and the drift velocity are modulated when the material carries an acoustic wave. On p. 337 we saw that this component is equal to \(-q \theta_1 \theta_{41} \), where \( \theta_1 \) and \( \theta_{41} \) are actual physical modulations; using complex wave variables \( n \) and \( \nu_d \), the d.c. component is \(-i\nu_{dc} \). We thus have an acousto-electric d.c. current source

\[
J_{ac} = -i\nu_{dc} \theta_{41} \theta_1.
\]

This is equivalent to a ‘field source’

\[
E_{ac} = J_{ac}/\sigma_0 = \frac{1}{2} \Re(\sigma_0 \nu_{dc} \theta_1)/\sigma_0.
\]


This is the field that is equivalent to the acoustic wave in its effect on the charge carriers. This acoustic-electric field is directly related to the attenuation of an acoustic wave. The quantity \( \Re(\rho_0v_0E^*) \) is proportional to \( \frac{1}{2} \Re(JE^*) \), the ohmic heat developed (\( \rho_0 \propto J \) from (80) and \( v_0 \propto E \)) which must be supplied by the acoustic wave which is thus attenuated. If this ohmic heat is the only cause of attenuation, then

\[
\alpha P = \frac{1}{2} \Re(JE^*) = -\frac{1}{2}(v_0/\rho_0)\Re(\rho_0v_0E^*),
\]

where \( \alpha \) is the acoustic attenuation coefficient and \( P \) the energy flow density of the sound wave. (By definition, \( \alpha = -(1/P)\partial P/\partial z \); since \( P \) is proportional to the square of the wave variables, \( \alpha = -2\Delta k_z \).) In (86) we made use of \( J = v_0\rho_0 \) (eq. 80) and \( v_0 = -\mu_0 E \) (eq. I.2 in Table 1).

From (85) and (86) the very simple relation between \( \alpha \) and \( E_{ae} \) first given by G. Weinreich follows directly:

\[
\alpha = -nqv_0E_{ae}/P.
\]

The very general reasoning used by Weinreich was as follows. A travelling acoustic wave represents not only a flow of energy but also a flow of momentum. Attenuation of the wave therefore implies a diminishing momentum flow: momentum is transferred to the ‘cause of the attenuation’ and hence a force acts on this ‘cause’, i.e. the charge carriers. The field equivalent to this force is the acousto-electric field. This reasoning shows that we are concerned here with a kind of radiation pressure exerted by the acoustic wave on the charge carriers.

Let us now make one or two further comments about Weinreich’s relation.

Assuming that charge carriers of only one kind, of mobility \( \mu_0 \), are responsible for the attenuation, the relation is very generally valid, even when the charge carriers have a non-zero constant drift velocity \( v_0 \). This means that, as \( v_0 \) increases, \( E_{ae} \) changes its sign exactly at the instant when attenuation changes to amplification.

If the charge carriers do have a non-zero drift velocity, several field and current components come into play, as discussed on p. 337. It is not perhaps immediately clear whether only the part \( \frac{1}{2} \Re(JE^*) \) of the ohmic heat is due to the acoustic wave (\( J \) and \( E \) are the wave variables) as is assumed in (86) (and not, for example, \( J_{ae}E_0 \)). If this is not the case, then the derivation of (87) would no longer be valid. However, (86) can also be proved explicitly starting from the expression (78) for the attenuation and from the fact (not proved here) that the acoustic-energy flow density \( P \) can be represented by \( \pm v_0cSS^* \). Since \( \Re(JE^*) \) is equal to \( \sigma_{ee}EE^* \), (86) gives

\[
\alpha = \frac{\sigma_{ee} EE^*}{v_0c SS^*}.
\]

Substituting here the value for \( E/S \) from (77b), we arrive, using \( \Delta k_z = -\alpha \), directly at (78). This is an indirect proof of (86).

Much experimental work has been done on the acousto-electric effect, acoustic attenuation and the relationship between them, especially in CdS. In spite of what has been said above, Weinreich’s relation is often not satisfied. For one thing, as \( v_0 \) increases, \( E_{ae} \) and \( \alpha \) often do not go through zero at the same point. The reason for this is that some of the charge carriers are trapped and thus are not mobile although they still contribute to the space charge. These charge carriers cause absorption but they do not contribute to the acousto-electric current. From this reasoning a generalization of Weinreich’s relation can be derived by taking the mobile charge for \( q_0 \) in (85) and the total charge for \( q_{ae} \) in (86). Further analysis shows that measurement of the frequency dependence of \( E_{ae} \) gives information about the trapping mechanism.

**Waves in two media with a common boundary**

In bounded media, the bulk waves studied above can, in general, no longer occur by themselves because alone they do not satisfy the conditions imposed by the boundaries. At the free surface of a solid, for example, the shear stress tangential to the surface and the tensile stress normal to it are of course zero, whereas almost every simple sound wave involves such stresses. Superposed simple sound waves are however possible if when taken together they cause the surface stress to be zero. This indicates the way to tackle the present problem — the problem of two adjacent media: we should try to combine the bulk waves in such a way that the boundary conditions at the interface are always satisfied at all points of the interface. In this way we can describe phenomena such as refraction and reflection of waves and — our special interest here — surface waves. We shall first illustrate this approach by some simple examples.

**Transmission, reflection, refraction**

*The junction of two transmission lines*

In the transmission line of fig. 2 waves can be propagated with a voltage-current ratio \( V/I \) given by \( \pm \sqrt{L/C} \). Or, more precisely, in the waves travelling to the right (\( v > 0 \)), \( V/I \) is equal to \( \pm \sqrt{L/C} \), the characteristic impedance \( Z \) of the line; in the waves travelling to the left (\( v < 0 \)), \( V/I \) is negative, \( V/I = -Z \). This follows directly from (3), (4) and (5). Let us now consider the reflection at the junction between two transmission lines of different characteristic impedances \( Z_1 \) and \( Z_2 \) (fig. 28), and consider what com-
Fig. 28. Reflection and transmission at the junction of two transmission lines of characteristic impedances $Z_1 = \sqrt{L_1/C_1}$ and $Z_2 = \sqrt{L_2/C_2}$; a incident wave, b reflected wave, c transmitted wave.

Fig. 29. Reflection and transmission of normally incident linearly polarized light at the interface between two media with intrinsic impedances $Z_1 = \sqrt{\mu_1/\epsilon_1}$ and $Z_2 = \sqrt{\mu_2/\epsilon_2}$; a incident wave, b reflected wave, c transmitted wave.

Fig. 30. Integration contour for $E$ along the surface, for the proof that $E$ is continuous at an interface (see text).

combinations of incident (a), reflected (b), and transmitted (c) waves are possible. In this one-dimensional example the interface has degenerated to a junction. The boundary conditions require continuity of the current and voltage at the junction. If we take the junction to be at $z = 0$, then we must have (see eq. 2):

$$V_a \exp j\omega t + V_b \exp j\omega t = V_c \exp j\omega t,$$

$$I_a \exp j\omega t + I_b \exp j\omega t = I_c \exp j\omega t.$$

Since these equations must always be satisfied it follows, first that $\omega_a = \omega_b = \omega_c$ and secondly that

$$V_a + V_b = V_c,$$

$$V_a/Z_1 - V_b/Z_1 = V_c/Z_2. \quad (88)$$

From (88), the voltage reflection coefficient $K_R$ (the ratio $V_b/V_a$) and the voltage transmission coefficient $K_T$ (the ratio $V_c/V_a$) are, respectively:

$$K_R = \frac{Z_2 - Z_1}{Z_2 + Z_1}, \quad K_T = \frac{2Z_2}{Z_1 + Z_2}. \quad (89)$$

When the two lines are matched, i.e. $Z_1 = Z_2$, there is no reflection and complete transmission ($K_R = 0$, $K_T = 1$). When the first transmission line is open-circuited ($Z_2 = \infty$) or short-circuited ($Z_2 = 0$) there is complete reflection ($K_R = \pm 1$).

Normally incident light

When light falls normally on the interface between two isotropic transparent media (fig. 29) we again have the three waves incident (a), reflected (b) and transmitted (c). Suppose the light is linearly polarized. The boundary conditions in this case are that $E$ and $H$ are continuous at the interface. This can be shown for $E$ by integrating $E$ along an extended contour as in fig. 30. The result is curl $E$ integrated over the enclosed area. If the loop of the contour is made infinitely thin, this surface integral is zero (curl $E$ does not become infinite, see eq. 16) and therefore the contour integral is also zero, so that $E$ must be the same on either side of the interface. The continuity of $H$ is shown in a similar way.

The ratio $E_x/H_y$ in each of the waves has an absolute value equal to the value of $\sqrt{\mu/\epsilon}$ in the corresponding medium (see p. 317); the sign is positive for waves travelling to the right ($k > 0$), negative for waves

travelling to the left \((k < 0)\); see (17a). Therefore

\[
E_a + E_b = E_e,
\]

\[
E_a/\sqrt{\mu_1/\varepsilon_1} - E_b/\sqrt{\mu_2/\varepsilon_2} = E_e/\sqrt{\mu_2/\varepsilon_2}.
\]

These equations have exactly the same form as (88). With the definition of the intrinsic impedance of a medium given on p. 317, \(Z = \sqrt{\mu/\varepsilon}\), the reflection and transmission are given again by (89).

**Normally incident sound**

For sound in two adjoining elastic isotropic media (fig. 31) the situation is analogous to the foregoing. It is clear that, at the interface, \(T_{zz}, T_{yy}, T_{xx}\) and \(u_x, u_y, u_z\) must be continuous (interface perpendicular to the \(z\)-axis). Let us assume that in each medium only one component of \(T\) and \(u\) is involved. For waves travelling to the right we have (with \(\omega/k = \sqrt{c/\rho_m}\)):

\[
T = cS = cuu = -jkcu = -j ou/\sqrt{\rho_m}.
\]

and for waves travelling to the left:

\[
T = +j ou/\sqrt{\rho_m}.
\]

Therefore

\[
T_a + T_b = T_e,
\]

\[
T_a/\sqrt{\rho_{m1}} - T_b/\sqrt{\rho_{m1}} = T_e/\sqrt{\rho_{m2}}.
\]

The quantity \(\sqrt{\rho_m}\) is called the mechanical impedance. The reflection and the transmission are again given by (89). The reasoning is valid for both transverse and longitudinal waves. Which component of \(T\) and of \(u\) and which constant \(c\) are relevant depends on the type of wave considered.

The close analogy between the above three cases might suggest that for normal incidence (89) is of very general validity. This is not the case. To obtain a result such as (89) with one transmitted and one reflected wave, these waves and their polarizations have to be matched to the incident wave. An example of a simple situation where this is no longer possible is the case of a linearly polarized beam of light that is normally incident on the interface between free space and an optically anisotropic uniaxial crystal when the optical axis lies \(\perp\) the interface but is not parallel or perpendicular to the plane of polarization of the light beam. Two beams of different velocities and different ratios \(E/H\) then arise in the crystal (see p. 316) so that (89) is no longer applicable. Of course, in this simple case, that incident beam can be resolved into two components with polarizations parallel and perpendicular to the optical axis, and (89) can then be applied to each of the two problems thus obtained.

**Obliquely incident waves**

In the three foregoing cases, the wave variables depended on \(z\) and \(t\) and thus, in the interface, only on \(t\). Because the boundary conditions have to be satisfied at all times the waves must all have the same frequency. This virtually self-evident requirement was mentioned explicitly only in the first example. With obliquely incident waves, the wave variables also depend on the coordinates in the interface. Since the boundary conditions must be satisfied at all times everywhere in the interface, all waves combining at an interface must have the same periodicity both in time and in place along the interface. In the following we shall take the \(y\)-axis to be normal to the interface. Then all the waves of a combination must have the same \(\omega, k_x\), and \(k_z\). Once these three quantities are given, e.g. by the incident wave, then \(k_y\) is determined for each other wave of the combination by its dispersion relation (which is a relation between \(\omega, k_x, k_y\), and \(k_z\) for that wave), and thus its direction of propagation is also determined. This is illustrated by the following example.

**The laws of optical reflection and refraction**

At the interface of two optically isotropic media, incident light is reflected and refracted (fig. 32). The velocity of light in medium 1 will be denoted by \(v_1\), that in medium 2 by \(v_2\). As stated above, the \(y\)-axis is taken to be perpendicular to the interface; the \(x,y\)-plane is taken as the plane of incidence (the plane
containing $k_a$ and the $y$-axis). Then $k_{za} = 0$. As a result $k_{ab}$ and $k_{za}$ are also zero: the reflected beam ($b$) and the refracted beam ($c$) lie in the plane of incidence. The law of reflection, $\chi = \theta$, follows from fig. 32 by observing that not only are $k_{ab}$ and $k_{za}$ equal, but $k_b$ and $k_a$ also, because the waves $a$ and $b$ are in the same isotropic medium and have the same frequency. Finally, Snell's law of refraction follows from the dispersion relations $k_c^2 = \omega^2/v_1^2$ and $k_c^2 = \omega^2/v_2^2$ for the waves $a$ and $c$ and from the fact that $k_{zc} = k_{za}$:

$$\sin \varphi = \frac{k_{zc}k_c}{k_{za}k_a} = \frac{v_2}{v_1}.$$

These laws are thus a direct consequence of the requirement for equal frequencies and equal wave-vector components along the interface. They do not of course represent all the information contained in the boundary conditions: as in the previous three cases, it is also possible to calculate how much light is reflected and how much refracted. This leads to Fresnel's laws, but we shall not consider these here.

In fig. 32 we assumed that the boundary conditions can be satisfied by one incident wave, one reflected wave and one refracted wave. For light waves in optically isotropic media, further investigation shows that this is indeed the case, but it is by no means a general rule. For longitudinal sound waves, for example, incident at a certain oblique angle, one reflected and one refracted longitudinal wave are not sufficient — it is also necessary to introduce transverse waves with different velocities and directions: mode conversion takes place at the interface. However, the wave vector of each of the waves is always completely determined by its dispersion relation and by the 'interface component' of the wave vector of the incident wave.

**Total reflection; surface waves**

Whatever the value of the angle of incidence $\theta$ (fig. 32), the number of variables and the number of boundary conditions remains the same, so that the number of waves necessary to satisfy the boundary conditions remains unchanged. In particular, when the angle of incidence is increased so far that total reflection occurs, three waves are still present. What then happens to the refracted wave can be seen by calculating $k_{yc}$ from

$$k_{zc}^2 + k_{ye}^2 = k_c^2,$$

in which $k_c$ is given by the dispersion relation for waves in medium 2:

$$k_c^2 = \omega^2/v_2^2,$$

and $k_{zo}$ is given by $k_{za}$:

$$k_{za}^2 = k_{zo}^2 = (\omega^2/v_1^2) \sin^2 \theta.$$

If $\sin \theta$ becomes larger than $v_1/v_2$ — only possible for $v_2 > v_1$ — then $k_{zo}^2$ becomes larger than $k_c^2$. Hence $k_{ye}$ is imaginary. The wave that was previously refracted now propagates along the surface ($k_{zo} = \text{real}$) but its amplitude in medium 2 decays exponentially at right angles to the surface. This means that the refracted wave has become a surface wave. This is why the reflection is total: no energy is carried away from the interface in medium 2. Total reflection thus implies a surface wave in medium 2.

It is clear that an imaginary $k_{ye}$ (positive imaginary in fig. 32) implies that wave $c$ transports no energy away from the boundary: no power is transmitted downwards through medium 2 because at large distances from the interface the wave has zero amplitude, and no energy is dissipated in the medium because our assumption was that it is lossless. We can calculate the situation explicitly as follows. The mean energy flow in the $y$-direction (see note [10]) is

$$\bar{S}_y = \frac{1}{2} \Re(E_x H_x^* - E_x H_x^*).$$

From Maxwell's equations

$$\text{curl} \, H = -\dot{D}, \quad \text{curl} \, E = -\dot{B},$$

with $\dot{d}_t = 0$, $\dot{d}_y = -j k_y$, $\dot{d}_z = j \omega$, $B = \mu H$, $D = \varepsilon E$, it follows that

$$H_x = k_y E_z / \omega \varepsilon \quad \text{and} \quad E_x = -k_y H_z / \omega \mu.$$

Substituting these values in the expression for $\bar{S}_y$ gives:

$$\bar{S}_y = \frac{1}{2} \Re \{k_y^2 E_z E_z^*/ \omega \varepsilon + k_y H_z H_z^*/ \omega \mu \} = (k_{zo}^2 / \omega \varepsilon) \{1 + \mu \mu_s / \omega \varepsilon \}.$$

Since the expression inside the curly brackets is real, $k_y$ would have to have a non-zero real part if there is to be a mean energy flow in the $y$-direction.

In this demonstration it is assumed that the material is isotropic ($D = \varepsilon E$), that there are no losses ($\varepsilon$ and $\mu$ real) and that $\omega$ is real. Under these conditions, the same is true for every other direction. For anisotropic media, on the other hand, it cannot be concluded that $S$ has no component in a given direction if $k$ has no real component in that direction. In fig. 6b, for example, $S$ does have a component in the $x$-direction even though $k_e$ is zero.

With an 'ordinary' wave, with real wave-vector components, the phase velocity $\omega/k_e$ in any direction ($x$) other than the direction of propagation is greater than the velocity of propagation $\omega/k$, since $k_e \leqslant k$. For a surface wave, on the other hand, as follows from the foregoing, $k_e > k$: the surface wave is propagated more slowly than the corresponding bulk wave. Surface waves are therefore said to be 'slow'.

The phenomenon of total reflection can thus be summarized as follows. As the angle of incidence $\theta$ is increased, the phase velocity along the surface of all the waves involved decreases. When this velocity becomes smaller than the velocity of bulk waves in medium 2, the wave in medium 2 becomes a surface wave.

The surface wave is excited by the incident wave. In certain circumstances, however, independent freely
propagating surface waves are possible. This is the case when all the waves necessary to satisfy the boundary conditions are surface waves, i.e. waves that have a real wave-vector component along the surface (the same for all of them) and for which the wave variables decay exponentially with distance from the surface. Such waves form the subject of the next section.

Surface waves

As was mentioned in the introduction, acoustic surface waves on the free surface of a piezoelectric material offer interesting possibilities in certain fields of electronics. Such waves carry an electric field that extends beyond the boundary of the medium so that they can be generated, detected, amplified or otherwise processed electrically anywhere on the surface.

Surface waves form the subject of the next section.

![Diagram of surface waves](image)

Let us for a moment consider isotropic solid media that are not piezoelectric. The well known Rayleigh wave can be propagated on the free surface of such a medium (fig. 33a). In a coordinate system in which the surface is perpendicular to the y-axis, and the surface wave propagates in the x-direction, the particles move in ellipses in the x,y-plane for a Rayleigh wave. The Rayleigh wave is a special case of the Stoneley wave (fig. 33b). This can occur at the common boundary of two elastically different solid media bonded to each other — at least, when the elastic moduli and the densities satisfy certain conditions. Here again the particle movement is in the x,y-plane. If the density of medium 1 tends to zero, the Stoneley wave reduces to the Rayleigh wave.

At the interface of two bonded media yet another type of surface wave is possible, the Love wave (fig. 33c). The particle movement is confined here to the z-direction. In this case, however, if medium 1 is made less dense, the penetration depth of the waves in medium 2 becomes greater, and in the limit the Love wave does not remain a surface wave but degenerates to a bulk wave, propagating parallel to the surface without being perturbed by it.

In piezoelectric media the situation differs from the foregoing only in this last respect. Rayleigh-like, Stoneley-like and Love-like waves are all possible, now carrying electric polarizations and fields. The Rayleigh waves can again be regarded as a special case of the Stoneley waves. But the Love wave no longer reduces to a bulk wave when the density of one medium becomes zero as it does in the case of a non-piezoelectric medium. Instead, it remains a surface wave — at least
movement confined to the $z$-direction, is a relatively simple example of an acoustic surface wave on a piezoelectric medium. We shall now consider this wave in somewhat more detail. The treatment is different from that usually given but is perhaps more instructive.  

**Bleustein-Gulyaev waves**

Let us consider a piezoelectric medium bounded by the plane $y = 0$, outside which there is free space (fig. 34; $y < 0$ medium, $y > 0$ free space). We shall show that under certain conditions combinations of waves exist in the piezoelectric medium and in the free space that consist entirely of surface waves and in combination satisfy the boundary conditions. These combinations are called Bleustein-Gulyaev waves. The common $\omega$ and $k_x$ of the waves are assumed real and positive, so that the waves in fig. 34 are propagated to the right.

The calculation is given in condensed form in Table II (opposite p. 348). The approach and the result of these calculations are roughly as follows. We put certain restrictions on the piezoelectric medium and on the waves to be considered. An introductory calculation first shows that the boundary conditions can be satisfied under these restrictions by the superposition of four waves — not all surface waves as yet — of given real positive $\omega$ and $k_x$ (see fig. 35, upper diagram). These four waves are: $a$) a surface wave in free space; $b$) a surface wave in the piezoelectric medium; $c$) an incident wave and $d$) a reflected wave of ‘stiffened sound’ (see p. 336), at least, if the given $k_x$ is smaller than the wave vector $k_s$ of ‘stiffened sound’ in the piezoelectric material. The waves $c$ and $d$ have equal but opposite real $k_y$. Denoting the amplitudes of the potentials of these four waves by $\phi_a$, $\phi_b$, $\phi_c$ and $\phi_d$, respectively, then $\phi_a$, $\phi_b$, and $\phi_d$ can be expressed in terms of $\phi_c$ because there are three independent boundary conditions. The problem of the reflection of ‘stiffened sound’ at a free surface is thus solved. However, this was not our problem: we were in search of a surface wave.

As for the refracted wave in the refraction of light (p. 343), we may now enquire what happens to the waves $c$ and $d$ if for any reason $k_s$ should become larger than $k_s$. The wave-vector components $k_y$ of $c$ and $d$ then become imaginary. They remain however of opposite sign and thus we do find a ‘well behaved’ surface wave propagating in the $x$-direction whose amplitude falls off exponentially with distance from the surface (let this be $c$); however, we also find a wave ($d$) that would again propagate in the $x$-direction, but whose amplitude would *increase* exponentially with distance from the surface, and which is therefore unacceptable.

The combination of waves $a$, $b$, $c$ and $d$ necessary to satisfy the boundary conditions therefore no longer forms an acceptable solution.

Nevertheless a freely propagating surface wave can be found, since for any real positive $\omega$ we can find a real, positive $k_x$ that is larger than $k_s$ and makes $\phi_d$ equal to zero. The unacceptable wave thus vanishes from the scene (fig. 35, lower diagram). In other

---

Fig. 34. Coordinate system used for the derivation of the Bleustein-Gulyaev wave; $y = 0$ is the interface between the piezoelectric medium $PE$ ($y < 0$) and free space $FS$ ($y > 0$). The wave propagation is in the $x,y$-plane, the particle movement in the $z$-direction. In this coordinate system the piezoelectric material must have an array of coefficients of the form $M$ in Table II (opposite p. 349).

Fig. 35. Above: a wave of stiffened sound ($c$) incident from the piezoelectric material on the interface gives rise to a reflected wave ($d$) and two surface waves, one in the free space ($a$) and the other in the piezoelectric medium ($b$). The common $k_y$ of the four waves is smaller than the wave vector $k_s$ of stiffened sound in the given piezoelectric material. If $k_s$ becomes larger than $k_s$ then the wave-vector components $k_y$ of ($c$) and ($d$) become imaginary; they remain opposite in sign. Both of these waves thus propagate along the surface, and while the amplitude of one of them decreases, that of the other increases exponentially away from the surface. At a certain value of $k_s$ ($> k_s$), however, the unacceptable wave ($d$), whose amplitude increases away from the surface, vanishes from the solution. The solution is then the Bleustein-Gulyaev wave (lower diagram).

---


[38] A more comprehensive survey of possible surface and interface waves and further literature references are given in the article by R. M. White [3].
words, when we attempt to satisfy the boundary conditions for a given \( \omega \) with the three waves \( a, b \) and \( c \), we find that this is possible with one value of \( k_x \) that is larger than \( k_s \), so that \( c \) is also a surface wave. This combination of three surface waves is the Bleustein-Gulyaev wave, and the relation found between \( k_x \) and \( \omega \) is its dispersion relation.

In our description of the Bleustein-Gulyaev wave we thus allow an incident wave and a reflected wave (real \( k_y \)) to change into two surface waves (imaginary \( k_y \)), one decreasing and the other increasing exponentially in amplitude with distance from the surface. We then cause the unacceptable increasing wave to vanish by choosing a suitable value \( k_x(\omega) \) for \( k_x \). We note that this approach is of general application to surface waves: the Bleustein-Gulyaev wave merely serves here as an example.

We shall now look into one or two details of the calculation. We impose the following restrictions (see R1-R4 in Table II): the piezoelectric material must have a certain symmetry and a certain orientation (R4); the waves must be 'slow' (R2), must propagate in the \( x,y \)-plane (R1) and must involve particle displacement in the piezoelectric medium in the \( z \)-direction only (R3). The waves must be slow in the sense that only electrostatic effects and no electrodynamic effects arise (see also p. 336); this means that the electric field can be derived from a potential \( \phi \) (\( E = -\text{grad} \phi \)).

For the waves indicated in Table II as FSL, which can occur in free space under the restrictions R1 and R2, \( k^2 = 0 \) for all \( \omega \). The wave variables thus satisfy the two-dimensional Laplace equation, for \( k^2 = k_x^2 + k_y^2 = -\Delta = -\nabla^2 \). These are in fact static field distributions which can propagate at an arbitrary (but not too high) velocity, provided that the boundary conditions at the surface are satisfied. For \( k_x \) real, \( k_y \) is evidently imaginary: \( k_y = \pm jk_x \). The wave to be used for combination (wave \( a \) in list \( L \), Table II) must decrease exponentially with distance from the surface (i.e. upwards in fig. 34) so that for the given \( k_x \), the \( k_y \) of the wave must be equal to \( -jk_x \). Lines of force and equipotential surfaces of this Laplace wave are given in fig. 36.

Restrictions on the symmetry and orientation (R4) of the piezoelectric medium are imposed in the array marked \( M \) in Table II, which gives the coefficients by means of which the variables \( T_1, T_2, \ldots D_x, D_y, D_z \) are expressed in terms of the variables \( S_1, S_2, \ldots E_x, E_y, E_z \). The many zeros imply a high symmetry and a simple orientation; the coefficients indicated by points are not relevant to the present discussion. Materials with a six-fold or rotational axis of symmetry are examples in which the array of coefficients can take the form \( M \). The restrictions R1, R2 and R3 have already limited the components of \( S \) and \( E \) to be taken into account here to \( S_4, S_5, E_x \) and \( E_y \) and now because of the form of the array the components of \( T \) and \( D \) are limited to \( T_4, T_5, D_x \) and \( D_y \).

As a result of the foregoing, there only remain two differential equations for the waves in the piezoelectric material: a mechanical equation of motion and one of Maxwell's equations. The resultant dispersion relation represents two waves — again, Laplace waves (PL), with \( k^2 = 0 \) and further 'stiffened sound' (PS), with the dispersion relation (DS). This reduction of the possible waves to two simple types is of course a consequence of the simplifications introduced by (M). The Laplace wave to be used for combination (wave \( b \) in list \( L \), Table II) must decrease with distance from the surface, i.e. downwards in fig. 34, and thus has \( k_y = jk_x \). For the incident sound wave (\( c \) in list \( L \)), \( k_y > 0 \); the reflected wave (\( d \)) has a negative \( k_y \). In the list \( L \) of possible waves, only wave variables are included that enter into the boundary conditions.

The boundary conditions B1, B2, B3 express, firstly, that the boundary \( y = 0 \) is a mechanically free surface. The shear stress \( T_{xy} = T_5 \) and \( T_{yx} = T_4 \) along the surface and the tensile stress \( T_{yy} = T_3 \) normal to it are thus zero. Two of the boundary conditions \( T_2 = 0, T_6 = 0 \) are automatically satisfied in all the waves considered, and therefore only one mechanical boundary condition remains (B1): \( T_4 = 0 \). The electrical boundary conditions (B2, B3) stating that \( \phi \) and \( D_y \) must be continuous at the surface, follow because at the surface the field must be derivable from a potential and there is no source of \( D \), i.e. no charge. We note that before the introduction of the boundary conditions, the factor exp \( j(\omega t - k_x x - k_y y) \) in each wave plays no part — it cancels out in all equations. However, when various waves are combined, different factors exp \( (-jk_y y) \) are involved. At the boundary these factors again vanish since \( y = 0 \) there.

Substitution of the wave variables from list \( L \) in the boundary conditions (B) yields, after eliminating \( \phi_a \) and \( \phi_b \), the relation (A) between \( \phi_e \) and \( \phi_0 \). Therefore we arrive at the two conclusions mentioned earlier:

**Fig. 36.** Lines of force (solid curves) and cross-sections of equipotentials (dashed) of the Laplace wave which decays exponentially for \( y \to + \infty \).
I. A wave of 'stiffened sound' incident at a certain angle imposes a $k_x$ and thus determines indirectly all the $k_y$'s; since $k_x$ is smaller than $k_y$, $k_{y0}$ and $k_{yd}$ turn out to be real. Using $(A)$ in Table II, all the wave variables can be expressed in terms of $\phi_0$. The angle of reflection is equal to the angle of incidence ($k_{yd} = -k_{y0}$). The reflected wave has the same amplitude as the incident wave ($|\phi_d| = |\phi_0|$), because the coefficients of $\phi_0$ and $\phi_d$ in $(A)$ are the complex conjugates of one another.

II. From $(A)$ we can produce a pure surface wave, the Bleustein-Gulyaev wave, because wave $d$ vanishes ($\phi_d = 0$) for the value of $k_{y0}$ given in $(K)$ (Table II); this value of $k_{y0}$ is positive imaginary and thus $c$ becomes a 'well behaved' surface wave. Substituting this value of $k_{y0}$ in the identity $k_x^2 + k_y^2 = k_0^2$ yields expression (DBG) in Table II, the dispersion relation for the Bleustein-Gulyaev wave. The velocity $\omega/k_x$ is independent of $\omega$: the Bleustein-Gulyaev wave is therefore dispersionless.

Notes on the Bleustein-Gulyaev wave

In the foregoing derivation of the Bleustein-Gulyaev wave it was not necessary — as it was in other problems discussed in this article — to assume that $e^2/\varepsilon c$ was much less than unity. This underlines once more the relative simplicity of the Bleustein-Gulyaev wave. We assumed only that the wave was 'slow' and the dispersion relation (DBG) shows that this is always the case except when $e^2/\varepsilon c \gg 1$, a situation that in fact never arises.

From $(K)$ it can be seen that the penetration depth increases as $e^2/\varepsilon c$ decreases and becomes infinite, as mentioned earlier, in the limit $e^2/\varepsilon c \rightarrow 0$; the surface wave then degenerates to a bulk wave. In contrast to this, the penetration depth of a Rayleigh-like wave always remains of the same order as the wavelength along the surface. Bearing in mind that, in practical applications, the great attraction of surface waves lies in their small penetration depth, it can be seen that the Bleustein-Gulyaev wave has its limitations; only when $e^2/\varepsilon c$ is approximately unity can it match the Rayleigh wave in this respect.

The factor $\varepsilon_0/(\varepsilon_0 + \varepsilon)$ in $(K)$ also tends to make $|k_{y0}|$ small and hence the penetration depth large because $\varepsilon$ is often much larger than $\varepsilon_0$ in piezoelectric materials. In the above the empty space behaves only as a medium of permittivity $\varepsilon_0$ with no mechanical effect on the surface. If the empty space is replaced by another medium with a large permittivity but still with no mechanical effect on the surface, then the penetration depth can be reduced. For this reason the surface is sometimes coated with a thin film of metal ($\varepsilon \approx \infty$). Further analysis shows that the wave $a$ then vanishes ($\phi_a = 0$); there is virtually no penetration in the metal so that the metal film can be so thin that its mechanical effect is negligible. Strictly speaking, we are then dealing with a wave problem in a single bounded medium, the piezoelectric material, but with different electrical boundary conditions.

Amplification of surface waves; the effect of a transverse magnetic field

An acoustic surface wave on a piezoelectric material is accompanied by an electric field that extends beyond the surface and propagates there as a Laplace wave. The wave can be excited, detected, directed, amplified and its dispersion relation changed by means of this electric field. Here we shall only examine the amplification of such waves and we shall show that with a transverse magnetic field the amplification can be enhanced.

From the foregoing it is fairly evident how to set about amplifying a surface wave on a piezoelectric material: a semiconductor is placed against the piezoelectric material and a current is passed through it of such a value that the drift velocity of the electrons is higher than the wave velocity. We shall see presently that this should work; that it does work was first shown experimentally by J. H. Collins et al. (9). Compared with the bulk-wave acoustic amplifier (see p. 337) the present configuration has the advantage that the piezoelectric material and the semiconductor can be selected independently, so that an optimum choice can be made. This is one of the reasons why the surface-wave amplifier seems to be a step nearer to practical applications than the bulk-wave amplifier. Nevertheless, there are still great difficulties associated with the surface-wave amplifier. In the first place, the required drift field is very high (this point is illustrated in fig. 27); this means that it is generally necessary to dissipate undesirable large amounts of power unless certain precautions are taken that present practical difficulties.

Secondly there is the problem of the electric coupling between the piezoelectric material and the semiconductor. In a Rayleigh wave, the surface particles move primarily in the $y$-direction (see fig. 33). This wave is therefore very sensitive to mechanical contact with the surface. To avoid this difficulty it can be arranged, for example, to have a gap between the surfaces of the piezoelectric material and the semiconductor. This gap, however, must be very small (a small fraction of the wavelength), for otherwise there will be no electric coupling. The Bleustein-Gulyaev wave, in which there is particle movement in the $z$-direction only, is better in this respect: it is very little affected by the presence of a substance such as a liquid (which attenuates a Rayleigh wave strongly), and a dielectric...
liquid of high permittivity in the gap results in a strong electric coupling. This, however, may give rise to new difficulties, such as corrosion of the surfaces by the liquid.

We shall now indicate briefly how acoustic surface waves are amplified and how this effect can be enhanced by a transverse field \(^{38}\). In fig. 37a, 1 denotes the semiconductor and 2 the piezoelectric material. In the semiconductor the slow wave consists of a Laplace wave (see fig. 36); associated with the field \(E\), there is the bulk current \(J_b = \sigma_0 E\). In a Laplace wave there can be no charge fluctuations in the material since in the Laplace wave div \(D = -\epsilon \nabla \phi = 0\). The current that flows towards and away from the surface gives, however, an alternating surface charge. In fig. 37a it is assumed that the d.c. drift velocity of the electrons is zero and that the Laplace wave, including the surface-charge pattern, travels to the right at velocity \(v_s\). The maxima of positive and negative surface charge must assume, with respect to the field pattern, the phases as shown: on the right (front) of the positive charge maxima, the charge must increase, so the current must be directed towards these points. In fig. 37a the field and the current are in phase: energy is therefore dissipated.

If the electron gas as a whole is now made to move through the semiconductor at a drift velocity \(v_{d0}\), the pattern of alternate surface charges itself represents — quite independently of the alternating field \(E\) — an a.c. surface current \(J_s\) (fig. 37b). A smaller bulk current will now maintain the same surface charge; or the same alternating field \(E\) and bulk current \(J_b\) will now give rise to a surface-charge wave of larger amplitude. When \(v_{d0}\) is equal to \(v_s\) no bulk a.c. current will be required to maintain a surface-charge wave. (Diffusion and trapping of charge carriers are neglected here.) If, finally, \(v_{d0}\) becomes larger than \(v_s\), charge has to be removed from the front of each positive charge maximum by the bulk a.c. current (fig. 37c). Field and current now have opposite phase and hence energy is supplied to the wave.

It will be clear that the operation of the surface-wave amplifier is rather similar to that of the bulk-wave amplifier. In the surface-wave amplifier, however, advantage can be taken of a transverse magnetic field in a way that has no parallel in the case of bulk acoustic waves. We shall attempt to explain this by means of the equation

\[
\frac{\Delta k_1}{k} = -\frac{e^2}{4\pi \epsilon \sigma} \frac{\epsilon}{\epsilon + \sigma_1/\omega} \left( \frac{\epsilon + \sigma_1/\omega}{\epsilon + (\sigma_1/\omega)^2} + \frac{\sigma_1/\omega}{\sigma} \right) (90)
\]

which is simply (78) in a slightly different form. This relation derived for phenomena in the bulk, is not of course exactly valid in this form for surface waves, but a very similar relation is valid and we use (90) to indicate qualitatively the effect of a transverse magnetic field \(^{39}\). The last factor in (90) is of the form \(\rho q (\rho^2 + q^2)\) and it therefore has values between \(+1/2\) and \(-1/2\). The 'best' value, \(-1/2\), can always be achieved, for arbitrary values of \(\epsilon\) and \(\sigma_1\), by giving \(\sigma_1\) via the drift velocity \(v_{d0}\), a suitable (negative) value. From the other factors in (90) it can be seen that the maximum amplification obtainable in this way could be increased if \(\sigma_1\) could be made negative, i.e. if the phase difference between \(J\) and \(E\) could be changed in a certain way. In the bulk-wave amplifier there is no way of doing this. In the surface-wave amplifier, however, it can be done by means of a transverse magnetic field. To make this clear, fig. 37d shows the extreme case of a magnetic field so large that there is a Hall angle of 90° between \(J\) and \(E\). The figure shows that for a travelling Laplace wave (fig. 37a) the operation of the acoustic surface-wave amplifier and the effect of a transverse magnetic field. 1 semiconductor, 2 piezoelectric material. In the semiconductor, the field \(E\) of the Laplace wave (velocity \(v_s\)) carries with it a bulk a.c. current \(J_b\) and an alternating surface charge. For a given alternating field, the amplitude and phase of the surface-charge wave are dependent on the mean drift velocity \(v_{d0}\) of the electrons in the semiconductor (see text). a) \(v_{d0} = 0\); b) \(0 < v_{d0} < v_s\); c) \(v_{d0} > v_s\). In case (c) amplification occurs. \(J_b\) is the surface a.c. current directly associated with the movement of the surface charges with the electron gas as a whole at velocity \(v_{d0}\). d) When a large magnetic field is present \(E\) and \(J_b\) are at right angles to one another which means, for the travelling Laplace wave, a phase difference of 90° between \(J\) and \(E\). Such a phase difference can lead to greater amplification.

Fig. 37. Operation of the acoustic surface-wave amplifier and the effect of a transverse magnetic field. 1 semiconductor, 2 piezoelectric material. In the semiconductor, the field \(E\) of the Laplace wave (velocity \(v_s\)) carries with it a bulk a.c. current \(J_b\) and an alternating surface charge. For a given alternating field, the amplitude and phase of the surface-charge wave are dependent on the mean drift velocity \(v_{d0}\) of the electrons in the semiconductor (see text). a) \(v_{d0} = 0\); b) \(0 < v_{d0} < v_s\); c) \(v_{d0} > v_s\). In case (c) amplification occurs. \(J_b\) is the surface a.c. current directly associated with the movement of the surface charges with the electron gas as a whole at velocity \(v_{d0}\).
wave this corresponds to a phase difference of 90° between \( J \) and \( E \). It is also clear that smaller phase differences are introduced by smaller fields and that the sign of the phase difference is determined by the polarity of the magnetic field.

The effect of a transverse field on the amplification has been experimentally confirmed for Rayleigh waves by J. Wolter \(^{[40]}\). Some particulars concerning this experiment are given in fig. 38.

**Summary.** A survey is given of freely propagating electromagnetic, elastic and electro-elastic waves, the accent falling on certain types of wave that have attracted attention in electronics and solid-state physics during the last decade or so. These include helicon waves, amplifying acoustic waves and electromagnetics and solid-state physics during the last decade or so. These include helicon waves, amplifying acoustic waves and electro-

**Fig. 38.** Effect of a transverse magnetic field on the amplification and attenuation of Rayleigh waves, \( a \) experimental \( b \) theoretical, after J. Wolter \(^{[40]}\). The amplification in dB is plotted in \( a \) as a function of the applied drift field \( E_d \) for various values of the transverse magnetic flux density \( B \); in \( b \) the amplification is plotted as a function of the drift velocity \( v_{ac} \) for various values of \( \mu B \) (\( u = \) mobility of the electrons in the semiconductor). The Rayleigh waves have a frequency of 50 MHz and are propagated on the surface of a plate of the piezoelectric material \( \text{LiNbO}_3 \). They are amplified or attenuated by the electrons moving in a silicon plate (type \( N \), 150 \( \Omega \)cm, 2 mm long\( \times \)200 \( \mu \)m thick) on that surface. The values of \( \mu B \) and \( v_{ac} \) in \( b \) are matched (i.e. chosen so that \( b \) fits \( a \) as well as possible). The silicon was pressed against the \( \text{LiNbO}_3 \) so that the two media were in intimate mechanical contact at one or two places (the Rayleigh wave is therefore somewhat attenuated). It was established optically that the gap was less than 0.1 \( \mu \)m over a length of about 0.4 mm. Only over this length was the silicon effective. The calculated amplification and attenuation were in reasonable agreement with experiment.

In a freely propagating surface wave, each component wave is a surface wave. As an example, the Bleustein-Gulyaev wave, known since 1968, is described. This is a surface wave by virtue of the piezoelectric property of the medium. This wave is closely related to the reflection of 'stiffened sound' at the surface of a piezoelectric medium. Acoustic surface waves on a piezoelectric medium can be amplified by a drift current in an adjacent semiconductor. A transverse magnetic field can enhance this amplification in an interesting way.

---

\(^{[38]}\) A fuller discussion of these and related subjects and further references are given in: C. A. A. J. Greebe, P. A. van Dalen, T. J. B. Swansenburg and J. Wolter, Electric coupling properties of acoustic and electric surface waves, Physics Reports 1C, 235-268, 1971.

\(^{[39]}\) In many problems the description can be made in terms of either a complex \( \sigma \) or a complex \( \varepsilon \) (see for example eq. 33). Here we use a complex \( \sigma \). In the article of note \(^{[38]}\) a complex \( \varepsilon \) was used.

Recent scientific publications

These publications are contributed by staff of laboratories and plants which form part of or co-operate with enterprises of the Philips group of companies, particularly by staff of the following research laboratories:

- Philips Research Laboratories, Eindhoven, Netherlands
- Mullard Research Laboratories, Redhill (Surrey), England
- Laboratoires d'Electronique et de Physique Appliquée, 3 avenue Descartes, 94450 Limeil-Brévannes, France
- Philips Forschungs laboratorium Aachen GmbH, Weißhausstraße, 51 Aachen, Germany
- Philips Forschungs laboratorium Hamburg GmbH, Vogt-Kölln-Straße 30, 2000 Hamburg 54, Germany
- MBE Laboratoire de Recherches, 2 avenue Van Becelaere, 1170 Brussels (Boisfort), Belgium
- Philips Laboratories, 345 Scarborough Road, Briarcliff Manor, N.Y. 10510, U.S.A. (by contract with the North American Philips Corp.)

Reprints of most of these publications will be available in the near future. Requests for reprints should be addressed to the respective laboratories (see the code letter) or to Philips Research Laboratories, Eindhoven, Netherlands.

**V. Belevitch:** On the realizability of non-rational positive real functions.

**H. J. van den Berg & A. J. Luitingh:** Reproducibility and irreproducibility of etching time in freeze-etch experiments.
Cytobiologie **7**, 101-104, 1973 (No. 1).

**J. Bloem** (Philips Semiconductor Development Laboratory, Nijmegen): High chemical vapour deposition rates of epitaxial silicon layers.
J. Crystal Growth **18**, 70-76, 1973 (No. 1).

**P. W. J. M. Boumans:** Spektralanalysen. Optische Atomspektroskopie.

**K. H. J. Buschow, A. M. van Diepen & H. W. de Wijn** (Staat University of Utrecht): Evidence for RKKY-type interaction in intermetallics, as derived from magnetic dilution of GdPdIn with Y or Th.

**M. C. W. van Buul & L. J. van de Polder:** Standards conversion of a TV signal with 625 lines into a video- phone signal with 313 lines.

**K. L. Bye, P. W. Whippes & E. T. Keve:** High internal bias fields in TGS (l-alanine).
Ferroelectrics **4**, 253-256, 1972/73 (No. 4).

**F. M. A. Carpay:** Theory of and experiments on aligned lamellar eutectoid transformation.

**F. M. A. Carpay:** Aligned composite materials obtained by solid state decomposition.

**V. Chalmeton:** Neutron radiography.
Acta Electronica **16**, 73-84, 1973 (No. 1). *(Also in French.)*

**J. W. Chamberlayne & B. Gibson:** Magnetic materials for integrated cores.


**J. A. Clarke & E. C. Yeadon** (Mullard Ltd, Mitcham, Surrey): Measurement of the modulation transfer function of channel image intensifiers.
Acta Electronica **16**, 33-41, 1973 (No. 1). *(Also in French.)*

**G. Clément & C. Loty:** The use of channel plate electron multipliers in cathode-ray tubes.
Acta Electronica **16**, 101-111, 1973 (No. 1). *(Also in French.)*

**J. Cornet & D. Rossier:** Phase diagram and out-of-equilibrium properties of melts in the As-Te system.

**C. Z. van Doorn:** On the magnetic threshold for the alignment of a twisted nematic crystal.

**H. Dormont:** Modèle théorique de système à avalanche conduisant à une étude du bruit.

**D. L. Emberson** (Mullard Ltd, Mitcham, Surrey) & R. T. Holmshaw: The design and performance of an inverting channel image intensifier.
Acta Electronica **16**, 23-32, 1973 (No. 1). *(Also in French.)*
Table II. Derivation of the Bleustein-Gulyaev wave

R1. Restriction to waves propagating in the x,y-plane: \( \mathbf{\hat{z}} = 0 \)

R2. Restriction to slow waves: \( E = -\text{grad} \phi \to \begin{cases} E_x = jk_x \phi \\ E_y = jk_y \phi \\ E_z = 0 \end{cases} \)

<table>
<thead>
<tr>
<th>Waves in free space</th>
<th>Waves in piezoelectric medium</th>
</tr>
</thead>
</table>
| \( D = \varepsilon_0 E \) | \begin{align*}
    R3. & \text{Restriction to displacement in the z-direction:} \\
    \begin{cases}
        u_x = 0 \\
        u_y = 0 \\
        u_z = 0 \\
    \end{cases} \\
    \text{other } S \text{ are zero}
\end{align*} |
| Maxwell’s equation: \( \text{div} \ D = 0 \to k^2 \phi = 0 \) | \begin{align*}
    R4. & \text{Piezoelectric coefficients:} \\
    \begin{array}{cccccccc}
    S_1 & S_2 & S_3 & S_4 & S_5 & S_6 & S_7 & S_8 \\
    T_1 & & & & & & & \\
    T_2 & & & & & & & \\
    T_3 & & & & & & & \\
    T_4 & & & & & & & \\
    T_5 & & & & & & & \\
    T_6 & & & & & & & \\
    T_7 & & & & & & & \\
    T_8 & & & & & & & \\
    \end{array} \\
\end{align*} |
| Dispersion relation \( k^2 = 0 \) | \begin{align*}
    \text{Laplace waves} & \quad \begin{cases}
        k_x = 0 \\
        D_x = jk_x \phi_x \\
        D_y = jk_y \phi_y \\
        D_z = 0 \\
    \end{cases} \\
    \text{Dyson equation:} & \quad \begin{cases}
        k_x = k_x^0, \quad \text{with } \frac{\omega^2}{k_x^2} = \left( 1 + \frac{\varepsilon_0}{\varepsilon} \right) \frac{c}{\delta} \\
        u_x = \frac{e\phi}{\varepsilon} \\
        T_x = -jk_x(\varepsilon_0 + e\phi) \\
        T_y = -jk_y(1 + \frac{\varepsilon_0}{\varepsilon}) \phi \\
        T_z = -jk_x(1 + \frac{\varepsilon_0}{\varepsilon}) \phi \\
        D_x = jk_x \phi_x \\
        D_y = jk_y \phi_y \\
        D_z = 0 \\
    \end{cases} \\
    \text{Mech. eq. of motion:} & \quad \begin{cases}
        e_x \partial_t^2 \phi_x = \partial_x T_x + \partial_y T_y \\
        e_x \partial_t^2 \phi_y = \partial_x T_x + \partial_y T_y \\
        \text{div} \ D = 0 \to k^2 \phi_x = 0 \\
        \text{Dispersion relation } k^2 \left[ (\varepsilon + e^2)k^2 - \varepsilon \omega^2 \right] = 0 \\
    \end{cases} \\
\end{align*} |

\[ (FSL) \quad (PL) \quad (PS) \]

Waves with real positive \( \omega \) and \( k_x \) to be used for superposition

a) The Laplace wave in free space which decays exponentially for \( y \to +\infty \):
\[ k_y = -jk_x \]
\[ \phi_a = \phi_0 \]
\[ T_{ab} = k_x \phi_0 \]
\[ D_{ab} = k_x \phi_0 \]

b) The Laplace wave in the piezoelectric medium which decays exponentially for \( y \to -\infty \):
\[ k_y = +jk_x \]
\[ \phi_b = \phi_0 \]
\[ T_{tb} = k_x \phi_0 \]
\[ D_{tb} = -k_x \phi_0 \]

\[ \text{Boundary conditions} \]
\[ T_{ab} + T_{ac} + T_{ad} = 0 \]
\[ D_{ab} = D_{tb} \]
\[ \phi_a + \phi_c + \phi_d = \phi_0 \]

\[ \left[ k_x + jk_x \left( 1 + \frac{\varepsilon}{\varepsilon_0} \right) \frac{c}{\varepsilon_0} + jk_x \right] \phi_c + \left[ k_x - jk_x \left( 1 + \frac{\varepsilon}{\varepsilon_0} \right) \frac{c}{\varepsilon_0} + jk_x \right] \phi_d = 0 \]
\[ (A) \]

\[ \text{Solutions} \]

I. Reflection of ‘stiffened sound’ at surface:
\[ k_x^2 < k_y^2 \to k_y \text{ real.} \]
\[ \phi_a, \phi_c, \phi_d \text{ to be calculated from } \phi_c. \]
\[ \left| \phi_c \right| = \left| \phi_a \right| \]
\[ k_y = -k_{yc} \]

II. Bleustein-Gulyaev wave
\[ k_{yc} = \frac{e_0 \varepsilon}{e_0 + \varepsilon + 1 + e^2/\varepsilon} \]
\[ \phi_a = 0 \]
\[ k_x^2 + k_y^2 = k_{yc}^2 \]
\[ \frac{\omega^2}{k_{yc}^2} = \left( 1 + \frac{\varepsilon_0}{\varepsilon_0 + e + 1 + e^2/\varepsilon} \right) \]
\[ \left( K \right) \quad \left( DBG \right) \]


J. R. Mansell: A study of an experimental TV pick-up tube incorporating a channel plate. Acta Electronica 16, 113-122, 1973 (No. 1). (Also in French.)


M. J. Sparnaay: Surface analysis with the aid of low-energy electrons. The case of semiconductor surfaces compared with other methods. J. radioanal. Chem. 12, 101-114, 1972 (No. 1).


