Chapter 4 EM Applications

4.1 DIFFRACTION

4.1.1 Geometrical Optics and Wave Optics

In elementary studies of wave motion, the ray is used as a convenient way to represent the motion of a train of waves; the ray is perpendicular to the wavefronts and indicates the direction of travel of the wave.

A ray is a convenient geometrical construction that is often helpful in studying the optical behaviour of a system such as a lens. A ray is not a physical entity, however, and it is not possible to isolate one.

Consider a train of light waves of wavelength λ incident on a barrier in which there is a slit of width *a.* As suggested by Figure 1, if $\lambda << a$, the waves pass through the slit, and the barrier forms a sharp 'shadow'.

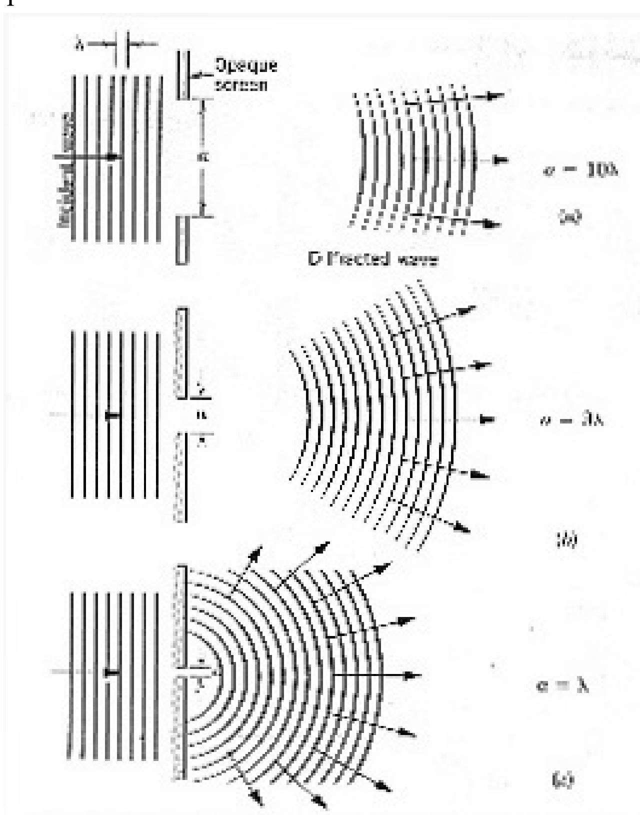

Figure 1 An attempt to isolate a ray by reducing the slit width fails because of diffraction, which becomes more pronounced for a fixed wavelength λ as the slit width *a* is reduced.

Figure 1

As we make the slit smaller, we find that the light flares out into what was formerly the shadow of the barrier, as shown in Figure 2.

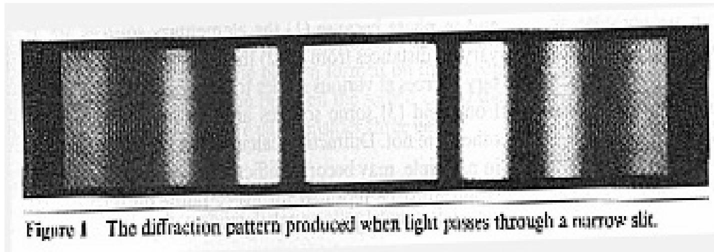

Figure 2

This phenomenon, which is called *diffraction,* occurs when the size of the slit or an obstacle in the path of the wave is comparable to the wavelength. Note that diffraction becomes more pronounced as the slit width becomes smaller; thus an attempt to isolate a single ray will be futile.

Diffraction is not the exclusive property of light waves. For example, when you shout through an open doorway, the sound waves are diffracted (the wavelength being comparable to the size of the doorway opening), and a friend can hear you even though he/she cannot see you (light waves having too small a wavelength to be diffracted noticeably upon passing through the opening).

If a is a measure of the smallest transverse dimension of a slit or obstacle, then the effects of diffraction can be ignored if the ratio of a/λ is large enough. In this case, the light appears to travel in straight-line paths that we can represent as rays. This is the condition for *geometrical optics,* also known as *ray optics*. When a light beam encounters such obstacles as mirrors, lenses, or prisms, whose lateral size is much greater than the wavelength of light, we are safe in using the equations of geometrical optics.

If the condition for geometrical optics is not met, we cannot describe the behaviour of light rays but must take its wave nature specifically into account. In this case we are in the realm of *physical optics* or *wave optics,* which includes geometrical optics as a limiting case.

4.1.2 Diffraction and the Wave Theory of Light

When light passes through a narrow slit (of width comparable to the wavelength of the light), the light beams not only flare out far beyond the geometrical shadow of the slit; they also give rise to a series of alternating light and dark bands.

The figure below shows a representation in the form of ray diagrams. The pattern formed on the screen depends on the separation between the screen C and the aperture B. In general, we can consider three cases:

1. *Very small separation:*
   When C is very close to $B,$ the waves travel only a short distance after leaving the aperture, and the rays diverge very little. The effects of diffraction are negligible, and the pattern on the screen is a geometrical shadow of the aperture.

2. *Very large separation:*
   We can regard the rays as parallel, or equivalently, the wavefronts as planes. (In this case, we also assume the source to be far from the aperture, so that the incident wavefronts are also planes. The same effect can be achieved by illuminating the aperture with a laser.) This is called *Fraunhofer Diffraction.*

3. *Intermediate separation:*
   The screen can be at any distance from the aperture, and the rays entering and leaving the aperture are not parallel. This general case is called *Fresnel Diffraction.*

Fraunhofer Diffraction is a special limiting case of the more general *Fresnel Diffraction*. In this, we shall only consider *Fraunhofer Diffraction*.

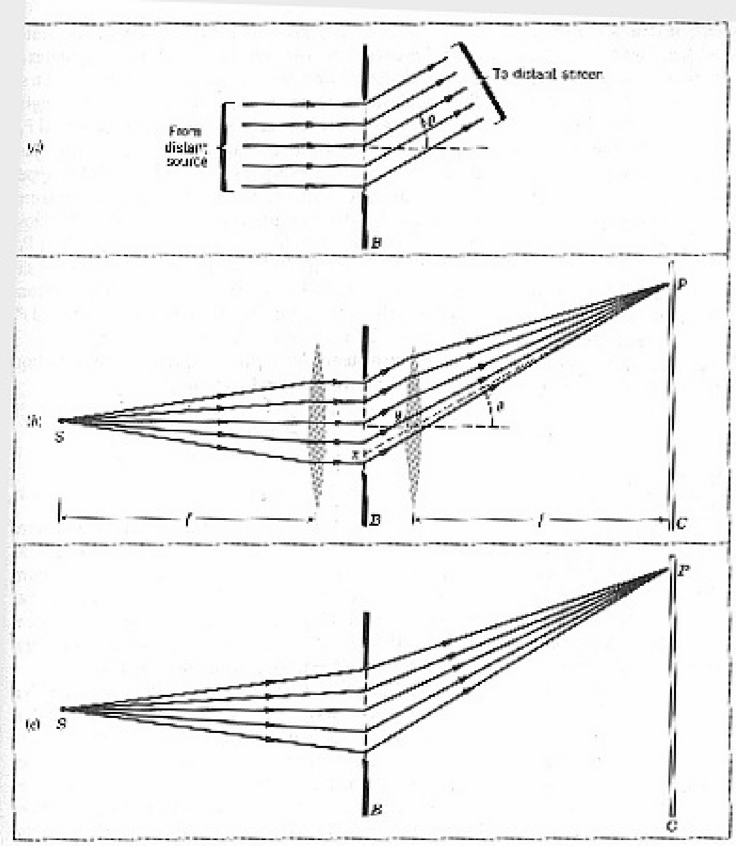

Figure 5 Light from point source S illuminates a slit in the opaque barrier B. The slit extends a long distance above and below the plane of the figure; this distance is much greater than the slit width a. The intensity at point P on screen C depends on the relative phases of the light received from various parts of the slit. (*a*) If source S and screen C are moved to large distances from the slit, both the incident and emergent light at B consist of nearly parallel rays.
(*b*) Rather than using large distances, the source and the screen can each be placed in the focal plane of a lens; once again, parallel light rays enter and leave the slit. (*c*) Without the lens, the rays are not parallel.

Figure 3

4.1.3 Single Slit Diffraction

The simplest diffraction pattern is that produced by a long narrow slit. Figure 4 shows a plane wave falling at normal incidence on a slit of width *a*. Let us first consider the central point P_0. Rays that leave the slit parallel to the central horizontal axis are brought to a focus at P_0. These rays are certainly in phase at the plane of the slit, and they remain in phase as they are brought to a focus by the lens. Since all rays arriving at P_0 are in phase, they interfere constructively and produce a maximum of intensity at P_0.

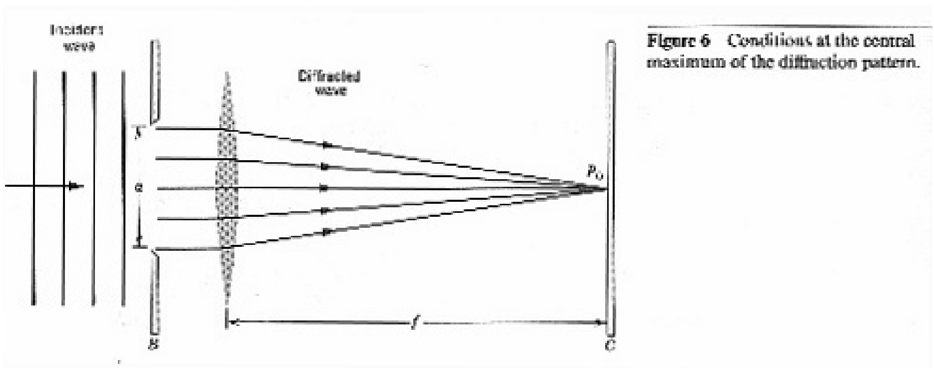

Figure 4

We now consider another point on the screen. Light rays that reach P_1 as shown in the figure below. The ray xP_1 passes undeflected through the centre of the lens and therefore determines θ. Ray r_1 originates at the top of the slit and ray r_2 at its centre. If θ is chosen so that the distance *bb'* in the figure is one-half wavelength, r_1 and r_2 are out of phase and interfere destructively at P_1. The same is true for a ray just below r_1 and another just below r_2.

In fact, for every ray passing through the upper half of the slit, there is a corresponding ray passing through the lower half, originating at a point *a/2* below the first ray, such that the two rays are out of phase at P_1. Every ray arriving at P_1 from the upper half of the slit interferes destructively with one coming from the bottom half of the slit. The intensity at P_1 is therefore zero, and P_1 is the first minimum of the diffraction pattern.

Since the distance *bb'* equals $\frac{a}{2}\sin\theta$, the condition for the first minimum can be written:

$$\frac{a}{2}\sin\theta = \frac{\lambda}{2}$$

or

$$a\sin\theta = \lambda \qquad (1)$$

Equation (1) shows that the central maximum becomes wider as the slit is made narrower. If the slit width is as small as one wavelength ($a=\lambda$), the first minimum occurs at $\theta = 90°$ ($\sin\theta = 1$), which implies that the central maximum fills the entire forward hemisphere.

In Figure 5 (b) below, the slit is divided into four equal zones, with a ray leaving the top of each zone. Let θ be chosen so that the distance *bb'* is one-half wavelength. Rays r_1 and r_2 then cancel at P_2. Rays r_3 and r_4 are also one-half wavelength out of phase and also cancel. Consider four other rays, emerging from the slit a given distance below these four rays. The two rays below r_1 and r_2 cancel, as do the two rays below r_3 and r_4. We can proceed across the entire slit and conclude again that no light reaches P_2; we have located a second point of zero intensity. The condition described requires that

$$\frac{a}{4}\sin\theta = \frac{\lambda}{2}$$

or

$$a \sin\theta = 2\lambda \qquad (2)$$

For a given slit width *a* and wavelength λ, equation (2) gives the position on the screen of the second minimum in terms of the angle θ. By extension of Eqs. (1) and (2), the general formula for the minima in the diffraction pattern on the screen *C* is:

$$a \sin\theta = m\lambda \qquad (3)$$

where *m*=1, 2, 3, … (minima)

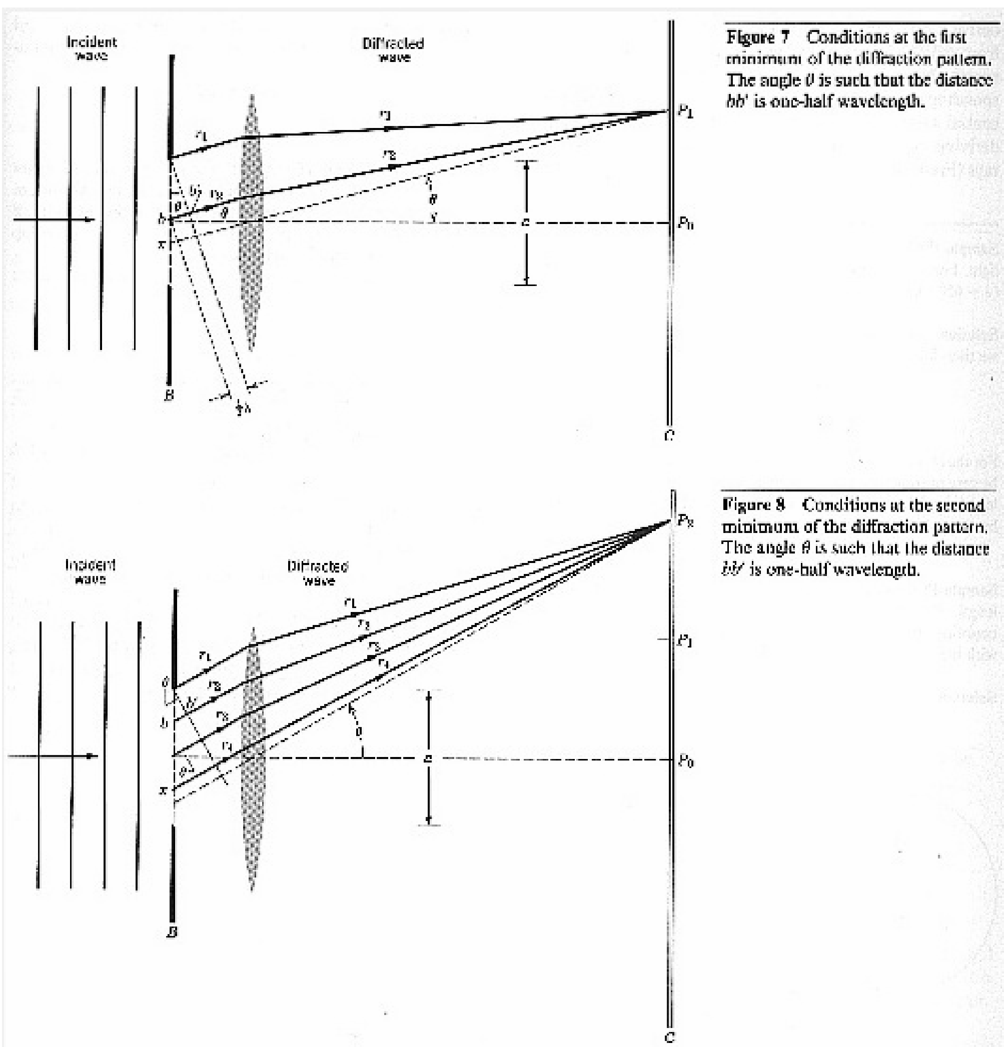

Figure 5 (a) and (b)

There is a maximum approximately halfway between each adjacent pair of minima. Note that Equation (3) suggests *two* minima (and corresponding maxima) for each *m,* one at an angle θ above the central axis and one below (corresponding to $m< 0$).

Example:
A slit of width *a* is illuminated by white light. For what value of *a* does the first minimum for red light ($\lambda=650n$m) fall at $\theta = 15°$?
Solution:
At the first minimum, $m=1$ in Equation (3). Solving for *a,* we then find

$$a = \frac{m\lambda}{\sin\theta} = 2510n\text{m}.$$

4.1.4 Diffraction at a Circular Aperture

In focusing an image, a lens passes only the light that falls within its circular perimeter. From this point of view, a lens behaves like a circular aperture in an opaque screen. Such an aperture forms a diffraction pattern analogous to that of a slit. Diffraction effects often limit the ability of telescopes and other optical instrument to form precise images.

The image formed by a lens can be distorted by other effects, including chromatic and spherical aberrations. These effects can be substantially reduced or eliminated by suitable shaping of the lens surfaces or by introducing correcting elements into the optical system. However, no amount of clever design can eliminate the effects of diffraction, which are determined only by the size of the aperture (the diameter of the lens) and the wavelength of the light. In diffraction, nature imposes a fundamental limitation on the precision of our instruments.

The mathematical analysis of diffraction by a circular aperture shows that, under *Fraunhofer* conditions, the first minimum occurs at an angle from the central axis given by

$$\sin\theta = 1.22\frac{\lambda}{d}$$

where *d* is the diameter of the aperture.

4.2 HOLOGRAPHY

Holography is a technique for recording and reproducing an image of an object through the use of interference effects. Unlike the two-dimensional images recorded by an ordinary photograph or television system, a holographic image is truly three-dimensional. Such an image can be viewed from different directions to reveal changing perspective. If you had never seen a hologram, you wouldn't believe it was possible.

The basic procedure for making a hologram is shown in Figure 6. We illuminate the object to be 'holographed' with monochromatic light, and we place a photographic film so that it is struck by scattered light from the object and also by direct light from the source. In practice, the light source must be a laser, for reasons to be discussed later. Interference between the direct and scattered light leads to the formation and recording of a complex interference pattern on the film.

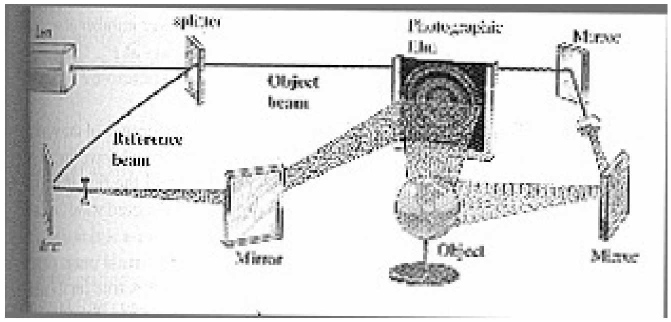

Figure 6

To form the image, we simply project light through the developed film. Two images are formed, a virtual image on the side of the film nearer the source and a real image on the opposite side.

A complete analysis of holography is beyond the scope of this course, but we can gain some insight into the process by looking at how a single point is 'holographed' and imaged. Consider the interference pattern that is formed on a sheet of photographic negative film by the superposition of an incident plane wave and a spherical wave, as shown in the diagram. The spherical wave originates at a point source *P* at a distance b_0 from the film; *P* may in fact be a small object that scatters part of the incident plane wave. We assume that the two waves are monochromatic and coherent and that the phase relation is such that constructive interference occurs at point *O* on the diagram.

Then constructive interference will *also* occur at any point *Q* on the film that is farther from *P* than *O* is by an integer number of wavelengths. That is, if $b_m - b_0 = m\lambda$, where *m* is an integer, then constructive interference occurs. The points where this condition is satisfied form circles on the film centred at *O,* with radii r_m given by

$$b_m - b_0 = ({b_0}^2 + {r_m}^2)^{1/2} - b_0 = m\lambda$$

where m=1, 2, 3, …

Solving this for ${r_m}^2$, we find

$${r_m}^2 = \lambda(2mb_0 + m^2\lambda)$$

Ordinarily, b_0 is very much larger than λ, so we neglect the second term in parentheses and obtain

$$r_m = (2m\lambda b_0)^{1/2}$$

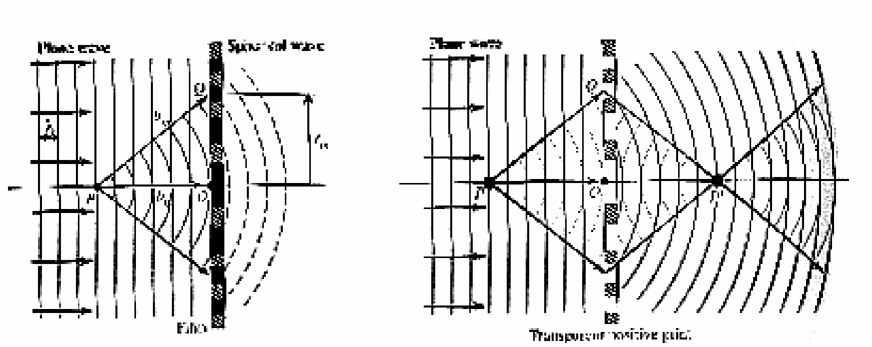

Figure 7

The interference pattern consists of a series of concentric bright circular fringes with radii given by the equation above.

Now we develop the film and make a transparent positive print, so the bright-fringe areas have the greatest transparency on the film. Then we illuminate it with monochromatic plane-wave light of the same wavelength λ that we used initially. In Figure 7, consider a point P' at a distance b_0 along the axis from the film. The centres of successive bright fringes differ in their distances from P' by an integer number of wavelengths, and therefore a strong *maximum* in the diffracted wave occurs at P'. That is, light converges to P' and then diverges from it on the opposite side. Therefore, P' is a *real image* of point *P*.

However, this is not the entire diffracted wave. The interference of the wave that spread out from all the transparent areas forms a second spherical wave that is diverging rather than converging. When this wave is traced back behind the film, it appears to be spreading out from point *P*. Thus the total diffracted wave from the hologram is a superposition of a spherical wave converging to form a real image at P' and a spherical wave that diverges as though it had come from the virtual image point *P*.

Because of the principle of superposition for waves, what is true for the imaging of a single point is also true for the imaging of any number of points. The film records the superposed interference pattern from the various points, and when the light is projected through the film, the various image points are reproduced simultaneously. Thus the images of an extended object can be recorded and reproduced just as for a single point object.

In making a hologram, we have to overcome several practical problems. First, the light used must be coherent over distances that are large in comparison to the dimensions of the object and its distance from the film. Ordinary light sources *do not* satisfy this requirement. Therefore, laser light is essential for making a hologram. Second, extreme mechanical stability is needed. If any relative motion of source, object, or film occurs during exposure, even by as much as a quarter of a wavelength, the interference pattern on the film is blurred enough to prevent satisfactory image formation. These obstacles are not insurmountable, however, holography has become important in research, entertainment, and a wide variety of technological applications.

4.3 FARADAY ROTATION

If a magnetic field is applied to a medium in a direction parallel with the direction in which light is passing through the medium, the result is a rotation of the polarisation direction of whatever is the light's polarisation state; in general, the polarisation ellipse is rotated. The phenomenon, known as the *Faraday Magneto-Optic Effect,* normally is used with a linearly polarised input, so that there is a straightforward rotation of a single polarisation direction. The magnitude of the rotation due to a field H, over a path length L, is given by:

$$\rho = V\int_0^L Hdl$$

where V is a constant known as the Verdet constant: V is a constant for any given material, but is wavelength dependent. Clearly, if H is a constant over the optical path, we have:

$$\rho = VHL$$

4.3.1 Physical Basis of Faraday Rotation

The physical reason for the effect is easy to understand in qualitative terms. When a magnetic field is applied to a medium the atomic electrons find it easier to rotate in one direction around the field than in the other: the Lorentz force acts on a moving charge in a magnetic field, and this will act radially on the electron as it circles the field. The consequent electron displacement will lead to two different radii of rotation and thus two different rotational frequencies and electric permittivities. Hence the field will result in two different refractive indices. Light which is circularly polarised in the 'easy' (say clockwise) direction will travel faster than that polarised in the 'hard' direction (anti-clockwise), leading to the observed effect.

Another important aspect of *the Faraday Magneto-Optic Effect* is that it is *'non-reciprocal'*. This means that linearly polarised light (for example) is always rotated in the same absolute direction in space, independently of the direction of propagation of the light. For an optically active crystal this is not the case: if the polarisation direction is rotated from right to left (say) on forward passage (as viewed by a fixed observer), it will be rotated from left to right on backward passage (as viewed by the same observer), so that back-reflection of light through an optically active crystal will result in light with zero final rotation, the two rotations having cancelled out. This is called a *reciprocal rotation* because the rotation looks the same for an observer who always looks in the direction of propagation of the light.

For the *Faraday Magneto-Optic* case, however the rotation always takes place in the same direction *with respect to the magnetic field* (not the propagation direction) since it is this which determines the 'easy' and 'hard' directions. Hence, an observer always looking in the direction of light propagation will see different directions of rotation since he/she is, in one case, looking along the field and, in the other, against it. It is a *non-reciprocal effect*.

4.3.2 Applications

The most popular application of Faraday Rotation is optical isolators. In these devices, light from a source passes through a linear polariser and then through a magneto-optic element which rotates the polarisation direction through 45°. Any light which is back-reflected by the ensuing optical system suffers a further 45° rotation during the backward passage, and in the same rotational direction, thus arriving back at the polariser rotated through 90°; it is thus blocked by the polariser.

Hence the source is isolated from back-reflections by the magneto-optic element/polariser combination which is thus known as a *Faraday Magneto-optic Isolator*. This is very valuable for devices whose stability is sensitive to back-reflection, such as lasers and optical amplifiers, and it effectively protects them from feedback effects.

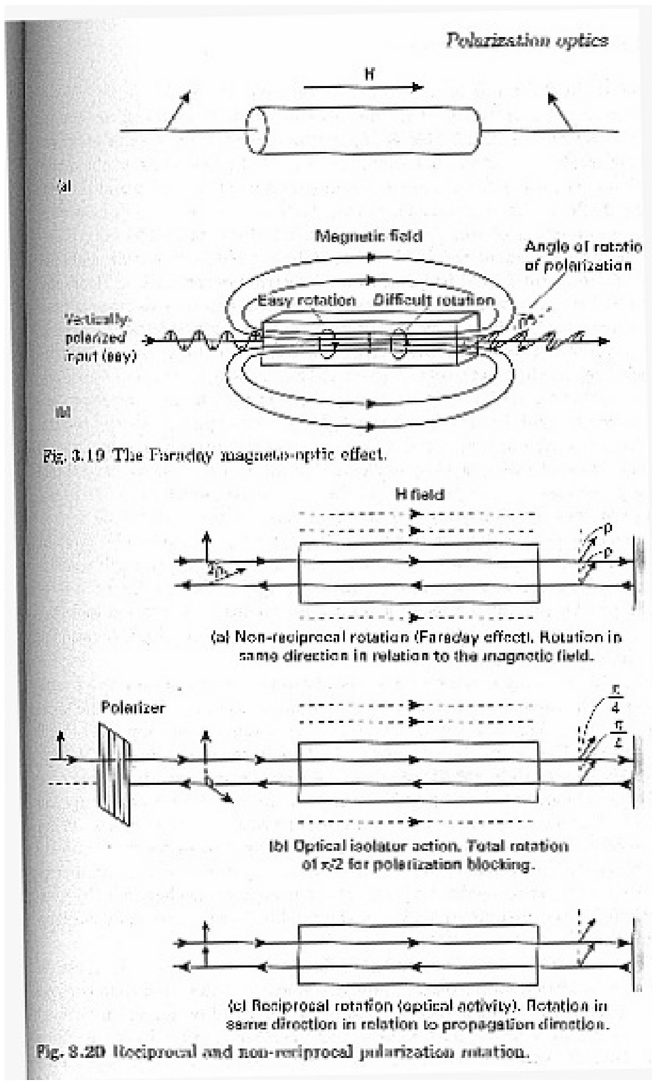

Figure 8

4.4 IONISED GASES (WAVES IN PLASMA)

In the earth's upper atmosphere, roughly from 50 km to 500 km in altitude, there are layers of ionised gases called ionosphere. The ionosphere contains free electrons and positive ions. These electrons and ions are generated when the atoms and molecules in the upper atmosphere absorb the ultraviolet radiation from the sun.

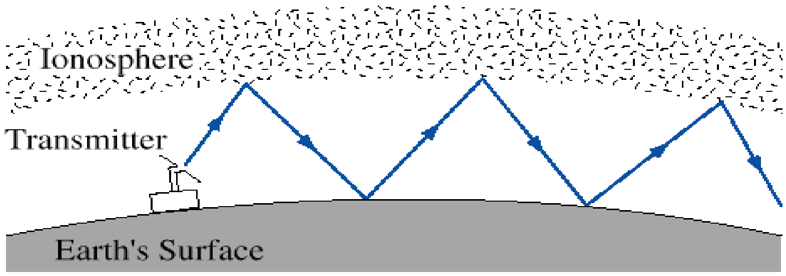

Figure 9

Depending on the time of the day, the thickness of each ionosphere layer varies. Ionised gases with equal electron and ion densities are called plasmas. Examples of plasmas are the ionosphere, the surface of the sun, hydrogen in a fusion reactor and those electrons in a highly conducting metal.

The ionosphere plays an important role in the propagation of electromagnetic waves and affects telecommunication. In the analysis, we shall ignore the motions of the ions and regards the ionosphere as a free electron gas.

Given an electron of charge *-e* and mass *m* in a time-harmonic electric field $\tilde{\mathbf{E}}$ at angular frequency ω (in *x* direction), this electron will experience a force of $-e\tilde{\mathbf{E}}$, that displaces it from a positive ion by a distance $\tilde{x}$. Thus, the expression of the balance of forces is

$$-e\tilde{E} = m\frac{d^2\tilde{x}}{dt^2} = -m\omega^2\tilde{x} \quad \Rightarrow \quad \tilde{x} = \frac{e}{m\omega^2}\tilde{E}$$

Then, the electric dipole moment is given by

$$\tilde{p} = -e\tilde{x}$$

and the polarisation (volume density of electric dipole moment) is given by

$$\tilde{P} = N\tilde{p} = -\frac{Ne^2}{m\omega^2}\tilde{E}$$

where *N* is the total number of electrons per unit volume. Thus,

$$\tilde{D} = \varepsilon_o\tilde{E} + \tilde{P} = \varepsilon_o(1 - \frac{Ne^2}{m\omega^2\varepsilon_o})\tilde{E} = \varepsilon_o(1 - \frac{\omega_p^2}{\omega^2})\tilde{E}$$

where $\omega_p = \sqrt{\frac{Ne^2}{m\varepsilon_o}}$ (rad/s) is called the plasma angular frequency, and the plasma frequency, f_p is then given by

$$f_p = \frac{\omega_p}{2\pi} = \frac{1}{2\pi}\sqrt{\frac{Ne^2}{m\varepsilon_o}} \text{ (Hz)}$$

The equivalent permittivity of the plasma is then

$$\varepsilon_p = \varepsilon_o \left(1 - \frac{\omega_p^2}{\omega^2}\right) = \varepsilon_o \left(1 - \frac{f_p^2}{f^2}\right) \qquad \text{(F/m)}$$

and the propagation constant is

$$\gamma = j\beta = j\omega\sqrt{\mu\varepsilon_o}\sqrt{1 - \left(\frac{f_p}{f}\right)^2}$$

and the intrinsic impedance is

$$\eta_p = \frac{\eta_o}{\sqrt{1 - \left(\frac{f_p}{f}\right)^2}}, \qquad \text{where } \eta_o = \sqrt{\frac{\mu_o}{\varepsilon_o}} = 120\pi\ \Omega.$$

- If $f < f_p$, then γ will become real, which means that the propagating wave will experience attenuation.

- If $f > f_p$, then γ becomes imaginary, and the wave will propagate through the ionosphere. Thus, the frequency f_p can be considered as the cut-off frequency.

Substituting the values of *e, m* and ε_o, we get

$$f_p \cong 9\sqrt{N} \qquad \text{(Hz)}$$

The electron density of the ionosphere varies with the time of the day, the season and other factors. It ranges from 10^{10} m^{-3} in the lowest layer to 10^{12} m^{-3} in the highest layer, this gives the range of f_p from 0.9 to 9 MHz.

For satellite communication, f must be higher than 9 MHz to ensure wave penetration through the ionosphere layer. For f < 0.9 MHz, the signal cannot penetrate the layer, but it can travel around the earth by way of multiple reflection.

4.5 DOPPLER RADAR

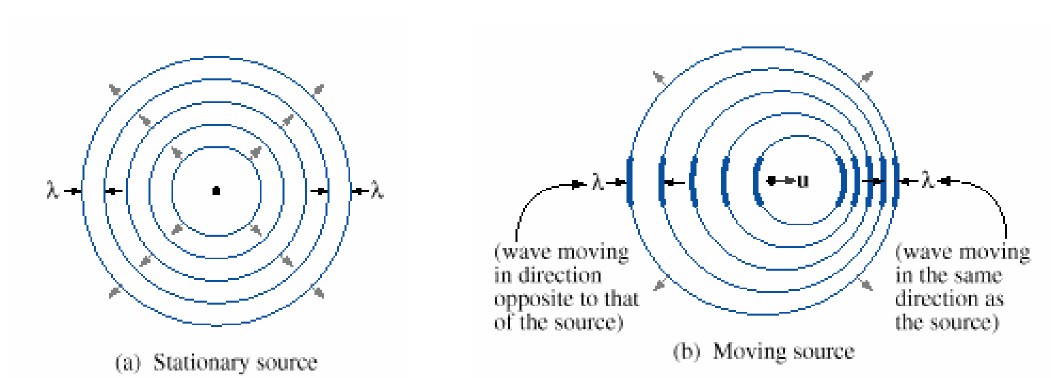

Figure 10

The Doppler Effect is a shift in the frequency of a wave caused by the motion of the transmitting source, the reflecting object, or the receiving system.

As illustrated in Figure 10, a wave radiated by a stationary isotropic point source forms equally spaced concentric circles as a function of time travel from the source. In contrast, a wave radiated by a moving source is compressed in the direction of motion and is spread out in the opposite direction. Compressing a wave shortens its wavelength, which is equivalent to increase its frequency. Conversely, spreading it out decreases its frequency.

The change in frequency is called the *Doppler Frequency Shift*, f_d. That is, if f_t is the frequency of the wave radiated by the moving source, then the frequency f_r of the wave that would be observed by a stationary receiver is

$$f_r = f_t + f_d$$

The magnitude and sign of f_d depend on the direction of the velocity vector relative to the direction of the range vector connecting the source to the receiver.

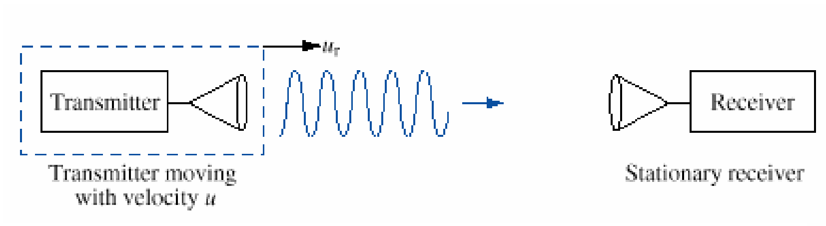

Figure 11

Consider a source transmitting an electromagnetic wave with frequency f_t as shown in Figure 11. At a distance R from the source, the electric field of the radiated wave is given by

$$E(R) = E_0 e^{j(\omega_t t - kR)} = E_0 e^{j\phi}$$

where E_0 is the wave's magnitude, $\omega_t = 2\pi f_t$, and $k = 2\pi/\lambda_t$, and λ_t is the wavelength of the transmitted wave.

The magnitude depends on the distance R and the gain of the source antenna, but it is not of concern as far as the Doppler Effect is concerned. The quantity ϕ is the phase of the radiated wave relative to its phase at $R = 0$ and to a reference time $t = 0$.

$$\phi = \omega_t t - kR = 2\pi f_t t - \frac{2\pi R}{\lambda_t}$$

If the source is moving toward the receiver or vice versa, as in the Figure 12, at a radial velocity u_r, then

$$R = R_0 - u_r t$$

where R_0 is the distance between the source and the receiver at $t = 0$.

Hence, $$\phi = 2\pi f_t t - \frac{2\pi}{\lambda_t}(R_0 - u_r t).$$

This is the phase of the signal detected by the receiver. The frequency of a wave is by definition equal to the time derivative of the phase ϕ divided by 2π. Thus,

$$f_r = \frac{1}{2\pi}\frac{d\phi}{dt} = f_t + \frac{u_r}{\lambda_t}$$

Thus, this leads to $f_d = u_r/\lambda_t$. For radar, the Doppler Shift happens twice, once for the wave from the radar to the target and again for the wave reflected by the target back to the radar. Hence, $f_d = 2u_r/\lambda_t$. The dependence of f_d on the direction is given by the dot product of the velocity and range unit vectors, which leads to

$$f_d = -2\frac{u_r}{\lambda_t} = -2\frac{u}{\lambda_t}\cos\theta$$

where u_r is the radial velocity component of u, and θ is the angle between the range vector and the velocity vector as shown in the Figure 12 with the direction of the range vector defined *from* the radar *to* the target. For a receding target (relative to the radar), $0 \le \theta \le 90°$, and for an approaching target, $90 \le \theta \le 180°$.

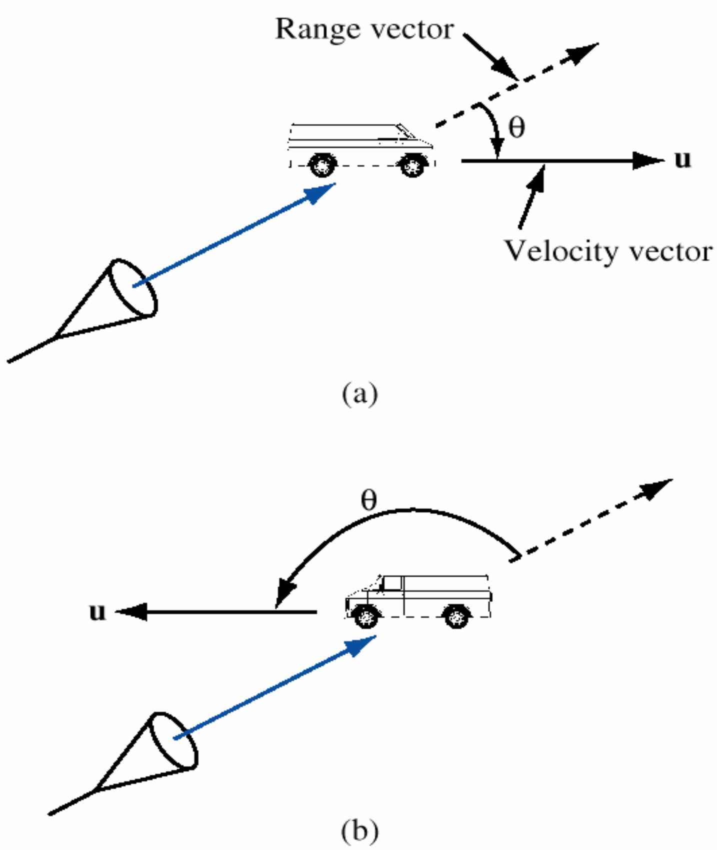

Figure 12

4.6 RAYLEIGH SCATTERING

Scattering is one of the two basic mechanisms that cause attenuation of light in glass fibre, the other being absorption. This is due to imperfections in fibre. When light is transmitted down the fibre, a portion of light is reflected back or back scattered to the source. This phenomenon is called *Rayleigh Scattering*.

It is due to changes in refractive index on a macroscopic level. Because light is reflected back power is lost from original input pulse. *Rayleigh Scattering* is reason for the sky to be blue. Molecules and small particles in air scatter blue light more than red and it is this blue light deflected from sun that we see when we look up into sky.

Effect of *Rayleigh Scattering* varies as $\frac{1}{\lambda^4}$. The attenuation is around 3dB/km at λ =0.7 micro metres for silica fibre and reduces rapidly for longer wavelengths.

Even if material is extremely pure and manufacturing process carefully controlled, solid glass will always have small fluctuations in refractive index and therefore *Rayleigh Scattering* is present in all glass materials. This is fundamental limit to attenuation of silica fibres at low optical wavelengths. At longer wavelengths, *Rayleigh Scattering* can be neglected