LIGHT: ITS NATURE AND ITS PROPERTIES

Over time, different theories have been formulated to explain light and its behavior. The corpuscular theory, developed by Newton, is based on the hypothesis that light consists of extremely small particles, emitted by luminous bodies, subject to the laws of mechanics. The wave theory, due to Huygens, explains the luminous phenomena as having a wave nature, that is light is a wave that propagates in a medium.

The electromagnetic theory, formulated by Maxwell, holds that light waves, already hypothesized by Huygens, are electromagnetic and do not require a transmission medium. According to this theory, still widely used today, light is a part of the electromagnetic spectrum. Quantum theory, due to Planck and Einstein, hypothesizes that the emission and absorption of light occurs through "wave packets" with well-defined energy and duration, called quanta of light or *photons*. In most cases light behaves according the electromagnetic theory but in certain phenomena the corpuscular properties of the photons are predominant. This is known as wave-particle duality. Electromagnetic and quantum theory are in perfect agreement if we assume that the amplitude of a field (electric or magnetic) at a certain point and at a certain instant is proportional to the probability (wave function) of finding a photon.

Electromagnetic spectrum

Figure 1 shows the electromagnetic spectrum with reference both to the wavelength and frequency. The field of interest of Photonics is that of the wavelengths between ultraviolet and far infrared. Note that the visible field occupies a very limited portion of the optical field. Often, even if improperly, we speak of "light" even when the electromagnetic wave is not visible but within the range indicated above.

Speed of light and refractive index

The wavelength in vacuum λ_0 and the frequency ν (often indicated as "*f*" in Engineering) are related by the speed of light in vacuum *c*:

$$\lambda_0 \, \nu = c \tag{1}$$

It is experimentally observed, and it can be proved, that when the propagation of light occurs in a material other than vacuum (commonly called "medium"), the speed decreases.

We define *refractive index*, indicated with *n*, the ratio between the speed of light in vacuum and the speed in the medium:

$$n = c \, / \, v \, ,$$

where *v* is the speed of light in the medium (not confound with the Greek letter ν!). Therefore, when propagation occurs in a medium with refractive index *n*, the (1) becomes:

$$\lambda_n \, \nu = v = c \, / \, n \tag{2}$$

and from the ratio between (2) and (1):

$$\lambda_n = \lambda_0 / n ,$$

that is, the wavelength decreases when light travels in a medium other than vacuum, since the dielectrics are characterized by a refractive index higher than 1.

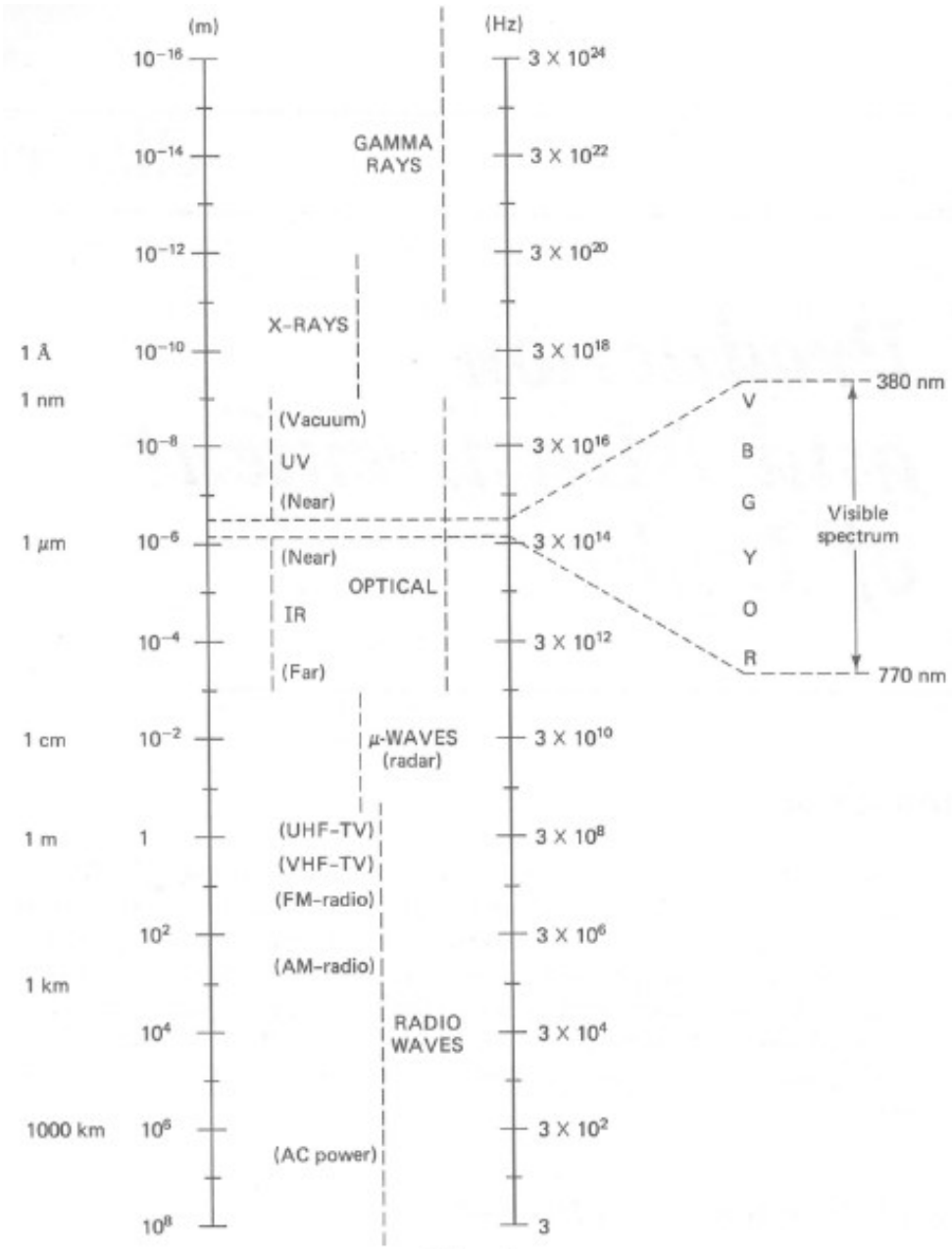


Fig. 1

Table 1 shows the refractive indices (in the visible range) of some materials.

Material	n	Material	n
air	1,0003	sodium chloride	1,54
water	1,33	"light flint" glass	1,57
methanol	1,33	carbon disulfide	1,62
ethanol	1,36	"medium flint" glass	1,63
magnesium fluoride	1,38	"dense flint" glass	1,66
fused silica	1,46	sapphire	1,77
pyrex glass	1,47	"very dense" flint glass	1,73
benzene	1,50	"extremely dense" flint glass	1,89
xilene	1,50	zinc sulfide (thin film)	2,3
"crown" glass	1,52	titanium dioxide (thin film)	2,4 – 2,9
Canada balsam (optical resin)	1,53	silicon carbide (6H-form)	2.65

Tab. 1

Dispersion

The refractive index varies slowly (if no resonant phenomena occur) with the wavelength, as (see Figure 2 for three different types of glass). This phenomenon is called *dispersion*.

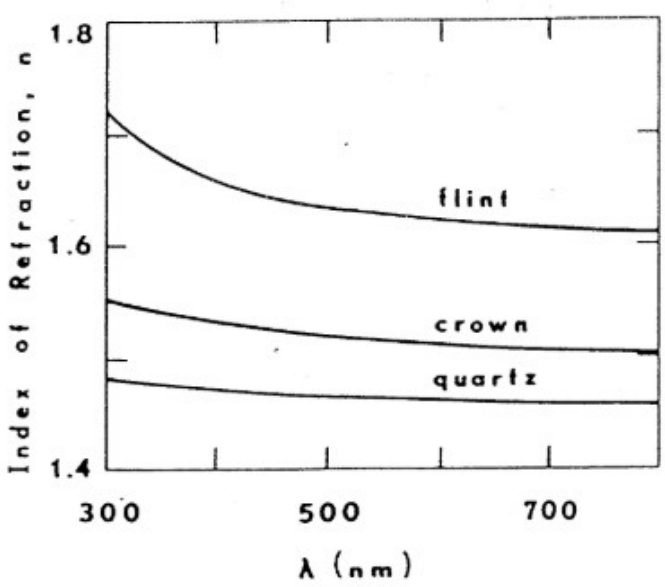


Figure 2

For most transparent and non-colored materials, the following relationship occurs:

$$n(\lambda) = a + \frac{b}{\lambda^2},$$

with *a* and *b* coefficients depending on the material.

Photon energy

For the quantum theory, the energy associated with the single photon is:

$$E = h\nu = \frac{hc}{\lambda_0} = \frac{1{,}98 \cdot 10^{-25}}{\lambda_0}[J] = \frac{1{,}24}{\lambda_0\,[\mu m]}[eV]$$

assuming that λ_0 is expressed in micrometers and knowing that:

h (Planck's constant) ≈ 6.626×10^{-34} J s,

c (speed of light in vacuum) ≈ 3×10^{8} m / s,

1 eV ≈ $1{,}6 \times 10^{-19}$ J.

Time and spatial coherence

Time coherence is defined as the average time range in which it is probable that a wave will not undergo a random phase change. In this time range the wave travels a certain path, called *spatial coherence*.

Within the spatial coherence field it is possible to establish a phase relationship between waves at different points at a certain time.

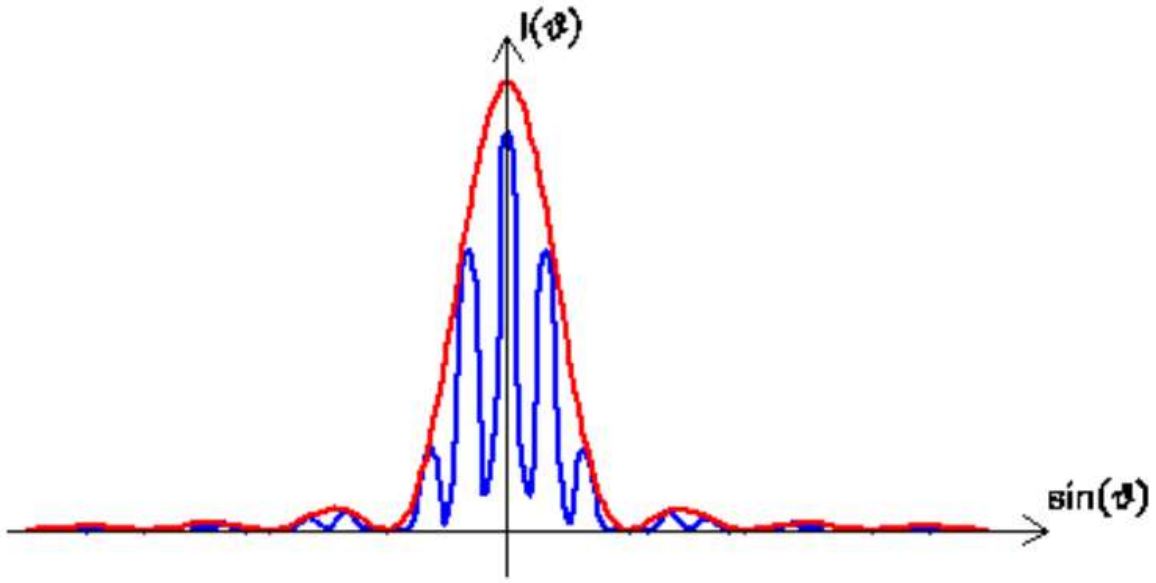
Conversely, within the time coherence field it is possible to establish a phase relationship between waves at a certain point at different times.

The spatial coherence of the light generated by optical oscillators (lasers) can reach several tens of meters, contrary to what happens for the light generated by conventional sources (including LEDs).

The finite space-time coherence explains the limitation of the phenomena of interference.

GRATINGS

Let us study now the phenomenon of diffraction of light waves. What happens when the incident plane wave finds as an obstacle an absorbing screen in which there are two parallel slits of equal width d, spaced by a quantity p (generally greater than d). If we consider the distribution of light intensity on the screen placed behind the obstacle, we find a situation similar to that shown in the figure below.



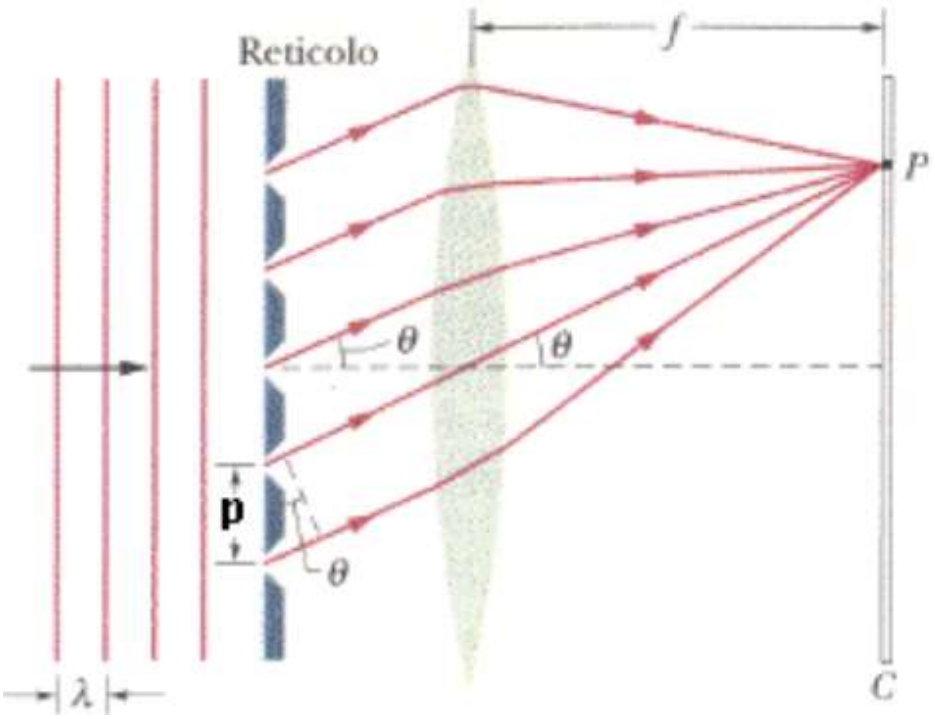
We assumed that the distance p between the centers of the two slits is equal to $p = 3d$; the red curve represents the intensity distribution due to a single slit, while the blue curve represents the case of two slits, with amplitude divided by a factor of 4. The peaks identify the angles at which illuminated areas are observed on the screen, while the points of the graph in which the curve drops to 0 identify angles for which dark areas are observed on the screen.

It can be easily observed that while in the case of a single slit there was a rather large central maximum, three much narrower 'central' maxima can now be clearly seen, i.e. better defined and therefore more easily identifiable. Furthermore, the height of these peaks turns out to be about 4 times greater than that of the peak due to the single slit.

This effect is produced by the simultaneous action of the two slits: the waves generated by them are added, by the superposition principle, generating this particular structure.

A logical generalization of what we have just seen leads to an increase in the number of slits in order to obtain an ever sharper diffraction phenomenon, using a device that takes the name of *grating*.

A diffraction grating consists of a very large number N of thin, parallel, equidistant slits of equal width d. The distance p between the central points of two contiguous slits is the grating *pitch*. A sketch of such a grating is shown in the figure below.



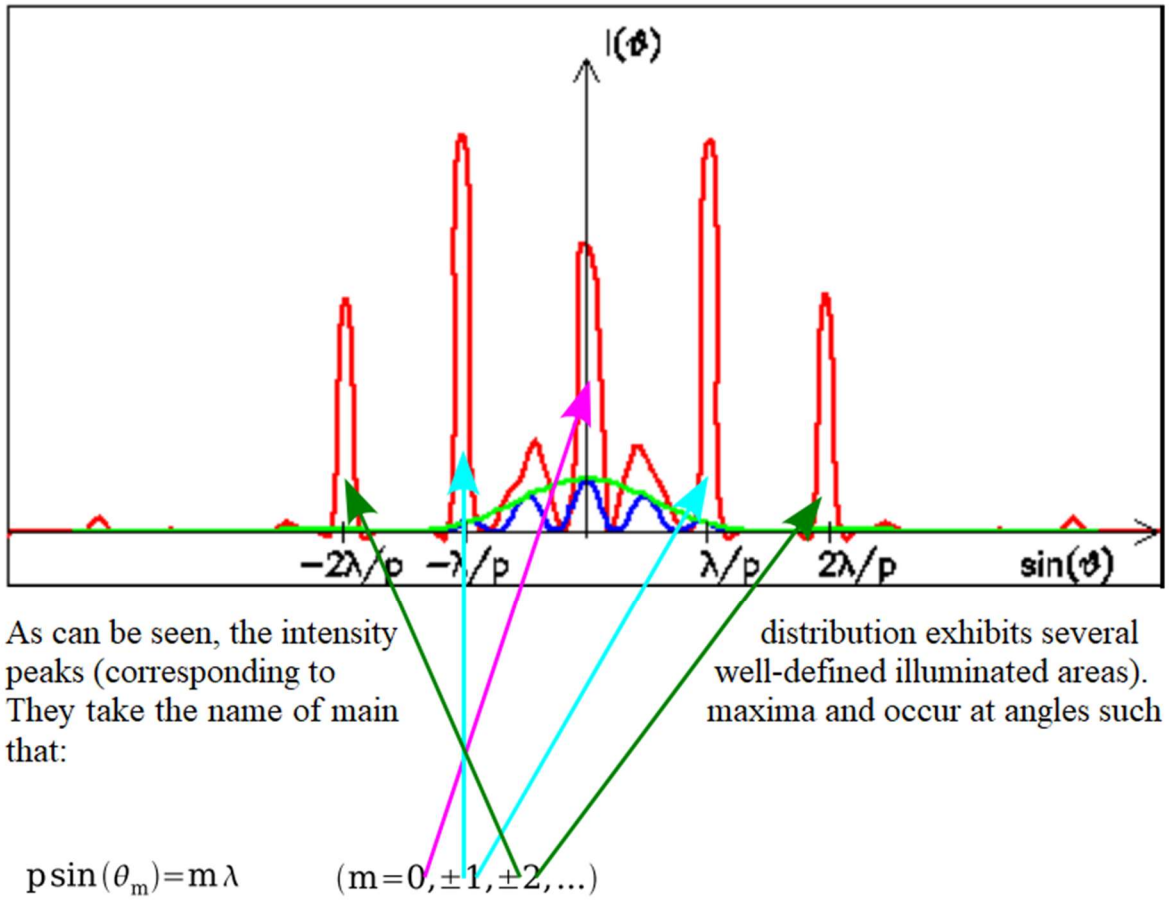
A plane wave is incident on the grating and the distribution of the light intensity is studied on a screen behind it, making the light rays emitted at equal angles converge by means of a converging lens.

The gratings are fabricated by engraving parallel grooves on a glass or metal plate by means of suitable machines that use diamond points. These grooves are then filled with an absorbent substance. We refers to "*transmission gratings*" when we consider devices in which light can pass through the slits, as in the case of grids obtained with glass plates; instead, we refers to "*reflection gratings*" when the light does not pass through the slits but is reflected above them, as in the case of gratings engraved on metal plates. In this last case the intensity distribution is studied on a screen placed to the left of the grating, in a direction different from that of the incident light beam.

In the following we will consider a grating that works in transmission.

The number of slits per unit of length is called the *grating constant*; it is generally of a few thousand per cm.

The light is diffracted from every single slit according the laws of diffraction; the light intensity distributions generated by each slit are added to those produced by all the other slits present in the grating (overlapping principle). The following figure shows the case in which there are $N = 1000$ slits (red curve) and, for comparison, the cases of single slit and two slits are repeated, (not in scale).



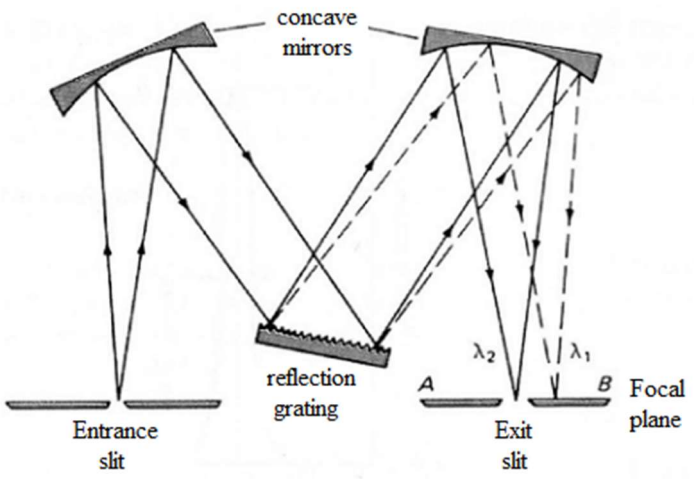
For small angles the main maxima are angularly equispaced and their position depends only on the grating pitch p and the wavelength of the incident light. The number m is called the *order of the main maximum*. In the main maxima, the luminous intensity is N^2 times higher than that provided by a single slit (green curve), as an effect of the sum of the intensities due to the N slits illuminated by the same wave front. The fringe corresponding to $m = 0$ is the *central fringe*, or *white fringe*, because its position does not depend on the wavelength; the other fringes are called *fringes* (or *maxima*) *of order m*.

MONOCHROMATOR

The grating is the main component of the *monochromator*, whose function is to permit a precise wavelength to pass through along a certain direction. To do this, instead of changing the direction of the white light with respect to the grating, the grating is rotated by a small fraction of an angle, so as to return to the condition of phase agreement with a different wavelength.

The monochromator has an entrance slit, where the white light (or the one to be analyzed) is sent, and which acts as a point source of the system. It illuminates a concave mirror, which transforms the diverging beam arriving from the slit into a parallel beam, which illuminates uniformly the reflection grating.

The reflected light follows an analogous but inverse optical path (concave mirror that converges the previously parallel rays), in an exit direction that is selected by another small slit.



The light that pass through the slit is only that component which have a wavelength such as to generate, for that angle, a constructive interference between all the thousands of reflected rays. The other wavelengths will make constructive interference at different angles, and therefore before and after the slit, at different points of the focal plane AB. However, by rotating the grid around an axis perpendicular to the sheet, they are gradually aligned with the slit, and can be used later.

Obviously the monochromator is the more effective the more it is able to separate two neighboring wavelengths; this *resolving power* must be chosen on the ground of the spectral range of interest, i.e. ultraviolet, visible, near infrared (UV-VIS-NIR) .

Indeed, if a spectrophotometer (i.e. the set consisting of the white light source, the monochromator and a photodetector) covers a range from 200 nm to 2000 nm (standard UV-VIS-NIR), 2 gratings are enough: one used from 200 nm to 800 nm, and one used from 800 nm to 2000 nm with 1200 and 600 lines/nm respectively. The reason why an additional 2400 line/nm grating is not used is that working in UV below 200 nm requires special materials; therefore most spectrophotometers cannot work below 200 nm.

The number of total lines, on the other hand, affects the quality of the interference, that is, the angular width and the intensity of the maximum interference for a given wavelength and therefore, with the same dispersion, allows to better separate two neighboring lines ("*resolving power*") and to have brighter spectra. However, as the size of the reticle increases, the focal length of the concave mirrors must also increase, and consequently the quality of the system structure (and, obviously, the cost).

Optical Activity

Certain crystals (and liquids) have the ability to rotate the plane of polarization of light passing through them; that is, they are *optically active*. Thus, for example, when a beam of plane polarized light is incident normally on a crystal plate of quartz cut perpendicular to the optic axis, it is found that the emergent beam is also plane polarized but that its electric vector vibrates in a different plane from that of the incident light. The plane of vibration may be rotated in a clockwise sense looking against the oncoming light by *right-handed* or *dextrorotatory* crystals, or in a counterclockwise sense by *left-handed* or *laevorotatory* crystals. Quartz exists in both forms. It is found that the rotation depends on the thickness of the crystal plate and the wavelength. The rotation produced by a quartz plate 1 mm thick for sodium light is 21.7° while it is 3.67° for 1 mm of sodium chlorate.

Optical activity can be explained by assuming that in optically active crystals the velocity of propagation of circularly polarized light is different for different directions of rotation, that is the crystal has refractive indices n_r and n_l for right and left circularly polarized light. It is easy to show (see Problem 3.3) that a plane polarized wave can be resolved into two circularly polarized waves with opposite directions of rotation. If these travel through the crystal at different speeds, a phase difference will be introduced between them at different distances through the crystal. This corresponds to a rotation of the plane of the plane polarized wave which results from the recombination of the two circularly polarized waves.

3.9.1 *The electro-optic effect*

When an electric field is applied to an optical medium the electrons suffer restricted motion in the direction of the field, when compared with that orthogonal to it. Thus the material becomes linearly birefringent in response to the field. This is known as the electro-optic effect.

Consider the arrangement of Fig. 3.18. Here we have incident light which is linearly polarized at 45° to an electric field and the field acts on a medium transversely to the propagation direction of the light. The field-induced linear birefringence will cause a phase displacement between components of the incident light which lie, respectively, parallel and orthogonal to the field; hence the light will emerge elliptically polarized.

A (perfect) polarizer placed with its acceptance direction parallel with the input polarization direction will of course, pass all the light in the absence of a field. When the field is applied, the fraction of light power which is passed will depend upon the form of the ellipse, which in turn depends upon the phase delay introduced by the field. Consequently, the field can be used to modulate the intensity of the light, and the electro-optic effect is, indeed, very useful for the modulation of light (see section 7.3.1).

The phase delay introduced may be proportional either to the field (Pockels effect) or to the square of the field (Kerr effect). All materials manifest a transverse Kerr effect. Only crystalline materials can manifest any kind of Pockels effect, or longitudinal (E field parallel with propagation direction) Kerr effect. The reason for this is physically quite clear. If a material is to respond linearly to an electric field the effect of the field must change sign when the field changes sign. This means that the medium must be able to distinguish (for example) between 'up' (positive field) and 'down' (negative field). But it can only do this if it possesses some kind of directionality in itself, otherwise all field directions must be equivalent in their physical effects. Hence, in order to make the necessary distinction between up and down, the material must possess an intrinsic asymmetry, and hence must be crystalline. By a similar argument a longitudinal E field can only produce a directional effect orthogonally to itself (i.e., in the direction of the optical electric field) if the medium is anisotropic (i.e., crystalline) for otherwise all transverse directions will be equivalent. In addition to the modulation of light (phase or intensity/power) it is clear that the electro-optic effect could be used to measure an electric field and/or the voltage which gives rise to it.

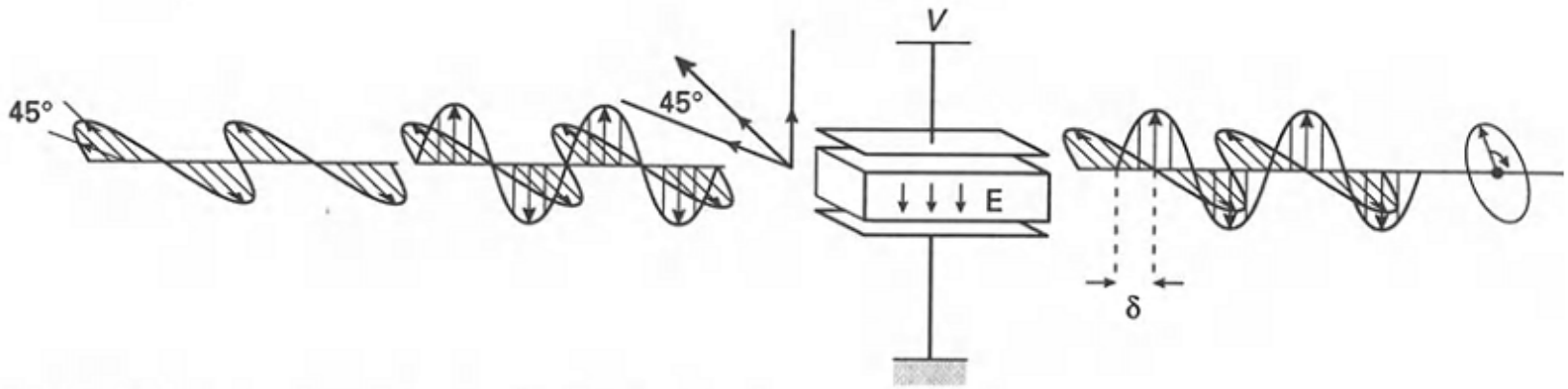


Fig. 3.18 The electro-optic effect.

7.3 OPTICAL MODULATORS

Thus far in this chapter we have been studying optical sources. It has been noted several times in the preceding text that the usefulness of optical sources in practical optoelectronics relies upon our ability to impress information upon, and subsequently to extract that information from, the light which they produce. The device which performs this function is the optical modulator.

Our next task then is to deal with the most important types of optical modulator. Fortunately, we have already covered almost all of the physical ideas which are needed for an understanding of these.

A purely sinusoidal optical wave carries no information at all for it extends over all space and all time. If it suddenly ceases to exist (either spatially or temporally), then we have some information conveyed (the time or position of the cessation) but it then ceases also to be a pure sine wave, since a truncation must introduce other frequencies (Fourier!). Clearly, then, there must be a relationship between coherence and modulation (which we shall not explore in detail).

In order to convey information by means of an optical signal we must modify a characterizing parameter of the wave in a way which allows a definite, deterministic relationship to exist between the modification and the information. Thus we might modify at the transmitter end the amplitude, intensity, frequency, phase, polarization state, direction or coherence of the wave. Which one of these we choose to modify depends very much on the nature of the information, the nature of the source, the nature of the environment and how much money we are allowed to spend! The criteria are complex and very applications dependent.

Clearly, the first consideration must be the extent to which the conveyed information can be corrupted by the environment: it may be, for example, that we wish to communicate through cloud, or dust, or a particular gas/liquid. In such circumstances environmental amplitude perturbations will, perhaps, be stronger than phase perturbations, so we would not chose amplitude modulation: the received signal probably would be too noisy. Again, it is necessary to extract the information at the receiving end, and this must be done as efficiently and as cheaply as possible. Hence we must choose a modulation scheme which allows this. Amplitude (power) detection is (as we shall see) the easiest and cheapest, but if this proves too noisy, we turn perhaps to phase or polarization detection with its increased

complexity and cost. System design is the art of optimization and compromise. Clearly a range of modulators is needed to give us the flexibility to effect these optimizations. In the next few sections the most important and prevalent of these are described.

7.3.1 The electro-optic modulator

The electro-optic effect was introduced briefly in section 3.9(a). It is necessary now to investigate it in more detail in order to understand its special usefulness in electro-optic modulators.

When an electric field is applied to a medium the effect is to induce a linear birefringence, so that there will be two linear, orthogonal optical polarization states which propagate without change of form, but at different velocities. If the medium is naturally isotropic (i.e., non-crystalline or symmetrically crystalline), then these two eigenstates lie respectively parallel and orthogonal to the electric field direction. In a crystalline material the eigenstate polarization directions will depend on the particular symmetry possessed by the crystal structure.

Two different velocities in a material are, of course, characterized by a difference in refractive index, Δn. The two most important electro-optic effects are distinguished by the dependence of Δn on the electric field in each of their cases: in one case the dependence is linear, in the other, quadratic, i.e.,

$$\Delta n = PE \quad \text{Pockels effect} \tag{7.2a}$$

$$\Delta n = KE^2 \quad \text{Kerr effect} \tag{7.2b}$$

where E is the applied electric field and P and K are the Pockels and Kerr coefficients, respectively. (K is often replaced by λB in the literature, where λ is the wavelength of the light and B is then called the Kerr coefficient; the justification for this seems to be that a phase change over a distance l can then be written $(2\pi/\lambda)\lambda BE^2 l$ and thus become independent of λ, to first order (B still depends on λ, however, to some extent, depending on the dispersion of the medium at the wavelength used)).

How can we use this modification of Δn to design a modulator?

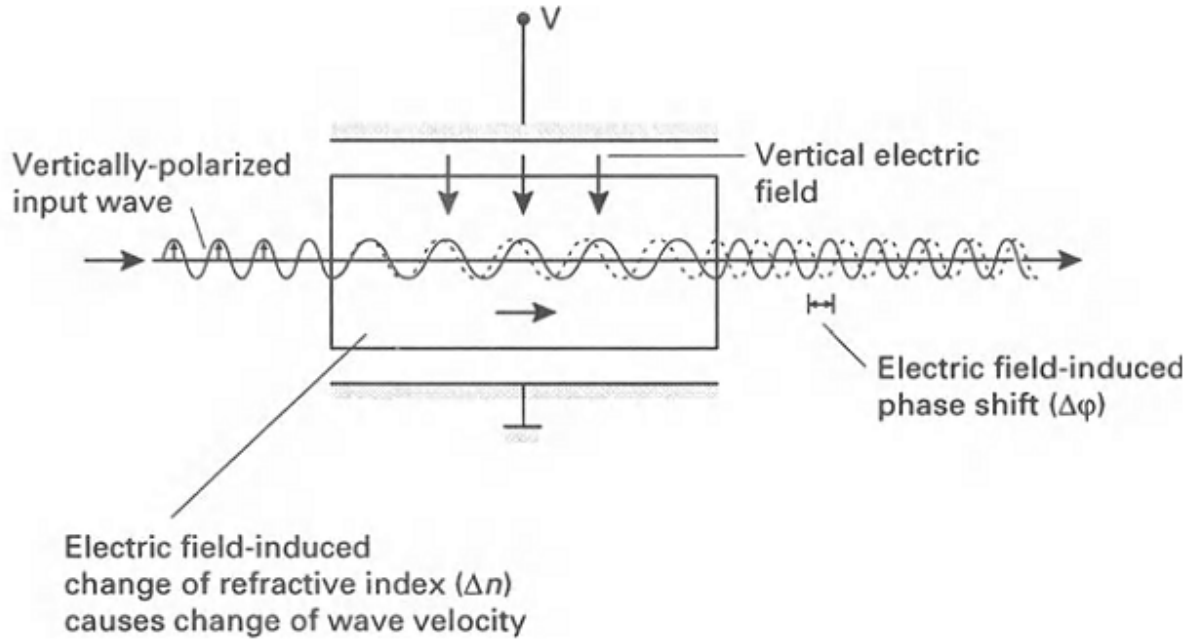


Fig. 7.12 Phase modulation via the electro-optic effect.

In the first place it is clear that a phase modulation is very easy. Fig. 7.12 shows how it can be done. Linearly polarized light is launched into a material to which a transverse electric field is applied. The polarization direction lies parallel to one of the birefringence eigenaxes produced by the applied electric field. The variation in optical phase caused by the field is given by:

$$\Delta\varphi = \frac{2\pi}{\lambda}\Delta n\, l$$

where l is the length of the optical path in the material and λ is the wavelength of the light.

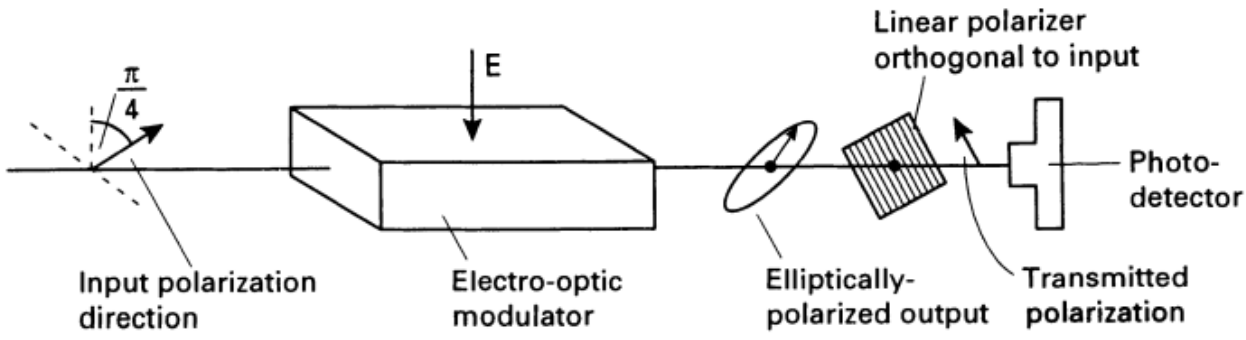
Hence, from equation (7.2a) we see that the Pockels effect will then provide

$$\Delta\varphi = \frac{2\pi}{\lambda} lPE$$

and if E is due to a voltage V being applied across a width d of material

$$\Delta\varphi = \frac{2\pi}{\lambda}\frac{l}{d}PV$$

and the phase change is proportional to the applied voltage. It is, of course, quite common for the information we wish to transmit to be in the form of a voltage waveform.



(a) Electro-optic amplitude modulation

Fig. 7.13 Linear electro-optic amplitude modulation.

It is also possible, however, to use the electro-optic effect to modulate the amplitude of the wave. To see this, consider Fig. 7.13(a). Here, the linearly polarized wave is launched at 45° to the eigenaxes produced by the applied field. The result now is that a phase difference is inserted between the two polarization components corresponding to the directions of the induced birefringence eigenaxes, so that the linearly polarized wave now emerges elliptically (in general) polarized. The light has been polarization modulated!

Suppose that a polarization analyser now is placed so that the emerging light has to pass through it before falling on a photodetector. Suppose also that the acceptance direction of this analyser is placed at 90° to the original, input, polarization direction so that, in the absence of an applied field, no light passes through the analyser to the detector. On application of the electric field the polarization becomes elliptical, and some light must now pass through the analyser, since there will be a component in the acceptance direction (a true ellipse has a non-zero amplitude in any given direction).