
Electrodynamics and Optics

Handout 1

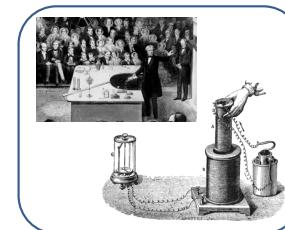
Prof. Paul Alexander

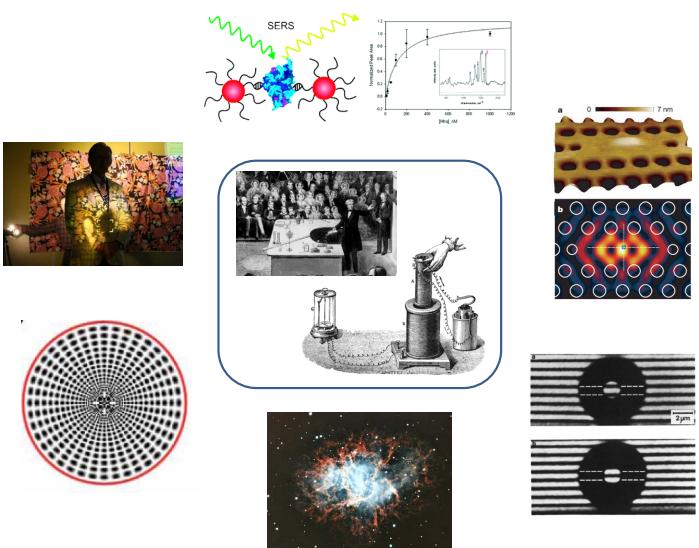
Michaelmas Term 2021

- 1806/1833 - Humphrey Davy/Michael Faraday: Electrochemical decomposition of potash and soda / Law of electrochemical equivalents.
- 1819 - Hans Christian Ørsted: Electric current in wire causes magnetic needle to deflect.
- 1821 - André-Marie Ampère: Forces between two current carrying wires.
- 1835 - Joseph Henry / Michael Faraday: Transformer, electromagnetic induction.
- 1864 - James Clerk Maxwell: Displacement current, electromagnetic theory of light.
- 1887 - Michelson-Morley experiment: No aether, speed of light the same in all inertial frames.
- 1897 - JJ Thomson: Discovery of electron.
- 1905 - Albert Einstein: Theory of special relativity.
- 1927 - Paul Adrien Maurice Dirac: Quantum theory of the emission and absorption of radiation → Quantum electrodynamics.

Electrodynamics and Optics: Some historic milestones

- Awareness of electrical and magnetic natural phenomena since prehistoric days; Magnetic needle compass (1088 - Shen Kuo, China)
- 1600 - William Gilbert: De Magnete (discovery of earth magnetism).
- 1675 - Robert Boyle: Electric repulsion and attraction can act across a vacuum.
- 1752 - Benjamin Franklin: proposal of kite experiment (charging a Leyden jar through kite); electricity generated by rubbing/friction same as lightning.
- 1771 - Henry Cavendish: electrostatic repulsion/attraction "with less power to the distance than cube".
- 1784 - Charles-Augustin de Coulomb: devised torsion balance with whom he discovered inverse square law.
- 1800 - Alessandro Volta: First battery.





Electrodynamics and Optics: Synopsis

- **Electromagnetic Waves:** Revision of Electromagnetism; Maxwell's equations; EM energy; vector potential; boundary conditions; EM waves; polarization.
- **Optics:** Polarized light; partial polarization. Light in media; anisotropic media; polarizers; waveplates; optical activity, Faraday rotation. Coherence: power-spectrum and examples; partial coherence; temporal coherence; fourier-transform spectroscopy; spatial coherence; imaging interferometry.
- **Electrodynamics:** Vector potential \mathbf{A} ; calculation of \mathbf{A} in simple cases; Aharonov-Bohm effect; flux quantization. Maxwell's equations in terms of \mathbf{A} and ϕ ; choice of gauge; wave equations for \mathbf{A} and ϕ ; general solution; retarded potentials.

- **Radiation:** Time-varying fields; radiation. Hertzian dipole; power radiated; angular distribution; magnetic dipoles; electric quadrupole.
- **Antennas:** Effective area; radiation resistance; power-pattern. Antenna arrays in astrophysics, radar.
- **Scattering:** Light scattering; cross-section Thomson and Rayleigh scattering; denser media.
- **Relativistic Electrodynamics:** Charges and currents; 4-current; 4-potential; transformation of \mathbf{E} and \mathbf{B} ; covariance of Maxwell's equations; magnetism as a relativistic effect; Relationship with other field theories
- **Radiation and Relativistic Electrodynamics:** Fields of a uniformly moving charge; Čerenkov radiation; accelerated charges; Larmor and Liénard formulæ; cyclotron and synchrotron radiation

BOOKS

- *Optics*, Hecht E (4th edn Addison Wesley 2002)
- *Optical Physics*, Lipson A, Lipson SG and Lipson H (4th edn CUP 2011)
- *Electromagnetic Fields and Waves*, Lorrain P, Corson DR and Lorrain F (3rd edn Freeman 1988)
- *Classical Electrodynamics*, Jackson JD (3rd edn Wiley 1998)

§1 Electromagnetic Waves

§1.1 Revision of Pt IB

- Electrostatic and magnetostatic *experiments* lead to the concepts of the electric field \mathbf{E} and the magnetic induction (or magnetic flux density) \mathbf{B} in vacuum.
- The *Lorentz force* on a point charge q moving with velocity \mathbf{v} is, *experimentally*,

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$$

and suggests that:

- \mathbf{E} is “force per unit charge”.
- \mathbf{B} is “force per unit current”.
- More generally, the charge is *distributed*, with $\rho_T(\mathbf{r})$ and $\mathbf{J}_T(\mathbf{r})$ the *total* charge and current density *distributions*.
- Continuity requires:

$$\int_S \mathbf{J}_T \cdot d\mathbf{S} = - \int_V \frac{\partial \rho_T}{\partial t} dV$$

In *static* circumstances and in vacuum various *experimental observations* are summarized by the integral equations:

$$\text{Gauss (ES): } \int_S \mathbf{E} \cdot d\mathbf{S} = \int_V \frac{\rho_T}{\epsilon_0} dV \quad (1.1)$$

$$\text{Gauss (MS): } \int_S \mathbf{B} \cdot d\mathbf{S} = 0 \quad (1.2)$$

$$\oint_C \mathbf{E} \cdot d\mathbf{l} = 0 \quad (1.3)$$

$$\text{Ampère: } \oint_C \mathbf{B} \cdot d\mathbf{l} = \int_S \mu_0 \mathbf{J}_T \cdot d\mathbf{S} \quad (1.4)$$

with the surface S bounding the volume V , or the contour C bounding the surface S , as appropriate.

In differential form (using Gauss's and Stokes's theorems):

$$\nabla \cdot \mathbf{E} = \frac{\rho_T}{\epsilon_0} \quad (1.5)$$

$$\nabla \cdot \mathbf{B} = 0 \quad (1.6)$$

$$\nabla \times \mathbf{E} = 0 \quad (1.7)$$

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J}_T \quad (1.8)$$

$\rho_T(\mathbf{r})$ and $\mathbf{J}_T(\mathbf{r})$ are *sources* for \mathbf{E} and \mathbf{B} respectively.

Eqn 1.3 implies that in electrostatics \mathbf{E} is a *conservative* field, and can be written in terms of a *scalar potential*:

$$\mathbf{E} = -\nabla \phi(\mathbf{r}) \quad (1.9)$$

Eqn 1.4 implies that in magnetostatics \mathbf{B} is a *conservative* field in regions where $\mathbf{J}_T = 0$, and can be written in terms of a *magnetic scalar potential*:

$$\mathbf{B} = -\mu_0 \nabla \phi_m(\mathbf{r}) \quad (1.10)$$

Eqns 1.9 and 1.5 give Poisson's Equation:

$$\nabla^2 \phi = -\frac{\rho_T}{\epsilon_0} \quad (1.11)$$

for the electrostatic potential. Similarly Eqns 1.10 and 1.6 give

$$\nabla^2 \phi_m = 0 \quad (1.12)$$

for regions where $\mathbf{J}_T = 0$.

Note that the equations for \mathbf{E} and \mathbf{B} are *independent* – electrostatics and magnetostatics are separate theories.

Faraday's experimental results on the *induction* of e.m.f.s in circuits produced by the *relative motions* of circuits and magnets, and Maxwell's insights on *displacement current*, both of which involve *time-varying fields*, allow electricity and magnetism to be *unified*.

This also marks out the unified theory of Electromagnetism as fully *relativistic*, as will become clear.

§1.2 Maxwell's Equations

Including the time-variations produces **Maxwell's Equations** for EM fields in vacuum:

$$\nabla \cdot \mathbf{E} = \frac{\rho_T}{\epsilon_0} \quad (1.13)$$

$$\nabla \cdot \mathbf{B} = 0 \quad (1.14)$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \quad (1.15)$$

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J}_T + \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \quad (1.16)$$

The equations for \mathbf{E} and \mathbf{B} are now interlinked by the *time-dependent* inductive and displacement current terms – **Electromagnetism**.

Media

In the presence of a medium

$$\left. \begin{array}{l} \mathbf{E} \\ \mathbf{B} \end{array} \right\} \rightarrow \left. \begin{array}{l} \text{electric} \\ \text{magnetic} \end{array} \right\} \text{ dipole moments per unit volume} \left. \begin{array}{l} \mathbf{P} \\ \mathbf{M} \end{array} \right\}$$

The electric polarization \mathbf{P} :

$$\mathbf{P} \rightarrow \left\{ \begin{array}{l} \text{volume charge density} \quad \rho_P = -\nabla \cdot \mathbf{P} \\ \text{surface charge density} \quad \sigma_P = \mathbf{P} \cdot \hat{\mathbf{n}} \end{array} \right. \quad (1.17)$$

where $\hat{\mathbf{n}}$ is the surface normal unit vector.

If \mathbf{P} varies with time, so does ρ_P , and there is an associated volume current density \mathbf{J}_P given by the continuity equation for the polarization charge:

$$\nabla \cdot \mathbf{J}_P = -\dot{\rho}_P = \nabla \cdot \dot{\mathbf{P}} \quad \rightarrow \quad \mathbf{J}_P = \dot{\mathbf{P}} \quad (1.18)$$

The magnetization \mathbf{M} :

$$\mathbf{M} \rightarrow \begin{cases} \text{volume current density} & \mathbf{J}_M = \nabla \times \mathbf{M} \\ \text{surface current density} & \mathbf{j}_M = \mathbf{M} \times \hat{\mathbf{n}} \end{cases} \quad (1.19)$$

Eqns 1.13 to 1.16 involve the *total* charge and current volume distributions.

$$\rho_T = \rho + \rho_P = \rho - \nabla \cdot \mathbf{P} \quad (1.20)$$

$$\mathbf{J}_T = \mathbf{J} + \mathbf{J}_P + \mathbf{J}_M = \mathbf{J} + \dot{\mathbf{P}} + \nabla \times \mathbf{M} \quad (1.21)$$

where ρ and \mathbf{J} refer to the “real”, free charges and currents present.

Substituting for ρ_T and \mathbf{J}_T in Eqns 1.13 to 1.16:

$$\nabla \cdot \mathbf{E} = \frac{\rho - \nabla \cdot \mathbf{P}}{\epsilon_0} \quad (1.22)$$

$$\nabla \cdot \mathbf{B} = 0 \quad (1.23)$$

$$\nabla \times \mathbf{E} = -\dot{\mathbf{B}} \quad (1.24)$$

$$\nabla \times \mathbf{B} = \mu_0 (\mathbf{J} + \dot{\mathbf{P}} + \nabla \times \mathbf{M}) + \mu_0 \epsilon_0 \dot{\mathbf{E}} \quad (1.25)$$

It is now convenient to define the *auxiliary* fields, \mathbf{D} the electric *displacement* and \mathbf{H} the magnetic *intensity*:

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P} \quad (1.26)$$

$$\mathbf{H} = \frac{\mathbf{B}}{\mu_0} - \mathbf{M} \quad (1.27)$$

Eqns 1.22 to 1.25 then become the standard form for MAXWELL'S EQUATIONS:

$$\nabla \cdot \mathbf{D} = \rho \quad (\text{ME1})$$

$$\nabla \cdot \mathbf{B} = 0 \quad (\text{ME2})$$

$$\nabla \times \mathbf{E} = -\dot{\mathbf{B}} \quad (\text{ME3})$$

$$\nabla \times \mathbf{H} = \mathbf{J} + \dot{\mathbf{D}} \quad (\text{ME4})$$

ρ and \mathbf{J} are *free* charge and current densities, and the response of the medium is included *via* the auxiliary fields \mathbf{D} and \mathbf{H} .

In integral form:

$$\text{Gauss (ES): } \int_S \mathbf{D} \cdot d\mathbf{S} = \int_V \rho dV \quad (1.28)$$

$$\text{Gauss (MS): } \int_S \mathbf{B} \cdot d\mathbf{S} = 0 \quad (1.29)$$

$$\text{Faraday, Lenz: } \oint_C \mathbf{E} \cdot d\mathbf{l} = -\frac{\partial}{\partial t} \int_S \mathbf{B} \cdot d\mathbf{S} \quad (1.30)$$

$$\text{Ampère: } \oint_C \mathbf{H} \cdot d\mathbf{l} = \int_S \left(\mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \right) \cdot d\mathbf{S} \quad (1.31)$$

Constitutive Relations

For linear, isotropic media:

$$\mathbf{P} = \epsilon_0 \chi \mathbf{E} \quad \text{and} \quad \mathbf{M} = \chi_m \mathbf{H} \quad (1.32)$$

where χ is the dielectric susceptibility and χ_m is the magnetic susceptibility.

Defining:

the relative permittivity (or dielectric "constant") $\epsilon = 1 + \chi$

the relative magnetic permeability $\mu = 1 + \chi_m$

gives the *constitutive relations*:

$$\mathbf{D} = \epsilon \epsilon_0 \mathbf{E} \quad \text{and} \quad \mathbf{B} = \mu \mu_0 \mathbf{H} \quad (1.33)$$

ϵ and μ encapsulate the *responses* (polarization and magnetization) of the medium to the EM fields.

§1.3 Electromagnetic Energy

The energy U of the EM fields in a volume V bounded by the surface S is:

$$U = \int_V \left(\frac{1}{2} \mathbf{E} \cdot \mathbf{D} + \frac{1}{2} \mathbf{B} \cdot \mathbf{H} \right) dV \quad (1.34)$$

and from the MEs it can be shown that:

$$\frac{dU}{dt} = - \int_S \mathbf{E} \times \mathbf{H} \cdot d\mathbf{S} - \int_V \mathbf{J} \cdot \mathbf{E} dV \quad (1.35)$$

the sum of a power flux and a dissipative term. So the *energy density* at a point is:

$$u = \frac{1}{2} \mathbf{E} \cdot \mathbf{D} + \frac{1}{2} \mathbf{B} \cdot \mathbf{H} \quad (1.36)$$

and the *energy flux* is given by the **Poynting Vector**:

$$\mathbf{N} = \mathbf{E} \times \mathbf{H} \quad (1.37)$$

§1.4 The Vector Potential

Note from Eqn 1.30 that when there is a time-dependent magnetic induction \mathbf{B} , \mathbf{E} is no longer in general a conservative field and $\mathbf{E} \neq -\nabla\phi$.

Now $\nabla \cdot (\nabla \times \mathbf{F}) = 0$ for any vector field \mathbf{F} , so ME2 implies that there is a vector field \mathbf{A} – the *magnetic vector potential* – of which the curl is \mathbf{B} .

$$\boxed{\mathbf{B} = \nabla \times \mathbf{A}}$$

(1.38)

From ME3:

$$\nabla \times \mathbf{E} = -\dot{\mathbf{B}} = -\nabla \times \dot{\mathbf{A}}$$

and integrating:

$$\boxed{\mathbf{E} = -\dot{\mathbf{A}} - \nabla\phi}$$

(1.39)

.... more on the vector potential in §??....

§1.5 Boundary Conditions

From the integral forms of Maxwell's Eqns (1.28 to 1.31) it is easy to derive conditions for the continuity of components of the EM fields at interfaces between two media:

- \mathbf{D}_\perp is conserved in the absence of any free charges at the interface
- \mathbf{B}_\perp is conserved
- \mathbf{E}_\parallel is conserved
- \mathbf{H}_\parallel is conserved in the absence of any free current at the interface

§1.6 EM Waves

For any vector field \mathbf{F} , $\nabla \times (\nabla \times \mathbf{F}) = \nabla(\nabla \cdot \mathbf{F}) - \nabla^2 \mathbf{F}$.

So, in regions with no free charge or current ($\rho = 0$, $\mathbf{J} = 0$), taking the curl of ME3 and substituting $\nabla \times \mathbf{H}$ from ME4:

$$\nabla \times (\nabla \times \mathbf{E}) = -\nabla \times \dot{\mathbf{B}}$$

$$\nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} = -\mu\mu_0 \frac{\partial}{\partial t}(\nabla \times \mathbf{H}) = -\mu\mu_0 \frac{\partial \dot{\mathbf{D}}}{\partial t} = -\epsilon\epsilon_0\mu\mu_0 \ddot{\mathbf{E}}$$

$$\boxed{\nabla^2 \mathbf{E} = \epsilon\epsilon_0\mu\mu_0 \ddot{\mathbf{E}}} \quad \text{and similarly} \quad \nabla^2 \mathbf{B} = \epsilon\epsilon_0\mu\mu_0 \ddot{\mathbf{B}} \quad (1.40)$$

i.e. both fields obey wave equations with the same speed v given by:

$$v = \frac{1}{\sqrt{\epsilon\epsilon_0\mu\mu_0}} = \frac{c}{\sqrt{\epsilon\mu}} = \frac{c}{n} \quad \text{with} \quad c = \frac{1}{\sqrt{\epsilon_0\mu_0}} \quad (1.41)$$

where c is the speed of EM waves in vacuum, and $n = \sqrt{\epsilon\mu}$ is the refractive index of the medium.

It is natural to look for harmonic-wave-like solutions to Eqn 1.40:

$$\mathbf{E} = \mathbf{E}_0 e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)} \quad ; \quad \mathbf{B} = \mathbf{B}_0 e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)}$$

where the amplitudes \mathbf{E}_0 \mathbf{B}_0 may be complex to include phase.

The wavevector can also be complex $\mathbf{k} \rightarrow \mathbf{k} + i\boldsymbol{\kappa}$ to include damped waves.

If

$$\begin{cases} \mathbf{k} \parallel \boldsymbol{\kappa} \\ \mathbf{k} \not\parallel \boldsymbol{\kappa} \end{cases} \quad \text{the wave is } \begin{cases} \text{homogeneous} \\ \text{inhomogeneous} \end{cases}.$$

If there are no free charges or currents, Maxwell's Equations take the form for these plane wave solutions:

$$\nabla \cdot \mathbf{D} = 0 \quad \rightarrow \quad \mathbf{k} \cdot \mathbf{D} = 0 \quad (1.42)$$

$$\nabla \cdot \mathbf{B} = 0 \quad \rightarrow \quad \mathbf{k} \cdot \mathbf{B} = 0 \quad (1.43)$$

$$\nabla \times \mathbf{E} = -\dot{\mathbf{B}} \quad \rightarrow \quad \mathbf{k} \times \mathbf{E} = \omega \mathbf{B} \quad (1.44)$$

$$\nabla \times \mathbf{H} = \dot{\mathbf{D}} \quad \rightarrow \quad \mathbf{k} \times \mathbf{H} = -\omega \mathbf{D} \quad (1.45)$$

From Eqns 1.44 and 1.45: (for real fields and \mathbf{k}):

$$\mathbf{B} \perp \mathbf{k} \text{ and } \mathbf{E} \quad : \quad \mathbf{D} \perp \mathbf{k} \text{ and } \mathbf{H} \quad (1.46)$$

while \mathbf{E} and \mathbf{H} are *not necessarily* \perp to \mathbf{k} .

For isotropic materials: ϵ and μ are simple scalars (see §2.3) so that $\mathbf{B} \parallel \mathbf{H}$ and $\mathbf{D} \parallel \mathbf{E}$.

Then $(\mathbf{E}, \mathbf{H}, \mathbf{k})$ are mutually perpendicular, and form a right-handed set with $\mathbf{E} \times \mathbf{H} = \mathbf{N} \parallel \mathbf{k}$. These are the familiar transverse EM waves.

The Poynting vector is in the direction of the wave *phase propagation*. Also

$$\frac{|\mathbf{E}|}{|\mathbf{B}|} = \frac{\omega}{k} = v = \frac{1}{\sqrt{\epsilon\epsilon_0\mu\mu_0}} = \frac{c}{n} \quad (1.47)$$

Since $\mathbf{B} = \mu\mu_0\mathbf{H}$:

$$\frac{|\mathbf{E}|}{|\mathbf{H}|} = \sqrt{\frac{\mu\mu_0}{\epsilon\epsilon_0}} = Z = \sqrt{\frac{\mu}{\epsilon}} Z_0 \quad (1.48)$$

is the *characteristic impedance*, with Z_0 the impedance of free space.

§1.6.1 Longitudinal waves

If there is free charge present, $\nabla \cdot \mathbf{D} \neq 0 \rightarrow \mathbf{k} \cdot \mathbf{D} \neq 0$, so \mathbf{D} and \mathbf{E} can have components parallel to \mathbf{k} .

For the special case of $\epsilon = 0$ and in the absence of free charge, Eqn. 1.42 is satisfied for $\mathbf{k} \not\perp \mathbf{E}$, and indeed for $\mathbf{k} \parallel \mathbf{E}$. – *Plasma waves* ($\epsilon(\omega_p) = 0$) are longitudinal.

§1.7 Linear Polarization

Arbitrarily taking $\mathbf{E}_1 = E_0(1, 0, 0)$ and $\mathbf{B}_1 = B_0(0, 1, 0) \rightarrow \mathbf{k} = k(0, 0, 1)$

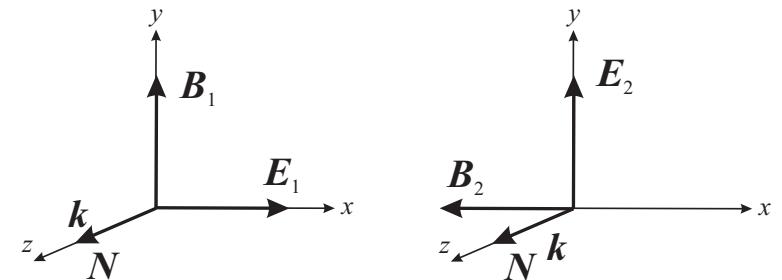


Figure 1.1: The vector orientations for x and y linearly-polarized EM waves travelling along Oz.

This represents a *plane-polarized* EM wave, conventionally “ x -polarized”.

$\mathbf{E}_2 = E_0(0, 1, 0)$, $\mathbf{B}_2 = B_0(-1, 0, 0)$, $\mathbf{k} = k(0, 0, 1)$ is a y -polarized wave.

At interfaces, plane-polarized EM waves are reflected and transmitted differently depending on whether \mathbf{E} is in the plane of incidence (“p”), or perpendicular to it (“s”):

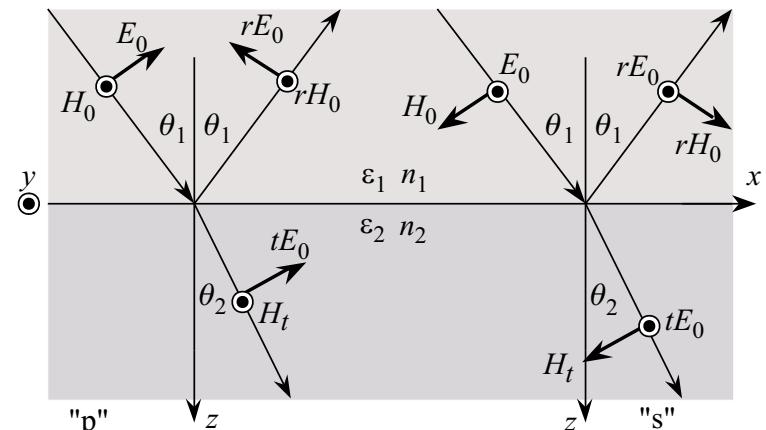


Figure 1.2: Geometry for calculating the Fresnel reflection and transmission coefficients r_p , r_s , t_p and t_s .

$$r_p = \frac{n_2 \cos \theta_1 - n_1 \cos \theta_2}{n_2 \cos \theta_1 + n_1 \cos \theta_2} \quad t_p = \frac{2n_1 \cos \theta_1}{n_2 \cos \theta_1 + n_1 \cos \theta_2} \quad (1.49)$$

$$r_s = \frac{n_1 \cos \theta_1 - n_2 \cos \theta_2}{n_1 \cos \theta_1 + n_2 \cos \theta_2} \quad t_s = \frac{2n_1 \cos \theta_1}{n_1 \cos \theta_1 + n_2 \cos \theta_2} \quad (1.50)$$

Note:

- $r_p = -r_s$ and $t_p = t_s$ at normal incidence, $\theta_1 = 0$.
- $r_p, r_s \rightarrow 1$ as $\theta_1 \rightarrow \frac{\pi}{2}$.
- $r_p = 0$ at $\theta_1 = \theta_B = \tan^{-1} \left(\frac{n_2}{n_1} \right)$, the Brewster angle.
- Reflection at the Brewster angle can be used in a *polarizing device*.

..... which concludes a rapid revision of relevant material from Pt IB.

§2 Optics

“Optics”: the material in this section is often encountered in phenomena at optical frequencies, i.e. in the wavelength (frequency) range of 400-800 nm (430-790 THz) in which our eyes are most sensitive. Much of the relevant wave physics does of course apply to other frequency ranges, such as X-rays ($\lambda < 1\text{nm}$), infrared ($\lambda = 1\text{ mm} - 800\text{ nm}$), microwaves ($\lambda = 1\text{ mm} - 1\text{m}$, $\nu = 0.3 - 300\text{ GHz}$) and radiowaves ($\lambda = 1\text{m} - 100\text{ km}$, $\nu = 3\text{kHz} - 0.3\text{ GHz}$).

At optical frequencies, in most circumstances $\mu \rightarrow 1$ and will be neglected. So:

$$n = \sqrt{\epsilon\mu} \rightarrow \sqrt{\epsilon}$$

Let's first have a look at more general polarisation states of light. For this let's consider *superpositions* of the *plane-wave solutions* (§1.7) of the wave equations derived from Maxwell's Equations...

Two perpendicularly plane-polarized waves $\mathbf{E}_1 = ia_1 E_0 e^{i(kz-\omega t)}$ and $\mathbf{E}_2 = ja_2 E_0 e^{i(kz-\omega t)}$ can be superposed to form other polarized waves:

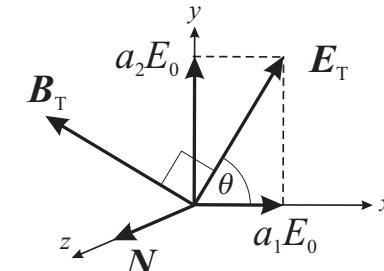


Figure 2.1: $\mathbf{E}_T = a_1 \mathbf{E}_1 + a_2 \mathbf{E}_2 = (a_1 \mathbf{i} + a_2 \mathbf{j}) E_0 e^{i(kz-\omega t)}$.

If a_1 and a_2 are both real, the x and y components of \mathbf{E}_T oscillate *in phase* to produce a wave *plane-polarized* at an angle:

$$\theta = \tan^{-1}(a_2/a_1) \text{ to Ox}$$

and with amplitude

$$E_T = E_0 \sqrt{a_1^2 + a_2^2}$$

§2.1 Circular and Elliptical Polarization

If a_1 and a_2 are *not* both real, the x and y components of \mathbf{E}_T oscillate with a fixed *phase difference* $\delta \neq 0$.

e.g. take $a_1 = 1$ and $a_2 = i = e^{i\frac{\pi}{2}}$: i.e. E_y lags behind E_x by $\delta = \frac{\pi}{2}$:

$$\mathbf{E}_T = E_0 (\mathbf{i} e^{i(kz-\omega t)} + \mathbf{j} e^{i(kz-[{\omega t - \frac{\pi}{2}}])})$$

At $z = 0$, and taking the real parts:

$$\mathbf{E}_T = E_0 \mathbf{i} \cos(\omega t) + E_0 \mathbf{j} \cos\left(\omega t - \frac{\pi}{2}\right) = E_0 \mathbf{i} \cos(\omega t) + E_0 \mathbf{j} \sin(\omega t)$$

At $t = 0$, and taking the real parts:

$$\mathbf{E}_T = E_0 \mathbf{i} \cos(kz) + E_0 \mathbf{j} \cos\left(kz + \frac{\pi}{2}\right) = E_0 \mathbf{i} \cos(kz) - E_0 \mathbf{j} \sin(kz)$$

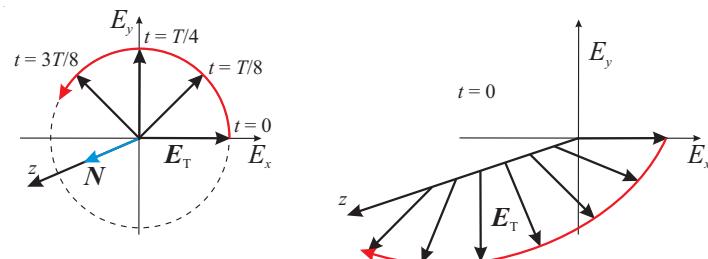


Figure 2.2: $a_1 = 1, a_2 = i$: \mathbf{E}_T at $z = 0$ rotates anti-clockwise. ($T = 2\pi/\omega$.)

Counter-intuitively (?), this is defined (Hecht p. 327) as a Left-Hand Circularly Polarized wave – LCP. An observer towards whom the light is propagating sees $\mathbf{E}_T(z = 0)$ rotate *anti-clockwise*, while the instantaneous field sweeps out a left-handed helix through space. (Electrical engineers use the opposite convention.)

For $a_1 = 1$ and $a_2 = -i$, $\delta = -\frac{\pi}{2}$ and RCP results. An observer towards whom the light is propagating sees $\mathbf{E}_T(z = 0)$ rotate *clockwise*.

For $|a_1| \neq |a_2|$ and $\delta \neq \pm\frac{\pi}{2}$, \mathbf{E} is *elliptically polarized* with the major/minor axes along directions in the xy -plane determined as follows. With $a_1 = a, a_2 = b e^{i\delta}$ (with a, b real):

$$E_x = a \cos \omega t \quad \text{and} \quad E_y = b \cos(\omega t - \delta)$$

$$\frac{E_y}{b} = \cos \omega t \cos \delta + \sin \omega t \sin \delta = \frac{E_x}{a} \cos \delta + \left(1 - \frac{E_x^2}{a^2}\right)^{\frac{1}{2}} \sin \delta$$

giving
$$\frac{E_x^2}{a^2} + \frac{E_y^2}{b^2} - 2 \cos \delta \frac{E_x}{a} \frac{E_y}{b} = \sin^2 \delta \quad (2.1)$$

the equation for an ellipse with axes at an angle α to the E_x, E_y directions:

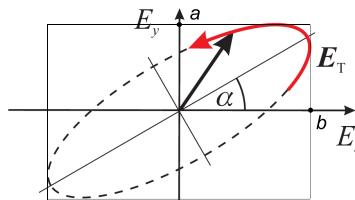


Figure 2.3: $a_1 = a, a_2 = b e^{i\delta}$: \mathbf{E}_T at $z = 0$ traces an ellipse. $\tan 2\alpha = \frac{2ab \cos \delta}{a^2 - b^2}$

§2.2 Jones notation

The complex amplitudes a_1 and a_2 of the two x and y linearly polarized waves can be used as the basis for a useful matrix approach for handling the effects of various optical devices on the polarization state – **Jones algebra**.

| Jones vector | x -pol | y -pol | θ -pol | RCP | LCP | General elliptical |
|--|--|--|--|--|---|--|
| | \underline{L}_x | \underline{L}_y | \underline{L}_θ | \underline{C}_R | \underline{C}_L | |
| $\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$ | $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ | $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ | $\begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix}$ | $\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix}$ | $\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix}$ | $\begin{pmatrix} a \\ b e^{i\delta} \end{pmatrix}$ |

The factors of $1/\sqrt{2}$ are for normalization.

[$\underline{L}_x, \underline{C}_R$ etc. are non-standard notation.]

Linear combinations, with appropriate phases, of the various polarization states can be used to form other polarization states. e.g.

$$\frac{\underline{C}_R + \underline{C}_L}{\sqrt{2}} = \underline{L}_x \quad (2.2)$$

The effects of various components of an optical system – polarizers, phase plates etc. can be represented by 2×2 Jones matrices, as will be seen below.

Polarised light is used in many scientific contexts, here are two examples from the fields of biophysics and spintronics:

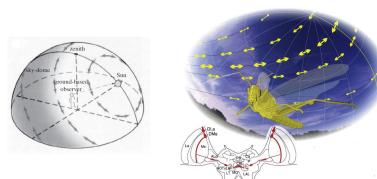


Figure 2.4: Illustration of polarisation vision used by certain insects (Cronin, Current Biology 21, R102 (2011))

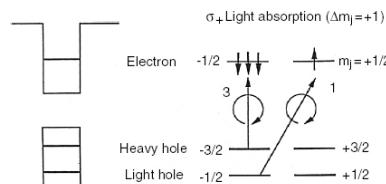


Figure 2.5: Generation of spin-polarised electrons in GaAs quantum wells by absorption of circularly polarised light

§2.3 Anisotropic Media

§2.3.1 Dichroism

Dichroic materials absorb light linearly polarized in one direction more than light polarized in the other, just like the wire grid polarizer but at molecular level.

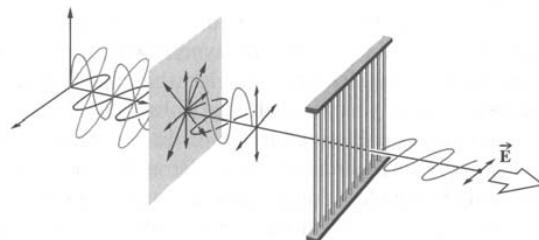


Figure 2.6: A wire grid polarizer operating at microwave frequencies.

e.g. Polaroid film — a plastic containing *conducting polymeric chains* aligned by stretching (along Oy say). The sheet is *anisotropic*, conducting along Oy but not along Ox. Light with $\mathbf{E} \parallel y$ is absorbed, and light with $\mathbf{E} \parallel x$ is not.

So when light passes through the sheet, only the E_x component is transmitted.

This is included in the Jones algebra as a *Jones matrix*:

$$\underline{\underline{J}}_x = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad (2.3)$$

$$\underline{\underline{J}}_x \underline{\underline{L}}_x = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \underline{\underline{L}}_x$$

$$\underline{\underline{J}}_x \underline{\underline{L}}_y = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} = 0$$

A polaroid with transmitting axis oriented at θ to Ox is represented by the matrix (see Examples Sheet):

$$\underline{\underline{J}}_\theta = \begin{pmatrix} \cos^2 \theta & \sin \theta \cos \theta \\ \sin \theta \cos \theta & \sin^2 \theta \end{pmatrix} \quad (2.4)$$

For initially linearly polarized light $\underline{\underline{L}}_x$ of intensity I_0 which then passes through polarizer $\underline{\underline{J}}_\theta$ the output is given by:

$$\underline{\underline{J}}_\theta \underline{\underline{L}}_x = \begin{pmatrix} \cos^2 \theta & \sin \theta \cos \theta \\ \sin \theta \cos \theta & \sin^2 \theta \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \times \sqrt{I_0} = \begin{pmatrix} \cos^2 \theta \\ \sin \theta \cos \theta \end{pmatrix} \times \sqrt{I_0}$$

so the transmitted intensity is:

$$I(\theta) = I_0 (\cos^4 \theta + \sin^2 \theta \cos^2 \theta) = I_0 \cos^2 \theta \quad (2.5)$$

...Malus's Law. For $\theta = \pi/2$: $I(\frac{\pi}{2}) = 0$.

Note: for light passing through a sequence of optical elements A,B,C the overall Jones matrix is

$$\underline{\underline{J}} = \underline{\underline{J}}_C \cdot \underline{\underline{J}}_B \cdot \underline{\underline{J}}_A \quad (2.6)$$

since $\underline{\underline{J}}_A$ is applied to the Jones vector first. So for *crossed polarizers*:

$$\underline{\underline{J}} = \underline{\underline{J}}_\theta \cdot \underline{\underline{J}}_{\frac{\pi}{2}+\theta} = 0$$

§2.3.2 Birefringence

Optical anisotropy also occurs naturally – e.g. in materials which are **birefringent** because they are structurally anisotropic — e.g. calcite.

It has been assumed so far that the medium is *isotropic*:

$$\mathbf{P} = \epsilon_0 \chi \mathbf{E} \quad \mathbf{D} = \epsilon \epsilon_0 \mathbf{E} \quad \mathbf{M} = \chi_m \mathbf{H} \quad \mathbf{B} = \mu \mu_0 \mathbf{H}$$

irrespective of the directions of \mathbf{E} and \mathbf{H} . i.e. that χ , ϵ , and χ_m are *scalars*.

For materials with an anisotropic crystal structure this is not so, and the relative permittivity (and therefore also the refractive index) is a *tensor of the second rank*, conveniently written in matrix form:

$$\mathbf{D} = \epsilon_0 \underline{\epsilon} \cdot \mathbf{E} \quad D_i = \epsilon_0 \sum_j \epsilon_{ij} E_j \quad \begin{pmatrix} D_x \\ D_y \\ D_z \end{pmatrix} = \epsilon_0 \begin{pmatrix} \epsilon_{xx} & \epsilon_{xy} & \epsilon_{xz} \\ \epsilon_{yx} & \epsilon_{yy} & \epsilon_{yz} \\ \epsilon_{zx} & \epsilon_{zy} & \epsilon_{zz} \end{pmatrix} \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix}$$

So in general $\mathbf{D} \nparallel \mathbf{E}$.

An important property of the dielectric tensor $\underline{\epsilon}$ can be deduced from energy flow considerations and the vector identity

$$\nabla \cdot (\mathbf{a} \times \mathbf{b}) = \mathbf{b} \cdot (\nabla \times \mathbf{a}) - \mathbf{a} \cdot (\nabla \times \mathbf{b}) \quad (2.7)$$

$$\mathbf{N} = \mathbf{E} \times \mathbf{H} \quad \text{and} \quad u = \frac{1}{2} \mathbf{E} \cdot \mathbf{D} + \frac{1}{2} \mathbf{B} \cdot \mathbf{H}$$

$$\begin{aligned} \frac{du}{dt} &= \frac{1}{2} (\dot{\mathbf{E}} \cdot \mathbf{D} + \mathbf{E} \cdot \dot{\mathbf{D}} + \dot{\mathbf{H}} \cdot \mathbf{B} + \mathbf{H} \cdot \dot{\mathbf{B}}) \quad \text{differentiating directly} \\ &= \dagger - \nabla \cdot \mathbf{N} - \mathbf{J} \cdot \mathbf{E} \quad \dagger \text{from energy conservation} \\ &= -[\mathbf{H} \cdot (\nabla \times \mathbf{E}) - \mathbf{E} \cdot (\nabla \times \mathbf{H})] - \mathbf{J} \cdot \mathbf{E} \\ &= \mathbf{H} \cdot \dot{\mathbf{B}} + \mathbf{E} \cdot (\mathbf{J} + \dot{\mathbf{D}}) - \mathbf{J} \cdot \mathbf{E} \quad \text{using ME3 and ME4} \\ &= \mathbf{H} \cdot \dot{\mathbf{B}} + \mathbf{E} \cdot \dot{\mathbf{D}} \end{aligned} \quad (2.8)$$

So comparing the first and last lines above:

$$(\dot{\mathbf{E}} \cdot \mathbf{D} - \mathbf{E} \cdot \dot{\mathbf{D}}) + (\dot{\mathbf{H}} \cdot \mathbf{B} - \mathbf{H} \cdot \dot{\mathbf{B}}) = 0$$

The dielectric and magnetic responses can (usually) be taken to be independent, so each of these bracketed terms must be zero, giving:

$$\mathbf{E} \cdot \dot{\mathbf{D}} = \mathbf{D} \cdot \dot{\mathbf{E}} \quad (2.9)$$

[ϵ is not a scalar, so this is not as obvious as it might look.]

When \mathbf{E} and \mathbf{D} might be complex, in any calculation of a physical quantity like u it is necessary to take the real parts first.

$$\Re(\mathbf{E}) \cdot \Re(\dot{\mathbf{D}}) = \Re(\mathbf{D}) \cdot \Re(\dot{\mathbf{E}})$$

Without loss of generality, it can be assumed that \mathbf{E} and \mathbf{D} are varying time-harmonically. For any complex numbers a and b varying time-harmonically, the time-averaged product of the real parts is:

$$\langle \Re(a) \Re(b) \rangle = \frac{1}{2} \Re(a^* b)$$

$$\begin{aligned} \text{So } \Re(\mathbf{E}^* \cdot \dot{\mathbf{D}}) &= \Re(\mathbf{D}^* \cdot \dot{\mathbf{E}}) = \Re(\dot{\mathbf{E}} \cdot \mathbf{D}^*) \\ \Re \sum_{ij} E_i^* \epsilon_{ij} \dot{E}_j &= \Re \sum_{ij} \dot{E}_i \epsilon_{ij}^* E_j^* \underset{(i \Rightarrow j)}{=} \Re \sum_{ij} \dot{E}_j \epsilon_{ji}^* E_i^* \\ &\longrightarrow \epsilon_{ij} = \epsilon_{ji}^* \end{aligned}$$

$$\boxed{\epsilon_{ij} = \epsilon_{ji}^*}$$

$$(2.10)$$

The dielectric tensor must be Hermitian.

For lossless media (and in the absence of “optical activity” (see §2.7)) the ϵ_{ij} are real, and then

$\underline{\epsilon}$ is symmetric

$\underline{\epsilon}$ can therefore be diagonalized by a rotation of the Cartesian axes.

So for such a material there is a set of orthogonal axes $\{\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3\}$ – the **principal axes** – such that:

$$\underline{\epsilon} = \begin{pmatrix} \epsilon_1 & 0 & 0 \\ 0 & \epsilon_2 & 0 \\ 0 & 0 & \epsilon_3 \end{pmatrix} = \begin{pmatrix} n_1^2 & 0 & 0 \\ 0 & n_2^2 & 0 \\ 0 & 0 & n_3^2 \end{pmatrix}$$

where n_1 , n_2 and n_3 are the *principal refractive indices*.

If $n_1 \neq n_2 \neq n_3$, the material is **biaxial**.

Materials (such as calcite) which have two of the principal refractive indices equal are called **uniaxial**.

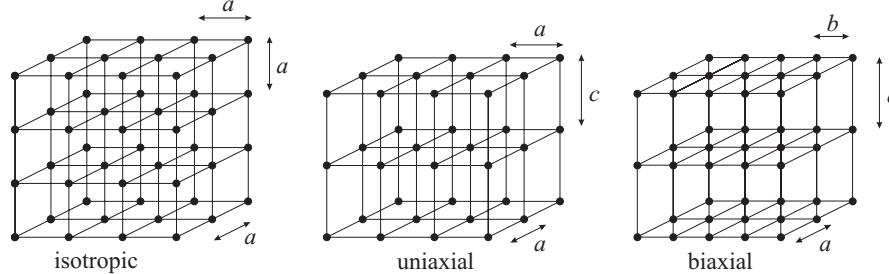


Figure 2.7: Simple crystal structures representing isotropic (cubic), uniaxial (tetragonal) and biaxial (orthorhombic) structures.

$$\underline{\epsilon} = \begin{pmatrix} \epsilon_1 & 0 & 0 \\ 0 & \epsilon_1 & 0 \\ 0 & 0 & \epsilon_1 \end{pmatrix} \quad \underline{\epsilon} = \begin{pmatrix} \epsilon_1 & 0 & 0 \\ 0 & \epsilon_1 & 0 \\ 0 & 0 & \epsilon_3 \end{pmatrix} \quad \underline{\epsilon} = \begin{pmatrix} \epsilon_1 & 0 & 0 \\ 0 & \epsilon_2 & 0 \\ 0 & 0 & \epsilon_3 \end{pmatrix}$$

Calcite (CaCO_3) is a naturally occurring mineral that crystallizes in a trigonal crystal structure with space group $R\bar{3}2/c$. The crystal plane perpendicular to the optical axis has three-fold symmetry. The refractive index depends on the whether the direction of the electric field is in the plane of the triangular CO_3 clusters or perpendicular to them.

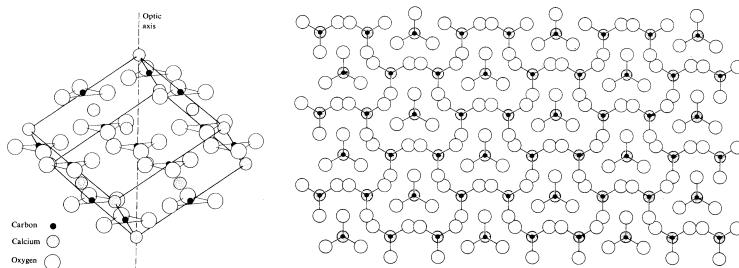


Figure 2.8: Crystal structure of calcite (CaCO_3) showing the triangular CO_3 clusters oriented with the plane perpendicular to the optical axis. The right view shows the atomic arrangement looking down along the optical axis.

For uniaxial systems like calcite it is conventional to take $n_1 = n_2 \neq n_3$ then:

$$\underline{\epsilon} = \begin{pmatrix} \epsilon_1 & 0 & 0 \\ 0 & \epsilon_1 & 0 \\ 0 & 0 & \epsilon_3 \end{pmatrix} = \begin{pmatrix} n_o^2 & 0 & 0 \\ 0 & n_o^2 & 0 \\ 0 & 0 & n_e^2 \end{pmatrix}$$

where the “o” denotes the “ordinary” directions, and “e” the “extraordinary” direction – the **optic axis**.

The **birefringence** for a uniaxial material is defined as $\Delta n = n_e - n_o$, and can be positive or negative.

Calcite has negative birefringence, with $\Delta n = n_e - n_o = -0.172$.

Note:

- (a) If \mathbf{E} lies along one of the principal axes of a *uniaxial* or *biaxial* medium, $\mathbf{D} \parallel \mathbf{E}$.
- (b) If \mathbf{E} is *perpendicular to the optic axis* of a *uniaxial* medium, $\mathbf{D} \parallel \mathbf{E}$.

§2.4 Linearly Polarized EM Waves in Anisotropic Materials

$$\text{Eqn 1.46: } \mathbf{B} \perp \mathbf{k} \text{ and } \mathbf{E} \quad : \quad \mathbf{D} \perp \mathbf{k} \text{ and } \mathbf{H}$$

For a non-magnetic (or magnetically isotropic) material μ is a scalar so $\mathbf{B} \parallel \mathbf{H}$. Then $\mathbf{D}, \mathbf{H} (\parallel \mathbf{B})$ and \mathbf{k} are mutually orthogonal.

§2.4.1 Special symmetry cases

If \mathbf{D} lies along a principal axis (say \hat{e}_1), $\mathbf{E} \parallel \mathbf{D}$. Then $\mathbf{E} \times \mathbf{H} = \mathbf{N} \parallel \mathbf{k}$.

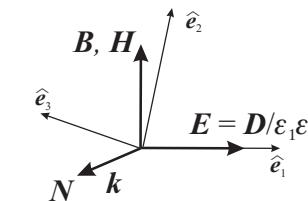


Figure 2.9: For $\mathbf{D} \parallel \hat{e}_1$ the wave propagates with velocity c/n_1 .

The wave equation is then identical to that for an isotropic medium with an ϵ corresponding to that for the axis along which \mathbf{D} and \mathbf{E} are directed.

2.4.2 Uniaxial Materials

For a *uniaxial* material:

$$\mathbf{D} = \epsilon_0 \underline{\epsilon} \cdot \mathbf{E} = \epsilon_0 \begin{pmatrix} \epsilon_1 & 0 & 0 \\ 0 & \epsilon_1 & 0 \\ 0 & 0 & \epsilon_3 \end{pmatrix} \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix} \quad (2.11)$$

So even if $\mathbf{D} \nparallel \hat{\mathbf{e}}_1$ or $\hat{\mathbf{e}}_2$, but lies in the $\hat{\mathbf{e}}_1\text{-}\hat{\mathbf{e}}_2$ plane for which $\epsilon_1 = \epsilon_2 = n_o^2$, $\mathbf{D} \parallel \mathbf{E}$.

So for $\mathbf{D} \perp \mathbf{Oz}$, the optic axis, $\mathbf{E} \parallel \mathbf{D}$ whatever its direction in this plane, and the wave velocity is c/n_o .

But if \mathbf{D} is *not* in the $\hat{\mathbf{e}}_1\text{-}\hat{\mathbf{e}}_2$ plane, it cannot be assumed that $\mathbf{E} \parallel \mathbf{D}$.

The Poynting vector $\mathbf{N} = \mathbf{E} \times \mathbf{H}$ is therefore *not necessarily parallel* to \mathbf{k} . – the phase and the energy may propagate in different directions.

Consider the formal equation $\epsilon_0 \mathbf{D} \cdot \mathbf{E} = 1$ or equivalently $\mathbf{D} \cdot \epsilon^{-1} \cdot \mathbf{D} = 1$ that must be obeyed for waves traveling in different directions as defined by the wavevector \mathbf{k} but with a given energy density. This defines an ellipsoid, the so-called *optical indicatrix*.

$$\frac{D_x^2}{\epsilon_1} + \frac{D_y^2}{\epsilon_2} + \frac{D_z^2}{\epsilon_3} = 1 \quad (2.12)$$

For each polarisation of the vector \mathbf{D} the corresponding \mathbf{E} vector can easily be shown to be normal to the ellipsoid surface at the tip of \mathbf{D} .

Interpretation: The refractive index experienced by a wave with polarisation vector \mathbf{D} is given by:

$$n^2 = \frac{c^2 \mu_0 D}{E \cos \alpha} = \frac{\epsilon_0 c^2 \mu_0 D^2}{\epsilon_0 E D \cos \alpha} = D^2 \quad (2.13)$$

Here we have used from ME that $|\mathbf{k} \times \mathbf{k} \times \mathbf{E}| = k^2 E \cos \alpha = \mu_0 \omega^2 D$ and $v^2 = \frac{c^2}{n^2} = \frac{\omega^2}{k^2}$. α is the angle between \mathbf{E} and the plane perpendicular to \mathbf{k} .

The length of the radius vector of the ellipsoid in each particular direction equals the refractive index for a wave with polarisation vector \mathbf{D} in that direction.

It is the polarisation direction, not the propagation direction that determines the wave velocity

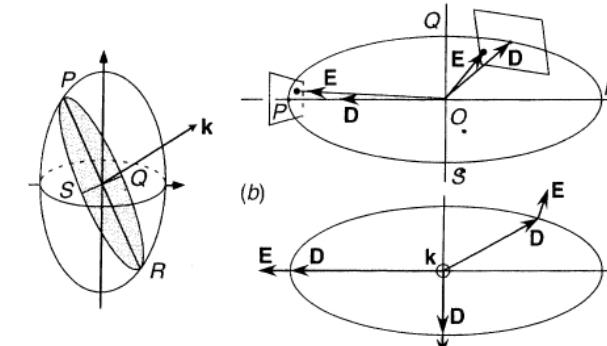


Figure 2.10: Illustration of the optical indicatrix. Here $n_e > n_o$ (i.e. the material has positive birefringence) so that $v_e < v_o$.

This can equivalently be represented in terms of the speed of Huygen's wavelets emanating from a point in the crystal and traveling in a particular propagation direction.

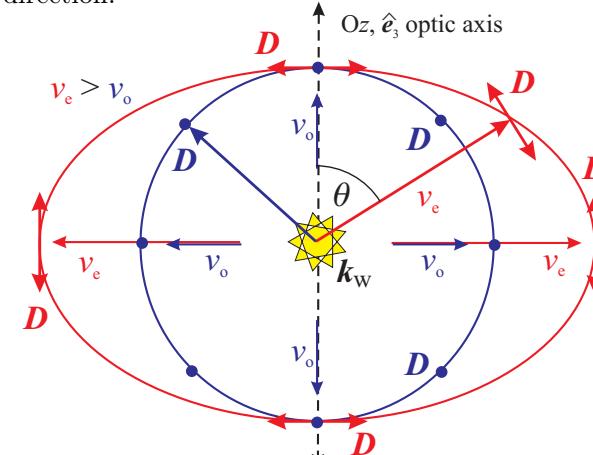


Figure 2.11: The wavelet's speed depends on the propagation direction and polarization of the wave. Here $n_e < n_o$ (i.e. the material has negative birefringence) so that $v_e > v_o$.

(a) For wavelets with $\mathbf{D} \perp \mathbf{Oz}$, the optic axis, $v_o = c/n_o$, independent of their direction (as shown on the previous slide).

These are “ordinary” wavelets, and form spherical wavefronts.

(b) For the linear polarizations orthogonal to (a), \mathbf{D} lies in the $\mathbf{k}_w\text{-}\hat{\mathbf{e}}_3$ plane and in general $\mathbf{E} \nparallel \mathbf{D}$.

The wavelet speed is $v_e = c/n_b$, where the effective refractive index n_b is (see Examples Sheet):

$$\frac{(n_b \sin \theta)^2}{n_e^2} + \frac{(n_b \cos \theta)^2}{n_o^2} = 1 \quad (2.14)$$

where θ is the angle between the wavelet direction and $\hat{\mathbf{e}}_3$.

These are “extra-ordinary” wavelets, and form ellipsoidal wavefronts since the system has cylindrical symmetry around Oz.

These wavelets are used in the Huygens construction to provide some understanding of ...

§2.4.3 Double Refraction

Consider linearly-polarized light normally incident on a surface S of a **uniaxial** crystal (e.g. calcite): $\mathbf{k}_{\text{inc}} \parallel \hat{\mathbf{n}}_S$, the surface normal. Take the optic axis $\hat{\mathbf{e}}_3$ to be at an angle θ to $\hat{\mathbf{n}}_S$ in the plane of the figure:

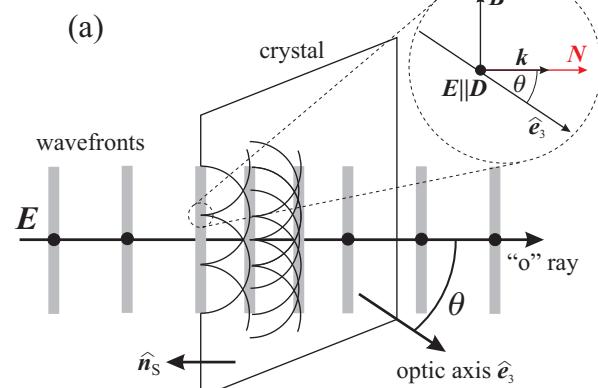


Figure 2.12: Light linearly polarized perpendicular to the plane of the diagram.

Inside the crystal, \mathbf{k} is the wavevector for the *transmitted ray* formed from the superposition of many *Huygens wavelets* propagating in all directions \mathbf{k}_w .

$\mathbf{k} \parallel \hat{\mathbf{n}}_S$, since $\mathbf{k}_{\parallel}(=0)$ is conserved at any interface. So \mathbf{k} makes an angle θ with $\hat{\mathbf{e}}_3$.

Case (a): $\mathbf{D} \perp \hat{\mathbf{e}}_3$, \mathbf{D} lies in the $\hat{\mathbf{e}}_1\text{-}\hat{\mathbf{e}}_2$ plane again, so $\mathbf{E} \parallel \mathbf{D}$ whatever its direction in this plane.

The wavelets for the Huygens construction have speed c/n_o , independent of direction, and are therefore *spherical* and

$$\mathbf{E} \times \mathbf{H} = \mathbf{N} \text{ is parallel to } \mathbf{k}$$

This is the “*ordinary*” ray. See Fig. 2.12.

At non-normal incidence (in the plane of Fig. 2.12 so that \mathbf{D} remains perpendicular to the optic axis) the ordinary ray would refract in the ordinary way corresponding to a medium with refractive index n_o .

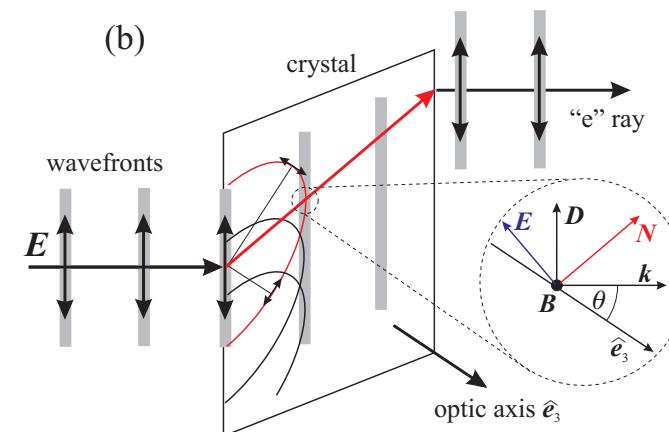


Figure 2.13: Light linearly polarized perpendicular to the plane of the diagram.

Case (b): For the linear polarization orthogonal to (a), \mathbf{D} lies in the plane including $\hat{\mathbf{e}}_3$, and in general $\mathbf{E} \nparallel \mathbf{D}$. The wavelet speed is c/n_b , where n_b is given by Eqn 2.14, and the Huygens wavelets are *ellipsoidal*.

The tangent planes to the superposition of these ellipsoidal wavelets give the overall wavefronts for the propagating ray, and the direction of \mathbf{k} for this ray remains normal to S.

\mathbf{D} is necessarily perpendicular to \mathbf{k} as shown, BUT in general $\mathbf{E} \nparallel \mathbf{D}$, so

$$\mathbf{E} \times \mathbf{H} = \mathbf{N} \text{ is NOT parallel } \mathbf{k}$$

So while the phase again propagates along the surface normal $\hat{\mathbf{n}}_S$, the energy propagates at an angle to the normal: the ray – the “extraordinary” ray – is therefore *laterally shifted* when it emerges from the crystal.

So an object viewed through a uniaxial crystal produces two images, one for the ordinary rays and one for the extraordinary rays – *double refraction*. The images obviously have different polarization properties. Double refraction in calcite was an early pointer to the polarization properties of light.

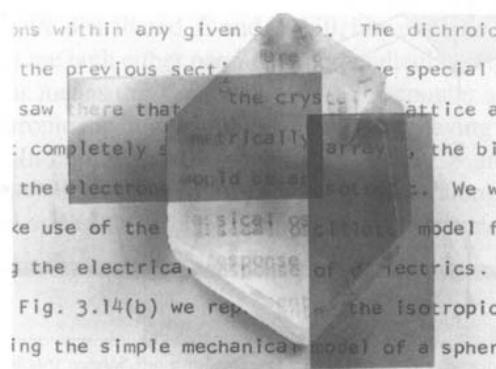


Figure 2.14: Double refraction in calcite.

§2.5 Optical Elements: Waveplates (or Retarders)

The considerable simplification when $\mathbf{E} \parallel$ one of the principal axes (§2.4.1) can be exploited in the construction of *waveplates* (*retarders*).

e.g. for normal incidence on a plate cut as shown:

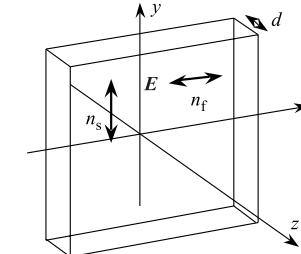


Figure 2.15: Here the principal axes are along Ox, y, z and $n_x = "n_f" < n_y = "n_s"$. Ox, y are the fast and slow axes.

A plane-polarized EM wave $e^{i(kz-\omega t)}$ travels along Oz at different speeds c/n_f or c/n_s depending on whether $\mathbf{E} \parallel$ Ox or Oy.

For $L_x : e^{ik(z=0)} \rightarrow e^{ik_f(z=d)}$ as it traverses the plate, where $k_f = \frac{wn_f}{c}$. So the plate applies phase terms depending on the different *optical thicknesses*:

$$\left. \begin{array}{l} e^{i\omega n_f d/c} \\ e^{i\omega n_s d/c} \end{array} \right\} \text{to} \left\{ \begin{array}{l} L_x \\ L_y \end{array} \right\}$$

So the Jones matrix for the plate can be written as

$$\begin{pmatrix} e^{i\omega n_f d/c} & 0 \\ 0 & e^{i\omega n_s d/c} \end{pmatrix} \propto \begin{pmatrix} e^{i\omega(n_f - n_s)d/c} & 0 \\ 0 & 1 \end{pmatrix} \propto \begin{pmatrix} e^{-i\Delta\phi/2} & 0 \\ 0 & e^{i\Delta\phi/2} \end{pmatrix} \quad (2.15)$$

where $\Delta\phi = \omega \frac{(n_s - n_f)d}{c}$ is the phase difference induced by the plate for waves polarized along the fast and slow axes.

$$\left\{ \begin{array}{l} \Delta\phi = \pi/2 \\ \Delta\phi = \pi \end{array} \right\} \rightarrow \left\{ \begin{array}{l} \lambda/4 \\ \lambda/2 \end{array} \right\} \text{in vacuum} - \text{a} \left\{ \begin{array}{l} \text{quarter-wave plate} \\ \text{half-wave plate} \end{array} \right\}$$

So for a $\lambda/4$ plate with fast axis along Ox the Jones matrix is:

$$\underline{\underline{J}}_{\lambda/4,x} = \begin{pmatrix} e^{-i\pi/4} & 0 \\ 0 & e^{i\pi/4} \end{pmatrix} = e^{-i\pi/4} \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix} \quad (2.16)$$

For the fast axis along Oy :

$$\underline{\underline{J}}_{\lambda/4,y} = \begin{pmatrix} e^{i\pi/4} & 0 \\ 0 & e^{-i\pi/4} \end{pmatrix} = e^{i\pi/4} \begin{pmatrix} 1 & 0 \\ 0 & -i \end{pmatrix} \quad (2.17)$$

(Note the prefactors have no net effect on the polarization state.)

So $\lambda/2$ plates:

$$\underline{\underline{J}}_{\lambda/2,x} = \begin{pmatrix} e^{-i\pi/2} & 0 \\ 0 & e^{i\pi/2} \end{pmatrix} = \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix} = e^{-i\pi/2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (2.18)$$

and

$$\underline{\underline{J}}_{\lambda/2,y} = \begin{pmatrix} e^{i\pi/2} & 0 \\ 0 & e^{-i\pi/2} \end{pmatrix} = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} = e^{i\pi/2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (2.19)$$

Quarter- and half-wave plates can, with linear polarizers, be used to manipulate and analyze the polarization state of light in detail. e.g.:

Suppose a plane polarized wave is incident on a wave plate (fast axis along Ox) with \mathbf{E} -vector at angle θ to Ox . The incident wave has Jones vector $\begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix}$, so the transmitted wave is:

$$\begin{pmatrix} e^{-i\Delta\phi/2} & 0 \\ 0 & e^{i\Delta\phi/2} \end{pmatrix} \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} = \begin{pmatrix} e^{-i\Delta\phi/2} \cos \theta \\ e^{i\Delta\phi/2} \sin \theta \end{pmatrix}$$

(i) if $\Delta\phi = \pi/2$ — a *quarter-wave plate*: $\rightarrow \begin{pmatrix} \cos \theta \\ i \sin \theta \end{pmatrix}$. (Dropping prefactor.)

From p. 26 this is seen to be *elliptically polarized* light with $\alpha = 0$. i.e. axes of the ellipse lie along Ox and Oy and have lengths $\cos \theta$ and $\sin \theta$.

If $\theta = 45^\circ$ this is circularly polarized – LCP.

If $\theta = 0, 90^\circ$ this is linearly polarized.

(ii) if $\Delta\phi = \pi$ — a *half-wave plate*: $\rightarrow \begin{pmatrix} \cos \theta \\ -\sin \theta \end{pmatrix}$.

Plane polarized light with the \mathbf{E} -vector directed at $-\theta$ to Ox . The direction of plane polarization is *rotated*.

If also $\theta = 45^\circ$, the plane of polarization becomes perpendicular to the original.

(iii) if $\theta = 0$ or $\pi/2$, the incident plane polarized wave is unaffected whatever the value of the plate thickness d , since the incident \mathbf{E} is parallel to one of the principal axes of the plate.

Summarizing:

| | | |
|----------------------|---|--|
| x -polarizer: | $\underline{\underline{J}}_x$ | $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ |
| y -polarizer: | $\underline{\underline{J}}_y$ | $\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$ |
| θ -polarizer: | $\underline{\underline{J}}_\theta$ | $\begin{pmatrix} \cos^2 \theta & \sin \theta \cos \theta \\ \sin \theta \cos \theta & \sin^2 \theta \end{pmatrix}$ |
| $\lambda/4$ plate: | $\underline{\underline{J}}_{\lambda/4,x}$ | $\begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}$ |
| $\lambda/2$ plate: | $\underline{\underline{J}}_{\lambda/2,x}$ | $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ |

§2.6 Induced Birefringence

Optical anisotropy can be induced in otherwise isotropic materials. e.g.:

§2.6.1 Photoelasticity

Photoelasticity (or stress birefringence) is the birefringence induced when an otherwise isotropic material is subjected to *stress*.

The corresponding distortion of the material, at molecular level, changes the dielectric response, producing an anisotropic permittivity tensor.

A transparent *isotropic* object placed between crossed linear polarizers would not change the initial polarization, so no light should be transmitted.

But if stressed to induce birefringence, linearly polarized light passing through the material has its polarization state affected in complex ways depending on the stress field, and some light is transmitted, allowing patterns of stress in transparent mechanical structures placed between crossed polarizers to be visualized.

§2.6.2 The Kerr and Pockels Effects

In an applied electric field E_0 an otherwise isotropic material can become uniaxially birefringent, with the optic axis along E_0 . In liquids and gases, this can be understood as arising from the alignment of anisotropic molecules by the field.

Since in an otherwise isotropic liquid or gas the optical properties cannot be sensitive to the sign of the field the change in the refractive index must be quadratic in the electric field to lowest order: $\Delta n = \lambda_0 K E_0^2$. K is the Kerr constant. The Kerr effect is an example of a non-linear optical phenomenon.

In solids a similar effect, the so-called Pockels effect, is associated with the lowering of the crystal symmetry by the induced *macroscopic* dielectric polarization. Crystals that do not have a centre of inversion symmetry could distinguish between positive and negative fields. Therefore, a linear electric field dependence is possible for the Pockels effect.

Suitable materials can therefore be used to make *voltage-controlled wave-plates*.

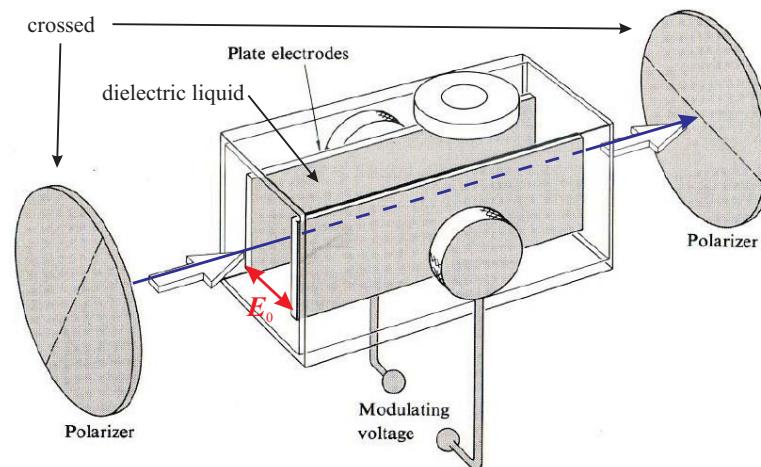


Figure 2.16: Schematic of a Kerr cell, after Hecht Fig. 8.56. The applied electric fields can switch the properties of the dielectric at high frequency, forming the basis of fast optical modulators.

§2.7 Optical Activity

§2.7.1 Chiral Materials

Some materials, while isotropic, have molecules with a *chiral* structure — a *handedness* built in at molecular level. The defining characteristic is that the *mirror image* of the molecular structure cannot be superposed on the original. An important biological example would be an α -amino acid of the form $\text{H}_2\text{NCHRCOOH}$.

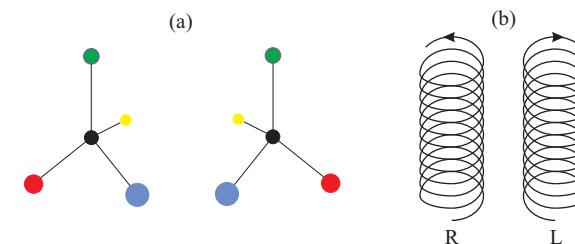


Figure 2.17: (a) the molecule and its mirror image cannot coincide. (b) Right and Left handed helices. For natural materials (e.g. dextrose, quartz) one type, right or left handed is usually prevalent.

This *chirality* is in-built, even in liquid form with no molecular organization or crystal structure.

A chiral material is “*optically active*”, or “*circularly birefringent*”, and responds differently to LCP and RCP waves. These then are the natural polarization states to use – the characteristic waves for the medium – with two refractive indices, n_L and n_R (just as for birefringence and plane polarization).

A chiral wave plate introduces different phases for LCP and RCP waves, just as a birefringent plate does for \underline{L}_x and \underline{L}_y .

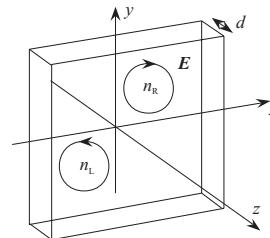


Figure 2.18: A slab of chiral material different optical thicknesses $n_L d$ and $n_R d$, for LCP and RCP light.

A plane polarized wave can be written as the sum of two counter-rotating circularly polarized waves (Eqn 2.2):

$$\underline{L}_x = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{\sqrt{2}} \left[\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix} + \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix} \right] = \frac{\underline{C}_R + \underline{C}_L}{\sqrt{2}}$$

A chiral plate of thickness d applies a phase term of $e^{i\omega n_{R,L} d/c}$ to $\underline{C}_{R,L}$ (cf. Eqn 2.15)

Using $\Delta\phi = \frac{\omega(n_L - n_R)d}{c}$ as the relative phase, the \underline{L}_x wave above becomes:

$$\frac{1}{\sqrt{2}} \left[\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix} e^{-i\Delta\phi/2} + \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix} e^{i\Delta\phi/2} \right] = \begin{pmatrix} \cos(\Delta\phi/2) \\ -\sin(\Delta\phi/2) \end{pmatrix}$$

– another *plane polarized* wave, but with its plane *rotated clockwise* by

$$\Delta\phi/2 = \frac{\omega(n_L - n_R)d}{2c} = \frac{\pi(n_L - n_R)d}{\lambda}$$

where λ is the wavelength *in air*.

The rotation per unit length – the *specific rotatory power* – is:

$$\alpha = \frac{\pi(n_L - n_R)}{\lambda} = \frac{1}{2}(k_L - k_R) = \frac{\omega}{2c}(n_L - n_R) \quad (2.20)$$

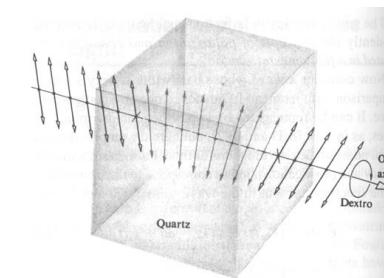


Figure 2.19: The orientation of plane-polarized light is continuously rotated as the light passes through an optically active dextrorotatory medium.

If the plane of polarization has rotated *clockwise* ($\alpha > 0$, $n_L > n_R$), the medium is said to be “*dextrorotatory*”, or *d-rotatory*; if anticlockwise ($\alpha < 0$, $n_L < n_R$), “*levorotatory*”, or *l-rotatory*.

§2.7.2 The Faraday Effect

Birefringence can be introduced by applying an electric field to an isotropic material. Likewise chirality can be introduced to a non-chiral system by an applied magnetic field \mathbf{B}_0 . Faraday first demonstrated this with glass.

\mathbf{B} alters the response of the electrons in the system to the optical fields, modifying the dielectric constant. But \mathbf{B}_0 is *directional*, so the medium is *not isotropic* as optically active media are; the geometry matters.

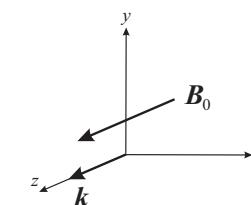


Figure 2.20: The Faraday geometry.

The basic points can be illustrated with a simple example: EM waves with $\mathbf{k} \parallel z$ in a *plasma*, with $\mathbf{B}_0 \parallel z$ – the “*Faraday geometry*”.

For each electron the equation of motion is (neglecting any scattering and the negligible effect of the magnetic field of the EM wave):

$$m\ddot{\mathbf{r}} = -e(\mathbf{E} + \dot{\mathbf{r}} \times \mathbf{B}_0) \quad (2.21)$$

where $\mathbf{E} = \begin{pmatrix} E_x \\ E_y \end{pmatrix} e^{-i\omega t}$ is the electric field of the incident EM wave. \mathbf{r} will also vary harmonically: $\mathbf{r} = \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} e^{-i\omega t}$ then:

$$m\ddot{x} = -eE_x - eB_0\dot{y} \rightarrow -\omega^2x_0 = -\frac{e}{m}E_x + i\omega\frac{eB_0}{m}y_0 \quad (2.22)$$

$$m\ddot{y} = -eE_y + eB_0\dot{x} \rightarrow -\omega^2y_0 = -\frac{e}{m}E_y - i\omega\frac{eB_0}{m}x_0 \quad (2.23)$$

so, with $\frac{eB_0}{m} = \omega_c$, the *cyclotron frequency*:

$$-\omega^2(x_0 + iy_0) = -\frac{e}{m}(E_x + iE_y) + \omega\omega_c(x_0 + iy_0) \quad (2.24)$$

$$-\omega^2(x_0 - iy_0) = -\frac{e}{m}(E_x - iE_y) - \omega\omega_c(x_0 - iy_0) \quad (2.25)$$

If $|E_x| = |E_y| = E$ then $|x_0| = |y_0| = a$ say, and Eqn 2.25 corresponds to an oscillator driven by a LCP EM wave $\underline{C}_L = E \begin{pmatrix} 1 \\ i \end{pmatrix}$, with a LCP circular displacement $a \begin{pmatrix} 1 \\ i \end{pmatrix}$ as the response.

So (with n the electron number density) there is a *circular polarization*:

$$P_L = -ena \begin{pmatrix} 1 \\ i \end{pmatrix} = \epsilon_0 \chi_L E \begin{pmatrix} 1 \\ i \end{pmatrix}$$

where the effective susceptibility for \underline{C}_L is

$$\chi_L = -\left(\frac{ne^2}{\epsilon_0 m}\right) \frac{1}{\omega^2 - \omega\omega_c} = -\frac{\omega_p^2}{\omega^2 - \omega\omega_c}$$

with ω_p the *plasma frequency*. Similarly for \underline{C}_R waves from Eqn 2.24:

$$\chi_R = -\frac{\omega_p^2}{\omega^2 + \omega\omega_c}$$

So:

$$\epsilon_{\frac{L}{R}}(\omega) = 1 - \frac{\omega_p^2}{\omega(\omega \mp \omega_c)} \quad (2.26)$$

If $B_0 = 0$, $\omega_c = 0$ and the familiar result for a simple plasma is recovered.

So n_L and n_R are different, just as for optical activity, and the plane of polarization of a plane polarized wave is steadily rotated as it passes along Oz.

From p. 63, the angle of rotation is:

$$\begin{aligned} \theta &= \frac{\Delta\phi}{2} = \frac{\omega(n_L - n_R)d}{2c} \\ &= \frac{\omega d}{2c} \left[\left(1 - \frac{\omega_p^2}{\omega(\omega - \omega_c)}\right)^{1/2} - \left(1 - \frac{\omega_p^2}{\omega(\omega + \omega_c)}\right)^{1/2} \right] \\ &\approx -\frac{\omega_p^2 \omega_c d}{2c \omega^2 \sqrt{1 - \frac{\omega_p^2}{\omega^2}}} \quad \text{if } B_0, \text{ and hence } \omega_c, \text{ is small} \end{aligned} \quad (2.27)$$

The **Verdet coefficient** V is defined from

$$\theta = VB_0d \quad (2.28)$$

so for the plasma in a weak field B_0 :

$$V = -\frac{e\omega_p^2}{2mc\omega^2 \sqrt{1 - \frac{\omega_p^2}{\omega^2}}} \quad (2.29)$$

The Dielectric Tensor:

Solving Eqns 2.24 and 2.25 for x_0 and y_0 :

$$x_0 = \frac{e}{m} E_x - \frac{i\omega_c}{\omega} \frac{e}{m} E_y \quad (2.30)$$

$$y_0 = \frac{e}{m} E_y + \frac{i\omega_c}{\omega} \frac{e}{m} E_x \quad (2.31)$$

with a corresponding polarization density

$$\mathbf{P} = -ne \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} e^{-i\omega t} = \epsilon_0 \underline{\underline{\chi}} \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix} e^{-i\omega t}$$

$\mathbf{B}_0 \parallel z$ so the z -motion is unaffected by the magnetic field and the z susceptibility is as for an unmagnetized plasma.

So

$$\underline{\underline{\epsilon}} = \underline{\underline{1}} + \underline{\underline{\chi}} = \begin{pmatrix} 1 - \frac{\omega_p^2}{\omega^2 - \omega_c^2} & \frac{i\omega_c\omega_p^2}{\omega(\omega^2 - \omega_c^2)} & 0 \\ -\frac{i\omega_c\omega_p^2}{\omega(\omega^2 - \omega_c^2)} & 1 - \frac{\omega_p^2}{\omega^2 - \omega_c^2} & 0 \\ 0 & 0 & 1 - \frac{\omega_p^2}{\omega^2} \end{pmatrix} \quad (2.32)$$

is the dielectric matrix for a plasma in the presence of $\mathbf{B}_0 \parallel z$ in the absence of damping.

The *off-diagonal terms* reflect the *magnetically-induced chirality* of the system.

The dielectric matrix is Hermitian, as required from Eqn 2.10.

If $\mathbf{B}_0 \rightarrow 0$, $\omega_c \rightarrow 0$, and the familiar result for an isotropic plasma emerges.

§2.8 Interference and Partial Polarization**§2.8.1 Interference of polarized waves**

- Interference: two waves of the same frequency are superposed and the *amplitudes* added to produce the net intensity. Unlike acoustics, QM, Pt IB Optics ..., where the amplitudes are taken to be scalars, for EM waves the amplitudes are *vectors*.

Consider the superposition of two waves along Oz perpendicularly plane polarized with arbitrary phase difference δ (i.e. an elliptically polarized wave as in §2.1).

- The net Poynting vector is:

$$\begin{aligned} \mathbf{N} &= [\hat{\mathbf{x}}E_{1x} \cos \omega t + \hat{\mathbf{y}}E_{2y} \cos(\omega t + \delta)] \\ &\quad \times [\hat{\mathbf{x}}H_{2x} \cos(\omega t + \delta) + \hat{\mathbf{y}}H_{1y} \cos \omega t] \\ &= \hat{\mathbf{z}}[E_{1x}H_{1y} \cos^2 \omega t - E_{2y}H_{2x} \cos^2(\omega t + \delta)] \end{aligned} \quad (2.33)$$

identical to the result taking the two plane polarized waves *independently*. (Remember that H_{2x} is negative.)

So the *intensity at a point* produced by *two superposed perpendicularly plane polarized waves* is the *sum of the intensities* of the two waves — perpendicularly plane polarized waves *do not interfere*.

§2.8.2 Unpolarized and Partially Polarized Light

Light from most natural sources has the *direction of the \mathbf{E} -vector changing randomly* with time and space (as well as variations in its amplitude and phase as discussed later for COHERENCE in §2.9).

It is *unpolarized*, and time-variations of the x and y components of \mathbf{E} are *uncorrelated*.

Plane polarized beams produced by passing the same beam of unpolarized light through x and y polarizers are therefore *mutually incoherent*. There is no well-defined phase difference — so they *cannot interfere*.

A beam which includes both polarized and unpolarized light (with intensities I_{pol} and I_{unpol}) is *partially polarized*, with a *degree of polarization*:

$$V = \frac{I_{\text{pol}}}{I_{\text{pol}} + I_{\text{unpol}}} \quad (2.34)$$

§2.8.3 The Fresnel-Arago Laws

Summarizing the conclusions of §2.8.1 and §2.8.2: the **Fresnel-Arago laws**:

- (i) Two beams, plane polarized parallel, interfere (if coherent).
- (ii) Two beams, plane polarized perpendicularly, cannot interfere (even if perfectly coherent).
- (iii) Two plane polarized beams cannot interfere (even if polarized *parallel*) if they are *derived* from perpendicularly polarized components of *unpolarized* light since these must be *mutually incoherent*.

Coherence will be discussed further in §2.9.

§2.9 Coherence

For simplicity, in this discussion the waves will be taken to have scalar amplitudes.

Interference phenomena, diffraction etc, rely on the *well defined phase* between wavelets (determined by optical path lengths etc.) which are eventually summed for the overall wave amplitude.

This can be exact only for *purely monochromatic* waves – a single, well-defined frequency.

However, Fourier Theory means that only *infinitely long* (in time and space) waves can be purely monochromatic. Obviously:

NO SUCH WAVES ACTUALLY EXIST.

Real sources and waves (even lasers) are at best “*quasi-monochromatic*”....

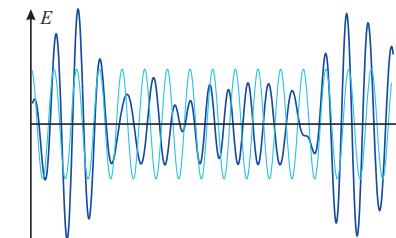


Figure 2.21: A quasi-monochromatic waveform deviates in amplitude and phase from the pure reference wave.

The formal theory of *coherence* was developed by Zernike in 1938 building on earlier work of Michelson and Fizeau. It describes non-monochromatic wavefields *quantitatively* and formed the basis of many modern experimental techniques in radio and optical astronomy and spectroscopy for materials characterisation. Examples to be discussed include:

- Fourier transform spectroscopy as used in modern vibrational infrared spectroscopy.
- Aperture synthesis in radio and optical telescopes

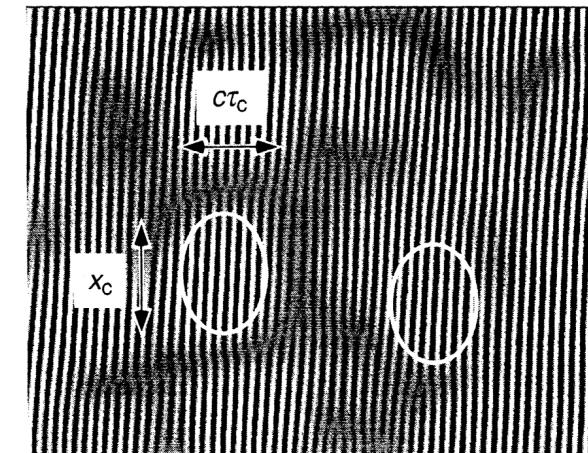


Figure 2.22: A highly schematic representation of a partially coherent wavefield. **Phase registration** is lost over a distance $c\tau_c - \tau_c$ is the (*temporal*) *coherence length* along the direction of propagation – and over a distance x_c – the (*spatial*) *coherence width* – perpendicular to the direction of propagation.

§2.9.1 The Power Spectrum

Some basic Fourier Theory: using time as the variable, the Fourier Transform relates the *time domain* t to the *frequency domain* ω .

A time-dependent function $f(t)$ can be described in terms of its *temporal harmonics* $F(\omega)e^{i\omega t}$. [Note the sign convention is not that used for waves earlier, but that of your *Maths Formulae* booklet.]

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{i\omega t} d\omega \quad : \quad F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt \quad (2.35)$$

The power in the frequency range ω to $\omega + d\omega$ – the Power Spectrum – is:

$$P(\omega)d\omega \sim |F(\omega)|^2 d\omega \quad (2.36)$$

(a) Lasers:

For a pure harmonic wave of frequency ω_0 : $f(t) \sim \cos(\omega_0 t + \alpha)$:

$$\begin{aligned} F(\omega) &= \text{FT}[f(t)] \\ &= \frac{1}{2} \int \left(e^{i(\omega_0 t + \alpha)} + e^{-i(\omega_0 t + \alpha)} \right) e^{-i\omega t} dt \\ &\propto e^{i\alpha} \delta(\omega - \omega_0) + e^{-i\alpha} \delta(\omega + \omega_0) \end{aligned} \quad (2.37)$$

So the power spectrum for a pure harmonic wave is a pair of δ -fns at $\pm\omega_0$.

This is the case (almost) for a *laser*, where stimulated emission produces an (almost) perfectly harmonic wave – an (almost) ideal *line source*.

(b) Spectral lines:

- gas discharge lamps, astrophysics.

(i) Lifetime broadening: Unstimulated emission from an isolated, stationary atom can be represented semi-classically as a decaying harmonic wave, beginning at $t = 0$ and characterized by a decay time τ_s or a scattering frequency $\omega_s = 1/\tau_s$:

$$F(\omega) = \int_0^\infty e^{i\omega_0 t} e^{-\omega_s t} e^{-i\omega t} dt = \frac{1}{\omega - \omega_0 - i\omega_s} \quad (2.38)$$

$$P(\omega) \sim |F(\omega)|^2 = \frac{1}{(\omega - \omega_0)^2 + \omega_s^2} \quad (2.39)$$

$P(\omega)$ is now a *Lorentzian peak* centred on ω_0 and with a linewidth (FWHM) of $2\omega_s$, determined by the decay time τ_s .

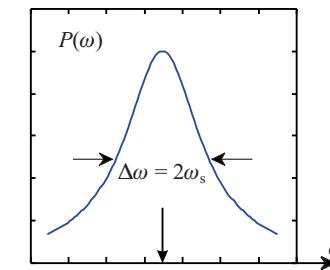


Figure 2.23: The Lorentzian lineshape (\equiv the response of a damped oscillator).

(ii) Thermal broadening: Radiation from atoms moving along the line of sight (say the x -axis) with velocity v_x will be *Doppler-shifted* in frequency:

$$\omega - \omega_0 = \frac{\omega_0 v_x}{c} \quad (2.40)$$

From kinetic theory, the distribution of atomic/molecular velocities along Ox is *Gaussian*:

$$f(v_x) dv_x \sim \exp\left(-\frac{mv_x^2}{2k_B T}\right) dv_x \quad (2.41)$$

so that the observed frequency spectrum will also be a Gaussian:

$$P(\omega) = C \exp\left(-\frac{m(\omega - \omega_0)^2 c^2}{2\omega_0^2 k_B T}\right) = C \exp\left(-\frac{(\omega - \omega_0)^2}{2\sigma^2}\right) \quad (2.42)$$

So, even neglecting the natural linewidth, if the atoms form a gas there will be Doppler broadening and the observed spectrum is again not a δ -fn but a narrow line in ω -space of linewidth (FWHM):

$$\text{FWHM} = 2.36 \sigma = 2.36 \omega_0 \left(\frac{k_B T}{mc^2} \right)^{\frac{1}{2}} \quad (2.43)$$

Not surprisingly, this depends on the *temperature* T of the gas.

(iii) Pressure broadening: In a gas, an individual atom is subject to collisions with other atoms, which at the very least perturb the phase correlation of the emitted wave before and after each collision.

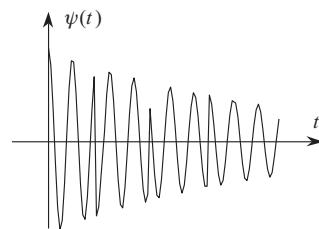


Figure 2.24: Amplitude profile for atom subjected to collisions.

The mean time τ_1 between collisions is:

$$\tau_1 = \frac{1}{4N\bar{v}A} = \frac{b\sqrt{T}}{p} \quad (2.44)$$

where N is the number density of atoms of collision cross-section A and with mean velocity \bar{v} , and p is the pressure.

Without detailed mathematics, this also produces (irrespective of the natural lifetime τ) another Lorentzian frequency profile:

$$P(\omega) = |\Psi(\omega)|^2 \sim \frac{1}{(\omega - \omega_0)^2 + 1/\tau_1^2} \quad (2.45)$$

At fixed T (i.e. fixed \bar{v}) the collision rate obviously increases with p , and so does the observed linewidth — *pressure broadening*. So usually try to run lamps at low pressure to give sharpest lines.

(iv) Overall: So for a gas discharge lamp, there are several line broadening mechanisms, with either Lorentzian or Gaussian profiles, and the overall lineshape is some mixture of the two depending on the gas conditions.

For a gas discharge lamp, the output is the superposition of large numbers of independent photons from individual similar atoms.

This is essentially harmonic with frequency ω_0 say, but with an amplitude and phase which have some random fluctuations — *quasi-monochromatic light*:

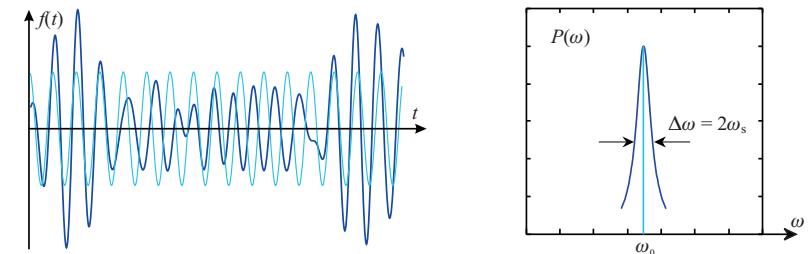


Figure 2.25: A quasi-monochromatic waveform and the corresponding frequency spectrum.

ω_0 is the underlying harmonic wave, and the linewidth $\Delta\omega$ can be related to some overall broadening equivalent to a lifetime τ_s .

At the extreme, a large number of atoms of different emission frequencies, or oscillators in the surface of an incandescent black body, result in **WHITE LIGHT** with a very broad power spectrum covering the visible:

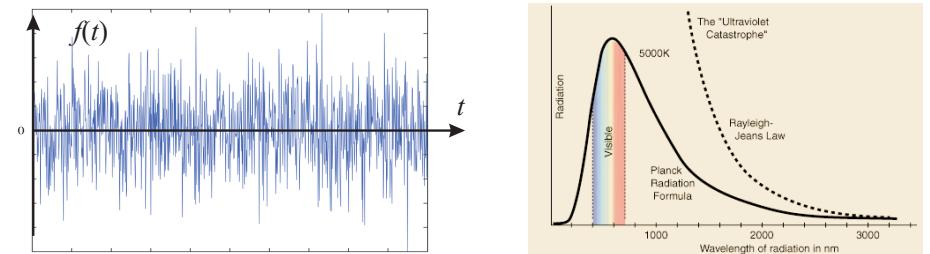


Figure 2.26: The amplitude and frequency spectrum for a “white” light source.

Totally irregular profile \Rightarrow many Fourier frequencies \Rightarrow wide power spectrum \Rightarrow zero coherence.

§2.9.2 Coherence and Interference

If one illuminated a Michelson interferometer or a double slit experiment with incoherent, white light the waveform at one point in time/space is not correlated / not coherent with that at another so there are generally *no interference effects*.

But obviously for a strongly correlated – strongly *coherent* – waveform (a laser) interference effects are very clear.

But what about diffraction and interference using an intermediate waveform – quasi-monochromatic – *partially coherent* — light?

It turns out that interference ideas provide a useful quantified description for the *degree of correlation*, or *degree of coherence*, of the partially coherent wavefield arising from a light source.

§2.9.3 A Partially Coherent Wavefield

Coherence means that a recipe exists that allows the prediction of the amplitude and phase of the wavefield at some location \mathbf{r}_2 and time t_2 from the knowledge of the amplitude and phase of the wavefield at location \mathbf{r}_1 and time t_1 .

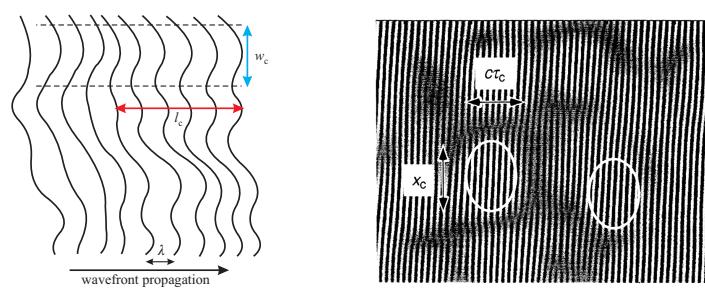


Figure 2.27: A highly schematic representation of a partially coherent wavefield. **Phase registration** is lost over a distance $l_c = c\tau_c - \tau_c$ is the *temporal coherence length* – along the direction of propagation, and over a distance w_c – the *spatial coherence width* – perpendicular to the direction of propagation.

§2.9.4 The “Optical Stethoscope”

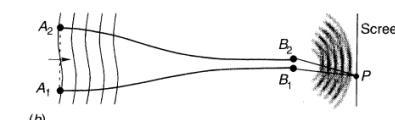
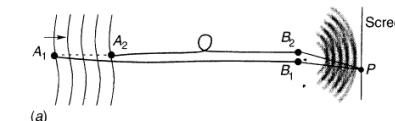


Figure 2.28: The optical stethoscope (after Lipson et al.) is an imaginary device for investigating the time and spatial variation of wavefields and their *temporal* (a) and *spatial coherence* (b). Two identical optical fibres sample the wavefield at points A_1 and A_2 and transfer the amplitudes of the wavefield to the closely spaced points B_1 and B_2 which act as point sources to generate an interference pattern on a screen P . The fibres are lossless and introduce identical phase shifts which can be ignored.

The *time average* for any function $g(t)$ is defined as:

$$\langle g(t) \rangle = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T g(t) dt \quad (2.46)$$

If one brings the ends of the two fibres, B_1 and B_2 , close together into a point the resultant intensity at this point is:

$$I \sim \langle (A_1 + A_2)(A_1 + A_2)^* \rangle \sim |A_1|^2 + |A_2|^2 + \langle A_1 A_2^* \rangle + \langle A_2 A_1^* \rangle$$

where the cross-terms determine any *interference effects*.

Suppose the stethoscope samples the wavefield f at two different points at different times:

$$f_1 \text{ at } \mathbf{r}_1 \text{ at } t \quad f_2 \text{ at } \mathbf{r}_2 \text{ at } t - \tau$$

the (complex) *mutual coherence function* Γ is defined as:

$$\Gamma(\mathbf{r}_1, \mathbf{r}_2, \tau) = \langle f_1(\mathbf{r}_1, t) f_2^*(\mathbf{r}_2, t - \tau) \rangle \quad (2.47)$$

Intensity-normalizing gives the complex *degree of mutual coherence* as:

$$\gamma(\mathbf{r}_1, \mathbf{r}_2, \tau) = \frac{\Gamma(\mathbf{r}_1, \mathbf{r}_2, \tau)}{\sqrt{I_1 I_2}} \quad (2.48)$$

where I_1 and I_2 are the mean intensities at \mathbf{r}_1 and \mathbf{r}_2 :

$$\begin{aligned} I_1 &= \langle f_1(\mathbf{r}_1, t) f_1^*(\mathbf{r}_1, t) \rangle = \Gamma(\mathbf{r}_1, \mathbf{r}_1, 0) \\ I_2 &= \langle f_2(\mathbf{r}_2, t) f_2^*(\mathbf{r}_2, t) \rangle = \Gamma(\mathbf{r}_2, \mathbf{r}_2, 0) \end{aligned} \quad (2.49)$$

But what does $\gamma(\mathbf{r}_1, \mathbf{r}_2, \tau)$ mean *physically*? How is it useful....?

$\gamma(\mathbf{r}_1, \mathbf{r}_2, \tau)$ determines how effectively the disturbances (wavelets) originating from \mathbf{r}_1 and \mathbf{r}_2 can interfere.... If I_1 and I_2 are equal and $|\gamma| \sim 1$, then I can vary down to zero, giving good fringe contrast, as will become clear.

Consider two limiting cases as in Figs. 2.28 (a) and (b):

In Fig. 2.28 (a): the system examines how the wavefield differs *along the direction of propagation*. It compares the wavefield $A_1 = f(t)$ at one time with the wavefield $A_2 = f(t - \tau)$ at some earlier time, at the “same point” on the wavefront: $\tau \neq 0$, $\mathbf{r}_1 = \mathbf{r}_2$.

The configuration is an example of *amplitude division*. i.e. it principally examines the *time dependence* of the field, and therefore its (self rather than mutual):

Temporal (or longitudinal) **Coherence** §2.9.5

In Fig. 2.28 (b): the system examines how the wavefield differs *across the wavefront*. It compares the wavefield A_1 at one point \mathbf{r}_1 in space with the wavefield A_2 at some other point \mathbf{r}_2 , at the same time: $\tau = 0$, $\mathbf{r}_1 \neq \mathbf{r}_2$.

The configuration is an example of *wavefront division*. i.e. it principally examines the *space dependence* of the field, and its:

Spatial (or transverse, or lateral) **Coherence** §2.9.9

§2.9.5 Temporal coherence

A practical arrangement for the “stethoscope” might be:

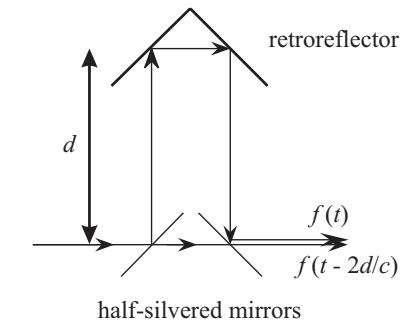


Figure 2.29: A set-up for examining longitudinal coherence.

The *time delay* between the two rays $\tau = 2d/c$ can be altered *spatially* by moving the retro-reflector. This is *amplitude division*, so $\mathbf{r}_1 \equiv \mathbf{r}_2$ and the spatial coordinates will be omitted for convenience in the following. The divided amplitudes are taken to be equal: $A_1 = f(t)$, $A_2 = f(t - \tau)$.

The output intensity depends on the *spatially introduced time interval* τ :

$$\begin{aligned} I(\tau) &= \langle [A_1 + A_2][A_1^* + A_2^*] \rangle = \langle [f(t) + f(t - \tau)][f^*(t) + f^*(t - \tau)] \rangle \\ &= \langle f(t)f^*(t) \rangle + \langle f(t - \tau)f^*(t - \tau) \rangle + \langle f(t)f^*(t - \tau) \rangle + \langle f(t - \tau)f^*(t) \rangle \\ &= 2I_0 + \Gamma(\tau) + \Gamma^*(\tau) \end{aligned}$$

Defining the (complex) *temporal coherence function* for the wavefield:

$$\Gamma(\tau) = \langle f(t)f^*(t - \tau) \rangle \quad (2.50)$$

and where

$$\langle f(t)f^*(t) \rangle = \langle f(t - \tau)f^*(t - \tau) \rangle = I_0 \quad (2.51)$$

is the intensity of each beam measured independently, then:

$$I(\tau) = 2I_0 + 2\Re[\Gamma(\tau)] \quad (2.52)$$

... $I(\tau)$ experimentally quantifies $\Gamma(\tau)$.

Because Γ is complex:

$$\Gamma = |\Gamma(\tau)| e^{i\Delta(\tau)} \quad (2.53)$$

so

$$I(\tau) = 2I_0 + 2\Re[\Gamma(\tau)] = 2I_0 + 2|\Gamma(\tau)|\cos(\Delta(\tau)) \quad (2.54)$$

i.e. $\Gamma(\tau)$ has an *amplitude* and a *phase term* which both depend on the *path difference* (Fig. 2.29) which determines τ

Now IF the waveform is *quasi-monochromatic*, $f(t) \sim e^{-i\omega_0 t} = e^{-ik_0 ct}$:

$$\begin{aligned} \Gamma(\tau) &= \langle f(t)f^*(t-\tau) \rangle \\ &\sim \frac{1}{T} \int_0^T e^{-i\omega_0 t} e^{i\omega_0(t-\tau)} dt \\ &\sim e^{-i\omega_0 \tau} = e^{-i2k_0 d} \end{aligned} \quad (2.55)$$

The phase term *oscillates* with changes in d on the order of the wavelength of the light – *rapidly* compared with the variation in the amplitude $|\Gamma(\tau)|$.

So if $I(\tau)$ is measured as $\tau = 2d/c$ is varied, “*fringes*” are observed.

The *fringe contrast*, or *fringe visibility*, is:

$$V = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} \quad (2.56)$$

$$I_{\max} = 2I_0 + 2|\Gamma(\tau)| \quad : \quad I_{\min} = 2I_0 - 2|\Gamma(\tau)|$$

so...

$$V(\tau) = \frac{|\Gamma(\tau)|}{I_0} = |\gamma(\tau)| \quad (2.57)$$

where $\gamma(\tau)$ is the normalized temporal coherence function – the

Degree of Temporal Coherence

$$\Gamma(\tau) = \langle f(t)f^*(t-\tau) \rangle \text{ so } \Gamma(0) = I_0 \text{ and } V(0) = 1.$$

As τ increases, any element of de-correlation in $f(t)$ will cause $\Gamma(\tau)$, and therefore $V(\tau)$, to decrease.

§2.9.6 Temporal Coherence and the Power Spectrum

Consider an irregular normalized waveform $f(t)$ and its auto-correlation function $h(\tau)$:

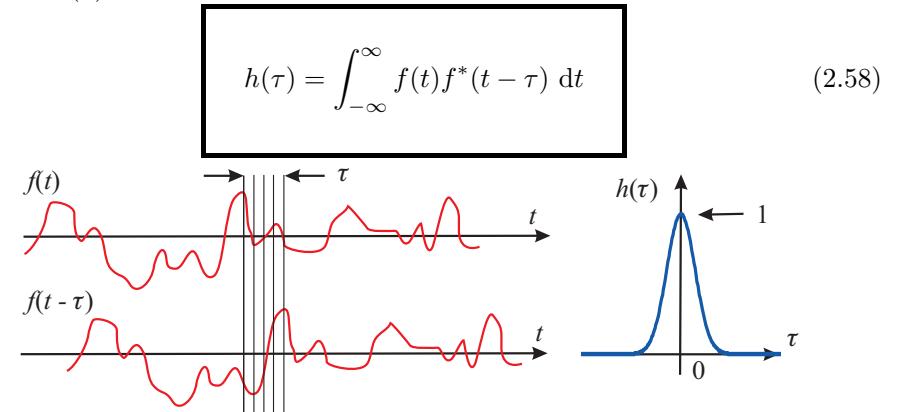


Figure 2.30: $f(t)$ and its auto-correlation function $h(\tau)$ – its self-convolution. $h(\tau)$ is a measure of how f at some time t is *correlated* with f at some other time. $h(\tau)$ must peak for $\tau = 0$.

Consider the FT $H(\omega)$ of $h(\tau)$ for $f(t)$, which itself has FT $F(\omega)$:

$$\begin{aligned} \text{FT}[h(\tau)] &= H(\omega) = \int e^{-i\omega\tau} h(\tau) d\tau \\ &= \iint e^{-i\omega\tau} f(t) f^*(t-\tau) dt d\tau \\ &= \iint e^{-i\omega\tau} f(t+\tau) f^*(t) dt d\tau \\ &= \iint e^{-i\omega(\tau+t)} f(t+\tau) e^{i\omega t} f^*(t) dt d\tau \\ &= \text{FT}[f(t)] \times (\text{FT}[f(t)])^* = F(\omega)F^*(\omega) \end{aligned} \quad (2.59)$$

$$H(\omega) = F(\omega)F^*(\omega) = |F(\omega)|^2 \quad (2.60)$$

.... the **WIENER-KHINCHINE THEOREM**:

“THE FT OF THE AUTO-CORRELATION FUNCTION OF A FUNCTION IS THE SQUARE MODULUS OF THE FT OF THE FUNCTION ITSELF.”

Apart from their normalization, $h(\tau)$ and $\gamma(\tau)$ have the same form, and the Wiener-Khinchine Theorem shows that the fringe visibility $V(\tau)$ for $f(t)$ is directly related to its Power Spectrum $P(\omega)$:

$$P(\omega) \sim \text{FT}[\gamma(\tau)] \Leftrightarrow V(\tau) \quad (2.61)$$

For broadband lightsources, no fringes are observable, just a smoothly varying $I(\tau)$. But nevertheless.... Eqn 2.54:

$$I(\tau) = 2I_0 + 2\Re[\Gamma(\tau)]$$

or, when expressed relative to the intensity ($2I_0$) for zero correlation:

$$I_r(\tau) = 1 + \Re[\gamma(\tau)] \quad (2.62)$$

— the relative intensity measured if a time difference τ is introduced between two beams of equal intensity I_0 obtained by amplitude division.

Now assuming for technical simplicity that $f(t)$ (and therefore $\gamma(\tau)$) is *real*:

$$\gamma(\tau) = I_r(\tau) - 1$$

Therefore:

$$P(\omega) \sim \text{FT}[\gamma(\tau)] = \text{FT}[I_r(\tau) - 1] \quad (2.63)$$

The Power Spectrum of a light beam can be obtained by FT-ing the relative intensity $I_r(\tau)$ observed by introducing a time delay τ between two beams obtained by amplitude division.

This is the basis of a powerful spectroscopic technique – *Fourier Transform Spectroscopy* – based upon ...

§2.9.7 Fourier Transform Spectroscopy - The Michelson (Spectral) Interferometer

Highly schematically:

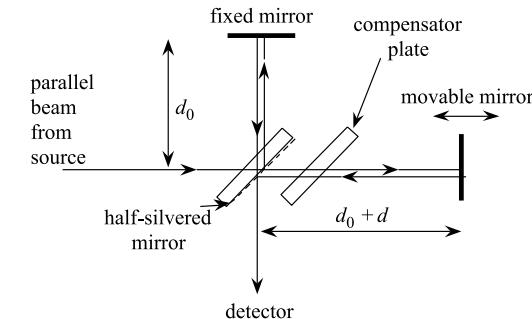


Figure 2.31: Schematic of the Michelson interferometer.

The optical path of one beam can be varied w.r.t. the other introducing a measurable time delay $\tau = 2d/c$.

[The compensator plate, identical to that forming the beamsplitter, ensures that the both optical paths are equivalent when the movable mirror is the same distance d_0 from the beamsplitter as is the fixed mirror. One beam is reflected *internally* at the beamsplitter, and one *externally*, which introduces a phase difference of π which can be ignored.]

Summary of the basic results:

$$\text{Source output} \quad f(t) \quad (2.64)$$

$$\text{Temporal coherence function} \quad \Gamma(\tau) = \langle f(t)f^*(t-\tau) \rangle \quad (2.65)$$

$$\text{Measured intensity} \quad I(\tau) = 2I_0 + 2\Re[\Gamma(\tau)] \quad (2.66)$$

$$\text{Time interval} \quad \tau = 2d/c \quad (2.67)$$

$$\text{Visibility} \quad V = \frac{|\Gamma(\tau)|}{I_0} = |\gamma(\tau)| \quad (2.68)$$

$$\text{Power Spectrum} \quad P(\omega) \sim \text{FT}[\gamma(\tau)] \quad (2.69)$$

Some examples follow....

(i) A broadened spectral line:

Suppose the power spectrum $P(\omega)$ has a Gaussian profile:

$$P(\omega) = C \exp\left(-\frac{(\omega - \omega_0)^2}{2\sigma^2}\right)$$

Working backwards from the Wiener-Khinchine Theorem:

$$\begin{aligned} \Gamma(\tau) &\sim \text{FT}^{-1}[P(\omega)] \\ &\sim \int \exp(+i\omega\tau) \exp\left(-\frac{(\omega - \omega_0)^2}{2\sigma^2}\right) d\omega \\ &= \exp(i\omega_0\tau) \int \exp(iu\tau) \exp\left(-\frac{u^2}{2\sigma^2}\right) du \quad (u = \omega - \omega_0) \\ &= \exp(i\omega_0\tau) \int \exp\left(-\left(\frac{u}{\sigma\sqrt{2}} - \frac{i\sigma\tau}{\sqrt{2}}\right)^2\right) \exp\left(-\frac{\sigma^2\tau^2}{2}\right) du \\ &\sim \exp(i\omega_0\tau) \exp\left(-\frac{\sigma^2\tau^2}{2}\right) \end{aligned} \quad (2.70)$$

$I(\tau) = 2I_0 + 2|\Gamma(\tau)| \cos(\omega_0\tau)$ so:

$$I_r(\tau = 2d/c) = 1 + \exp\left(-\frac{\sigma^2\tau^2}{2}\right) \cos(\omega_0\tau) \quad (2.71)$$

so that the visibility is:

$$V(\tau) = \exp\left(-\frac{\sigma^2\tau^2}{2}\right) \quad (2.72)$$

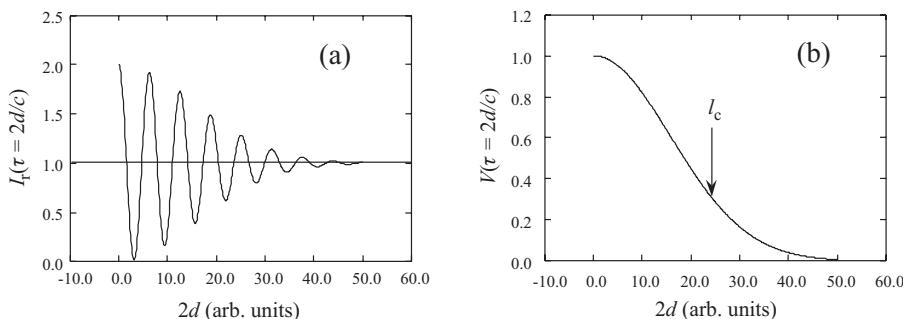


Figure 2.32: An interferogram (a) observed with a Michelson interferometer for a quasi-monochromatic source and the corresponding visibility (b).

The interferogram shows essentially the same fringes as for a truly monochromatic source, but the fringe visibility falls as d increases.

The spacing of the fringes ($I \sim \cos(2d\omega_0/c)$) gives the central frequency ω_0 of the line.

The visibility gradually decreases as the d increases because the two beams are becoming less mutually coherent:

| | | |
|------------------------------|--------------------|-------------|
| $d = 0$ | coherent | $V = 1$ |
| $d \rightarrow \text{large}$ | incoherent | $V = 0$ |
| d intermediate | partially coherent | $0 < V < 1$ |

Define the **coherence length** l_c as the path difference ($= 2d$) at which the visibility falls to about $1/e$.

From Eqn 2.72:

$$V = \exp\left(-\frac{2\sigma^2 d^2}{c^2}\right) \quad (2.73)$$

So:

$$l_c \approx \frac{c\sqrt{2}}{\sigma} \quad (2.74)$$

σ determines the *frequency linewidth* $\delta\omega$ (FWHM, $= 2.36\sigma$) of the spectral line. So l_c , again *directly measurable from the interferogram*, can be used to obtain the linewidth $\delta\omega$.

Note that the preceding discussion has been of *temporal coherence*, and that l_c is really a measure of a **coherence time**

$$\tau_c (= l_c/c \sim \sigma^{-1}) \quad (2.75)$$

expressed as a length. The Michelson interferometer introduces a time shift using the finite velocity of light and a path length.

Consider the Kr⁸⁴ line: $\lambda = 560 \text{ nm}$; $\delta\lambda \sim 0.0002 \text{ nm}$.

With $\delta\omega$ the FWHM of the Gaussian profile $P(\omega)$:

$$2.36\sigma \equiv \delta\omega \sim c\delta k = 2\pi c \delta \left(\frac{1}{\lambda} \right) \sim 2\pi c \frac{\delta\lambda}{\lambda^2}$$

$$l_c \approx \frac{c\sqrt{2}}{\sigma} \sim \left(\frac{2.36}{\pi\sqrt{2}} \right) \frac{\lambda^2}{\delta\lambda} \quad (2.76)$$

$$l_c \sim " \frac{\lambda^2}{2\delta\lambda} "$$

(2.77)

So for Kr⁸⁴: $l_c \sim 0.78 \text{ m}$.

This easy to measure! Very narrow linewidths can be measured with the Michelson since the very short lifetime is effectively being converted to a distance by multiplying by c .

(ii) More complex line structures:

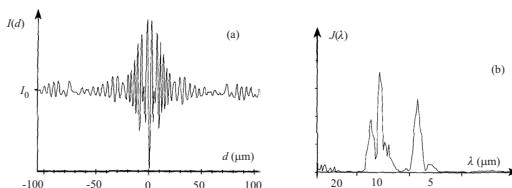


Figure 2.33: (a) An interferogram for a complex source; (b) the corresponding power spectrum $J(\lambda)$.

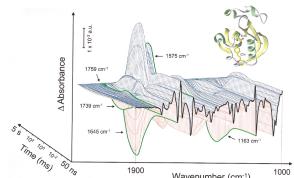


Figure 2.34: FTIR spectrum of photoactive yellow protein as a function of time after photoexcitation (Brudler et al., Nature Structural Biology 8, 265 (2001))

Fourier transform spectroscopy is particularly powerful in the infrared spectral range. FTIR (Fourier transform infrared spectroscopy) has two important advantages over conventional infrared spectrometers that use monochromators / diffraction gratings:

(a) The *radiation throughput is higher* because it is not limited by the input slit width; the size of the detector in an FTIR can be as large as the central interference ring (Jacquinot advantage)

(b) The *signal-to-noise ratio is higher* because as the mirror separation is scanned the signal is continuously falling on the detector. in contrast in a grating spectrometer only background signal is falling on the detector while the monochromator switches between wavelengths (Fellgett advantage).

The spectral resolution of a Fourier transform spectrometer is given by $\frac{\lambda}{\delta\lambda} = 2m$, where m is the number of fringes recorded (limited by the size of the spectrometer).

§2.9.8 Finite coherence length/time: effect on Diffraction

A broadened spectral line source with coherence length $l_c = c\tau_c$ affects the fringes obtained with Young's slits:

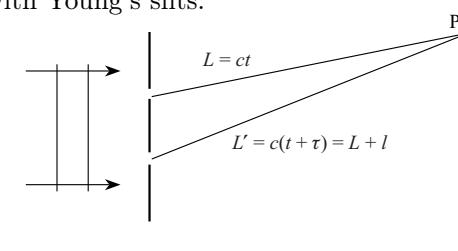


Figure 2.35: Young's slits.

Waves at P from slits illuminated with a parallel beam have path lengths which differ by l . They become increasingly mutually coherent as $l \rightarrow l_c$ and beyond, so the fringe visibility will decrease as P moves off-axis.

This is a *spatial* observation of a *temporal* coherence effect, and is to be contrasted with the following use of Young's slits to illustrate Case (a) of §2.9.4.

§2.9.9 Spatial Coherence

Young's slits produce two waves by *wavefront division* rather than by amplitude division as for Michelson's interferometer, and are useful to introduce the concept of *Spatial Coherence*.

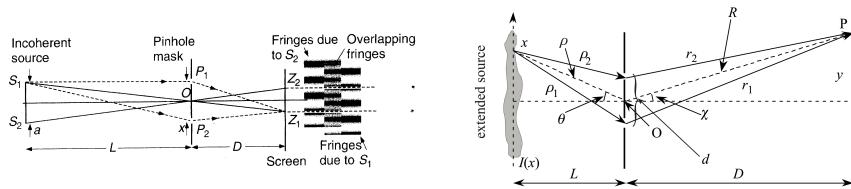


Figure 2.36: Young's slits illuminated not by a parallel beam but a beam of wavelength λ and *finite angular width* from an *extended source* with intensity variation $I(x)$ in its own plane as shown.

Qualitatively: Each point on the source produces a set of \cos^2 fringes angularly offset from each other because of the different angles the points on the source make at O. So the *fringe contrast* on the screen is *degraded*.

If the angular width of the source measured from O is α , then the fringes from the two edges will be offset by αD . The spacing of each set of fringes is $\lambda D/d$, so serious contrast degradation sets in if:

$$d > \frac{\lambda}{\alpha} \approx w_c$$

This degradation will be associated with the *degree of coherence* of the light from each slit, which is determined by their separation d and the angular size of the source.

It relates to coherence variations *across the wavefield* rather than along the rays, i.e. to *lateral*, or *spatial*, coherence rather than longitudinal, or temporal coherence as discussed earlier.

It is described by the *coherence width*, w_c , as opposed to the coherence length l_c .

More quantitatively: for the two rays shown:

$$\begin{aligned} r_1 + \rho_1 &= \rho + R + \frac{d}{2} \sin \theta + \frac{d}{2} \sin \chi \\ r_2 + \rho_2 &= \rho + R - \frac{d}{2} \sin \theta - \frac{d}{2} \sin \chi \end{aligned} \quad (2.78)$$

Now take: $\sin \theta \approx x/L$; $\sin \chi \approx y/D$ so the phase difference for the rays arriving at P (i.e. at y) from the point x is ($k = 2\pi/\lambda$):

$$\begin{aligned} k(r_1 + \rho_1 - r_2 - \rho_2) &= kd(\sin \theta + \sin \chi) \\ &= kd \left(\frac{x}{L} + \frac{y}{D} \right) \\ &= 2ks \end{aligned} \quad (2.79)$$

The amplitude $\psi(y)$ at P arising from the source region $x \rightarrow x + dx$ is then:

$$\begin{aligned} \psi(y) &\sim e^{ik(\rho+R)} \sqrt{I(x)} (e^{iks} + e^{-iks}) \\ &= e^{ik(\rho+R)} \sqrt{I(x)} 2 \cos ks \end{aligned} \quad (2.80)$$

where $I(x)$ is the *intensity profile* of the source.

If the extended source is *incoherent* — i.e. the radiation from each point on it is *uncorrelated* with that from any other point — the nett *intensity at P* as a *function of the slit separation d* is obtained by summing *intensities* $\sum |\psi(y)|^2$, rather than $|\sum \psi(y)|^2$:

$$\begin{aligned} I_y(d) &= \int |\psi(y)|^2 dx = 4 \int I(x) \cos^2 ks dx \\ &= 2 \int I(x)[1 + \cos 2ks] dx \\ &= 2 \int I(x) dx + 2 \int I(x) \cos \left[kd \left(\frac{x}{L} + \frac{y}{D} \right) \right] dx \\ &= 2I_0 + 2\Re \left[e^{-ikdy/D} \int I(x) e^{-ikdx/L} dx \right] \end{aligned} \quad (2.81)$$

Note that the form of this result reflects the earlier form for temporal coherence.

Changing variables: $x \approx L\theta$, $y \approx D\chi$ so $I(x)dx \rightarrow I(\theta)d\theta$ and writing $kd = u$, then the intensity on the screen as a function of y and of the slit separation d is:

$$I_y(u) = 2I_0 + 2I_0 \Re[e^{-iux}\gamma(u)] \quad (2.82)$$

where

$$\gamma(u) = \frac{1}{I_0} \int I(\theta)e^{-iu\theta} d\theta = \frac{\text{FT}[I(\theta)]}{I_0} \quad (2.83)$$

Writing $\gamma(u) = |\gamma(u)|e^{-i\beta}$, then, since $u\chi = kdy/D$:

$$I_y(u) = 2I_0 + 2I_0|\gamma(u)| \cos\left(\beta + \frac{kdy}{D}\right) \quad (2.84)$$

— “cos²”-fringes with spacing determined by kdy/D , as expected for two slits.

The offset, β , is determined by the source angular profile $I(\theta)$.

Symmetric source profiles:

When $I(x)$ is symmetric about the axis of the system, $\gamma(u)$ is *real* and

$$\beta = \begin{cases} 0 & \text{if } \gamma(u) > 0, \\ \pi & \text{if } \gamma(u) < 0. \end{cases}$$

The *visibility* of the fringes is then:

$$V = \gamma(u = kd) \quad (2.85)$$

and can be negative in the sense that a minimum occurs where otherwise a maximum might be expected, such as at $y = 0$ (see examples below).

Note: the slits are effectively sources giving rise to fringes on the screen; the visibility of the fringes depends on the degree of coherence of these two sources; varies with their separation d across the wavefront – an obvious *optical stethoscope*.

So, following Eqn 2.57 the

Degree of Lateral Coherence

in the plane of the slits is defined as (repeating Eqn 2.83):

$$\gamma(u) = \frac{1}{I_0} \int I(\theta)e^{-iu\theta} d\theta = \frac{\text{FT}[I(\theta)]}{I_0} \quad (2.86)$$

.... the VAN CITTERT-ZERNIKE THEOREM:

“THE DEGREE OF LATERAL COHERENCE (see below) IS THE FT OF THE ANGULAR INTENSITY DISTRIBUTION OF THE SOURCE, $I(\theta)$.”

$V = \gamma(u = kd)$ is determined by the FT of the source angular profile $I(\theta)$ and by the slit separation d , and can be used to define the *coherence width* w_c of the light in the plane of the slits. Examples follow

Example 1: suppose the slits are illuminated by a distant, uniform line source, (parallel to the slits) of *angular width* α :

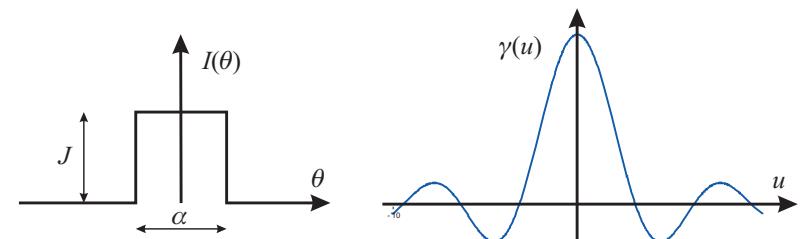


Figure 2.37: Angular profile of a line source.

$$\gamma(u) = \frac{1}{I_0} \int_{-\alpha/2}^{\alpha/2} J e^{-iu\theta} d\theta = \text{sinc}\left(\frac{u\alpha}{2}\right) \quad (2.87)$$

since $I_0 = \alpha J$. The observed fringe contrast depends on $u\alpha/2$ and falls to zero at

$$\frac{u\alpha}{2} = \frac{kda}{2} = \frac{\pi da}{\lambda} = \pi$$

i.e. at

$$w_c \sim d = \frac{\lambda}{\alpha} \quad (2.88)$$

beyond which V becomes *negative*, see below.

Example 2: a *disc source* of angular diameter α illuminating two *pinholes* separated by d , gives:

$$\gamma(u) = \frac{2J_1(u\alpha/2)}{(u\alpha/2)} \quad (2.89)$$

Very similar in form to the 1D sinc case discussed above.

For this circular symmetry, the fringe *visibility falls to zero* for $u\alpha/2 = 3.83$ (i.e. at $kda/2 = \pi da/\lambda = 3.83$), and the *coherence width* is then generally quoted as:

$$w_c = \frac{1.22\lambda}{\alpha} \sim \frac{\lambda}{\alpha} \quad (2.90)$$

In general, the *broader the source, the narrower the coherence width*.

The following Figures (Hecht) show how the fringes obtained for two pinhole apertures (hence the envelope profile) depend on their separation d for fixed source width. As d is increased, the fringe spacing falls, as does the fringe contrast, as the two pinholes become less mutually coherent. Because of the “sinc” form of the $\gamma(u)$ function (Eqn. 2.89), V in fact goes through zero and then rises slightly.

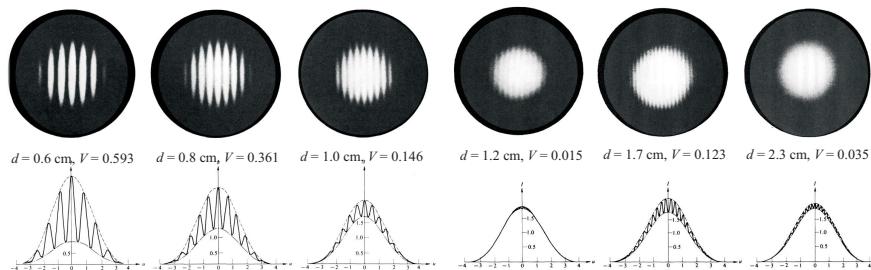


Figure 2.38: Variation of fringe visibility for two pinhole apertures as a function of their separation d .

Note that for $d = 1.7$ cm, $\gamma(u) < 0$, so there is an *intensity minimum* on-axis.

Or if d is fixed and α is varied:

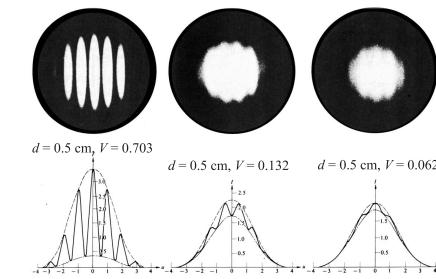


Figure 2.39: Variation of fringe visibility for two pinhole apertures at fixed separation d but increasing source width.

Again, note the intensity *minimum* on-axis for the middle case.

Coherence area/volume:

Clearly, in 2D there is a *coherence area*.

Temporal coherence effects have been ignored throughout. But generally both are present, and from the coherence area and length arises the idea of a *coherence volume*.

For sunlight ($\lambda \sim 500$ nm, $\delta\lambda \sim 500$ nm, $\alpha \sim 0.5^\circ \sim 0.01$ rad). So:

$$w_c \sim \frac{\lambda}{\alpha} \sim 5 \times 10^{-5} \text{ m} \quad : \quad \text{from Eqn 2.77} \quad l_c \sim \frac{\lambda^2}{\delta\lambda} \sim 500 \text{ nm}$$

These are much smaller than everyday objects and apertures, so it is not surprising that diffraction effects are not an everyday sight.

(Indeed it is worth thinking about how diffraction was effectively demonstrated 200 years ago, without lasers or discharge lamps.....)

For distant stars, $\alpha \ll 0.01$ radians, so their coherence widths at the Earth are much greater and can be exploited to measure their angular diameters...

§2.9.10 Michelson's Stellar Interferometer - Aperture synthesis

α is very small for stars, so w_c is several metres for visible wavelengths. So using a Young's slit set-up to observe the reduction in fringe contrast as d is increased is impossible, as fringes with very small spacing would result.
But...

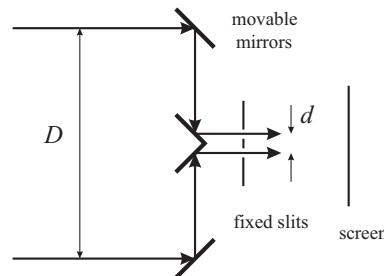


Figure 2.40: Michelson's Stellar Interferometer – the optical stethoscope at work.

d is fixed and small, producing well-spaced \cos^2 fringes. Light is gathered using mirrors with variable large spacing D . The fringe visibility as a function of D allows coherence widths on the scale of metres to be measured.

Michelson used this to look at Betelgeuse (α -Orionis), a red supergiant star in the constellation Orion, the ninth brightest in the night sky. For $\lambda = 570$ nm, the fringes vanished at $D = 3.07$ m, corresponding then to $\alpha = 22.6 \times 10^{-8}$ rad (0.047 sec of arc). Using the distance from the Earth determined independently by parallax measurements, this can be used to determine the diameter of the star, which is $950\text{--}1200 R_\odot$, the radius of the sun.

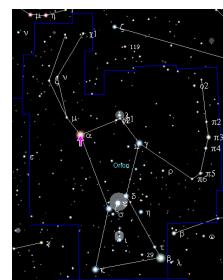


Figure 2.41: The constellation of Orion. Betelgeuse is marked with an arrow.

To create an image of the star one needs to know not the magnitude of the coherence function $|\gamma|$ but also its phase β . The difficulty in determining β is that it is affected also by atmospheric jitter that can cause shifts of the interference pattern with respect to the optical axis of the telescope. To overcome this more than two entrance apertures, at least three apertures A_1 , A_2 and A_3 can be used. If the apertures are selected such that the pairs A_1A_2 , A_2A_3 and A_1A_3 produce interference fringes with different periods that can be measured simultaneously, the atmospheric phase shifts can be eliminated and the real phase of γ can be determined (method of phase closure). This allows reconstructing an image of the star through Fourier transformation.



Figure 2.42: Cambridge Optical Aperture Synthesis Telescope (COAST)

Reconstruction of the phase of the coherence function using the phase closure/aperture masking method has allowed imaging of Betelgeuse, which has revealed hotspots in its atmosphere that can be resolved at certain wavelengths.

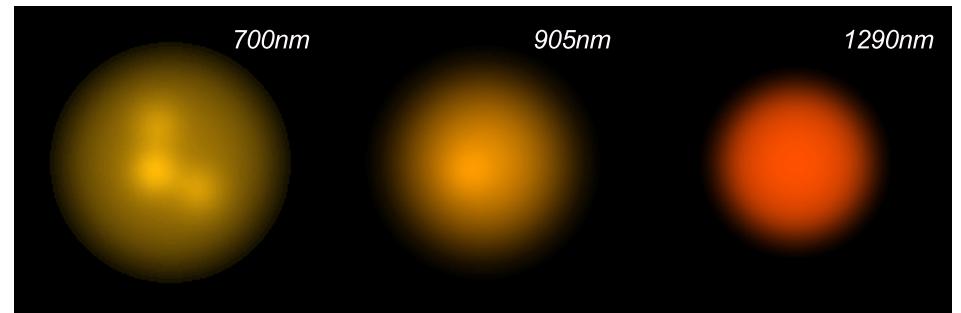


Figure 2.43: Images of Betelgeuse taken at different wavelengths by the Cambridge Optical Aperture Synthesis Telescope (COAST) (see Baldwin, Haniff, MNRAS 315, 635 (2000))