Abstract

The purpose of this note is to introduce the linear electro-optic effect (Pockels effect) and its applications for modulating the properties of a light wave, such as its phase, polarization and intensity. For this purpose, basic electromagnetic theory is first reviewed. Some elements of light propagation in an anisotropic medium are then introduced. The linear electro-optic effect is then presented and illustrated using the classic example of KDP crystals. Due to its importance for modulation applications in optical communication systems, the use of lithium niobate (LiNbO$_3$) crystals is also discussed in details. Typical electro-optic modulation configurations are introduced. Finally, transit time limitations and their implications for high speed modulation are also discussed.
1 Elements of electromagnetic theory

1.1 Maxwell’s equations

The electromagnetic field consists of coupled electric and magnetic fields that are described by the vectors \( \mathbf{E} \) and \( \mathbf{B} \), known as electric and magnetic induction vectors, respectively. In order to describe the effects of these two fundamental fields on matter, it is necessary to introduce the electric displacement and magnetic vectors, denoted by \( \mathbf{D} \) and \( \mathbf{H} \), respectively\(^1\). If the electric current density \( \mathbf{J} \) is also introduced, the four fields \( \mathbf{E}, \mathbf{B}, \mathbf{D} \) and \( \mathbf{H} \) are linked by Maxwell’s equations\(^2\)

\[
\nabla \times \mathbf{E}(\mathbf{r}, t) = -\frac{\partial \mathbf{B}}{\partial t}(\mathbf{r}, t), \tag{1}
\]
\[
\nabla \times \mathbf{H}(\mathbf{r}, t) = \mathbf{J}(\mathbf{r}, t) + \frac{\partial \mathbf{D}}{\partial t}(\mathbf{r}, t), \tag{2}
\]
\[
\nabla \cdot \mathbf{D}(\mathbf{r}, t) = \rho(\mathbf{r}, t), \tag{3}
\]
\[
\nabla \cdot \mathbf{H}(\mathbf{r}, t) = 0, \tag{4}
\]

where \( \rho \) is the electric charge density and where the \( \nabla \times \) and \( \nabla \cdot \) notations have been used for the vector operators curl and div, respectively. Equations (1), (2) and (3) are known as

\(^1\)Many texts introduce \( \mathbf{E} \) and \( \mathbf{H} \) as the fundamental fields, and \( \mathbf{D} \) and \( \mathbf{B} \) to describe their effect on matter. This distinction is unessential for the purpose of the present note. For a justification of this choice, one may refer to e.g. Born & Wolf [1]
\(^2\)James Clerk Maxwell, Edinburgh 1831 – Cambridge 1879.
Anisotropic Media and the Linear Electro-Optic Effect

The electric displacement can be written according to

\[ \mathbf{D}(\mathbf{r}, t) = \varepsilon_0 \mathbf{E}(\mathbf{r}, t) + \mathbf{P}(\mathbf{r}, t), \]

where \( \mathbf{P} \) is the electric polarisation, while the magnetic field is defined as

\[ \mathbf{H}(\mathbf{r}, t) = \frac{1}{\mu_0} \mathbf{B}(\mathbf{r}, t) - \mathbf{M}(\mathbf{r}, t), \]

where \( \mathbf{M} \) is the magnetisation. The constants \( \varepsilon_0 \) and \( \mu_0 \) are known as the vacuum permittivity and vacuum permeability, respectively. In order to solve Maxwell’s equations, it is necessary to relate \( \mathbf{D} \) and \( \mathbf{H} \) to \( \mathbf{E} \) and \( \mathbf{B} \). In a conducting medium, the current density will also need to be related to the fundamental fields. Those relations can be expressed under the most general form as

\[ \mathbf{D} = \mathbf{D}[\mathbf{E}, \mathbf{B}], \quad \mathbf{H} = \mathbf{H}[\mathbf{E}, \mathbf{B}], \quad \mathbf{J} = \mathbf{J}[\mathbf{E}, \mathbf{B}]. \]

This set of equations are known as constitutive relations since they depend on the material properties. Assumptions on the material are therefore necessary in order to proceed further in the resolution of Maxwell’s equations. In the following section, it will be shown step-by-step how the knowledge of the material properties can be used to derive useful relations between the electric displacement and the electric field vectors.

1.2 The constitutive relations

In this section, assumptions on the material properties are progressively introduced in order to derive explicit constitutive relations between the electric displacement and the electric field for the types of media pertaining to this note. This is achieved by considering properties of the electric polarisation \( \mathbf{P} \).

First, it is assumed that \( \mathbf{P} \) does not depend on an applied magnetic field. Dependence on \( \mathbf{B} \) would lead to magneto-optic effects such as the Faraday effect, magneto-optic Kerr effect and Cotton-Mouton (or Voigt) effect. The polarisation then expresses the response of the material to an applied electric field.

Second, the material is assumed to be linear, which means that the polarisation \( \mathbf{P} \) has a linear dependence on the electric displacement \( \mathbf{E} \). Assuming that \( \mathbf{P} \) can be expanded as a power series of the electric field would lead to the field of nonlinear optics (see for instance [2] for the derivation of constitutive relations that apply when nonlinear terms can no longer be neglected). Using the formalism of linear systems, each component of the polarisation vector on an orthonormal basis can be written

\[ P_i(\mathbf{r}, t) = \varepsilon_0 \int_{-\infty}^{+\infty} \int F_{ij}(\mathbf{r}, \mathbf{r}', t, t') E_j(\mathbf{r}', t') \, dt' \, d\mathbf{r}', \]

where \( F_{ij} \) is the response function of the material and the usual convention according to which repetition of the indices implies summation is used.

If furthermore the relation between \( \mathbf{P} \) and \( \mathbf{E} \) is local, i.e. \( \mathbf{P} \) at the position \( \mathbf{r} \) only depends on \( \mathbf{E} \) at the same position, then the material does not exhibit spatial dispersion and

\[ P_i(\mathbf{r}, t) = \varepsilon_0 \int_{-\infty}^{+\infty} G_{ij}(\mathbf{r}, t, t') E_j(\mathbf{r}, t') \, dt'. \]

\[^3\text{Michael Faraday, Newington 1791 – Hampton Court 1867.}\]
\[^4\text{André-Marie Ampère, Lyon 1775 – Marseille 1836.}\]
\[^5\text{Carl Friedrich Gauss, Braunschweig 1777 – Göttingen 1855.}\]
Spatial dispersion is encountered for instance in materials that are said to be optically active. One further assumption is that the material is homogeneous, so that the relation between \( \mathbf{P} \) and \( \mathbf{E} \) does not depend on the position, hence

\[
P_i (\mathbf{r}, t) = \epsilon_0 \int_{-\infty}^{+\infty} T_{ij} (t, t') \mathbf{E}_j (\mathbf{r}, t') \, dt'.
\]  

(10)

Finally, if time invariance is assumed,

\[
P_i (\mathbf{r}, t) = \epsilon_0 \int_{-\infty}^{+\infty} R_{ij} (t - t') \mathbf{E}_j (\mathbf{r}, t') \, dt' = \epsilon_0 \int_{0}^{+\infty} R_{ij} (\tau) \mathbf{E}_j (\mathbf{r}, t - \tau) \, d\tau,
\]  

(11)

where \( R_{ij} (t - t') = T_{ij} (t, t') = T_{ij} (t - t', 0) \) and the assumption of causality, according to which the effect can not be anterior to the cause, has been expressed by limiting the integration range in the second equality.

The convolution relation between \( \mathbf{E} \) and \( \mathbf{P} \) in the time domain can be expressed in a simpler form as a product in the frequency domain

\[
\tilde{P}_i (\mathbf{r}, \omega) = \epsilon_0 \chi_{ij} (\omega) \tilde{E}_j (\mathbf{r}, \omega),
\]  

(12)

where \( \tilde{P}_i (\mathbf{r}, \omega) \) and \( \tilde{E}_j (\mathbf{r}, \omega) \) are the Fourier transforms of \( P_i (\mathbf{r}, t) \) and \( P_i (\mathbf{r}, t) \) with respect to the time variable, and the linear susceptibility \( \chi_{ij} (\omega) \) is the Fourier transform of the response function \( R_{ij} (\tau) \). Here, the Fourier and inverse Fourier transforms between the time and angular frequency domains are defined according to

\[
\tilde{E}_i (\mathbf{r}, \omega) = \int_{-\infty}^{+\infty} E_i (\mathbf{r}, t) e^{-j\omega t} \, dt \quad E_i (\mathbf{r}, t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \tilde{E}_i (\mathbf{r}, \omega) e^{j\omega t} \, d\omega.
\]  

(13)

Finally, (5) can be used to obtain the constitutive relation between \( \mathbf{D} \) and \( \mathbf{E} \)

\[
\tilde{D}_i (\mathbf{r}, \omega) = \epsilon_0 (1 + \chi_{ij} (\omega)) \tilde{E}_j (\mathbf{r}, \omega),
\]  

(14)

which can also be expressed as

\[
\tilde{D}_i (\mathbf{r}, \omega) = \epsilon_{ij} (\omega) \tilde{E}_j (\mathbf{r}, \omega),
\]  

(15)

where the elements \( \epsilon_{ij} \) of the permittivity tensor [\( \epsilon \)] are related to the elements \( \chi_{ij} \) of the linear susceptibility tensor [\( \chi \)] according to

\[
\epsilon_{ij} (\omega) = \epsilon_0 (1 + \chi_{ij} (\omega)).
\]  

(16)

It is essential at this point to keep in mind the assumptions that have led to (15): the medium has been assumed to be linear, homogeneous and spatially non-dispersive. Note however that (15) applies to cases where the medium is dispersive and anisotropic. Its dispersive nature can be seen from the frequency dependence of the permittivity. This stems from the fact that the response of the medium is not instantaneous: \( \mathbf{P} (\mathbf{r}, t) \) depends on the electric displacement vector \( \mathbf{E} (\mathbf{r}, t') \) at times \( t' \) earlier than \( t \). The tensorial nature of (15) is a consequence of the anisotropy of the medium: \( \mathbf{P} \) depends on the direction of \( \mathbf{E} \).

If it is assumed that the medium is isotropic, then (15) can be replaced by the scalar relation

\[
\tilde{D} (\mathbf{r}, \omega) = \epsilon (\omega) \tilde{E} (\mathbf{r}, \omega) = \epsilon_0 \epsilon_r (\omega) \tilde{E} (\mathbf{r}, \omega),
\]  

(17)

where \( \epsilon_r \) is known as the relative permittivity of the material.
Finally, in a non dispersive medium, the permittivity does not depend on frequency and relations of the type (15) and (17) also hold in the time domain, depending on whether the medium is anisotropic or isotropic, respectively. For an isotropic and non dispersive medium, for instance

$$D(r, t) = \varepsilon E(r, t).$$  \hfill (18)

Furthermore, when concerned with harmonic waves, which will be the case for most of the following analysis, the frequency dependence of the material is of no importance since the wave is single frequency and the time domain relation (18) also holds at any given frequency.

At this point, a constitutive relation has been found between the electric displacement $D$ and the electric field $E$. The magnetic properties of the medium would need to be explored in order to derive a similar relation between the magnetic field $H$ and the magnetic induction $B$. Here, non magnetic materials for which $M = 0$ will be considered, hence

$$B(r, t) = \mu_0 H(r, t).$$  \hfill (19)

Otherwise, for magnetic, linear, non dispersive and isotropic materials, a more general relation of the type

$$B(r, t) = \mu_0 \mu_r H(r, t) = \mu H(r, t),$$  \hfill (20)

where $\mu_r$ is the relative permeability, could be substituted to (19).

In a conductor, the current density $J$ can be related to the electric field according to

$$J(r, t) = \sigma E(r, t),$$  \hfill (21)

which is a microscopic form of Ohm’s law. The proportionality factor $\sigma$ is known as the conductivity of the material. Note that (21) assumes the medium to be homogeneous and isotropic, otherwise a dependence on the position $r$ and a tensorial relation should be introduced in a similar fashion to what has been done in the earlier study of the constitutive relation between $E$ and $D$. In this note, the materials will be considered non-conducting, hence $J = 0$.

1.3 Electromagnetic energy

Using a classic vector identity\(^6\), the dot product between the electric field $E$ and $\nabla \times H$ can be calculated according to

$$E \cdot [\nabla \times H] = H \cdot [\nabla \times E] - \nabla \cdot [E \times H].$$  \hfill (22)

Introducing Maxwell’s equations (1) and (2) leads to

$$E \cdot \frac{\partial D}{\partial t} + H \cdot \frac{\partial B}{\partial t} + \nabla \cdot [E \times H] = -J \cdot E,$$  \hfill (23)

which can be further integrated over a volume ($V$) delimited by a closed surface ($\Sigma$). Using the divergence theorem (also known as Gauss’s theorem), which states that

$$\int_V (\nabla \cdot S) \ dV = \oint_{\Sigma} S \cdot d\Sigma,$$  \hfill (24)

the following relation is obtained

$$\int_V \left[ E \cdot \frac{\partial D}{\partial t} + H \cdot \frac{\partial B}{\partial t} \right] \ dV = - \int_V J \cdot E \ dV - \oint_{\Sigma} (E \times H) \cdot d\Sigma.$$  \hfill (25)

\(^6\nabla \cdot (A \times B) = B \cdot (\nabla \times A) - A \cdot (\nabla \times B)\)
Eq. (25) is interpreted as a conservation equation, where the left hand side term corresponds to the rate of change of the electromagnetic energy stored in the volume \( V \) and the first term to the right corresponds to the power lost from the electromagnetic field to the charges. In the case of a conductor where (21) applies, this resistive power dissipation due to conduction current is known as Joule\(^7\) effect.

The second term to the right then corresponds to the power flowing into the volume delimited by \( \Sigma \). It is equal to the flux over the closed surface \( \Sigma \) of a vector

\[
S = \mathbf{E} \times \mathbf{H}
\]

know as \textit{Poynting’s vector}\(^8\). The dimension of the magnitude of Poynting’s vector is therefore that of power per unit area. It is commonly interpreted as the electromagnetic power density carried by the field. However it should be pointed out that the definition of \( S \) is somehow arbitrary. Since \( \nabla \cdot [\nabla \times \mathbf{Q}] = 0 \), where \( \mathbf{Q} \) is an arbitrary vector field, any \( \mathbf{S} + \nabla \times \mathbf{Q} \) would lead to the same conservation equation (25).

The rate of change of electromagnetic energy density is thus

\[
\frac{\partial w}{\partial t} = \mathbf{E} \cdot \frac{\partial \mathbf{D}}{\partial t} + \mathbf{H} \cdot \frac{\partial \mathbf{B}}{\partial t},
\]

and the local form of Poynting’s theorem can be expressed as

\[
\frac{\partial w}{\partial t} + \nabla \cdot \mathbf{S} = -\mathbf{J} \cdot \mathbf{E}.
\]

The definition of the rate of change of electromagnetic energy density according to (27) is in line with thermodynamic arguments, according to which the variation of the internal energy density (per unit volume) of a dielectric is equal to

\[
dU = \mathbf{E} \cdot d\mathbf{D}.
\]

Similarly, under the influence of a magnetic field

\[
dU = \mathbf{H} \cdot d\mathbf{B}.
\]

Under some assumptions, it becomes possible to express the total electromagnetic energy density \( w \), as defined in (27), under a more usual form. If the material is assumed to be linear, non dispersive and isotropic, equations (18) and (20) can be used and

\[
\mathbf{E} \cdot \frac{\partial \mathbf{D}}{\partial t} = \frac{\partial \mathbf{E}}{\partial t} \cdot \mathbf{D} = \frac{\partial}{\partial t} \left( \frac{1}{2} \varepsilon \mathbf{E}^2 \right) = \frac{\partial}{\partial t} \left( \frac{1}{2} \mathbf{E} \cdot \mathbf{D} \right),
\]

provided the dielectric constant \( \varepsilon \) does not change with time. Similarly

\[
\mathbf{H} \cdot \frac{\partial \mathbf{B}}{\partial t} = \frac{\partial \mathbf{H}}{\partial t} \cdot \mathbf{B} = \frac{\partial}{\partial t} \left( \frac{1}{2} \frac{\mathbf{B}^2}{\mu} \right) = \frac{\partial}{\partial t} \left( \frac{1}{2} \mathbf{H} \cdot \mathbf{B} \right),
\]

where it is possible to identify more widespread expressions for the electric and magnetic energy densities, denoted as \( w_e \) and \( w_m \), respectively:

\[
w_e = \frac{1}{2} \mathbf{E} \cdot \mathbf{D}, \quad w_m = \frac{1}{2} \mathbf{B} \cdot \mathbf{H}.
\]

\(^7\)James Prescott Joule, Salford 1818 – Sale 1889.

\(^8\)John Henry Poynting, Monton 1852 – Birmingham 1914.
The electromagnetic energy density can therefore be written
\[ w = \frac{1}{2} [\mathbf{E} \cdot \mathbf{D} + \mathbf{B} \cdot \mathbf{H}], \]  
which is in total agreement with (27). It is however essential to keep in mind the assumptions that have led to (34), namely that the medium is isotropic and non dispersive. The case of dispersive media will be briefly mentioned in Sect. 1.4, while it will be seen in Sect. 2 how the results above can be extended to anisotropic media.

1.4 Propagation of electromagnetic waves in an isotropic medium

In this section, a linear, homogeneous and isotropic medium is considered and conditions so that harmonic plane waves of the type
\[ \mathbf{E}(\mathbf{r}, t) = \text{Re} \left\{ \mathbf{E}_0 e^{j(\omega t - \mathbf{k} \cdot \mathbf{r})} \right\}, \]  
are solutions of Maxwell’s equations are examined\(^9\). In (35), \( \mathbf{E}_0 \) is a complex vector and \( \text{Re} \) denotes the real part. For fields that satisfy equations of the type (35), the following substitutions can be made
\[ \nabla \rightarrow -j\mathbf{k}, \quad \frac{\partial}{\partial t} \rightarrow j\omega, \]  
and, in the absence of charges and current sources, Maxwell’s equations (1)-(4) can be written
\[ \mathbf{k} \times \mathbf{E}_0 = \omega \mathbf{B}_0, \]  
\[ \mathbf{k} \times \mathbf{H}_0 = -\omega \mathbf{D}_0, \]  
\[ \mathbf{k} \cdot \mathbf{D}_0 = 0, \]  
\[ \mathbf{k} \cdot \mathbf{H}_0 = 0. \]  
For an harmonic wave, \( \mathbf{D}_0 = \epsilon \mathbf{E}_0 \) and \( \mathbf{B}_0 = \mu \mathbf{H}_0 \). Using a well known vector identity\(^{10}\), the following relation, known as Helmholtz\(^{11}\) equation, can be derived for \( \mathbf{E}_0 \)
\[ \nabla^2 \mathbf{E}_0 + \omega^2 \mu \epsilon \mathbf{E}_0 = \mathbf{0}. \]  
An harmonic plane wave described by (35) is therefore solution of the propagation equation provided
\[ |\mathbf{k}| = \omega \sqrt{\mu \epsilon}. \]  
The surfaces of constant phase of the wave are defined by
\[ \omega t - \mathbf{k} \cdot \mathbf{r} = \text{constant}. \]

\(^9\)Note that it is somehow implicitly assumed here that harmonic plane waves of the type (35) could be solutions of Maxwell’s equations. Many texts derive the well-known propagation equation \( \nabla^2 \mathbf{E} - \mu_0 \epsilon_0 \frac{\partial^2 \mathbf{E}}{\partial t^2} = \mathbf{0} \) directly from Maxwell’s equations for a linear, homogeneous and isotropic medium in the absence of free charges and free currents, which is obviously satisfied by harmonic plane waves. However, in order to do so, the relation \( \mathbf{D}(\mathbf{r}, t) = \epsilon \mathbf{E}(\mathbf{r}, t) \) needs to be employed. It has been seen in Sect. 1.2 that this relation is correct only for a non-dispersive medium, or for an harmonic wave. Since no assumption is usually made on the dispersive nature of the medium at this stage, the harmonic nature of the solution of the propagation equation is also implicitly assumed in those derivations.

\(^{10}\)\( \nabla \times (\nabla \times \mathbf{A}) = \nabla (\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A} \), where \( \nabla^2 \) is the Laplacian operator.

\(^{11}\)Hermann von Helmholtz, Potsdam 1821 – Charlottenburg 1894.
and are therefore planes orthogonal to the wave vector \( \mathbf{k} \) that defines the direction of propagation. These surfaces travel at the velocity

\[ v_{\varphi} = \frac{\omega}{|\mathbf{k}|} \tag{44} \]

in the direction of \( \mathbf{k} \), which is known as the phase velocity of the wave. From (42), the phase velocity can be expressed as a function of the permittivity and permeability of the medium according to

\[ v_{\varphi} = \frac{1}{\sqrt{\varepsilon \mu}} \tag{45} \]

In vacuum, \( \varepsilon = \varepsilon_0, \mu = \mu_0 \) and the phase velocity can be expressed as

\[ c = \frac{1}{\sqrt{\mu_0 \varepsilon_0}} \tag{46} \]

which takes the well known value \( c = 299792458 \, \text{m} \cdot \text{s}^{-1} \). In a non-magnetic medium \( \mu = \mu_0 \) and \( \varepsilon = \varepsilon_0 \varepsilon_r = \varepsilon_0 n^2 \), where \( n = \sqrt{\varepsilon_r} \) is the refractive index of the medium, hence

\[ v_{\varphi} = \frac{c}{n} \tag{47} \]

Equations (39) and (40) show that both \( \mathbf{D} \) and \( \mathbf{H} \) are orthogonal to the direction of propagation \( \mathbf{k} \). Since in an isotropic medium such as the one considered here \( \mathbf{D} \) and \( \mathbf{E} \) are proportional, \( \mathbf{E} \) is also perpendicular to the direction of propagation. Such a wave where \( \mathbf{E} \) and \( \mathbf{H} \) lie in a plane orthogonal to the direction of propagation is said to be transverse. From (37), and since \( \mathbf{B}_0 = \mu \mathbf{H}_0 \), the vector triplet \((\mathbf{k}, \mathbf{E}_0, \mathbf{H}_0)\) is a right handed set of orthogonal vectors whose directions are identical to those of the orthogonal set \((\mathbf{k}, \mathbf{D}_0, \mathbf{B}_0)\). From (37) and (42) it can be easily established that the magnitudes of the vectors \( \mathbf{E}_0 \) and \( \mathbf{H}_0 \) satisfy the relation

\[ \frac{|\mathbf{E}_0|}{|\mathbf{H}_0|} = \sqrt{\frac{\mu}{\varepsilon}} \tag{48} \]

where the quantity \( \eta = \sqrt{\mu/\varepsilon} \) is known as the impedance of the medium. The impedance of vacuum is therefore equal to \( \eta_0 = \sqrt{\mu_0/\varepsilon_0} \approx 377 \, \Omega \).

From (31), (32) and (34), the electromagnetic energy density of the wave is given by

\[ w = \frac{1}{2} \left( \varepsilon \mathbf{E}^2 (\mathbf{r}, t) + \frac{1}{\mu} \mathbf{B}^2 (\mathbf{r}, t) \right) \tag{49} \]

For an harmonic wave, the electric field can be expressed according to (35) where the complex vector \( \mathbf{E}_0 \) can be decomposed into real and imaginary vectors

\[ \mathbf{E}_0 = \mathbf{E}_{0r} + j \mathbf{E}_{0i} \tag{50} \]

Similarly, for the magnetic induction

\[ \mathbf{B}_0 = \mathbf{B}_{0r} + j \mathbf{B}_{0i} \tag{51} \]

Using such decompositions, the electric field is given by

\[ \mathbf{E} (\mathbf{r}, t) = \mathbf{E}_{0r} \cos (\omega t - \mathbf{k} \cdot \mathbf{r}) - \mathbf{E}_{0i} \sin (\omega t - \mathbf{k} \cdot \mathbf{r}) \tag{52} \]
and a similar expression can be derived for $B(r, t)$. Reporting into (49) and using the fact that
\[
\langle \cos^2 (\omega t - \phi) \rangle = \langle \sin^2 (\omega t - \phi) \rangle = \frac{1}{2},
\]
(53)
\[
\langle \cos (\omega t - \phi) \sin (\omega t - \phi) \rangle = 0,
\]
(54)
where $\langle \cdot \rangle$ denotes time averaging, the time average of the electromagnetic density can be expressed as
\[
\langle w \rangle = \frac{1}{4} \left[ \epsilon (E_{0r}^2 + E_{0i}^2) + \frac{1}{\mu} (B_{0r}^2 + B_{0i}^2) \right],
\]
(55)
hence finally
\[
\langle w \rangle = \frac{1}{2} \epsilon |E_0|^2.
\]
(56)
From (48) it can easily be established that $|B_0|^2 = \mu \epsilon |E_0|^2$ and therefore
\[
\langle w \rangle = \frac{1}{2} \epsilon |E_0|^2.
\]
(57)
It is important to remember at this point that the starting point for the derivation of the electromagnetic energy density is (34), which assumes a non dispersive medium. In the case of a dispersive material, it can be shown (see for instance [3]) that, away from absorption resonances, the integration of (27) leads to
\[
\langle w \rangle = \frac{1}{4} \left[ \frac{\partial (\omega \epsilon)}{\partial \omega} |E_0|^2 + \frac{\partial (\omega \mu)}{\partial \omega} |H_0|^2 \right],
\]
(58)
which obviously simplifies to (56) when $\epsilon$ and $\mu$ are assumed to have no dependence on the frequency (non dispersive medium).

If the Poynting vector $S = E \times H$ is now considered, it can easily be shown using decompositions of $E_0$ and $H_0$ into real and imaginary parts that its time average can be expressed according to
\[
\langle S \rangle = \frac{1}{2} \text{Re} (E_0 \times H_0^*).
\]
(59)
From the definition of the Poynting vector $S = E \times H$, and since $E$ and $H$ lie in the plane orthogonal to $k$ and are furthermore orthogonal, it is obvious that $S$ is parallel to $k$. The average of the Poynting vector can also be calculated from (59) by substituting (37) into (59), leading to
\[
\langle S \rangle = \frac{1}{2} \text{Re} \left( E_0 \times \frac{1}{\omega \mu} (k \times E_0^*) \right)
\]
(60)
\[
= \frac{1}{2\omega \mu} \text{Re} \left[ (E_0 \cdot E_0^*) k - (E_0 \cdot k) E_0^* \right]
\]
(61)
\[
= \frac{1}{2\omega \mu} |E_0|^2 k,
\]
(62)
where the last equality stems from the facts that $E_0 \cdot k = 0$ according to (39) and $D_0 = \epsilon E_0$, as well as by definition $E_0 \cdot E_0^* = |E_0|^2$. Expressing the wave vector according to
\[
k = |k| u = \omega \sqrt{\epsilon \mu} u = \frac{\omega}{v_\varphi} u = \frac{\omega n}{c} u,
\]
(63)
where $u$ is an unit vector in the direction of propagation, finally leads to
\[
\langle S \rangle = \langle w \rangle v_\varphi u,
\]
(64)
which is in line with the interpretation of the Poynting theorem presented in Sec. 1.3 where it was stated that the Poynting vector corresponds to the flow of electromagnetic energy.

The following properties of a plane harmonic wave propagating into an isotropic medium should therefore be kept in mind:

- The electric field \( \mathbf{E} \) and electric displacement \( \mathbf{D} \) are parallel: \( \mathbf{D} = \varepsilon \mathbf{E} \).
- The wave is transverse: both \( \mathbf{E}/\mathbf{D} \) and \( \mathbf{B}/\mathbf{H} \) lie in a plane that is perpendicular to the direction of propagation \( \mathbf{k} \).
- \( (\mathbf{k}, \mathbf{D}_0, \mathbf{B}_0) \) forms a right handed orthogonal vector triplet
- So does \( (\mathbf{k}, \mathbf{E}_0, \mathbf{H}_0) \) since \( \mathbf{E}_0 \) and \( \mathbf{D}_0 \) are parallel, as well as \( \mathbf{H}_0 \) and \( \mathbf{B}_0 \).
- Poynting’s vector \( \mathbf{S} \), hence the direction of energy propagation, is oriented along the direction of propagation of the phase \( \mathbf{k} \).

These essential properties are summarised in Fig. 1. It will be seen in Sec. 2 that some of those properties no longer apply in anisotropic media.

### 2 Light propagation in an anisotropic medium

When the medium is anisotropic, the relation between the electric displacement and the electric field becomes tensorial, as expressed in (15). Therefore the vectors \( \mathbf{D} \) and \( \mathbf{E} \) are no longer parallel. In the following, it will be assumed that dispersive effects can be neglected, either because the medium is non-dispersive in the frequency range of interest, or because the fields involved are monochromatic, or of a limited spectral width. In this case, a tensorial relation also holds between \( \mathbf{D} \) and \( \mathbf{E} \) expressed in the time domain,

\[
\mathbf{D} (\mathbf{r}, t) = [\varepsilon] \mathbf{E} (\mathbf{r}, t),
\]
where the dielectric tensor is denoted by $[\epsilon]$. On an orthonormal basis $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$, this tensorial relation links the components of the vectors $\mathbf{D}$ and $\mathbf{E}$ and can be expressed in a matrix form according to

$$
\begin{pmatrix}
D_1 \\
D_2 \\
D_3 \\
\end{pmatrix} =
\begin{pmatrix}
\epsilon_{11} & \epsilon_{12} & \epsilon_{13} \\
\epsilon_{21} & \epsilon_{22} & \epsilon_{23} \\
\epsilon_{31} & \epsilon_{32} & \epsilon_{33} \\
\end{pmatrix}
\begin{pmatrix}
E_1 \\
E_2 \\
E_3 \\
\end{pmatrix}.
$$

(66)

Such a relation is customarily also expressed as $D_i = \sum_j \epsilon_{ij} E_j$ or using Einstein’s notation, according to which summation is performed over repeated indices, $D_i = \epsilon_{ij} E_j$.

In the case of non absorbing media that are magnetically isotropic (i.e. for which $\mathbf{B} = \mu \mathbf{H}$ where $\mu$ is a scalar quantity), the elements of the dielectric tensor are real quantities. Such materials will be considered in this note, unless explicitly stated.

An important property of the dielectric tensor that has implications on the propagation of electromagnetic waves in anisotropic media is its symmetry\(^{12}\), i.e. $\epsilon_{ij} = \epsilon_{ji}$.

### 2.1 Symmetry of the dielectric tensor

It appears that the demonstration of the symmetry of the dielectric tensor raises some issues in many of the popular texts, a point that has been discussed in [4]. In this section, a thermodynamic proof of this property is first given, followed by a discussion of an often reported attempt of derivation based on Poynting’s theorem.

#### 2.1.1 Thermodynamic arguments

In the case of static fields, thermodynamic arguments can be used to demonstrate the symmetry of the dielectric tensor, as reported in e.g. [5, 6]. The differential of the Helmholtz free energy $F$ of a dielectric is given by

$$
dF = - S \, dT - \mathbf{D} \cdot d\mathbf{E},
$$

(67)

where $S$ is the entropy and $T$ denotes the temperature. This relation can be further expanded into

$$
dF = - S \, dT - \sum_i D_i \, dE_i = - S \, dT - \sum_i \sum_j \epsilon_{ij} E_j \, dE_i,
$$

(68)

therefore

$$
D_i = - \frac{\partial F}{\partial E_i}.
$$

(69)

Since

$$
\epsilon_{ij} = \frac{\partial D_i}{\partial E_j},
$$

(70)

then

$$
\epsilon_{ij} = - \frac{\partial^2 F}{\partial E_i \partial E_j}.
$$

(71)

However $dF$ is a total differential. Therefore the value of $\epsilon_{ij}$ as expressed in (71) should not depend on the order of the differentiations. Consequently

$$
\epsilon_{ij} = - \frac{\partial^2 F}{\partial E_j \partial E_i} = - \frac{\partial^2 F}{\partial E_i \partial E_j} = \epsilon_{ji},
$$

(72)

and the dielectric tensor is symmetric.

\(^{12}\)In the case of media whose dielectric tensor is complex, and provided the complex nature of the elements is not due to absorption, it can be shown that the dielectric tensor exhibits Hermitian symmetry, i.e. $\epsilon_{ij} = \epsilon_{ji}$. This is for instance the case of media that present optical activity.
2.1.2 Principal axes of an anisotropic medium

The dielectric tensor \( [\varepsilon] \) has been shown above to be symmetric and real. Therefore there exists an orthonormal basis \((x, y, z)\) where the dielectric tensor is represented by a diagonal matrix

\[
\begin{pmatrix}
D_x \\
D_y \\
D_z
\end{pmatrix} = \begin{pmatrix}
\varepsilon_x & 0 & 0 \\
0 & \varepsilon_y & 0 \\
0 & 0 & \varepsilon_z
\end{pmatrix} \begin{pmatrix}
E_x \\
E_y \\
E_z
\end{pmatrix},
\]

(73)

where \( \varepsilon_x, \varepsilon_y \) and \( \varepsilon_z \) are the eigenvalues of \( [\varepsilon] \). The directions determined by \( x, y \) and \( z \) are known as the principal axes of the medium. In analogy with the isotropic case, refractive indices can be defined along the three principal axes according to \( n_i = \sqrt{\varepsilon_i/\varepsilon_0} \), where \( i = x, y, z \).

2.2 Propagation of a monochromatic plane wave in an anisotropic medium

In this section, the propagation of a monochromatic plane wave with frequency \( \omega \) and wave vector \( \mathbf{k} \) in an anisotropic medium defined by its dielectric tensor \( [\varepsilon] \) is considered. The electric displacement of the plane wave can be expressed in complex notation according to

\[
D(r, t) = \text{Re} \left\{ D_0 e^{j(\omega t - \mathbf{k} \cdot \mathbf{r})} \right\}
\]

(74)

where \( D_0 \) is a complex vector that describes the state of polarization of the wave. The fields \( \mathbf{E}, \mathbf{B} \) and \( \mathbf{H} \) can be expressed in a similar way. In analogy with the isotropic case, the wave vector can be written

\[
\mathbf{k} = \omega \frac{n}{c} \mathbf{u}
\]

(75)

where \( \mathbf{u} \) is a unit vector in the direction of the wave vector \( \mathbf{k} \) and the physical meaning of \( n \) will be clarified. As in the isotropic case, \( \mathbf{u} \) is orthogonal to the planes of constant phase and defines the direction of the wave normal.

In the following, the conditions for such a plane wave to propagate undistorted in the dielectric medium are examined.

2.2.1 Relative configuration of the fields

For such a monochromatic plane wave, the substitution \( \frac{\partial}{\partial t} \to j\omega \) and \( \nabla \to -j\omega \frac{n}{c} \mathbf{u} \) can be performed into Maxwell’s equations (1)-(4), leading to

\[
\frac{n}{c} \mathbf{u} \times \mathbf{E} = \mathbf{B}
\]

(76)

\[
\frac{1}{\mu_0} \frac{n}{c} \mathbf{u} \times \mathbf{B} = -\mathbf{D}
\]

(77)

\[
\mathbf{u} \cdot \mathbf{D} = 0
\]

(78)

\[
\mathbf{u} \cdot \mathbf{B} = 0
\]

(79)

In deriving (79) it has been further assumed that the medium is non magnetic, \( M = 0 \), hence \( \mathbf{B} = \mu_0 \mathbf{H} \). Under this assumption, the Poynting vector becomes

\[
\mathbf{S} = \mathbf{E} \times \frac{\mathbf{B}}{\mu_0}
\]

(80)

Considering eq. (76)-(80) above, it can easily be established that, at any given time

- \((\mathbf{u}, \mathbf{D}, \mathbf{B})\) is a right handed orthogonal vector triplet.
Figure 2  Illustration of the relative directions of the vectors and fields associated with the propagation of an electromagnetic wave in an anisotropic medium. Note that the relative magnitude of the vectors is arbitrary in this representation.

- (S, E, B) is also a right handed orthogonal vector triplet.
- However, the tensor relation (65) imposes that, contrarily to the isotropic case, D is not parallel to E, and consequently S is not parallel to u. Therefore the two right handed orthogonal vector triplets (u, D, B) and (S, E, B) are distinct.
- Consequently, the direction of the energy flow, as defined by the Poynting’s vector, is no longer parallel to u, direction of propagation of the phase.
- D and E are no longer parallel, as in the isotropic case. Hence u · E ≠ 0 and the E field is no longer transverse. However D and B are transverse.\(^\text{13}\)

The relative configuration of the fields and directions of propagation and energy flow is illustrated schematically in Fig. 2.

### 2.2.2 Energy considerations

From the standard definition \(w_e = \frac{1}{2} E \cdot D\) and performing a decomposition of \(E_0\) and \(D_0\) into real and imaginary parts, as was done in the case of isotropic media, it is immediate that

\[
\langle w_e \rangle = \frac{1}{4} (E_0 \cdot D_0^*).
\]

Similarly,

\[
\langle w_m \rangle = \frac{1}{4} (B_0 \cdot H_0^*).
\]

In the derivation of (81) and (82) it was furthermore taken into account the fact that \(D_0 = [\epsilon] E_0\), where the dielectric tensor [\(\epsilon\)] is real.

The expression of the electromagnetic energy density (58) can be extended to the general case of a dielectrically and magnetically anisotropic and dispersive medium, leading to (for both \([\epsilon]\) and \([\mu]\)

\[
\langle w \rangle = \frac{1}{4} \left[ E^* \cdot \frac{\partial (\omega [\epsilon])}{\partial \omega} E + H^* \cdot \frac{\partial (\omega [\mu])}{\partial \omega} H \right].
\]

\(^{13}\)This last point legitimates the choice of \(D_0\) to define the state of polarisation of a monochromatic plane wave since \(D_0\) is transverse in both isotropic and anisotropic cases, while \(E_0\) is, in general, not transverse in anisotropic media.
2.2.3 Eigenmodes of propagation

In this section, the solutions of the propagation equation in an anisotropic medium are studied in more details. In particular it will be shown that, for a given direction of propagation, two linearly polarised plane waves with well defined polarisations can maintain their state of polarisation over propagation in the medium. These so-called eigenmodes of propagation, and associated phase velocities, or equivalently refractive indices, will be derived in this section, and their main characteristics will be examined.

Substituting (76) into (77) enables to solve Maxwell’s equations for the electric displacement

\[ D = \frac{n^2}{c^2 \mu_0} [E - (\mathbf{u} \cdot \mathbf{E}) \mathbf{u}] . \]  

(84)

Therefore, two alternative expressions relating the electric displacement \( D \) to the electric field vector \( \mathbf{E} \) have been found. The first one stems from the constitutive relation (65), while the second one is a consequence of Maxwell’s equations. Those two relations can be expressed in a coordinate system that coincides with the principal axes of the dielectric tensor, leading to

\[ D_i = \epsilon_0 n_i^2 E_i , \]  

(85)

and

\[ D_i = \frac{n^2}{c^2 \mu_0} [E_i - (\mathbf{u} \cdot \mathbf{E}) u_i] . \]  

(86)

Eliminating \( D_i \) between the two equations and recalling that \( c = 1/\sqrt{\epsilon_0 \mu_0} \),

\[ E_i = \frac{n^2}{n^2 - n_i^2} (\mathbf{u} \cdot \mathbf{E}) u_i , \]  

(87)

hence

\[ \mathbf{u} \cdot \mathbf{E} = \sum_i u_i E_i = \sum_i \frac{n^2 (\mathbf{u} \cdot \mathbf{E})}{n^2 - n_i^2} u_i^2 , \]  

(88)

and finally

\[ \sum_i \frac{u_i^2}{n^2 - n_i^2} = \frac{1}{n^2} . \]  

(89)

Using the fact that \( \sum u_i^2 = 1 \), since \( \mathbf{u} \) is a unit vector, enables to obtain the alternative expression

\[ \sum_i \frac{n_i^2 u_i^2}{n^2 - n_i^2} = 0 , \]  

(90)

\[ \sum_i \frac{u_i^2}{n^2 - n_i^2} = 0 , \]  

(91)

and

\[ \sum_i \frac{v_i^2}{v^2 - v_i^2} = 0 , \]  

(92)

where \( v = c/n \) and \( v_i = c/n_i \). Eq. (89)-(92) are known as alternative forms of Fresnel’s equation of wave normal. They enable the determination of \( n \), or equivalently \( v \), once the direction of the wave normal \( \mathbf{u} \) is known.
2.2.4 The surface of the indices

For each direction of propagation \( \mathbf{u} = (u_x, u_y, u_z) \), (90) is a quadratic equation in \( n^2 \)

\[
\begin{aligned}
  u_x^2 n_x^2 (n^2 - n_y^2) (n^2 - n_z^2) + u_y^2 n_y^2 (n^2 - n_x^2) (n^2 - n_z^2) + u_z^2 n_z^2 (n^2 - n_x^2) (n^2 - n_y^2) = 0,
\end{aligned}
\]

which can be shown to have two positive solutions \( n' \) and \( n'' \), not necessarily distinct.

It is interesting to graphically visualise the solutions of (93) in a Cartesian coordinate system \((O, x, y, z)\) associated with the orthonormal basis in which the dielectric tensor is represented by a diagonal matrix. For this purpose, the surface of the indices, defined as the set of points \( M \) whose distance from the origin \( O \) in the direction of propagation \( \mathbf{u} \) is equal to \( n \) is considered,

\[
(\mathcal{I}) = \{ M, OM = n \mathbf{u} \}.
\]

Reporting \( n^2 = x^2 + y^2 + z^2 \) in (93) leads to the equation of the surface of the indices

\[
\begin{aligned}
  x^2 n_x^2 (x^2 + y^2 + z^2 - n_y^2) (x^2 + y^2 + z^2 - n_z^2) \\
  + y^2 n_y^2 (x^2 + y^2 + z^2 - n_x^2) (x^2 + y^2 + z^2 - n_z^2) \\
  + z^2 n_z^2 (x^2 + y^2 + z^2 - n_y^2) (x^2 + y^2 + z^2 - n_x^2) = 0.
\end{aligned}
\]

Considering the intersections of \( \mathcal{I} \) with the planes \( x = 0, y = 0 \) or \( z = 0 \) turns out to help visualising the rather complicated surface represented by (95). This can be done in a straightforward way from (95). For instance, the intersection with the plane \( y = 0 \) is given by

\[
\begin{aligned}
  x^2 + z^2 &= n_y^2, \\
  \frac{x^2}{n_x^2} + \frac{z^2}{n_z^2} &= 1.
\end{aligned}
\]

Eq. (96) represents a circle of radius \( n_y \), while eq. (97) is that of an ellipse with semi-axes \( n_z \) (along \( x \)) and \( n_x \) (along \( z \)). The intersection of the surface of the indices with the planes \( x = 0 \) or \( z = 0 \) will also lead to a circle and an ellipse whose radius and semi-axes can easily be found by proper permutation of the refractive indices \( n_x, n_y \) and \( n_z \). Those intersections are represented in Fig. 3 in the general case \( n_z > n_y > n_x \).

It can easily be shown that:

- If \( n_x = n_y = n_z \), the surface of the indices is a sphere of radius \( n = n_x = n_y = n_z \). This is the case for isotropic media.

---

\textsuperscript{14} The identity \( \mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = (\mathbf{A} \cdot \mathbf{C}) \mathbf{B} - (\mathbf{A} \cdot \mathbf{B}) \mathbf{C} \) has been used in order to derive (84).

\textsuperscript{15} Augustin Jean Fresnel, Broglie 1788 – Ville d’Avray 1827.
Figure 4  Representation of the surface of the indices for \( y \geq 0 \) and \( z \leq 0 \) in the arbitrary case \( n_x = 1, n_y = 2, n_z = 3 \). The two shells of the surface are clearly visible, as well as their intersection in the plane \( y = 0 \).

- If two out of the three indices are equal, by convention \( n_x = n_y \), the surface of the indices is constituted of a sphere and an ellipsoid of revolution around the axis \((O, z)\). The sphere and ellipsoid are tangent at their intersection with the \((O, z)\) axis. The medium is said to be uniaxial. For historical reasons, \( n_x = n_y \) is known as the ordinary index \( n_o \) and \( n_z \) is the extraordinary index \( n_e \). If furthermore \( n_o > n_e \) the medium is said to be negative, while it is described as positive when \( n_e > n_o \).

- In the more general case where all three indices are different, and using the convention \( n_z > n_y > n_x \), the surface is constituted of two shells whose intersections can be determined by considering their intersection with the \( y = 0 \) plane (since \( n_z > n_y > n_x \) the two shells do not intersect in the \( x = 0 \) and \( z = 0 \) planes, and have four intersection points in the \( y = 0 \) plane). In this case the medium is said to be biaxial. An example of surface of the indices for a biaxial medium is represented in Fig. 4. In order to emphasize its shape in the most general case, it has been represented for arbitrary values of the indices \( n_x = 1, n_y = 2 \) and \( n_z = 3 \). The intersection points between the two shells in the plane \( y = 0 \) can be clearly seen.

Consequently, it has been shown that, for any given direction of the wave normal \( \mathbf{u} \) there exists at most two allowed positive values for \( n \). Those values are found by considering the intersection of the direction \( \mathbf{u} \) with the two shells constituting the surface of the indices. In the case of propagation along the \( z \) direction in an uniaxial crystal, those two values reduce to a single index \( n_o \). This is also the case for specific propagation directions corresponding to the intersections between the shells of the surface of the indices in a biaxial crystal. The directions of propagation for which only a single value of \( n \) can be found are known as the optic axes. Hence the denomination of uniaxial and biaxial media. Apart in the case of propagation along an optic axe, two harmonic plane waves characterised by their displacement \( \mathbf{D}' \) and \( \mathbf{D}'' \) and their phase velocity \( v'_\varphi = c/n' \) and \( v''_\varphi = c/n'' \) can be associated with \( n' \) and \( n'' \), respectively. It will be shown in Sec. 2.3 that \( \mathbf{D}' \) and \( \mathbf{D}'' \) correspond to two orthogonal linearly polarised
waves.

The discussion above has been based on the use of the surface of the indices (94). Other relevant geometric constructions are often considered in the literature. The $k$-vector surface $(\mathcal{K}) = \{K, \mathbf{OK} = k\}$ is obviously homothetic to $(I)$. Another frequent choice\textsuperscript{16} is that of the \textit{normal surface}, defined as

\[(\mathcal{N}) = \{P, \mathbf{OP} = v_\phi \mathbf{u}\}, \quad (98)\]

which directly enables the determination of $v_\phi$ for a given direction of the wave normal $\mathbf{u}$.

In summary:

- For each direction of the wave normal $\mathbf{u}$ there exists two allowed refractive indices $n'$ and $n''$ that are solutions of Fresnel's equation and that can be found by considering the intersection of the surface of the indices with the direction of $\mathbf{u}$. These two indices reduce to one in some pathological cases.

- Two plane waves characterised by their displacement vectors $\mathbf{D}'$ and $\mathbf{D}''$ can be associated to $n'$ and $n''$, respectively. These two waves are known as the eigenmodes of propagation in the direction of the wave normal $\mathbf{u}$.

- Those two waves are linearly polarised and preserve their linear polarisation over propagation in the medium.

- Furthermore their displacement vectors $\mathbf{D}'$ and $\mathbf{D}''$ are orthogonal and are also orthogonal to the direction of propagation $\mathbf{u}$.

- Since both $\mathbf{D}'$ and $\mathbf{D}''$ are orthogonal to $\mathbf{u}$, the orthogonal vector triplet $(\mathbf{u}, \mathbf{D}', \mathbf{D}'')$ can be used to decompose arbitrary waves propagating along $\mathbf{u}$ in the medium.

So far the use of the surface of the indices has enabled the determination of the refractive indices $n'$ and $n''$, or equivalently the phase velocities $v'_\phi$ and $v''_\phi$, but does not provide directly usable information on the associated eigenmodes $\mathbf{D}'$ and $\mathbf{D}''$. It will be shown in the next section that the properties of the eigenmodes, including the orientation of their $\mathbf{D}$ vector, can be determined in a straightforward manner based on a powerful geometrical construction known as the index ellipsoid.

### 2.3 The index ellipsoid

The tensor relation between the electric field $\mathbf{E}$ and the displacement $\mathbf{D}$ in an anisotropic medium has been introduced in (65) and can be expressed in an alternative form as

\[\mathbf{D} = \varepsilon_0 [\varepsilon_r] \mathbf{E}, \quad (99)\]

where $[\varepsilon_r]$ is the relative dielectric tensor. It is convenient to introduce the impermeability tensor $[\eta]$ defined by

\[\varepsilon_0 \mathbf{E} = [\eta] \mathbf{D}. \quad (100)\]

The impermeability tensor is the inverse of the relative dielectric tensor

\[[\eta] = [\varepsilon_r]^{-1}. \quad (101)\]

\textsuperscript{16}Interestingly, the French literature seems to favour the use of the surface of the indices, while English texts tend to use the normal surface.
Since $[\varepsilon]$ is symmetric, $[\eta]$ is also symmetric and both tensors have the same principal axes. In the orthonormal basis $(x, y, z)$ where $[\varepsilon]$ is diagonal, $[\eta]$ is also diagonal and can be expressed as

$$[\eta] = \begin{pmatrix} \frac{1}{n_x^2} & 0 & 0 \\ 0 & \frac{1}{n_y^2} & 0 \\ 0 & 0 & \frac{1}{n_z^2} \end{pmatrix}. \quad (102)$$

Reporting (100) into (84)

$$[\eta] \mathbf{D} - \{ \mathbf{u} \cdot [\eta] \} \mathbf{D} \mathbf{u} = \frac{1}{n^2} \mathbf{D} \quad (103)$$

The left hand side of (103) represents the projection of the vector $[\eta] \mathbf{D}$ on a plane that is orthogonal to $\mathbf{u}$. The relative configuration of the vectors in (103) is illustrated in Fig. 5.

It is therefore natural to introduce a new orthonormal basis $(i, j, u)$ where $i$ and $j$ are two orthonormal vectors in the plane orthogonal to $\mathbf{u}$. Representing the impermeability tensor by $(\eta_{ij})$ with $i, j = 1, 2, 3$, in this basis and recalling that $\mathbf{D}$ is orthogonal to $\mathbf{u}$ immediately leads to

$$\begin{pmatrix} \eta_{ii} & \eta_{ij} & \eta_{ji} \\ \eta_{ji} & \eta_{jj} \end{pmatrix} \mathbf{D} = \frac{1}{n^2} \mathbf{D}. \quad (104)$$

This is a standard eigenvalue equation with $1/n^2$ as eigenvalue and $\mathbf{D}$ as eigenvector. Since $(\eta_{ij})$ is symmetric and real, its eigenvalues are real and its eigenvectors associated to two different eigenvalues $\mathbf{D}'$ and $\mathbf{D}''$ are real and orthogonal. Consequently, the solutions of the propagation equation in a linear anisotropic medium are two linearly polarised plane waves that are furthermore orthogonal. These solutions are known as the two normal modes of propagation, to which two refractive indices, $n'$ and $n''$ can be associated, respectively. $n'$ and $n''$ can be found by solving the eigenvalue equation

$$\det \left[ \eta_{ij} - \frac{1}{n^2} I \right] = 0, \quad (105)$$

where $\eta_{ij}$ is sometimes described as the transverse impermeability tensor, represented by $(\eta_{ij})$ (with $i, j = 1, 2$) in the basis $(i, j)$ of the plane orthogonal to $\mathbf{u}$, and $I$ is the identity matrix.

Not surprisingly, solving the eigenvalue equation (105) will lead to Fresnel’s equation already studied in Sec. 2.2.

A geometric construction that enables to find the normal modes and the associated refractive indices for a given propagation direction is now introduced. The surface $(\mathcal{E})$ defined by the points $\mathbf{M}$ such that $|\mathbf{OM}| = n$, where $n$ is the refractive index in the direction of propagation $\mathbf{u}$, and the direction of the vector $\mathbf{OM}$ is that of the vector $\mathbf{D}$ is considered. Mathematically, this surface is defined as

$$(\mathcal{E}) = \left\{ \mathbf{M}, \mathbf{OM} = n \frac{\mathbf{D}}{||\mathbf{D}||} \right\}. \quad (106)$$
Recalling equation (84)

\[ D = \frac{n^2}{c^2 \mu_0} [E - (u \cdot E) u], \tag{107} \]

immediately leads to

\[ D^2 = \frac{n^2}{c^2 \mu_0} E \cdot D \tag{108} \]

since \( u \) and \( D \) are orthogonal. In the orthonormal basis \((x, y, z)\) where \( [\epsilon] \) is diagonal, (85) also applies and

\[ D_i = \epsilon_0 n_i^2 E_i. \tag{109} \]

Hence

\[ |D|^2 = \frac{n^2}{c^2 \mu_0} \sum_i \frac{D_i^2}{\epsilon_0 n_i^2}, \tag{110} \]

and therefore

\[ \sum_i \frac{1}{n_i^2} \left( \frac{n D_i}{|D|} \right)^2 = 1, \tag{111} \]

where, according to the definition of \((\mathcal{E})\), the \( n D_i / |D| \) correspond to the coordinates of the vector \( OM = (X, Y, Z) \) in the orthonormal basis \((x, y, z)\). Eq. (111) can be expressed as a function of \(X, Y, Z\), leading to

\[ \frac{X^2}{n_x^2} + \frac{Y^2}{n_y^2} + \frac{Z^2}{n_z^2} = 1, \tag{112} \]

which is the equation of an ellipsoid expressed in its principal axes and whose semi-axes are of lengths \(n_x, n_y\) and \(n_z\). This ellipsoid is known as index ellipsoid or optical indicatrix and is schematically represented in Fig. 6.

It is now shown how the index ellipsoid can be used to determine the directions of the eigenpolarisations \(D'\) and \(D''\), as well as their associated refractive indices. The electric displacement \(D\) is known to lie in the plane orthogonal to the direction of propagation \(u\). Consequently, based on the definition of the surface \((\mathcal{E})\), the extremity \(M\) of the vector \(OM = (X, Y, Z)\) associated to \(D\) belongs to the intersection of the index ellipsoid with the plane orthogonal to \(u\), which is an ellipse (\(\Gamma\)) of center \(O\). Let \(n\) be the unit vector orthogonal to the ellipsoid \((\mathcal{E})\) in \(M\). \(n\) can be expressed in the basis \((x, y, z)\) by\(^{17}\):

\[ n \propto \begin{pmatrix} X/n_x^2 \\ Y/n_y^2 \\ Z/n_z^2 \end{pmatrix}, \tag{113} \]

In the basis \((x, y, z)\), \(E_i = D_i/\epsilon_0 n_i^2\) and, by definition, \(OM = nD/|D|\). Consequently,

\[ E = \frac{|D|}{n\epsilon_0} \begin{pmatrix} X/n_x^2 \\ Y/n_y^2 \\ Z/n_z^2 \end{pmatrix}, \tag{114} \]

and therefore \(n\) is parallel to \(E\). It is also known that \(B\) lies in the plane orthogonal to \(u\) and that \(B\) is orthogonal to \(E\). Since \(B\) is orthogonal to \(E\), \(B\) lies in the plane that is tangent to \((\mathcal{E})\) in \(M\). As \(B\) lies in the plane orthogonal to \(u\), necessarily \(B\) is tangent to the ellipse (\(\Gamma\)). From Maxwell’s equations, \(B\) is also known to be orthogonal to \(D\).

\(^{17}\)The normal to a surface defined by \(f(x, y, z) = 0\) is proportional to the vector \(\nabla f = \left( \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right)\)
Consequently it has been established that $\mathbf{B}$ is tangent to the ellipse $(\Gamma)$ and orthogonal to $\mathbf{OM}$, where $M$ belongs to $(\Gamma)$. The only points $M$ of an ellipse where its tangent is orthogonal to $\mathbf{OM}$ correspond to the extremities of its axes.

Hence the allowed directions of $\mathbf{D}$ correspond to the two axes of the ellipse $(\Gamma)$, intersection of the index ellipsoid $(\mathcal{E})$ with the plane orthogonal to the direction of propagation $\mathbf{u}$. The corresponding refractive indices are the semi-axes of $(\Gamma)$. It also becomes immediate that the eigenmodes $\mathbf{D}'$ and $\mathbf{D}''$ are linearly polarised and orthogonal.

In summary:

- Knowing the direction of the wave normal $\mathbf{u}$, consider the intersection of the index ellipsoid with the plane orthogonal to $\mathbf{u}$ and going through the origin. This intersection is an ellipse.

- The orientations of $\mathbf{D}'$ and $\mathbf{D}''$ correspond to the two axes of the ellipse.

- The corresponding refractive indices $n'$ and $n''$ correspond to the semi-axes of the ellipses.

Therefore a simple and elegant way to geometrically find the orientations of the eigenmodes of propagation and the associated refractive indices based on the index ellipsoid has been introduced.
2.4 Light rays in anisotropic media

So far it has been demonstrated that anisotropic media support two orthogonal eigenmodes, each with its own phase velocity and polarisation. The values of the phase velocities and the orientations of the $\mathbf{D}$ vectors of the eigenmodes can be easily retrieved, for a given direction of the wave normal $\mathbf{u}$. Historically, the observation of double images through a crystal of calcite has been reported as early as 1669 by Erasmus Bartholin\(^{18}\) in his famous book “Experimenta crystalli Islandici disdiaclastici quibus mira & insolita refractio detegitur” (Copenhagen, 1669). This effect, which is now known to be a manifestation of light propagation in an anisotropic medium, has been puzzling several generation of scientists including Newton, Huygens and Fresnel. It has also been instrumental in establishing the wave theory of light, including the transverse nature of light vibrations and the concept of polarisation.

The theory that has been developed so far in this note does not enable yet the explanation of the formation of double images (also known as double refraction) by an anisotropic crystal. For this purpose it is essential to clarify the nature and properties of rays of light propagating in an anisotropic medium. In this section, the concept of ray velocity will be introduced. A new surface enabling its determination for a given ray direction will be presented, and its relation with the surface of the indices will be analysed. In particular, it will be shown that the light rays, which follow the direction of Poynting’s vector, are orthogonal to the surface of the indices. Finally, the effect of double refraction will be clarified.

2.4.1 The ray velocity

In analogy with (64), derived in the case of isotropic media, the ray velocity $v_r$ is defined in the case of anisotropic media as

$$\langle \mathbf{S} \rangle = \langle w \rangle v_r \mathbf{s}, \quad (115)$$

where $\mathbf{s}$ is an unit vector in the direction of Poynting’s vector $\mathbf{S}$.

Based on the expressions of the electric and magnetic energy densities (33), as well as on Maxwell’s equations (76) and (77), and using the expression of the scalar triple product\(^{19}\), the following expressions are obtained:

$$\mathbf{E} \cdot \mathbf{D} = -\frac{1}{\mu_0 c} \mathbf{E} \cdot (\mathbf{u} \times \mathbf{B}) = \frac{1}{\mu_0 c} \mathbf{u} \cdot (\mathbf{E} \times \mathbf{B}), \quad (116)$$

and

$$\mathbf{H} \cdot \mathbf{B} = \frac{1}{\mu_0 c} \mathbf{B} \cdot (\mathbf{u} \times \mathbf{E}) = \frac{1}{\mu_0 c} \mathbf{u} \cdot (\mathbf{E} \times \mathbf{B}). \quad (117)$$

Hence the electromagnetic energy density

$$w = \frac{n}{c} \mathbf{u} \cdot \left( \mathbf{E} \times \frac{\mathbf{B}}{\mu_0} \right) = \frac{n}{c} \mathbf{u} \cdot \mathbf{S}. \quad (118)$$

From (115), and since $\mathbf{u} \cdot \mathbf{s} = \cos \alpha$, the ray velocity can be expressed as a function of the phase velocity $v_\varphi = c/n$ as

$$v_\varphi = v_r \cos \alpha. \quad (119)$$

It is also possible to define a ray index $n_r$ such that $v_r = c/n_r$, which can be related to $n$ according to

$$n_r = n \cos \alpha. \quad (120)$$

\(^{18}\)Erasmus Bartholin, Roskilde 1625 – Copenhagen 1698.

\(^{19}\)A $\cdot (\mathbf{B} \times \mathbf{C}) = \mathbf{B} \cdot (\mathbf{C} \times \mathbf{A}) = \mathbf{C} \cdot (\mathbf{A} \times \mathbf{B})$. 
Equation (119) shows that the phase velocity is the projection of the ray velocity on the direction of the wave normal.

Equation (115) has important implications in the context of geometrical optics. The light rays follow the direction of Poynting’s vector, hence that of the unit vector \( s \). The ray index is therefore the one that needs to be considered in the framework of ray optics. An important issue in the understanding of double refraction is the practical determination of the direction of the light rays in an anisotropic medium, a problem which will be addressed in Sec. 2.4.3.

It can furthermore be shown that, in case the medium is dispersive, the ray velocity can be identified with the group velocity. Starting from Maxwell’s equations

\[
\begin{align*}
\mathbf{k} \times \mathbf{E} &= \omega \mu_0 \mathbf{H}, \\
\mathbf{k} \times \mathbf{H} &= -\omega \mathbf{D},
\end{align*}
\]

infinitesimal variations of the wave vector \( \delta \mathbf{k} \) and of the frequency \( \delta \omega \) will result in variations of the fields \( \delta \mathbf{E} \), \( \delta \mathbf{H} \) and \( \delta \mathbf{D} \), that can be expressed from (121) and (122) as

\[
\delta \mathbf{k} \times \mathbf{E} + \mathbf{k} \times \delta \mathbf{E} = \delta \omega \mu_0 \mathbf{H} + \omega \mu_0 \delta \mathbf{H}
\]

and

\[
\delta \mathbf{k} \times \mathbf{H} + \mathbf{k} \times \delta \mathbf{H} = -\delta \omega \mathbf{D} - \omega \delta \mathbf{D}.
\]

Performing the scalar product of (123) by \( \mathbf{H} \) and using the properties of the triple scalar product (see footnote 19) leads to

\[
\delta \mathbf{k} \cdot (\mathbf{E} \times \mathbf{H}) + \mathbf{k} \cdot (\delta \mathbf{E} \times \mathbf{H}) - \omega \mu_0 \mathbf{H} \cdot \delta \mathbf{H} = \delta \omega \mu_0 \mathbf{H}^2,
\]

where, using Maxwell’s equation (121), the following substitution can be made

\[
-\omega \mu_0 \mathbf{H} \cdot \delta \mathbf{H} = - (\mathbf{k} \times \mathbf{E}) \cdot \delta \mathbf{H} = - \mathbf{k} \cdot (\mathbf{E} \times \delta \mathbf{H}),
\]

finally resulting in

\[
\delta \mathbf{k} \cdot (\mathbf{E} \times \mathbf{H}) + \mathbf{k} \cdot (\delta \mathbf{E} \times \mathbf{H}) - \mathbf{k} \cdot (\mathbf{E} \times \delta \mathbf{H}) = \delta \omega \mu_0 \mathbf{H}^2.
\]

Similarly, multiplying (124) by \( \mathbf{E} \) results in

\[
- \delta \mathbf{k} \cdot (\mathbf{E} \times \mathbf{H}) - \mathbf{k} \cdot (\delta \mathbf{E} \times \mathbf{H}) + \omega \delta \mathbf{D} \cdot \mathbf{E} = - \delta \omega \mathbf{D} \cdot \mathbf{E}.
\]

Furthermore, since the dielectric tensor is symmetric, it is straightforward that

\[
\delta \mathbf{D} \cdot \mathbf{E} = \mathbf{D} \cdot \delta \mathbf{E},
\]

hence

\[
\omega \delta \mathbf{D} \cdot \mathbf{E} = \omega \mathbf{D} \cdot \delta \mathbf{E} = - (\mathbf{k} \times \mathbf{H}) \cdot \delta \mathbf{E} = \mathbf{k} \cdot (\delta \mathbf{E} \times \mathbf{H})
\]

where Maxwell’s equation (122) and the usual scalar triple product have been used. Finally

\[
- \delta \mathbf{k} \cdot (\mathbf{E} \times \mathbf{H}) - \mathbf{k} \cdot (\delta \mathbf{E} \times \mathbf{H}) + \mathbf{k} \cdot (\delta \mathbf{E} \times \mathbf{H}) = - \delta \omega \mathbf{D} \cdot \mathbf{E}.
\]

Subtracting (131) from (127) finally leads to

\[
2 \delta \mathbf{k} \cdot (\mathbf{E} \times \mathbf{H}) = \delta \omega \left( \mu_0 \mathbf{H}^2 + \mathbf{D} \cdot \mathbf{E} \right),
\]
which can be simply expressed as
\[ \delta k \cdot S = \delta \omega \omega \]  
(133)

The group velocity of the wave packet is equal to\(^{20}\)
\[ v_g = \nabla_{k} \omega, \]  
(134)
hence
\[ \delta \omega = \nabla_{k} \omega \cdot \delta k = v_g \cdot \delta k. \]  
(135)
Using (115), (133) and (135), it is immediate that the group velocity is equal to the ray velocity\(^{21}\)
\[ v_g = v_r \cdot s. \]  
(136)

2.4.2 The ray surface

It has been shown in Sect. 2.2.3 how the existence of two alternative relations expressing the displacement \( D \) as a function of the electric field \( E \), one based on the constitutive relation (65), the other stemming from Maxwell’s equations (84), led to the definition of the surface of the indices, once expressed in a coordinate system coinciding with the principal axes of the dielectric tensor. This surface of the indices enables the geometric determination of the possible values of \( n \) for a given direction of propagation, as defined by the wave normal \( u \).

The problem of interest here is to determine the allowed values of the ray index \( n_r \) for a particular ray direction \( s \). It is now shown how a geometric construction of the same type as that of the surface of the indices can be introduced in order to solve this problem. The vector
\[ D_\perp = D - (s \cdot D) s \]  
(137)
is the projection of \( D \) on the plane orthogonal to \( s \). Based on the relative configuration of the fields represented in Fig. 2, this vector is parallel to \( E \). It can therefore also be expressed as a function of the unit vector in the direction of \( E \) as
\[ D_\perp = \left( D \cdot \frac{E}{|E|} \right) \frac{E}{|E|}, \]  
(138)
where the scalar product between \( D \) and \( E \) can be calculated from (84), leading to
\[ D \cdot E = \frac{n_r^2}{c^2 \mu_0} \left| E \right|^2 - (u \cdot E)^2 \]  
(139)
\[ = \frac{n_r^2}{c^2 \mu_0} \left| E \right|^2 \left[ 1 - \cos^2 (u, E) \right], \]  
(140)
which, after some further arithmetics leads to
\[ D \cdot E = \frac{n_r^2}{c^2 \mu_0} \left| E \right|^2 \cos^2 \alpha = \frac{n_r^2}{c^2 \mu_0} \left| E \right|^2. \]  
(141)
Therefore (138) becomes
\[ D_\perp = \frac{n_r^2}{c^2 \mu_0} E. \]  
(142)
\(^{20}\)Equation (134) is simply a generalisation of the well known expression \( v_g = \frac{\partial \omega}{\partial k} \) that applies to a wave packet of the type \( E(z, t) = \text{Re} \left\{ \int_{-\infty}^{+\infty} A(k) e^{j(\omega \tau - kz)} dk \right\} \) to the case of a 3 dimensional wave packet.

\(^{21}\)In the case of magnetic anisotropy, the same conclusion would have been reached. The starting point would have been \( k \times E = \omega B \) and the symmetry of the magnetic permeability tensor \(|\mu|\) would have been used.
The two alternative expression for $D_{\perp}$, (137) and (142) finally lead to

$$E = \frac{c^2 \mu_0}{n_r^2} [D - (s \cdot D) s]. \tag{143}$$

Another relation linking $D$ to $E$ is of course the constitutive relation

$$E = [\epsilon]^{-1} D, \tag{144}$$

which can be written in in a coordinate system coinciding with the principal axes of the dielectric tensor as

$$E_i = \frac{1}{\epsilon_0 n_i^2} D_i. \tag{145}$$

It can be observed that (84) and (143) are similar equations that can be obtained one from the other by making the substitutions

$$D \leftrightarrow E \quad u \leftrightarrow s \quad \mu_0 \leftrightarrow \frac{1}{\mu_0} \quad v_\phi \leftrightarrow \frac{1}{v_r} \quad n \leftrightarrow \frac{1}{n_r}. \tag{146}$$

Consequently, in the same way as Fresnel’s equations, which enable the determination of possible values of $n$ or $v_\phi$ for a given direction of the wave normal $u$, had been established from (84) and (85), (143) and (145) will lead to quadratic equations enabling the determination of $n_r$, or equivalently $v_r$, for a given ray direction $s$. Such relations can be expressed as

$$\sum_i \frac{s_i^2}{v_r^2} = 0, \tag{147}$$

$$\sum_i \frac{v_i^2 s_i^2}{v_r^2 - v_i^2} = 0, \tag{148}$$

and

$$\sum_i \frac{s_i^2}{n_r^2 - n_i^2} = 0. \tag{149}$$

Geometric constructions, which are the counterparts of the surface of the indices (94) and of the normal surface (98), can be introduced to visualise the solutions of the Fresnel equations for the ray indices or ray velocities. The ray surface is defined as

$$(R) = \{Q, OQ = v_r s\}. \tag{150}$$

It is also possible to define a surface of ray indices according to

$$(S) = \{N, ON = n_r s\}. \tag{151}$$

It will be shown in the next section how the ray surface and the surface of the ray indices relate to the normal surface and the surface of the indices.

### 2.4.3 Relation between the surfaces

The surface of the indices has been defined in (94) as the the locus of point $M$ such that $OM = nu$. Defining the vector $n$ such that $n = nu$, this surface can expressed by an implicit equation of the form $f(n) = 0$, or equivalently $g(k, \omega) = 0$ where $g(k, \omega) = f(ck/\omega)$. The
normal to the surface of the indices at a point \( M \) is proportional to the gradient of \( f \) with respect to \( n \) at this point, i.e. to \( \nabla_n f \). Since \( g(k, \omega) = 0 \), its differential \( dg \) is equal to zero

\[
dg = \nabla_k g \cdot dk + \frac{\partial g}{\partial \omega} d\omega = 0.
\] (152)

The group velocity is equal to

\[
v_g = \nabla_k \omega,
\] (153)

hence

\[
d\omega = \nabla_k \omega \cdot dk = v_g \cdot dk.
\] (154)

Reporting into (152), the following relation is obtained

\[
\left( \nabla_k g + \frac{\partial g}{\partial \omega} v_g \right) \cdot dk = 0.
\] (155)

Substituting

\[
\nabla_k g = \frac{c}{\omega} \nabla_n f,
\] (156)

and

\[
\frac{\partial g}{\partial \omega} = \nabla_n f \frac{\partial n}{\partial \omega} = -\frac{1}{\omega} n \cdot \nabla_n f
\] (157)

into (155) leads to

\[
\left( \frac{c}{\omega} \nabla_n f - \frac{1}{\omega} n \cdot \nabla_n f v_g \right) \cdot dk = 0,
\] (158)

from which the group velocity can be expressed according to

\[
v_g = c \frac{\nabla_n f}{n \cdot \nabla_n f}.
\] (159)

Since it has been shown in (136) that the group velocity vector is parallel to \( s \), it can be concluded that \( s \) is parallel to the gradient of \( f \) with respect to \( n \), hence to the normal to the surface of the indices.

The direction of the rays for a given direction of the wave normal \( u \) is parallel to the normal to the surface of the indices at its intersection point with the direction defined by \( u \).

This property is of a high practical importance since it enables the determination of the directions of the light rays in an anisotropic medium, once the direction of the wave normal is known. It will be shown that continuity conditions at the interface between two media enable the determination of the wave normal direction. However, this direction is only of theoretical interest since it cannot be observed directly in experiments. On the other hand, the direction of the light rays can be directly related to experimental observations.

Taking into account the duality property between the ray surface and the surface of the indices (146), it can also be shown that the direction of \( u \) is that of the normal to the ray surface at its intersection with the ray direction \( s \).

These important conclusions are illustrated in Fig. 7 for the case of a wave propagating in a biaxial medium with \( n_z > n_y > n_x \). In this particular example, it is assumed that the wave vector is in the plane \( y = 0 \). Fig. 7 (right) represents the intersection of the surface of the indices (\( I \)) and the surface of the ray indices (\( S \)) with the plane \( y = 0 \). As described previously, the intersection of the surface of the indices consist of a circle of radius \( n_y \) and an ellipse with semi-axes \( n_z \) and \( n_x \) along \( x \) and \( z \), respectively. The surface of the ray indices also consists of two shells whose intersections with the plane \( y = 0 \) reduce to the same circle of radius \( n_y \).
Figure 7  Relations between the normal surface and the ray surface (left) and between the surface of the indices and the surface of the ray indices (right). The general case of a biaxial medium is considered here. The intersections of the various surfaces with the plane \( y = 0 \) are represented.

and a more complex curve that also intersects the axes \((Ox)\) and \((Oy)\) at distances \(n_z\) and \(n_x\) from the origin, respectively. First, the focus is on the intersection of the wave normal with the ellipse. It has just been established that \(s\) is parallel to the normal to the surface of the indices at its intersection \(M\) with the direction \(u\). The vector \(s\) is therefore orthogonal to the tangent to the ellipse at \(M\). Hence the corresponding point \(N\) on the surface of the ray velocities, which is defined by its intersection with the tangent to the ellipse at \(M\). The ray direction is that of \(ON\) and the corresponding ray index is given by the distance \(|ON|\). It thus becomes possible to geometrically determine the ray direction and the ray index once the direction of wave normal is known. In the case of the intersection of the wave normal direction with the circle of radius \(n_y\), the directions defined by \(u\) and \(s\) are identical and \(n_r = n\).

In a similar fashion it is possible to determine the ray direction from the wave normal direction using the ray surface \((R)\) and the normal surface \((N)\), as represented in Fig. 7 (left). Since \(u\) is normal to \((R)\) at its intersection \(Q\) with the ray surface \(s\), the direction of \(s\) can be found by considering the intersection of the plane orthogonal to \(u\) at \(P\) and the ray surface \((R)\). This plane is tangent to \((R)\) at \(Q\), which therefore defines the ray direction and the ray velocity according to \(OQ = v_s s\). It is important to observe at this point that all this discussion applies to the general case of a biaxial medium provided the direction of the wave normal does not coincide with an optical axis (i.e. the vector \(u\) is not directed towards the intersection between the circle and the ellipse defining the intersection of the surface of the indices with the plane \(y = 0\)). This situation would give rise to a peculiar behaviour known as conical refraction.

2.4.4 Double refraction by an anisotropic medium

- Reminder of Snell’s law and how it is obtained for refraction at a boundary between 2 isotropic media: plane of refraction is the same as plane of incidence + angular relation \(n_1 \sin \theta_1 = n_2 \sin \theta_2\)
- Refraction of a plane wave at a boundary between isotropic and anisotropic medium:
Figure 8  Example of the determination of the rays refracted from an isotropic to an isotropic medium using Descartes’ construction. In this example the medium is taken as uniaxial with \( n_x = n_y = n_0 \) and \( n_z = n_e \). The plane of incidence corresponds to \( x = 0 \).

Based on continuity of tangential component of the \( \mathbf{k} \) vector.

- Descartes’ construction (see Fig. 8)
- Huygens’ construction

2.5 Summary

Throughout this section, the following important results have been established:

1. For a given direction of propagation, there exist two linearly polarised plane waves that are solutions of the propagation equation.

2. These two solutions are furthermore orthogonal to each other. They are known as the eigenmodes, or normal modes of propagation.

3. The two normal modes propagate with different phase velocities or, equivalently, experience different refractive indices, a concept known as birefringence.

4. The directions of the \( \mathbf{D} \) vectors of the normal modes can be found using the index ellipsoid construction. They correspond to the axes of the intersection ellipse between the index ellipsoid and the plane normal to the direction of propagation. The associated refractive indices correspond to the semi-axes of this ellipse.

5. Since the normal modes are linearly polarised and orthogonal, they can be used as a basis to decompose a plane wave of any polarisation that is incident on the anisotropic medium. As the polarisation of the normal modes is preserved and the refractive index they experience is known, it becomes possible, by linearity, to calculate the electromagnetic field associated to the wave anywhere within the medium.
3 The linear electro-optic effect

The isotropy or anisotropy of a medium can be modified under the action of external constraints. Some isotropic media can become anisotropic under these constraints, while the properties of some media that are already anisotropic can be modified. This induced anisotropy can arise from the application of an external electric field, an external magnetic field, or some mechanical constraints that will change the density of the material, hence its refractive index. The corresponding modification of the anisotropic properties of the medium are known as electro-optic, magneto-optic, and photoelastic or elasto-optic effects, respectively. The acousto-optic effect, which can be exploited to perform light modulation at relatively low speeds, is actually a photoelastic effect where the required changes of refractive index are induced by an acoustic wave propagating through the medium. In the following, the electro-optic effect will be considered since it is commonly used to perform high speed modulation of light with bandwidths of the order of several tens of gigahertz.

It has been seen in Sect. 2.3 that the anisotropy of a medium can be characterised through its index ellipsoid. Assuming a coordinate system that coincides with the principal axes of the dielectric tensor of the medium, the index ellipsoid can be written

\[
\frac{x^2}{n_x^2} + \frac{y^2}{n_y^2} + \frac{z^2}{n_z^2} = 1.
\]  

To simplify the notations, the equation of the index ellipsoid can also be expressed according to

\[
B_1^0 x^2 + B_2^0 y^2 + B_3^0 z^2 = 1,
\]  

where the superscript 0 denotes the initially unperturbed state and

\[
B_1^0 = \frac{1}{n_x^2}, \quad B_2^0 = \frac{1}{n_y^2}, \quad B_3^0 = \frac{1}{n_z^2}.
\]

Under the influence of an external applied electric field \(E_a\), the anisotropy of the medium is modified, resulting in a deformation of the index ellipsoid. In the most general case, the equation of the deformed index ellipsoid can be written

\[
B_1 x^2 + B_2 y^2 + B_3 z^2 + 2B_4 yz + 2B_5 xz + 2B_6 xy = 1.
\]  

Assuming that the variation of the \(B_i\) coefficients has a linear dependence on the applied electric field

\[
\begin{pmatrix}
B_1 - B_1^0 \\
B_2 - B_2^0 \\
B_3 - B_3^0 \\
B_4 \\
B_5 \\
B_6
\end{pmatrix} =
\begin{pmatrix}
r_{11} & r_{12} & r_{13} \\
r_{21} & r_{22} & r_{23} \\
r_{31} & r_{32} & r_{33} \\
r_{41} & r_{42} & r_{43} \\
r_{51} & r_{52} & r_{53} \\
r_{61} & r_{62} & r_{63}
\end{pmatrix}
\begin{pmatrix}
E_{ax} \\
E_{ay} \\
E_{az}
\end{pmatrix},
\]

where \([r]\) is known as the electro-optic tensor. This type of electro-optic effect having a linear dependence on the applied electric field is known as the Pockels effect\(^{22}\).

\(^{22}\)Friedrich Carl Alwin Pockels, Vicenza, 1865 – Heidelberg 1913.
Remark: on the notations

The relation between the change of the coefficients of the equation of the index ellipsoid and the applied external field has been written in (164) according to

\[ \Delta B = B(E_a) - B(E_a = 0) = [\eta] E_a. \]

Since the \( B_i \) are the coefficients of the quadratic form associated with the ellipsoid index, they can be identified with the coefficients of the impermeability tensor

\[ [\eta] = \begin{pmatrix} \eta_{11} & \eta_{12} & \eta_{13} \\ \eta_{21} & \eta_{22} & \eta_{23} \\ \eta_{31} & \eta_{32} & \eta_{33} \end{pmatrix}. \]

Indeed, the equation of the index ellipsoid is

\[ \eta_{ij} x_i x_j = 1, \]

where Einstein’s summation rule over repeated indices has been applied.

Without any prior knowledge of the physical phenomena involved, the tensor \( \Delta \eta_{ij} \) has 9 elements and the vector \( E_a \) has three components. Therefore the relation linking \( \Delta \eta_{ij} \) to \( E_a \) should be written

\[ \Delta \eta_{ij} = r_{ijk} E_k, \]

where the summation over \( k \) is implicit and where \( r_{ijk} \), which is known as a third rank tensor, should have 27 elements. However, the study of the physics of the problem leads to the conclusion that \( \Delta \eta \) is symmetric and therefore has only 6 distinct elements. Consequently \( r_{ijk} \) only possess 18 distinct elements (this number can be reduced further by introducing more physical considerations and examining the symmetries of the crystal). It then becomes possible to adopt the so-called contracted notations for \( ij \)

\( 11 \rightarrow 1, \quad 23 \rightarrow 4, \quad 22 \rightarrow 2, \quad 13 \rightarrow 5, \quad 33 \rightarrow 3, \quad 12 \rightarrow 6 \)

resulting in

\[ \Delta B_l = r_{lk} E_{ak}, \]

which corresponds to (164).

3.1 Determination of the refractive indices under an applied electric field

In this section, the general procedure to determine the directions of the normal modes and associated refractive indices is illustrated on a particular example. Here the classic case of a crystal of potassium dihydrogen phosphate (KH\(_2\)PO\(_4\)), known in short as KDP, is considered. KDP is a uniaxial crystal whose index ellipsoid can be written in its principal axes \( (x, y, z) \)

\[ \frac{x^2}{n_o^2} + \frac{y^2}{n_a^2} + \frac{z^2}{n_e^2} = 1, \]

\(23\) The coefficient 2 in front of the cross products \( x_i x_j \) where \( i \neq j \) in the equations of the index ellipsoid such as (163) appears due to the symmetry of \([\eta]\).
where \( n_o = 1.5074 \) and \( n_e = 1.4669 \) at \( \lambda = 633 \) nm. It belongs to the \( \bar{4}2m \) class for which symmetry considerations would enable to show that the electro-optic tensor can be expressed as

\[
\mathbf{r}_{\bar{4}2m} = \begin{pmatrix}
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & r_{41} \\
0 & r_{41} & 0 \\
0 & 0 & r_{63}
\end{pmatrix}.
\]  

(171)

In the case of KDP, the values of the electro-optic coefficients at 633 nm are \( r_{41} = 8 \times 10^{-12} \) m/V and \( r_{63} = 11 \times 10^{-12} \) m/V. Under an applied external electric field \( E_a \), the index ellipsoid becomes

\[
\frac{x^2}{n_o^2} + \frac{y^2}{n_o^2} + \frac{z^2}{n_e^2} + 2r_{41}E_{ax}yz + 2r_{41}E_{ay}xz + 2r_{63}E_{az}xy = 1.
\]  

(172)

Eq. (172) shows that the principal axes of the crystal have been rotated by applying an external electric field. In order to find out the new principal axes and associated refractive indices, it is necessary to diagonalise the quadratic form associated to the new index ellipsoid (172). For simplicity, the analysis is carried further in the case of an external electric field applied in the \( z \) direction. The index ellipsoid thus becomes

\[
\frac{x^2}{n_o^2} + \frac{y^2}{n_o^2} + \frac{z^2}{n_e^2} + 2r_{63}E_{az}xy = 1,
\]  

(173)

and the matrix of its associated quadratic form can be expressed in the basis \((x, y, z)\) as

\[
\mathbf{Q} = \begin{pmatrix}
B_1 & B_6 & 0 \\
B_6 & B_1 & 0 \\
0 & 0 & B_3
\end{pmatrix},
\]  

(174)

where \( B_1 = 1/n_o^2, B_3 = 1/n_e^2 \) and \( B_6 = r_{63}E_{az} \). The eigenvalues of \( \mathbf{Q} \) can be found by solving the equation

\[
\begin{vmatrix}
B_1 - \lambda & B_6 & 0 \\
B_6 & B_1 - \lambda & 0 \\
0 & 0 & B_3 - \lambda
\end{vmatrix} = 0,
\]  

(175)

leading to

\[
(B_3 - \lambda) \left[ (B_1 - \lambda)^2 - B_6^2 \right] = 0.
\]  

(176)

Hence the eigenvalues

\[
\lambda_1 = B_1 + B_6, \quad (177)
\]

\[
\lambda_2 = B_1 - B_6, \quad (178)
\]

\[
\lambda_3 = B_3. \quad (179)
\]

In order to define the directions of the axes of the deformed index ellipsoid, it is necessary to find the eigenvectors associated to the eigenvalues (177)-(179). This step is detailed below for each of the eigenvalues.

- \( \lambda_1 = B_1 + B_6 \)

The eigenvectors can be found by solving the system:
Anisotropic Media and the Linear Electro-Optic Effect

\[
\begin{align*}
\begin{cases}
    x - y &= 0 \\
    (B_3 - B_1 - B_6) z &= 0
\end{cases} \\
(180)
\end{align*}
\]

Hence an eigenvector associated with the eigenvalue \( \lambda_1 \) is \( \mathbf{x}' = \frac{1}{\sqrt{2}} (x + y) \).

- \( \lambda_2 = B_1 - B_6 \)

The eigenvectors can be found by solving the system:

\[
\begin{align*}
\begin{cases}
    x + y &= 0 \\
    (B_3 - B_1 + B_6) z &= 0
\end{cases} \\
(181)
\end{align*}
\]

Hence an eigenvector associated with the eigenvalue \( \lambda_2 \) is \( \mathbf{y}' = \frac{1}{\sqrt{2}} (-x + y) \).

- \( \lambda_3 = B_3 \)

The eigenvectors can be found by solving the system:

\[
\begin{align*}
\begin{cases}
    (B_1 - B_3) x + B_6 y &= 0 \\
    B_6 x + (B_1 - B_3) y &= 0 \\
    (B_3 - B_3) z &= 0
\end{cases} \\
(182)
\end{align*}
\]

Hence an eigenvector associated with the eigenvalue \( \lambda_3 \) is \( \mathbf{z}' = \mathbf{z} \).

The transformation matrix from the old basis \((\mathbf{x}, \mathbf{y}, \mathbf{z})\) to the new basis \((\mathbf{x}', \mathbf{y}', \mathbf{z}')\) can be obtained by expressing the components of the vectors of the new basis on the old basis, leading to

\[
P = \begin{pmatrix}
\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\
\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\
0 & 0 & 1
\end{pmatrix},
\]

which corresponds to the matrix of a rotation of angle \( \pi/4 \) around \( \mathbf{z} \).

In the new coordinate system, the equation of the deformed ellipsoid index is

\[
\lambda_1 x'^2 + \lambda_2 y'^2 + \lambda_3 z'^2 = 1,
\]

which can also be expressed in terms of the refractive indices along the axes of the deformed ellipsoid

\[
\frac{x'^2}{n_{x'}^2} + \frac{y'^2}{n_{y'}^2} + \frac{z'^2}{n_{z'}^2} = 1,
\]

where \( n_{x'} \), \( n_{y'} \) and \( n_{z'} \) are obtained through the relations

\[
\begin{align*}
\frac{1}{n_{x'}^2} &= \frac{1}{n_o^2} + r_{63} E_{az} \\
(186) \\
\frac{1}{n_{y'}^2} &= \frac{1}{n_o^2} - r_{63} E_{az} \\
(187) \\
\frac{1}{n_{z'}^2} &= \frac{1}{n_e^2} \\
(188)
\end{align*}
\]

From (186)

\[
n_{x'} = \frac{n_o}{(1 + n_o^2 r_{63} E_{az})^{1/2}} \approx n_o \left( 1 - \frac{1}{2} n_o^2 r_{63} E_{az} \right)
\]

(189)

07/10/2010
where a first order Taylor expansion has been performed since the order of magnitude of \( r_{63} \) is \( 10^{-12} \). In a similar fashion, it is possible to derive an expression for \( n_y' \). Hence the values of the semi-axes of the deformed index ellipsoid:

\[
\begin{align*}
\text{and with an external applied field parallel to the } \mathbf{z} \text{ direction.}
\end{align*}
\]

The configuration of the axes of the index ellipsoid of KDP with and without an external applied field are summarised in Fig. 9. In the following sections, it will be investigated how the refractive index experienced by a light wave propagating in a KDP crystal can be controlled by the applied external electric field.

### 3.2 Longitudinal configuration

First, the configuration depicted in Fig. 10 is examined. The external electric field \( \mathbf{E}_a \) is applied along the \( \mathbf{z} \) direction of the crystal while the light wave propagates in the \( \mathbf{z} \) direction. Such a configuration is known as longitudinal since the external field is applied in the direction of light propagation. From (185), the intersection of the ellipsoid index with the plane orthogonal to \( \mathbf{z} \) is

\[
\begin{align*}
\frac{x'^2}{n_x'^2} + \frac{y'^2}{n_y'^2} &= 1, \\
\text{and the normal modes of propagation are oriented along } &\mathbf{x}' \text{ and } \mathbf{y}', \text{ with associated refractive indices given by (190) and (191), respectively. Consequently, a light wave propagating along } \mathbf{z} \text{ and linearly polarised along } \mathbf{x}' \text{ will remain linearly polarised along the same direction and will experience the refractive index}
\end{align*}
\]

while a light wave linearly polarised along \( \mathbf{y}' \) will remain linearly polarised along the same direction and experience the index

\[
\begin{align*}
\text{while a light wave linearly polarised along } &\mathbf{y}' \text{ will remain linearly polarised along the same direction and experience the index}
\end{align*}
\]
An arbitrary state of polarisation at the input of the KDP crystal can be decomposed along the \( x' \) and \( y' \) directions since those are orthogonal. For a monochromatic plane wave described by (74), the input electric displacement can be written
\[
D_0(0) = D_{0x'} x' + D_{0y'} y'.
\]  (196)

After propagation through a distance \( z \) in the crystal, the electric displacement becomes
\[
D_0(z) = D_{0x'} e^{-j\frac{2\pi}{\lambda} n_{x'} z} x' + D_{0y'} e^{-j\frac{2\pi}{\lambda} n_{y'} z} y'.
\]  (197)

The components of the displacement along \( x' \) or \( y' \) experience different phase shifts that depend on the refractive index associated with that particular normal mode. This means that the state of polarisation of the input light wave will be modified upon propagation in the crystal, unless it is aligned with \( x' \) or \( y' \). Hence an important quantity is the phase retardation through the crystal, defined as
\[
\Gamma(z) = \frac{2\pi}{\lambda} (n_{y'} - n_{x'}) z.
\]  (198)

The values of \( \Gamma(z) \) will determine the relative phase shifts accumulated by the components of \( D_0 \) along the \( x' \) and \( y' \) axes, hence the state of polarisation of the field at distance \( z \) within the crystal.

If \( l \) is the crystal length, the retardation at the output of the crystal is
\[
\Gamma(l) = \frac{2\pi}{\lambda} n_o^3 r_{63} E_{az} l.
\]  (199)

Since the external electric field is created by applying a voltage \( V \) to the electrodes that are separated with the crystal length, \( V = E_{az} l \), and therefore
\[
\Gamma(l) = \frac{2\pi}{\lambda} n_o^3 r_{63} V.
\]  (200)
which is customarily expressed as

$$\Gamma = \pi \frac{V}{V_\pi},$$

(201)

where the quantity

$$V_\pi = \frac{\lambda}{2n_o^3 r_{63}}$$

(202)

is known as the half wave voltage of the crystal under longitudinal configuration. Applying a voltage equal to $V_\pi$ to the electrodes induces an electro-optic retardation of $\pi$, or equivalently creates an optical path length difference of $\lambda/2$ between the two normal modes. It is important to notice that, under longitudinal configuration, $V_\pi$ only depends on the material parameters $n_o$ and $r_{63}$. At 633 nm, the half-wave voltage is therefore equal to $V_\pi = 8400$ V. This voltage is independent of the geometry of the crystal.

Since in this configuration the direction of the applied electric field is the same as that of light propagation, it should be ensured that the electrodes that are employed to apply the voltage that induces $E_a$ are transparent to the optical wave.

3.3 Transverse configuration

In the so-called transverse configuration the direction of light propagation is orthogonal to the direction of the applied electric field. Fig. 11 illustrates such a configuration where $E_a$ is applied along $z$ and the wave propagates in the $y'$ direction. In this case, waves linearly polarised along $x'$ or $z$ will remain linearly polarised along the same directions and will experience refractive indices equal to

$$n_{x'} = n_o - \frac{1}{2} n_o^3 r_{63} E_{az},$$

(203)

and

$$n_z = n_e,$$

(204)

respectively. Hence the electro-optic retardation at the crystal output

$$\Gamma(l) = \frac{2\pi}{\lambda} (n_z - n_{x'}) l = \frac{2\pi}{\lambda} (n_e - n_0) l + \frac{\pi n_o^3 r_{63} E_{az} l}{\lambda}$$

(205)
Anisotropic Media and the Linear Electro-Optic Effect

The first term corresponds to a fixed retardation induced by the birefringence of the crystal. It could in principle be compensated by a proper choice of the crystal length so that the fixed retardation is an integer multiple of $2\pi$ at the wavelength of interest, or by inserting an appropriate wave plate. In practice, the temperature dependence of the static birefringence $n_e - n_0$ might make such a compensation challenging. This term will nevertheless be ignored in the discussion that follows where the focus is on the induced birefringence.

In the transverse configuration, the external electric field is induced by a voltage applied across the crystal whose thickness is $e$, hence $V = E_{az}e$, and the induced retardation becomes

$$\Gamma = \frac{\pi n_0^3 r_{63} V}{\lambda} \left( \frac{l}{e} \right).$$

Similarly to the longitudinal case, it is possible to define a half-wave voltage

$$V_\pi = \frac{\lambda}{n_0^3 r_{63}} \left( e \right),$$

so that the induced retardation can be expressed according to (201). It is however important to note that, contrarily to the longitudinal case, both the induced retardation and the half-wave voltage depend on the form factor $e/l$. It is therefore possible to reduce $V_\pi$ (i.e. the voltage necessary to induce a retardation equal to $\pi$) by increasing the interaction length $l$ between the applied electric field and the optical wave, or by reducing the electrode spacing $e$. For optical communication applications, guided wave structures will be used to modify the properties of a light wave through the electro-optic effect. Therefore the required waveguide geometry will set a lower limit to $e$. Consequently, the product $V_\pi l$ can be used as a figure of merit for transverse electro-optic modulators. Furthermore, since the external electric field is applied orthogonally to the direction of light propagation, no difficulty arises due to the electrode configuration with respect to the optical path, as is the case for the longitudinal configuration.

Consequently the transverse configuration will be preferred for the practical application of the electro-optic effect to the modulation of light, as described in more details in the following section.

4 Electro-optic modulation

4.1 The electro-optic effect in lithium-niobate

Lithium niobate (LiNbO$_3$) is the material customarily used for light modulation in optical communication applications due to its relatively large electro-optic coefficient and the good control of the technology required to produce low loss optical modulators. LiNbO$_3$ is a uniaxial crystal whose unperturbed index ellipsoid is given, in the principal axes system, by

$$\frac{x^2}{n_0^2} + \frac{y^2}{n_0^2} + \frac{z^2}{n_e^2} = 1,$$

where commonly reported values at 633 nm for the ordinary and extraordinary refractive indices are $n_0 = 2.29$ and $n_e = 2.2$, respectively. It belongs to the 3m crystal point group and its dielectric tensor is therefore given by

$$r_{3m} = \begin{pmatrix}
0 & -r_{22} & r_{13} \\
0 & r_{22} & r_{13} \\
0 & 0 & r_{33} \\
r_{51} & 0 & 0 \\
r_{51} & 0 & 0 \\
r_{22} & 0 & 0
\end{pmatrix}.$$
Typical values for the electro-optic coefficients are\textsuperscript{24}:

\[ r_{22} = 3.4 \text{ pm/V}, \quad r_{13} = 8.6 \text{ pm/V}, \]
\[ r_{31} = 28.0 \text{ pm/V}, \quad r_{33} = 30.8 \text{ pm/V}. \]

Strictly speaking, those values apply at a wavelength of 633 nm\textsuperscript{25}. Even though those coefficients are known to be wavelength dependent, and in the absence of readily available data at the telecommunication wavelength of 1550 nm, these values will be used in the present discussion of LiNbO\textsubscript{4} modulators. More data on the physical properties of lithium niobate, including a comparison of values for the refractive indices and electro-optic coefficients reported by various authors, can be found in published reviews such as [8].

In any case, \( r_{33} \) is the largest coefficient of the electro-optic tensor. For practical modulation applications, it is desirable to induce the largest possible refractive index change for a given applied voltage. In order to exploit \( r_{33} \), the external modulating electric field \( E_a \) should therefore be aligned with the \( z \) direction: \( E_a = E_{az}z \). Under those conditions, the index ellipsoid becomes

\[ \left( \frac{1}{n_o^2} + r_{13}E_{az} \right) x^2 + \left( \frac{1}{n_e^2} + r_{13}E_{az} \right) y^2 + \left( \frac{1}{n_z^2} + r_{33}E_{az} \right) z^2 = 1. \]  \hspace{1cm} (210)

One important observation is that the principal axes of the index ellipsoid of LiNbO\textsubscript{3} are therefore unchanged when an electric field parallel to \( z \) is applied to the crystal.

For practical modulation applications the following external requirements should be satisfied:

1. The applied electric field should be parallel to the \( z \) direction in order to exploit the large \( r_{33} \) electro-optic coefficient.

2. The modulation should be transverse in order to benefit from the form factor \( e/l \) discussed in Sec. 3.3.

Those requirements lead to the three possible configurations depicted in Fig. 12. In those configurations, the light propagates along the longest dimension of the crystal in order to benefit from the form factor. They are customarily referred to as \( x \)-cut and \( z \)-cut. In the \( x \)-cut configuration, the plane from which the waveguide will be implanted and on which the electrodes will be deposited is perpendicular to the \( x \) crystal direction.

\textsuperscript{24} Those values result from high frequency (50–86 MHz) measurements performed at 633 nm and first reported in [7]. Slightly different values can be found in the literature, depending on the measurement conditions and experimental method. Nevertheless the values presented here will be considered sufficiently accurate for the purpose of this note.

\textsuperscript{25} The justification for this wavelength is that it corresponds to the line of helium-neon lasers.
4.1.1 \( x \)-cut configuration

For transverse modulation, the light has to propagate along the \( y \) direction. The refractive index along \( z \) is therefore given by

\[
\frac{1}{n'_z} = \frac{1}{n_z^2} + r_{33}E_{az} = \frac{1}{n_e^2} \left( 1 + r_{33}n_e^2E_{az} \right),
\]

hence, with the usual assumption that \( r_{33}n_e^2E_{az} \ll 1, \)

\[
n'_z = n_e - \frac{1}{2} n_e^3 r_{33}E_{az}.
\]

Similarly, the refractive index in the \( x \) direction can be shown to be

\[
n'_x = n_o - \frac{1}{2} n_o^3 r_{13}E_{az}.
\]

Since the phase shift experienced by a light wave propagating into a length \( l \) of a medium of refractive index \( n \) is equal to \( \phi = \frac{2\pi}{\lambda} nl \), it is clear that for efficient phase modulation the light wave should be polarised along the \( z \) direction. Consequently, in the \( x \)-cut configuration, efficient phase modulation is performed for the transverse electric (TE) mode.

4.1.2 \( z \)-cut configuration

In the \( z \)-cut configurations, the top plane of the crystal is orthogonal to the \( z \) direction, as illustrated in Fig. 12. Two possibilities exist for the direction of light propagation: either along \( x \) or along \( y \). In the case of propagation along \( x \), the principal refractive indices under an applied electric field along \( z \) are

\[
n'_y = n_o - \frac{1}{2} n_o^3 r_{13}E_{az},
\]

\[
n'_z = n_e - \frac{1}{2} n_e^3 r_{33}E_{az},
\]

while if propagation takes place in the \( y \) direction, the indices

\[
n'_x = n_o - \frac{1}{2} n_o^3 r_{13}E_{az},
\]

\[
n'_z = n_e - \frac{1}{2} n_e^3 r_{33}E_{az},
\]

should be considered. In both cases, it can be seen that efficient phase modulation takes place when the electric field of the light wave is polarised along the \( z \) direction, which corresponds to the transverse magnetic (TM) mode of the waveguide.

Fig. 13 shows cross sections of practical structures used for electro-optic modulation in LiNbO\(_3\). The light is confined to a waveguide that is realised by locally raising the refractive index of the lithium niobate substrate. This is usually achieved by selectively diffusing titanium into the crystal. Photolithographic techniques are used to deposite a Ti film whose width and length are similar to those of the final waveguide on top of the crystal. The sample is then heated around 1000°C for several hours, which allows the titanium to diffuse into the crystal and raises its refractive index. This process results in waveguides where the relative refractive index difference between the core and the cladding is of the order of \( \Delta \sim 0.5-1\% \) and the loss is less than 0.2 dB/cm. A buffer layer made of silica (SiO\(_2\)) is then deposited on top of the crystal. This layer prevents the optical field from extending into the metal electrodes, which
would significantly increase the propagation loss. It also enables to increase the characteristic impedance of the electrodes to which the radio-frequency (RF) modulating signal will be applied, so that impedance matching with the signal source can be improved. Finally, gold electrodes are deposited on top of the buffer layer using photolithographic patterning techniques.

The electrodes are configured in order to align the applied electric field with the \( z \) direction and their position with respect to the waveguide therefore depends on the orientation of the crystal, as represented schematically in Fig. 13. In the \( x \)-cut configuration, the electrodes should be placed on both sides of the waveguide so that the field lines of \( E_a \) are parallel to the \( z \) direction (i.e. horizontal in the figure). On the other hand, in the \( z \)-cut configuration, the hot electrode should be placed directly on top of the waveguide so that the field lines are vertical over the waveguide area. The higher concentration of the applied electric field over the waveguide in the case of a \( z \)-cut modulator, resulting in a better overlap with the optical field, is expected to lead to a lower required voltage for achieving a given refractive index change, as compared with the \( x \)-cut configuration.

More details on the technology of lithium niobate modulators can be found in a number of review articles such as [9, 10]. More advanced structures suitable for efficient high speed modulation are described in a more recent review [11].

4.2 Light modulation

The purpose of optical communication is to transmit information in analogue or digital form over a medium, which can be either an optical fibre or free space. This purpose is achieved by changing one or several of the properties of a light wave, such as its intensity, phase, frequency or polarisation, according to the information to be transmitted. A key functionality of an optical system is therefore the modulation operation, where the information signal that is available under electrical form is transferred to the optical domain by modifying one or several of the attributes of an incoming light wave.

A general expression for the electric field\(^{26}\) after the modulation operation is

\[
E(t) = \sqrt{P(t)} e^{-j\phi(t)} e^{j\omega_0 t} e(t),
\]

where the quantities that can be modulated are the power \( P(t) \), the phase \( \phi(t) \), or the polarisation, which is described by a complex vector \( e(t) \). The carrier frequency \( \omega_0 \) could also be modulated. It is possible to express this shift in carrier frequency through the time varying

\(^{26}\)Note that for simplicity the complex notation has been adopted without the explicit notation \( \text{Re}\{\ldots\} \). It is nevertheless implicitly assumed that the real part of the right hand side of expressions such as (218) should be considered.
Figure 14  Principle of an electro-optic phase modulator illustrated here for a z-cut LiNbO$_3$ crystal.

phase term $\phi (t)$, according to

$$\omega (t) = \omega_0 - \frac{\partial \phi}{\partial t} (t),$$  \hspace{1cm} (219)

where $\omega (t)$ is the instantaneous frequency.

It has been seen in the previous sections how the refractive index experienced by a light wave can be changed by applying an external field to an electro-optic material. This refractive index change has been shown to depend on the orientation of the applied external field with respect to the principal axes of the crystal, as well as on the polarisation of the optical field. In the following, it will be shown how this property can be exploited to modulated the phase, polarisation or intensity of a light wave.

4.3 Phase modulation

The application of the electro-optic effect to phase modulation is straightforward, as represented in Fig. 14 for a z-cut LiNbO$_3$ crystal. If the incoming light is linearly polarised along the $z$ direction, which is one of the two eigenpolarisations of the crystal, it will remain linearly polarised along the same direction over propagation through the crystal and will experience the phase shift

$$\phi = \frac{2\pi}{\lambda} n_z' l,$$ \hspace{1cm} (220)

where $l$ is the crystal length and $n_z'$ is given by (217). The alignment of the input state of polarisation with the direction $z$ can be ensured by using a polarisation controller, or by placing a polariser parallel to $z$ in front of the crystal, as represented in Fig. 14. Since $n_z'$ depends on the external electric field, which is itself created by a time varying voltage $V (t)$ applied across the crystal, the phase of the light wave will be modulated according to $V (t)$ after it has propagated through the crystal. Neglecting a constant phase shift, and since $V = -E_a e$, where $e$ is the crystal thickness, the electro-optically induced phase variation can be written

$$\phi (t) = \pi \frac{V (t)}{V_\pi},$$  \hspace{1cm} (221)

where the half-wave voltage for the present configuration depends on the electro-optic coefficient according to

$$V_\pi = \frac{\lambda}{n_z^2 r_{33}} \left( \frac{e}{l} \right).$$  \hspace{1cm} (222)

Note that (222) assumes a uniform distribution of the electric field over the waveguide and moreover does not take into account the transverse distribution of the optical mode in
that some trade-offs between voltage requirement and applied electric fields \[12, 13\]

\[
\gamma = \frac{\epsilon}{V} \int \frac{E^2(x, z)}{\int E^2(x, z) \, dx \, dz} \, dx \, dz,
\]

so that

\[
V_\pi = \frac{\lambda}{n_i^2 r_{33} \gamma} \left( \frac{\epsilon}{l} \right).
\]

Assuming \(\epsilon = 8 \mu m\) and \(\gamma = 0.5\), and using the values \(n_i = 2.14\) and \(r_{33} = 30.8 \text{ pm}/V\), leads to a figure of merit of \(V_\pi l = 8.2 \text{ V} \cdot \text{cm}\) for a \(z\)-cut LiNbO\(_3\) modulator at 1550 nm. In practice one would like to operate the modulator with as low a voltage as possible by increasing the device length. However, it will be shown in Sec. 5 that some trade-offs between voltage requirement and device length need to be found for high-frequency modulation applications.

### 4.4 Polarisation modulation

As discussed in Sec. 3.2, the state of polarisation of an optical field which is not linearly polarised along one of the principal axes can be controlled by applying an external electric field to an electro-optic crystal. For illustration, the case of a \(z\)-cut LiNbO\(_3\) crystal is considered in the configuration represented in Fig. 15. The incoming light wave is linearly polarised at an angle of \(45^\circ\) with the directions \(x\) and \(z\), so that its displacement vector can be expressed as

\[
D_0(0) = D_0 \hat{i},
\]

where \(\hat{i} = \frac{1}{\sqrt{2}} (\hat{x} + \hat{z})\) corresponds to its direction of polarisation. At the output of the crystal, the field can be written

\[
D_0(l) = \frac{D_0}{\sqrt{2}} e^{-j \frac{2 \pi}{\lambda} n'_i l} \hat{z} + \frac{D_0}{\sqrt{2}} e^{-j \frac{2 \pi}{\lambda} n'_i l} \hat{x},
\]

or equivalently

\[
D_0(l) = \frac{D_0}{\sqrt{2}} \left( e^{-j \frac{\pi}{\lambda} \hat{z}} + e^{j \frac{\pi}{\lambda} \hat{x}} \right) e^{-j \frac{\pi}{\lambda} (n'_z + n'_x) l},
\]

where the retardation \(\Gamma\) is

\[
\Gamma = \frac{2 \pi}{\lambda} (n'_z - n'_x) l.
\]
Eq. (227) enables to directly express the field at the output of the crystal by mean of its Jones vector, expressed in the \((z, x)\) basis according to
\[
\mathbf{J}_{(z, x)}(l) = \frac{1}{\sqrt{2}} \begin{pmatrix} e^{-j\frac{\Gamma}{2}} \\ e^{j\frac{\Gamma}{2}} \end{pmatrix},
\]  
(229)
from which its state of polarisation can be determined, depending on the value of \(\Gamma\). Using (216) and (217), the retardation can be expanded into
\[
\Gamma = \frac{2\pi}{\lambda} (n_e - n_o) l + \frac{\pi}{\lambda} \left( n_e^3 r_{33} - n_o^3 r_{13} \right) \left( \frac{l}{e} \right) V,
\]  
(230)
where the external field across the crystal has been related to the applied voltage according to \(V = -E_{ax} e\). As in the case of KDP studied in Sec. 3.3, the retardation consists of two terms. The first term is a static term, \(\Gamma_0\), that results from the natural birefringence of the crystal. The second term corresponds to the electro-optically induced birefringence and can be written under the usual form \(\Gamma_{eo} = \pi V/V_\pi\), where the half-wave voltage is equal to
\[
V_\pi = \frac{\lambda}{n_e^3 r_{33} - n_o^3 r_{13}} \left( \frac{e}{l} \right).
\]  
(231)
Eq. (230) shows that the natural birefringence can be compensated in practice by adding a constant d.c. bias to the modulating voltage. If the voltage applied to the crystal is expressed as \(V = V_0 + V_s\), where \(V_0\) is a d.c. offset and \(V_s\) is the modulating signal, the condition for compensation of the natural birefringence can be expressed as \(\Gamma_0 + \pi V_0/V_\pi = k2\pi\). When such a condition is satisfied, the state of polarisation of the incoming beam is not modified when \(V_s = 0\) V. On the other hand, if a modulating voltage equal to \(V_s = V_\pi\) is applied to the crystal, \(\Gamma_{eo} = \pi\), and the displacement at the output of the crystal becomes
\[
D_0(l) = \frac{D_0}{\sqrt{2}} (x - z) e^{-j\frac{\Gamma}{2} (n'_z + n'_x)} e^{j\frac{\Gamma}{2}},
\]  
(232)
so that the polarisation of the light wave at the output of the crystal is linear and parallel to \(j = \frac{1}{\sqrt{2}} (x - z)\), which is a vector orthogonal to \(i\). It is convenient to express the output field in the general case in the basis \((i, j)\). Starting from (227), one immediately obtains
\[
D_0(l) = D_0 \left( \cos \frac{\Gamma}{2} i + j \sin \frac{\Gamma}{2} j \right) e^{-j\frac{\Gamma}{2} (n'_z + n'_x)},
\]  
(233)
Consequently, provided the static birefringence has been adequately compensated:

- if \(V_s = 0\) V, the state of polarisation of the light at the output of the crystal is the same as the one at the input, i.e. linear and parallel to \(i\).
- if \(V_s = V_\pi\), the state of polarisation of the light at the output of the crystal is linear and orthogonal to \(i\).

Polarisation modulation between two orthogonal linear states parallel to \(i\) and \(j\), respectively, has therefore been achieved from a binary electrical signal taking the values \(\{0, V_\pi\}\). In the context of optical communications, such a modulation is known as polarisation shift keying (PolSK).

27This term could in principle be compensated by properly adjusting the length of the crystal so that the static retardation is made equal to \(k2\pi\), where \(k\) is an integer. However, this would require a precision of the order of the wavelength for the cleaving of the crystal.
In the basis \((i, j)\), the Jones vector at the output of the crystal can be written

\[
\mathbf{J}_{(i,j)}(l) = \begin{pmatrix}
\cos \frac{\Gamma}{2} \\
\sin \frac{\Gamma}{2}
\end{pmatrix},
\]

which corresponds to the expression of a general elliptical state of polarisation having an ellipticity \(\Gamma/2\) expressed in the coordinate system that corresponds to the axes of the ellipse. Therefore, depending on the value of the applied voltage, hence \(\Gamma\), all elliptical states of polarisation with axes \(i, j\) can be obtained at the output of the crystal starting from an input that is linearly polarised along \(i\).

### 4.5 Intensity modulation

So far it has been shown how the linear electro-optic effect can be used to modify the refractive index of a crystal according to an applied electric field. This property can be used in a straightforward way to realise a phase modulator. Utilising the induced birefringence of an electro-optic crystal can also lead to polarisation modulation. However, optical communication systems relying on either phase or polarisation modulation are difficult to implement in practice. At the receiver side, the most straightforward detection scheme, known as direct detection, consists of a photodiode that only responds to the power of the light that is incident upon it. A photodiode produces a current whose intensity \(i\) is proportional to the received optical power \(P\) according to \(i(t) = RP(t)\), where \(R\) is known as the responsivity of the photodiode. Consequently, all information contained in the phase or the polarisation of the incoming light wave is lost in such a detection process. The phase information can nevertheless be recovered by using a coherent detection scheme where the incoming optical signal beats with light emitted by a local oscillator laser located at the receiver. Such schemes have been the object of intense investigations in the late 80’s and early 90’s. They are however difficult to implement in practice, in part due to stability requirements and polarisation sensitivity. Nevertheless, thanks to the significant progress accomplished over the past decade, a revival of such techniques is currently observed in a number of research laboratories worldwide. It will be seen later in this course how the difficulty inherent to phase detection can be overcome by converting first the phase modulated signal into an intensity modulated signal, followed by direct detection.

Polarisation detection also requires some complex receiver structures, or the use of a polarisation discriminating element, such as a polariser, followed by direct detection. However, polarisation effects in optical fibres also make such schemes difficult to implement in practice. Consequently, the modulation of the intensity of a light wave, in conjunction with direct detection, is so far the preferred method for transmitting information in optical communication systems. Since the electro-optic effect only affects the phase of the signal, some techniques need to be found to achieve intensity modulation. This can be performed by converting either polarisation modulation or phase modulation into intensity modulation.

#### 4.5.1 Polarisation modulator followed by a polariser

The first scheme, represented in Fig. 16, simply consists of the polarisation modulator structure studied in Sec. 4.3, followed by a polariser aligned with the \(j\) direction. When no modulating voltage is applied to the crystal, and provided its natural birefringence is properly compensated, the state of polarisation of the light at the output of the crystal is parallel to \(i\), hence orthogonal to \(j\), and no light is transmitted by the polariser. On the other hand, when a modulating voltage equal to \(V_0\) is applied, the output state of polarisation becomes parallel to \(j\) and the integrality

---

\(^{28}\)See section on “Phase modulation”. 

---

07/10/2010
Anisotropic Media and the Linear Electro-Optic Effect

Intensity modulator based on a polarisation modulator followed by a polariser.

![Image](image1)

**Figure 16** Intensity modulator based on a polarisation modulator followed by a polariser.

![Image](image2)

**Figure 17** Transfer function of an intensity modulator based on the configuration represented in Fig. 16, and operating points for digital and analogue modulation.

of the light is transmitted through the polariser. In the most general case, an expression for the field at the output of the polariser can be immediately obtained from (233)

\[ D_{0j} = D_0 \sin \frac{\Gamma}{2} e^{-j\frac{\pi}{2}(n'_x + n'_y)} e^{j\frac{\pi}{2}j}, \]  

hence the power at the output of the polariser

\[ P_{out} = P_0 \sin^2 \frac{\Gamma}{2} = P_0 \sin^2 \left( \frac{\Gamma_0}{2} + \frac{\pi}{2} \frac{V}{V_r} \right), \]  

which is modulated according to the applied voltage \( V \). Fig. 17 shows the power versus voltage characteristic, also known as transfer function of such a modulator. If the natural birefringence of the crystal is exactly compensated, and if the applied voltage is a digital signal that takes the values \( \{0, V_r\} \), the optical signal will be intensity modulated and will take the power values \( \{0, P_0\} \). It is clear that the relation (236) between the optical power and the electrical signal voltage is not linear. This is usually not a concern for digital modulation applications, where
the modulator will be operated between a minimum and a maximum of its transfer function, but might however be a concern for analogue applications, since it will result in the creation of undesired harmonics. In this case, one can exploit the portion of the transfer function curve that can approximately be considered as linear by operating the modulator around a biasing point equal to $V_\pi/2$, also known as quadrature point, and by reducing the peak-to-peak modulation voltage so that the linear approximation remains valid. Since an applied voltage of $V_\pi/2$ corresponds to a retardation of $\pi/2$, adding a quarter wave plate whose axes are aligned to $x$ and $y$, which is known as optical biasing, is equivalent to applying a suitable electrical bias.

### 4.5.2 Interferometric structure

It is well known that the light intensity at the output of a two-beam interferometer depends on the relative phase shifts experienced by the two beams after they have been separated from a common primary beam. If those phases shifts are equal (modulo $2\pi$), then constructive interference occurs when the two beams are recombined. Out-of-phase beams will then result in destructive interference. The electro-optic effect enables to control the refractive index of a crystal by an external applied electric field. A change in refractive index results in a change of the phase shift experienced by a light beam propagating through the crystal, a principle that has been used to design a phase modulator in Sec. 4.3. It is therefore natural to use an interferometric structure to convert the electro-optically induced refractive index change into a modulation of the intensity of the incoming light beam.

A Mach–Zehnder interferometer structure such as the one represented in Fig. 18 is customarily used for this purpose. The primary beam $E_0$ is spatially separated by a semi-reflecting surface. The two beams experience different paths before being recombined at another semi-reflecting surface. One of the beams will propagate through an electro-optic crystal in a phase-modulator configuration such as the one represented in Fig. 14. If $E_1$ and $E_2$ denote the complex envelopes of the electric fields of the two beams after the semi-reflecting surface where they recombine, and assuming the fields have identical polarisation, the total envelope can be written

$$E(t) = E_1(t) + E_2(t),$$

(237)

---

29 Ludwig Mach (son of Ernst Mach, Chirlitz-Turas, now Chrliče-Turany, 1838 – Vatterstetten, 1916.)


31 The complex envelope of a field $\mathbf{E}(t) = E_0(t) e^{-j\phi(t)} e^{j\omega_0 t}$ is defined as the quantity $E(t) = E_0(t) e^{-j\phi(t)}$. 

---

07/10/2010
Anisotropic Media and the Linear Electro-Optic Effect

which can be expanded into

$$E(t) = E_{01} e^{-j\phi_1} + E_{02} e^{-j\phi_2(t)}, \quad (238)$$

where $\phi_1$ and $\phi_2(t)$ are the phase shifts experienced by the upper and lower beams, respectively. The amplitudes $E_{01}$ and $E_{02}$ can be related to the amplitude $E_0$ of the incoming beam once the loss and reflection coefficients of the various surfaces involved for the separation and recombination of the beams are known. Hence the power at the output of the interferometer

$$P_{\text{out}}(t) = |E(t)|^2 = P_1 + P_2 + 2\sqrt{P_1 P_2} \cos[\phi_1 - \phi_2(t)], \quad (239)$$

where $P_1 = E_{01}^2$ and $P_2 = E_{02}^2$. The output power therefore depends on the relative phase shift experienced by the two beams. This phase shift depends itself on the path length difference, reflections on the mirrors, as well as on the electro-optically induced refractive index change in the crystal. Assuming the constant phase shifts can be compensated, either by proper delay adjustment of by the use of compensators, the phase difference experienced by the two beams simply depends on the electro-optic effect through $\phi_2(t) - \phi_1 = \pi V(t)/V_\pi$, where $V_\pi$ is the half-wave voltage corresponding to the electro-optic configuration employed in the phase modulator. If furthermore the power of the two beams can be made equal when they recombine, i.e. $P_0 = P_1 = P_2$, then the power at the output of the interferometer becomes

$$P_{\text{out}}(t) = P_0 \cos^2\left(\frac{\pi V(t)}{V_\pi}\right), \quad (240)$$

which is similar to the expression obtained in (236) for the intensity modulator structure based on a polarisation modulator followed by a polariser. Consequently, the same issues regarding the linearity of the modulator will be raised when such an interferometric structure is used. In practice, Mach-Zehnder modulators are employed to perform intensity modulation of light for optical communication applications. Those devices are based on the Mach-Zehnder interferometer principle outlined here. They are however integrated on a lithium-niobate crystal substrate where the interferometer structure is made of waveguides and Y-junctions, hence enabling the realisation of relatively compact and stable devices. This type of modulators, as well as their use for the generation of various modulation formats in optical communication systems, will be described in detail in a separate section32.

5 High frequency electro-optic modulation

So far it has been assumed that the physical mechanisms leading to the electro-optic effect is instantaneous and that no bandwidth limitation arises from the application of the electric field. Currently, digital optical communication systems operate typically at 10 Gbit/s and next generation 40 Gbit/s systems are expected to be deployed soon. Future 100 Gbit/s solutions are also being investigated. At all those bit rates, the first approximation is fully justified. However, the electrodes that are used to apply the modulation voltage across the electro-optic crystal nevertheless limit the bandwidth of electro-optic modulators. The possible origin of such bandwidth limitations is discussed in this section.

A first obvious limitation results from the basic structure of electro-optic modulators considered so far. As illustrated for instance in Fig. 11, such a modulator consists of an electro-optic crystal, which exhibits insulating properties, sandwiched between two electrodes. Consequently it is essentially a parallel plate capacitor whose capacitance depends on the dielectric constant

32See section on “The Mach-Zehnder modulator”.

07/10/2010 45
of the electro-optic material $\varepsilon$, as well as on the surface and separation of the electrodes, $A$ and $\varepsilon$, respectively, according to $C = \varepsilon A/e$. In order to match the impedance of the electrodes to that of the source and the transmission line that will provide the modulating signal, a load resistance $R_L$ is used to connect the two electrodes. Consequently, seen from the source, the modulator is an $RC$ circuit whose $3$ dB bandwidth is therefore equal to

$$f_{3\,\text{dB}} = \frac{1}{2\pi R_L C}.$$  \hfill (241)

Therefore the capacitance of the modulator sets an upper limit to the frequency range at which it can be operated. It has been seen in Sec. 3.3 that one of the benefits of the transverse configuration for electro-optic modulators is that it is possible to reduce the voltage requirements in order to achieve a given phase retardation by increasing the electrodes length $l$. Increasing $l$ will also increase the capacitance $C$, hence decrease the modulator bandwidth. The $RC$ limitation described above is however only valid for relatively low modulation frequencies, when the value of the modulating signal can be considered constant over the duration of the propagation of the phase front through the crystal. At higher modulation frequencies, transit time effects will limit the modulator bandwidth.

### 5.1 Transit time limitation

In the case of a longitudinal configuration, it has been shown in Sec. 3.2 that the electro-optic retardation induced by a crystal could be expressed as (199)

$$\Gamma = \frac{\pi}{V_\pi} E_{az} l.$$  \hfill (242)

Eq. (242) assumes that the applied electric field $E_a$ is uniform over the $z$ direction in the crystal. If this is not the case, it is necessary to consider elementary retardations

$$d\Gamma = \frac{\pi}{V_\pi} E_{az} (z) dz,$$  \hfill (243)

and integrate over the crystal length, hence

$$\Gamma = \frac{\pi}{V_\pi} \int_0^l E_{az} (z) dz.$$  \hfill (244)

For a transverse configuration, the induced retardation can be expressed according to (205) by a relation of the type

$$\Gamma = \frac{\pi n_3 r_{63} E_{az} l}{\lambda}.$$  \hfill (245)

Hence, if the value of the applied field is not uniform along the direction of propagation $y'$, the elementary retardation is

$$d\Gamma = \frac{\pi n_3 r_{63} E_{az} (y') dy'}{\lambda}.$$  \hfill (246)

In both longitudinal and transverse cases, it can be seen that the electro-optic retardation induced by an electric field whose value depends on the position along the direction of light propagation in the crystal can be written

$$\Gamma = a \int_0^l E_a (\zeta) d\zeta$$  \hfill (247)

where $a$ is a proportionality factor that depends on the configuration, and $\zeta$ is the position along the longitudinal dimension of the crystal.
Anisotropic Media and the Linear Electro-Optic Effect

Figure 19 Illustration of transit time limitations. Since it takes a certain time for the phase front of an optical wave to propagate through a crystal subjected to a time varying applied electric field, it will experience different values of the applied field at different positions in the crystal, resulting in a reduced efficiency of the electro-optic modulation process.

So far, static applied electric fields have been assumed. However, in practice, for modulation applications, the applied electric field is also a function of time. Fig. 19 represents schematically the progression of a phase front of a light wave along an electro-optic crystal subjected to a time varying external electric field. It is assumed that the applied electric field is uniform all over the crystal length. While it propagates through the crystal, the phase front experiences different values of the external applied field. Consequently it does not experience the same value of the field over the interaction length that would be necessary to achieve a desired electro-optic retardation. This effect is expected to result in a reduced efficiency of the electro-optic modulation process, unless the variations of the external applied field are slow compared to the transit time of the light wave through the modulator. One might therefore expect a low-pass behaviour of electro-optic modulation due to this transit time effect. Consequently, the total retardation experienced by the optical wave will need to be corrected in order to take this effect into account.

In the following, the low-pass nature of the modulation process is confirmed analytically by calculating the retardation $\Gamma (t)$ at the output of the electro-optic crystal at time $t$ assuming that:

- The optical field propagates at the velocity $c/n$ in the crystal.
- At any time, the value of the applied electric field is constant all over the crystal.
- The applied electric field is sinusoidal of angular frequency $\omega_m$.

07/10/2010
The transit time $\tau$ through the crystal is equal to

$$
\tau = \frac{n l}{c}.
$$

(248)

The contribution to the total retardation of the slice of thickness $d\zeta$ positioned at distance $\zeta$ in the crystal is therefore

$$
d\Gamma = a \mathcal{E}_a (t_\zeta) d\zeta.
$$

(249)

Since

$$
\zeta = \frac{c}{n} (t_\zeta - t + \tau),
$$

(250)

the elementary retardation becomes

$$
d\Gamma = a \mathcal{E}_a (t_\zeta) \frac{c}{n} dt_\zeta,
$$

(251)

which can be integrated over the crystal length to provide the total retardation

$$
\Gamma (t) = \frac{ac}{n} \int_{t-\tau}^{t} \mathcal{E}_a (t_\zeta) dt_\zeta.
$$

(252)

Assuming a sinusoidal modulation,

$$
\mathcal{E}_a (t_\zeta) = \mathcal{E}_0 \cos (\omega_m t_\zeta + \varphi),
$$

(253)

it is convenient to introduce the complex notation

$$
\mathcal{E}_a (t_\zeta) = \text{Re} \left\{ \mathcal{E}_0 e^{j\omega_m t_\zeta} e^{j\varphi} \right\}
$$

(254)

in order to evaluate the integral (252), Hence

$$
\Gamma (t) = \text{Re} \left\{ \frac{ac}{n} \mathcal{E}_0 e^{j\varphi} \int_{t-\tau}^{t} e^{j\omega_m t_\zeta} dt_\zeta \right\},
$$

(255)

which can be easily evaluated, leading to

$$
\Gamma (t) = \Gamma_0 \left[ \frac{\sin \left( \frac{\omega_m \tau}{2} \right)}{\omega_m \tau} \right] \cos \left( \omega_m t - \omega_m \frac{\tau}{2} + \varphi \right)
$$

(256)
Anisotropic Media and the Linear Electro-Optic Effect

Figure 21  Reduction of the peak electro-optic retardation induced by transit time limitations: \( \omega_m \) is the modulating signal angular frequency and \( \tau \) is the transit time through the electro-optic medium.

where \( \Gamma_0 = a E_0 l \). The retardation obtained in (256) should be compared to the value that would have been obtained if the transit time effect had not been taken into account

\[
\Gamma(t) = \Gamma_0 \cos (\omega_m t + \phi).
\]  

(257)

The transit time effect results in a reduction of the peak retardation, as can be seen in Fig. 21 where this quantity is plotted as a function of the modulation frequency. The efficiency of the transfer of modulation from the electrical to the optical domains decreases with increasing modulation frequency. This effect can be ignored provided the modulation frequency is sufficiently low, i.e. \( \omega_m \tau/2 \ll \pi \). However, it is clear that the transit time effect sets serious limitations that need to be overcome if one wishes to achieve broadband modulation\(^{33}\). In the next section, it will be seen how a travelling wave modulator can prevent this reduction of peak retardation induced by the finite transit time through the crystal.

5.2 Travelling wave modulator

The transverse configuration discussed in Sec. 3.3 is now considered, since it enables to achieve the desired retardation for an available voltage by increasing the interaction length between the optical wave and the electrical modulating signal. It furthermore permits travelling wave electrode configurations, such as the one depicted in Fig. 23, to be implemented. In such a configuration, the modulating RF field is launched in a travelling wave electrode, so that it is co-propagating with the optical field. Since both optical and electrical fields propagate in the same direction, and provided they experience the same phase velocity, transit time effects are expected to be reduced, as shown in more details below\(^{14, 15}\).

Using the same conventions as in the study of the transit time effect in Sec. 5.1, the sinusoidal modulating electric field is launched at \( \zeta = 0 \) in the electrode and propagates with phase velocity \( v_{tf} \) in the same direction as the optical field. Its value at distance \( \zeta \), which corresponds to the point reached at time \( t_\zeta \) by the phase front of the optical field, is therefore

\[
E_a (\zeta, t_\zeta) = E_0 \cos (\omega t_\zeta - k_{tf} \zeta + \phi),
\]  

(258)

\(^{33}\)For digital communication applications, the modulating signal corresponds to the data stream that needs to be transmitted. This modulating signal has a baseband spectrum whose width depends on the line code being used and the bit rate. It should therefore be ensured that the bandwidth reduction resulting from transit time limitation still enables to preserve the integrity of the modulating signal.
Figure 22  Principle of travelling wave electrode used to reduce transit time limitation in electro-optic modulators.

where \( k_{rf} \) is the propagation constant of the RF field, which can be related to its phase velocity according to

\[
k_{rf} = \frac{\omega_m}{v_{rf}}. \tag{259}
\]

Furthermore, (250) still applies for this geometry and

\[
\zeta = \frac{c}{n} (t_\zeta - t + \tau) = v_\varphi (t_\zeta - t + \tau), \tag{260}
\]

where, as usual, \( v_\varphi = c/n \) denotes the phase velocity of the optical signal. Rearranging (258) and using (260) leads to

\[
E_a (\zeta, t_\zeta) = E_0 \cos \left[ \omega_m t_\zeta \left( 1 - \frac{v_\varphi}{v_{rf}} \right) + \omega_m \frac{v_\varphi}{v_{rf}} (t - \tau) + \varphi \right], \tag{261}
\]

which has a similar form as (253) provided the following substitutions are made

\[
\begin{align*}
\omega_m & \rightarrow \omega_m \left[ 1 - \frac{v_\varphi}{v_{rf}} \right], \tag{262} \\
\varphi & \rightarrow \varphi + \omega_m \frac{v_\varphi}{v_{rf}} (t - \tau). \tag{263}
\end{align*}
\]

Consequently, integrating (251) over the crystal length immediately leads to

\[
\Gamma (t) = \Gamma_0 \left[ \frac{\sin \left( \omega_m \left[ 1 - \frac{v_\varphi}{v_{rf}} \right] \tau \right)}{\omega_m \left[ 1 - \frac{v_\varphi}{v_{rf}} \right] \frac{\tau}{2}} \right] \cos \left[ \omega_m \left( 1 - \frac{v_\varphi}{v_{rf}} \right) t - \omega_m \left( 1 - \frac{v_\varphi}{v_{rf}} \right) \frac{\tau}{2} + \omega_m \frac{v_\varphi}{v_{rf}} (t - \tau) + \varphi \right]. \tag{264}
\]
Eq. (264) can be rearranged by substituting the value of the delay \( \tau = n l / c \) in the argument of the \( \sin x / x \) function,

\[
\omega_m \left[ 1 - \frac{v_\varphi}{v_{rf}} \right] \frac{\tau}{2} = \frac{\omega_m l}{2} \left( \frac{1}{v_\varphi} - \frac{1}{v_{rf}} \right),
\]

and performing a similar operation for the argument of the cosine function, finally leading to

\[
\Gamma (t) = \Gamma_0 \left[ \frac{\sin \left( \frac{\omega_m l}{2} \left( \frac{1}{v_\varphi} - \frac{1}{v_{rf}} \right) \right)}{\omega_m l \left( \frac{1}{v_\varphi} - \frac{1}{v_{rf}} \right)} \right] \cos \left[ \omega_m t - \frac{\omega_m l}{2} \left( \frac{1}{v_\varphi} + \frac{1}{v_{rf}} \right) + \varphi \right].
\]

If the phase velocities of the modulating RF field and optical wave are equal,

\[
v_\varphi = v_{rf},
\]

the retardation at the electro-optic crystal output becomes

\[
\Gamma (t) = \Gamma_0 \cos \left( \omega_m t - \omega_m \tau + \varphi \right),
\]

which, apart from a constant phase shift, is the same as the one that would have been obtained in the absence of transit time effect.\(^{34}\)

Consequently, the intuitive vision that, provided the RF and optical fields co-propagate along the waveguide with identical phase velocities, the wavefront of the light wave always experiences the same value of the electric field over propagation, which should lead to a cancellation of the transit time effect, has been confirmed. In case the phase velocities of the RF and optical signals are not equal, the peak retardation is reduced by a factor that depends on the phase velocity mismatch, as illustrated in Fig. 21 where the horizontal axis now corresponds to the quantity given in (265). A key point in engineering broadband modulators is therefore to ensure proper matching between the RF and optical signals velocities.

In the case of LiNbO\(_3\), the refractive indices are \( n_o = 2.21 \) and \( n_e = 2.14 \) at 1550 nm while the high frequency RF dielectric constants are \( \epsilon_{rf}^e = 44 \) and \( \epsilon_{rf}^i = 28 \). The RF phase velocity can be expressed in term of a microwave effective refractive index \( n_{rf} \) according to

\[
v_{rf} = \frac{c}{n_{rf}},
\]

\(^{34}\)Changing the origin of time by making the substitution \( t \rightarrow t + \tau \) in (266) leads to a cosine term in \( \cos \left[ \omega_m t + \frac{\omega_m l}{2} \left( \frac{1}{v_\varphi} - \frac{1}{v_{rf}} \right) + \varphi \right] \), for which the influence of the relative values of \( v_\varphi \) and \( v_{rf} \) becomes even more straightforward.

07/10/2010 51
and it can be shown that

\[ n_{rf} = \sqrt{\frac{\varepsilon_{rf}^2 + 1}{2}}, \]

(270)

where \( \varepsilon_{rf} = \sqrt{\varepsilon_{x}^r \varepsilon_{z}^r} \) is the relative dielectric constant of the material. In (270), the effective relative permittivity \( \varepsilon_{\text{eff}}^r = n_{rf}^2 \) can be seen as an average of the relative permittivity of the medium and air, where the RF field propagates. In the coplanar electrode is deposited directly on top of the LiNbO\(_3\) substrate, the value of the microwave effective index is then equal to 4.25. Therefore the RF modulating field is slower than the optical field by a factor 2. Consequently, even when a travelling wave electrode configuration is used, the phase velocity mismatch between the RF field and the light wave will severely limit the bandwidth of electro-optic modulators based on LiNbO\(_3\). Hopefully some techniques have been devised to overcome this inherent limitation. As already discussed in Sec. 4.1, using a silica buffer layer, whose dielectric constant is of the order of \( \varepsilon = 3.8 \), enables to reduce the velocity mismatch between the RF and optical fields.

An imperfect match of the optical and RF phase velocities, \( v_{\phi} \neq v_{rf} \), will result in walk-off between the phase fronts of the electrical and optical waves. This means that those two waves will oscillate from in-phase to out-of-phase conditions, depending on the distance in the crystal, causing a reduction of the peak retardation. One way to overcome this limitation, at least at some discrete design frequencies, is to engineer the interaction between the optical and RF waves so that they never become out of phase [16]. This can be achieved by a proper configuration of the electrodes so that the direction of \( E_a \) in the waveguide is periodically reversed, or by ensuring interaction between the optical and electrical fields only occurs close to the in-phase positions. However, such designs are inherently bandpass, since the walk-off relation between electrical and optical fields depends on the frequency of the modulating signal.

As discussed above, true velocity matching can be obtained provided the phase velocity of the RF wave can be increased by using a suitable buffer layer and electrode structure. The key to broadband operation for electro-optic modulators based on materials such as LiNbO\(_3\) is therefore the design of suitable structures that will furthermore present low microwave loss and impedance matching. From a practical point of view, it is also desirable to achieve low voltage operation, since it may be technically challenging, hence expensive, to design broadband high output voltage electrical amplifiers [17]. Low voltage operation can be achieved in principle by increasing the electrode length. There is however a trade off between high bandwidth and low voltage operation since long electrodes will result in high microwave loss. A good review of the issues and challenges associated with the design of broadband electro-optic modulators in LiNbO\(_3\) can be found in [18]. So far, record bandwidths above 100 GHz have been reported for carefully designed LiNbO\(_3\) modulators. For telecommunication applications, this is far beyond the current requirement of about 30 GHz for the next generation systems that will operate at 40 Gbit/s. The use of other material systems, such as polymers, which have the benefit of simultaneously presenting a relatively high electro-optic coefficient and a small microwave dielectric constant that eases velocity matching, is also considered a promising and cost effective way to overcome bandwidth limitations.

So far, bandwidth limitations have been discussed assuming the modulating signal is a single frequency RF field. This has lead to the conclusion that, in order to overcome transit time limitations, the phase velocities of the optical and RF fields should be matched. However, it is clear that once the modulation process has been initiated in the electro-optic crystal, the optical wave is no longer strictly harmonic since it has become phase modulated. Furthermore, for most applications such as digital transmission, the modulating signal is no longer an harmonic field, but a waveform where the information to be transmitted has been encoded under digital or analogue form. For such non-harmonic signals, the velocity matching condition should be...
based on the group velocity instead of the phase velocity, as discussed for instance in [19, 20].

Appendices

A Light intensity

In this section, the concept of light intensity is clarified starting from the energy considerations already discussed following the introduction of the Poynting vector in Sec. 1.4. Furthermore, a relation is established between the electromagnetic formalism and standard expressions employed for the electric field in the context of optical modulation and the study of optical communication systems in general.

A.1 Optical intensity

Starting from the the expression of the time average of the Poynting vector derived in (62),

$$\langle S \rangle = \frac{1}{2\omega \mu} |E_0|^2 k,$$

(271)

it is straightforward to derive and alternate expression

$$\langle S \rangle = \frac{1}{2\eta} |E_0|^2 u,$$

(272)

where $\eta$ is the impedance of the medium (48). The optical intensity is defined as the magnitude of the time averaged Poynting vector

$$I = \frac{1}{2\eta} |E_0|^2,$$

(273)

where the time averaging $\langle \rangle$ is performed over a duration that is long compared to an optical cycle, but short compared to the variations of other physical quantities of interest.

In the context of the modulation of a light wave for optical communication systems applications, wavelengths around 1550 nm are typically used, corresponding to frequencies of the order of 193 THz, hence to a period of the optical carrier frequency equal to about 5.2 as. This value has to be compared to the characteristic duration of the quantity being modulated, which corresponds to either the inverse of the symbol rate for non return-to-zero (NRZ) modulation, or to the pulse duration in case return-to-zero (RZ) modulation is used. For instance, in a standard 40 Gbit/s NRZ system, the symbol duration is 25 ps, meaning that as many as 4825 optical cycles are present in each symbol. The time averaging performed in order to define the intensity enables to follow the variations of the power envelope of the electric field, corresponding to the intensity modulation, but not those of the optical carrier.

Alternate expressions can be easily derived for the optical intensity. For instance, reporting (57) into (64) leads to

$$I = \frac{1}{2} \varepsilon v_r |E_0|^2,$$

(274)

from which it is straightforward to derive another frequently used alternate form

$$I = \frac{1}{2} \varepsilon_0 c n |E_0|^2.$$

(275)
A.2 From the electromagnetism to the communications formalisms

In this section, a plane wave of the type

$$\mathbf{E}(\mathbf{r}, t) = \text{Re}\left\{ \mathbf{E}_0(\mathbf{r}, t) \ e^{j(\omega t - \mathbf{k} \cdot \mathbf{r})} \right\},$$  \hspace{1cm} (276)

where $\mathbf{E}_0$ is a complex vector, is assumed to propagate in an isotropic medium. Defining an orthonormal basis ($\mathbf{x}, \mathbf{y}, \mathbf{z}$), where $\mathbf{z}$ is the direction of propagation, $\mathbf{E}_0$ is in the plane orthogonal to $\mathbf{z}$ due to the transverse nature of the wave, and (276) can be expressed as

$$\mathbf{E}(z, t) = \text{Re}\left\{ \mathbf{E}_0(z, t) \ e^{j(\omega t - kz)} \right\}. \hspace{1cm} (277)$$

The transverse vector $\mathbf{E}_0$ can be decomposed on ($\mathbf{x}, \mathbf{y}$) according to

$$\mathbf{E}_0(z, t) = \kappa [u_x(z, t) \mathbf{x} + u_y(z, t) \mathbf{y}], \hspace{1cm} (278)$$

where $u_x(z, t)$ and $u_y(z, t)$ are complex functions and $\kappa$ is a normalisation constant. It is clear that such an expression allows for the intensity, phase and polarisation of the light wave to be time dependent. The magnitude of $\mathbf{E}_0$ can be easily calculated

$$|\mathbf{E}_0|^2 = \mathbf{E}_0 \cdot \mathbf{E}_0^* = \kappa^2 \left[ |u_x(z, t)|^2 + |u_y(z, t)|^2 \right], \hspace{1cm} (279)$$

hence the light intensity

$$I(z, t) = \frac{1}{2} \epsilon_0 c n \kappa^2 \left[ |u_x(z, t)|^2 + |u_y(z, t)|^2 \right], \hspace{1cm} (280)$$

and the optical power

$$P(z, t) = \frac{1}{2} \epsilon_0 c n \kappa^2 A \left[ |u_x(z, t)|^2 + |u_y(z, t)|^2 \right], \hspace{1cm} (281)$$

where $A$ corresponds to the area over which the intensity is integrated. The arbitrary normalisation parameter $\kappa$ can now be chosen so that the optical power is simply expressed as

$$P(z, t) = |u_x(z, t)|^2 + |u_y(z, t)|^2. \hspace{1cm} (282)$$

Consequently,

$$\kappa = \left( \frac{2 \epsilon_0 c n A}{\lambda_0} \right)^{1/2}, \hspace{1cm} (283)$$

whose dimension is in $\sqrt{\Omega/m^2}$.

The coordinates of $\mathbf{E}_0$ along $\mathbf{x}$ and $\mathbf{y}$ in (278) can be further expressed in terms of their magnitudes and phases, so that

$$\mathbf{E}_0(z, t) = \kappa \left[ A_x(z, t) \ e^{-j\phi_x(z,t)} \mathbf{x} + A_y(z, t) \ e^{-j\phi_y(z,t)} \mathbf{y} \right], \hspace{1cm} (284)$$

which immediately leads to

$$\mathbf{E}_0(z, t) = \kappa e^{-j\phi(z,t)} \left[ A_x(z, t) \ e^{-j\frac{\phi_x(z,t)}{2}} \mathbf{x} + A_y(z, t) \ e^{j\frac{\phi_y(z,t)}{2}} \mathbf{y} \right], \hspace{1cm} (285)$$

where

$$\phi(z, t) = \frac{\phi_x(z, t) + \phi_y(z, t)}{2}. \hspace{1cm} (286)$$
Anisotropic Media and the Linear Electro-Optic Effect

is the common phase, and

$$\delta = \phi_x (z,t) - \phi_y (z,t)$$  \hspace{1cm} (287)

is the differential phase term, which determines the state of polarisation of the wave. Introducing the normalised Jones vector,

$$\mathbf{e} (z,t) = \frac{1}{\sqrt{A_x^2 (z,t) + A_y^2 (z,t) }} \left[ A_x (z,t) \ e^{-j \frac{\delta (z,t)}{2}} \mathbf{x} + A_y (z,t) \ e^{j \frac{\delta (z,t)}{2}} \mathbf{y} \right],$$  \hspace{1cm} (288)

enables to express the transverse field $\mathbf{E}_0$ as

$$\mathbf{E}_0 (z,t) = \kappa \ e^{-j \phi (z,t)} \sqrt{A_x^2 (z,t) + A_y^2 (z,t) } \ \mathbf{e} (z,t),$$  \hspace{1cm} (289)

and therefore

$$\mathbf{E}_0 (z,t) = \kappa \sqrt{P (z,t) } \ e^{-j \phi (z,t)} \ \mathbf{e} (z,t).$$  \hspace{1cm} (290)

$\mathbf{E}_0 (z,t)$ is therefore proportional to the complex envelope of the field, multiplied by a unit vector describing the state of polarisation. In the study of optical modulation formats, the explicit spatial dependence of the field is not always necessary, hence the following simplified notations will be adopted

$$\mathbf{E} (t) = \text{Re} \left\{ \kappa \sqrt{P (t) } \ e^{-j \phi (t)} \ e^{j \omega t} \ \mathbf{e} (t) \right\},$$  \hspace{1cm} (291)

from which the various quantities that can be modulated, namely the power $P (t)$, the phase $\phi (t)$, the frequency $\omega$ and the polarisation $\mathbf{e} (t)$, are clearly visible. In order to simplify the notations, the proportionality constant $\kappa$ is often neglected in the optical communication literature. This is of little consequence in practice. However, taking $\kappa$ into account is necessary in order to ensure that $\mathbf{E}$ has actually the dimension of an electric field ($\text{V} \cdot \text{m}^{-1}$) in equations of the type (291). It can also be remarked that in the derivation above, the light wave is assumed to be polarised at any time, i.e. its state of polarisation is $\mathbf{e} (t)$ at time $t$, even though it is not single frequency due to the time varying term $\mathbf{E}_0 (z,t)$ in (277), but has a narrow spectrum centred around the carrier frequency $\omega$.

B On the reduction of quadratic forms

We consider the second order surface $(\Sigma)$ defined by the following equation in a Cartesian coordinate system $(O, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$, where $\mathbf{e} = (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ is an orthonormal basis:

$$a_{11} x^2 + a_{22} y^2 + a_{33} z^2 + 2a_{23} yz + 2a_{13} xz + 2a_{12} xy + b_1 x + b_2 y + b_3 z + c = 0.$$  \hspace{1cm} (292)

Such a surface is known as a quadratic surface or quadric. Depending on the values of the coefficients, (292) can describe an ellipsoid, an hyperboloid of one or two sheets, an elliptic cone, an elliptic paraboloid, an hyperbolic paraboloid, an elliptic cylinder, an hyperbolic cylinder, a parabolic cylinder, one or two parallel planes, a line, a point, or the empty set. Quadrics can be classified according to the properties of the eigenvalues of the associated quadratic form. In our study of anisotropic media and the electro-optic effect, we will be primarily concerned with the case of ellipsoids. Here we show how the general equation (292) can be reduced to a more familiar canonical expression.

The quadratic form associated to $(\Sigma)$ is

$$q (x,y,z) = a_{11} x^2 + a_{22} y^2 + a_{33} z^2 + 2a_{23} yz + 2a_{13} xz + 2a_{12} xy,$$  \hspace{1cm} (293)

07/10/2010
whose matrix in the basis \( e \) is

\[
Q = \begin{pmatrix}
a_{11} & a_{12} & a_{13} \\
a_{12} & a_{22} & a_{23} \\
a_{13} & a_{23} & a_{33}
\end{pmatrix}.
\]

(294)

We can therefore express the quadratic form \( q \) according to

\[
q(x, y, z) = x^T Q x,
\]

(295)

where \( x \) is the column vector \( x = (x, y, z) \) and \( T \) denotes the transposition operation. To simplify the notations, the equation of the surface (\( \Sigma \)) can be written in vectorial form

\[
x^T Q x + b^T x + c = 0,
\]

(296)

where \( b^T = (b_1, b_2, b_3) \).

\( Q \) is a symmetric matrix whose elements are real. Hence there exists an orthonormal basis \( \bar{e} = (\bar{e}_1, \bar{e}_2, \bar{e}_3) \) where the quadratic form \( q \) is represented by a diagonal matrix

\[
\Lambda = \begin{pmatrix}
\lambda & 0 & 0 \\
0 & \mu & 0 \\
0 & 0 & \nu
\end{pmatrix}.
\]

(297)

such that \( \Lambda = P^T A P \), where \( P \) is the transformation matrix from the basis \( e \) to the new orthonormal basis \( \bar{e} \), and \( \lambda, \mu \) and \( \nu \) are the eigenvalues of \( Q \).

We can write the equation of (\( \Sigma \)) in the new basis by performing the transformation \( x = P \bar{x} \). We immediately obtain

\[
\bar{x}^T \Lambda \bar{x} + b^T P \bar{x} + c = 0.
\]

(298)

If we perform the extra transformation \( \bar{x} = \bar{x} - a \), which corresponds to a change of origin of the coordinate system (translation by \( a \)), the equation of (\( \Sigma \)) becomes

\[
\bar{x}^T \Lambda \bar{x} + (2a^T \Lambda + b^T P) \bar{x} + a^T \Lambda a + b^T P a + c = 0.
\]

(299)

Provided none of the eigenvalues of \( Q \) is equal to 0, the linear terms of the equation can be suppressed by a careful choice of \( a \) so that \( 2a^T \Lambda + b^T P = 0^T \). In this case, it is immediately found that (\( \Sigma \)) has an equation of the type

\[
\lambda \bar{x}^2 + \mu \bar{y}^2 + \nu \bar{z}^2 + \gamma = 0,
\]

(300)

where \( \gamma \) is a constant, in the new coordinate system. If the three eigenvalues \( \lambda, \mu \) and \( \nu \) have the same sign, then (\( \Sigma \)) is either the empty set or an ellipsoid. In the later case, believed to be of significantly higher interest, the equation of the surface can be further reduced to the well known canonical form

\[
\left( \frac{\bar{x}}{x_0} \right)^2 + \left( \frac{\bar{y}}{y_0} \right)^2 + \left( \frac{\bar{z}}{z_0} \right)^2 = 1.
\]

(301)

The following special cases may be further distinguished:

- If all three eigenvalues are equal, \( \lambda = \mu = \nu \), then (\( \Sigma \)) is a sphere.
- If \( Q \) has only 2 distinct eigenvalues, e.g. \( \lambda = \mu \neq \nu \), then (\( \Sigma \)) is an ellipsoid of revolution.

It can also easily be shown that the intersection of a quadric with a plane is a conic. In the case of an ellipsoid, this intersection is an ellipse. This property has been exploited in Sec. 2.3 for the geometrical determination of the normal modes of propagation from the index ellipsoid.
References


Problems

Problem 1

Derive the Helmholtz equation (41) for the electric field.

Problem 2

Derive the expression of the time averaged electric energy density (81) in the case of an anisotropic and non dispersive medium.

Problem 3

Bi₁₂SiO₂₀, also known as BSO, is an example of crystals that are naturally isotropic but that become anisotropic under the influence of an external applied electric field. Its electro-optic tensor is

\[
\begin{pmatrix}
0 & 0 & 0 \\
0 & 0 & 0 \\
r_{41} & 0 & 0 \\
0 & r_{41} & 0 \\
0 & 0 & r_{41}
\end{pmatrix}
\]

(302)

1. Write an expression for the index ellipsoid of BSO in an orthonormal basis \((x, y, z)\) in the absence of any external field.

2. An external electric field \(E_a\) is now applied to the crystal. Write how this field modifies the index ellipsoid of BSO.

3. It is assumed in particular that the applied electric field is parallel to \(x\). Find the directions of the new eigenmodes of propagation and associated refractive indices.

4. Assuming the crystal is used in the longitudinal configuration, find an expression for the half-wave voltage. Calculate the value of this voltage at 633 nm assuming \(n_0 = 2.53\) and \(r_{41} = 3.6 \times 10^{-12}\) m/V.
Problem 4

A travelling wave lithium-niobate Mach-Zehnder modulator designed for 10 Gbit/s operation is accidentally connected so that the optical wave and radio-frequency signal propagate in opposite directions. As shown in Fig. 24, the signal at the output of the modulator is completely distorted. However, when the bit rate of the electrical non return-to-zero driving signal is reduced to 5, then 2.5 Gbit/s, the eye diagram tends to open.

Figure 24  Eye diagrams recorded at the output of a travelling wave LiNbO₃ Mach-Zehnder designed for operation at 10 Gbit/s when the electrical and optical waves are counter propagating. The eye diagrams have been obtained for an NRZ electrical signal having a bit rate of 2.5 Gbit/s (left), 5 Gbit/s (middle) and 10 Gbit/s (right).

1. Explain the behaviour observed in Fig. 24.

2. Advanced topic (after reviewing the part of the course on generation of optical modulation formats): suggest a practical application for the observed effect and discuss its practical requirements.