# OPTICS COURSE

**Aims and Objectives:** This is a foundation Optics course. It aims to bring together Optics into a single coherent framework covering key optical concepts, optical elements, instruments, and light-based applications (e.g. imaging, interferometry, fibre optics, lasers).

The course builds on maths, physics, and Fourier Transform theory of waves taught in other parts of "Oscillations and Waves" module. It supports other activities: experimental laboratory and projects, and future research work; and underpinning concepts/techniques for some future course modules.

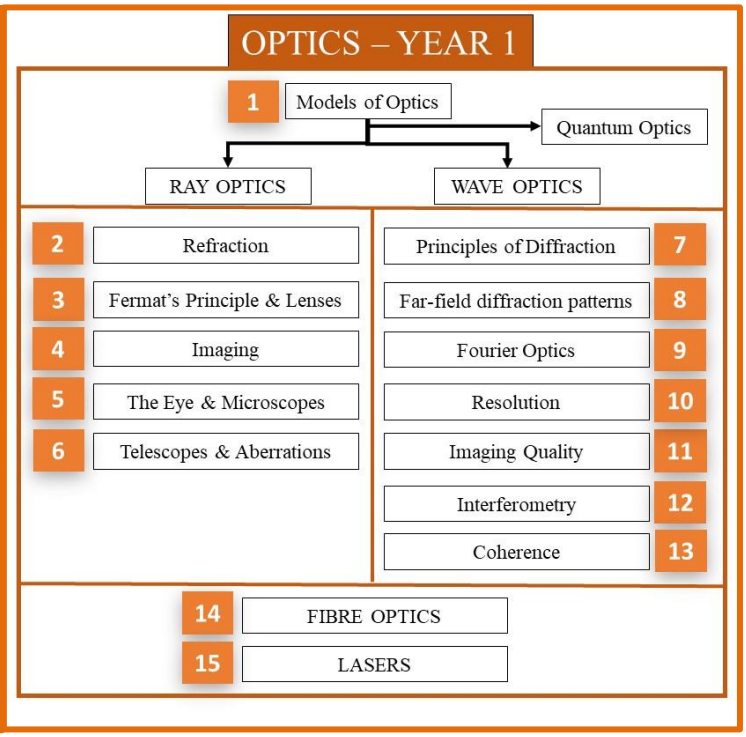
**What will this course teach you?**

- **Models for optics** (*ray*, *wave*, *quantum*): know when and how to use them.
- **Optics Laws and Principles** underpinning *reflection*, *refraction*, *interference*, *diffraction*.
- **Optical elements:** understand their form and function (e.g. *lenses*, *prisms*, *mirrors*, *gratings*).
- **Optical imaging instruments:** apply ray optics of lenses (and mirrors) for design of *eye; camera; magnifying lens; compound microscope; telescope*.
- **Wave theory of imaging** applied to understand *optical resolution*, *spatial frequencies*; *optical transfer function and image quality.*
- **Interferometry:** know types of *interferometers* and their applications; understand and apply optical coherence: *fringe visibility*; *spatial coherence; temporal coherence.*
- **Fibre optics:** understand in ray & wave model and use for light guiding and communications.
- **Lasers:** know fundamentals and some key properties.

**Delivery Format:** The course is delivered in 15 Lectures: each lecture generally consisting of 3 short video clips + some supplementary materials (demo videos + questions + quizzes).

**Professor Mike Damzen**
m.damzen@imperial.ac.uk

**I hope you enjoy the course!**




**Recommended Reading**

*Optics* by E. Hecht (*a good illustrated text to unlock further insight and practice of topics*)

Hyperphysics website http://hyperphysics.phy-astr.gsu.edu/ has online quick reference guidance of some basic ray and wave optics. Worth a quick look!

# OPTICS Year 1

## SYLLABUS (15 lectures)

### NATURE OF LIGHT and MODELS OF OPTICS (1 lecture)

**1** Brief history of our understanding of light. Models of optics: Ray optics (rays and wavefronts); wave optics (scalar waves and EM waves); quantum optics (photons)

### RAY (Geometrical) OPTICS (5 lectures)

**2 Refraction and Reflection**: Reflection Law; Snell's Law; critical angle; total internal reflection. Deviation, displacement, glass dispersion $n(\lambda)$ and deviation angle $\delta(\lambda)$ (+examples: prism; rainbow).

**3 Fermat's Principle & Lens Design**: principle of stationary ray paths; examples: derivation of Snell's Law; lens shape for imaging; Spherical lenses, Lens-Maker's Formula.

**4 Imaging**: Lens imaging: Ray diagrams; Thin lens formula; transverse magnification; real and virtual images; equivalence of a curved mirror to a lens; two-lens imaging; optical power of a lens.

**5 Optical Instruments 1:** The (human) Eye; near and far-points; angular size; magnifying glass; compound microscope.

**6 Optical Instruments 2**: (Astronomical) Telescopes; lens and mirror aberrations: spherical, coma, and chromatic aberration; methods to eliminate/minimise aberration; real-world lens imaging systems

### WAVE OPTICS (7 lectures)

**7 Principles of Diffraction**: Plane & spherical waves; fields and intensity; Principle of Superposition (Interference); Huygens-Fresnel Principle; near-field & far-field diffraction; aperture transmission function.

**8 Far-field Diffraction patterns:** diffraction from two-slits, single extended slit, "cosine" aperture.

**9 Fourier Optics**: far-field diffraction as a Fourier transform of aperture A(x); lens as a Fourier Transforming element. Double extended slit; diffraction grating.

**10 Optical Resolution:** 2-D rectangular aperture; circular aperture and Airy Pattern; point spread function; Rayleigh resolution criterion; resolution of eye and telescopes; atmospheric turbulence and adaptive optics.

**11 Imaging Quality**: Image as convolution of object and point spread function; image spatial frequency content; optical transfer function; resolvable feature size limit; Abbé's sine rule; imaging as double Fourier-Transform process.

**12 Interferometry:** Examples of Interferometers (Michelson; Mach-Zehnder; Fabry-Perot); thin-film interference and anti-reflection coatings.

**13 Coherence**: Fringe visibility; spatial coherence; temporal coherence; bandwidth theorem; optical coherence tomography (OCT).

### MODERN APPLICATION OF OPTICS (2 lectures)

**14 Fibre Optics**: acceptance angle for highly multimode fibres; wave concept of fibre modes; optical communications: fibre attenuation, dispersion, data rate limits.

**15 Lasers**: basic laser physics and concepts: stimulated emission; population inversion; laser threshold; laser spectral and spatial modes and Gaussian beams.

# Our Understanding & Models of Light

## 1. Brief history (from 17th Century) of our development & understanding of light (also watch video "**What is Light?**" with additional material and history timeline)

**Newton** developed a *particle (corpuscular) model* of light travelling in straight lines (e.g. beam of sunlight)

**Huygens** developed a *wave model* for light, with a geometrical construct using secondary waves (Huygens Principle of Secondary Waves) to predict light propagation, including bending of light through glass.

**Young** performed two-slit experiment showing light *interference* and providing clear experimental evidence that light has a *wave nature*.

**Fresnel** performed analytical theory incorporating Huygens secondary waves and account of amplitude and phase to make accurate predictions of diffraction. We will use the Huygens-Fresnel Principle (HFP) of secondary waves for diffraction in this course.

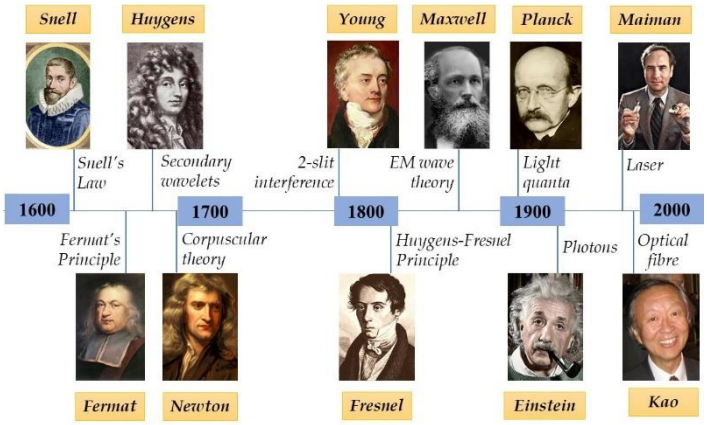
**Maxwell** completed the last term in the equations of electricity and magnetism and showed that *light can be fully describe as EM waves*: with speed of light $c = \sqrt{\varepsilon_0 \mu_0}$ predicted by known EM constants; and E and B fields are transverse oscillations at right angles to each other and the direction of propagation k. This was a tour de force in physics, unifying light with electricity and magnetism and the basis for a full *electromagnetic theory of light*.

[Speed of light, **c = 299,792,458 m/s** is now fixed by international agreement to define SI metre unit, but we will use **c = 3 $10^8$ m/s**, throughout this course.]

**Evidence for a photon model. Planck**'s explanation of blackbody radiation required EM wave energy to be quantised in units $E = h\upsilon$. **Einstein** developed a *photon* (particle) model for photoelectric effect with photon energy $E = h\upsilon$. **Compton** scattering showed (X-ray) photon-electron collision, as if photon was a particle with $E = h\upsilon$ and momentum $\underline{\boldsymbol{p}} = \hbar \underline{\boldsymbol{k}}$.

**Quantum Mechanics. Bohr** model of atom with quantised light transitions (absorption and emission) between discrete energy states $E_2 - E_1 = h\upsilon$. Electrons observed to diffract from crystals demonstrates *wave-like* behaviour of matter. **Schroedinger** developed equation describing particle as a *wavefunction* with a probability distribution of its location. **Heisenberg's** *Uncertainty Principle* describes fundamental limit to our simultaneous knowledge of particle (e.g. its position & momentum $\Delta x \Delta p_x \geq \hbar/2$).

**Modern Optics**: **Lasers** (1960); **Nonlinear Optics** (1961); **Quantum Optics** (quantum entanglement; qubits; secure quantum communication) provide enabling new technology but also further thinking and theory about the nature of light and the models we use. The application of light technology is increasingly enabling our modern society!

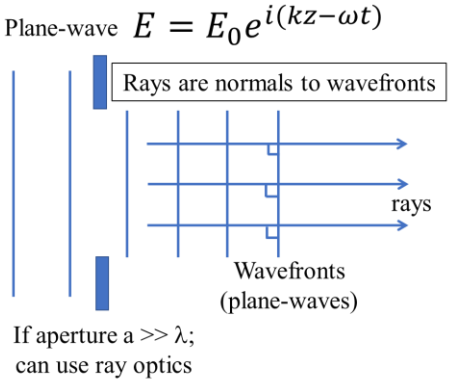


## 2. Models of Optics

All branches of physics use models with different levels of approximation. Model chosen is often the most convenient to achieve the result, prediction, or insight we need in each situation; and sometimes it is useful to use more than one model to analyse a situation. We overview here 3 models of light: Ray optics (rays and wavefronts); wave optics (scalar and EM waves); quantum optics (photons and quantum mechanics).

**Ray Optics** – light description in terms of rays travelling in straight lines & ray paths transformed by refraction and reflection (e.g. ray tracing in imaging systems).

- Ray Model useful when light relatively unrestricted with no small apertures.
- Light rays follow straight paths in the same medium. They are normals to wavefront (e.g. parallel rays normal to plane wavefront; radial rays normal to spherical wavefront of a point source).
- Wavefronts are contours of equal optical path length – rather than formally being phase fronts (i.e. we don't need to know or consider actual wavelength);



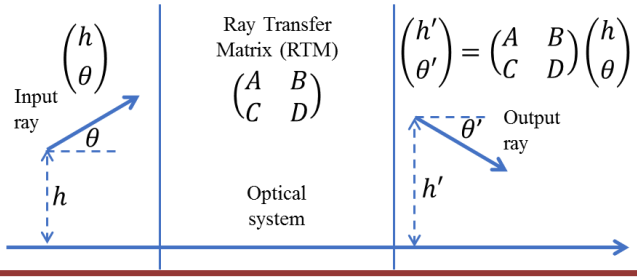

- Ray Optics considers "virtual" rays – ray projection of where the light appears to come from even though they don't exist but can locate "virtual" images.
- **No diffraction** - obstacles may block ray paths, but no creation of new diffracted ray paths is considered.
- **No interference** – rays may overlap don't carry amplitude or phase information.
- Whilst we don't need to know wavelength, ray optics can consider separate ray paths for different wavelengths (colours) due to their different refraction (e.g. due to wavelength dependence of refractive index $n(\lambda)$) e.g. prism/rainbow; chromatic lens aberration.

Ray optics can operate under two levels of approximation:

- **Geometrical Optics** (all angles): Snell's Law; Fermat's Principle; ray-tracing.
- **Paraxial Ray Optics** (small angles $\theta = sin\theta = tan\theta$): can get analytical solutions in this approximation e.g. thin lens formula

**Ray Tracing with computers**:

Describe rays as (2x1) vectors; optical elements as (2x2) ray transfer matrices (RTMs); use computers to map ray paths through optical system using professional software packages (or you can write your own Python computer coding).

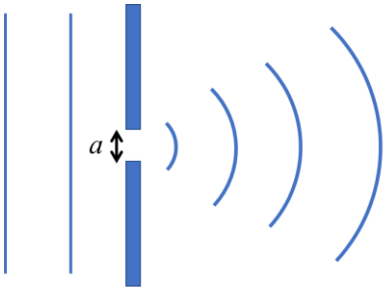


**Wave Optics** – describes interference and diffraction; and in fuller EM theory also describes polarisation and light-matter interaction.

Wave model is necessary to describe diffraction when light encounters small apertures and does not follow straight lines.

Wave model describes interference of light with amplitude and phase information. When two or more light "rays" cross their light fields add (Principle of Superposition) and their summation depends on their *amplitudes* and *phases*.

We need to know wavelength of light to quantify (calculate) interference and diffraction patterns.




If aperture $a \sim \lambda$ ;
must use wave optics

e.g. plane wave $E(z,t) = E_0 e^{i(kz-\omega t)}$

has amplitude $E_0$ and phase $\phi = (kz - \omega t)$.

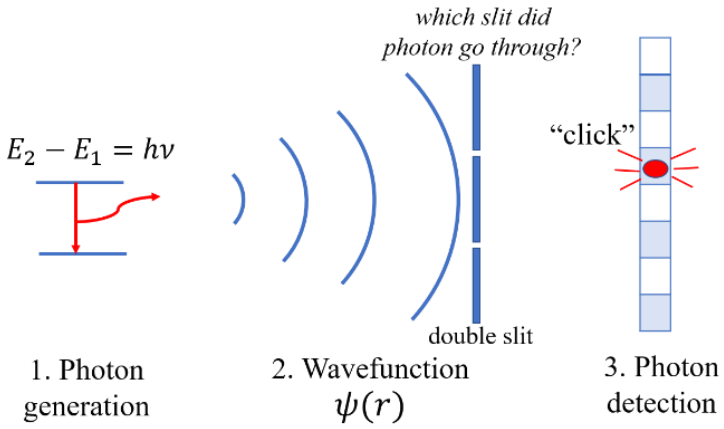
Wavefronts are actual depictions of contours of constant phase (e.g. wave crests); optical path length (OPL=nL) where n is refractive index determines phase change $\Delta\phi = \left(\frac{2\pi}{\lambda}\right)(OPL)$.

Wave optics can be considered under various levels of approximation:

- **Scalar Wave Optics**: basic wave description where detailed knowledge of EM theory is not necessary
    - Suitable for near & far-field diffraction and interference of light with Huygens-Fresnel Principle (of secondary wavelets)
    - It is a useful description in free-space and when simple approximation made about light-matter interaction (e.g. refractive index).
- **Vector Wave Optics**: full rigorous use of Maxwell's equations & EM theory for light-material interaction.
    - Incorporates polarisation state of light **E** (light is a vector)
    - Uses individual Maxwell Equations to get continuity conditions of **E** and **B** fields at boundaries of materials
    - Predicts reflection and transmission of light at refractive boundaries
    - Full light interaction in dielectrics, metals, plasmas even to sub-wavelength scales
    - Describes full mode theory in optical fibres

**Quantum Optics** – photon model and approaches using quantum mechanical description of light (e.g. wavefunctions; states; operators).

Quantum model is necessary to describe a single photon and correlated photons (e.g. "entangled" photon pair produced by using nonlinear optics).




- Single photon generation and detection has a particle nature.
- Quantised light generation results from a matter transition between two quantised energy states $E_2 - E_1 = h\nu$ (e.g. Bohr model of atom).
- (unobserved) propagation of "photon" (e.g. through slits) displays a *wave nature*. In Quantum Mechanics light (and particles) can be described by *wavefunctions* representing probability of location and other properties, allowing light to follow many (or all possible) paths leading to interference.
- Detection of light is a light-matter interaction that can be considered as the annihilation of a photon at a localised detector position – i.e. displaying particle nature. Observation can be said to "cause" the collapse of the wavefunction (that has all possible states) to a single observed state – the click or flash in a single detector element (pixel).
- The term photon should be treated with some care and considered a shorthand for a much more complex picture of light. You may describe light as a stream of photon "particles" travelling through space between source and detector - but it also leads to paradoxes (e.g. which slit did the photon go through?) showing just a simplistic particle picture may not be helpful.
- The description of light in both a wave and particle model almost at the same time has been sometimes termed as the wave-particle duality of light.

## Use of Models in this Course

- ***What model to use for light*?**

In this course, we will generally use the simplest model necessary to explain or suitably quantify the situation:

e.g. ray model for simple refraction, reflection

e.g. ray model for lens imaging and then we revisit when needed with a more sophisticated wave model to incorporate diffraction to analyse image resolution

e.g. optical fibres are described in both a ray and wave description: the ray picture is fine for large diameter (highly multimode) fibres; but wave and EM theory must be used for small diameter (e.g. single mode) fibres.

e.g. lasers are sometimes described as quantum devices, but we can mostly explain their operation in a basic wave and EM theory description.

- ***Do we need Quantum Model for this course*?**

Quantum model is needed when we consider single or small number of "photons". So, let's examine photon rates in normal light levels:

Energy of a visible photon (wavelength 500nm): $E = h\upsilon = hc/\lambda \sim 4.10^{-19} J$.

In a 1W (=1J/sec) light source there are ~ 2.5 $10^{18}$ photons/sec.

There are so many photons we can ignore quantum model at normal light levels. However, in this course we will make relevant quantum links and photon picture analogies where opportunities arise, e.g. describing laser physics with spontaneous and stimulated emission of radiation.

- ***What about using electromagnetic (EM) theory in this course*?**

EM theory will not be covered until Year 2, so, unfortunately, you do not have the background theory necessary to do rigorous EM, nor is there time to cover this level of detail in this short course. However, we have time to point to more complex EM theory effects where necessary e.g. evanescent waves at boundary between two dielectric materials where total internal reflection occurs, as in optical fibres, and the origin of refractive index.

# OPTICS: PART 1 - Ray Optics

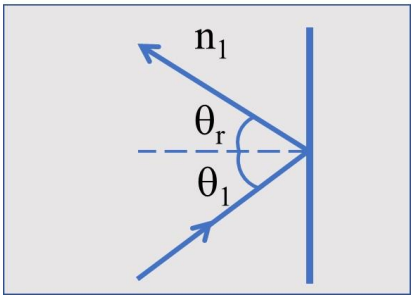
## REFLECTION and REFRACTION LAWS

These are the basic laws of how **rays** are transformed at reflecting surfaces or boundaries of different refractive index. They originate from experimental observations from hundreds of years ago. They can now be derived from first principles, but we will just state these here.

### Reflection Law:

**Angle of reflection = angle of incidence**

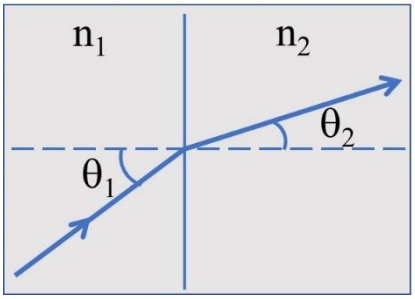
| | |
|---|---|
| i) | $\boldsymbol{\theta_r = \theta_1}$ (angles wrt surface normal) |
| ii) | Surface normal, incident and reflected rays all lie in the same plane |




### Refraction Law:

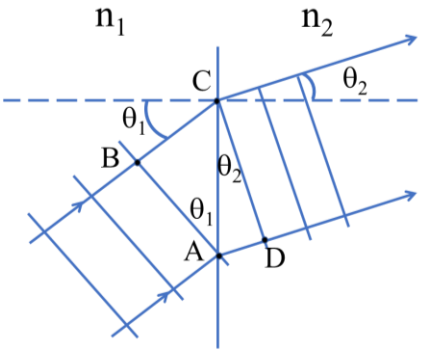
This is **SNELL'S LAW** applying to angles of light at boundary between two media of different refractive index $(n_1; n_2)$:

| | |
|---|---|
| i) | $\boldsymbol{n_1 sin\theta_1 = n_2 sin\theta_2}$ (angles wrt surface normal) |
| ii) | Surface normal, incident and refracted rays all lie in the same plane |




**Derivation of Snell's Law** using the known speed of light in medium: $v = c/n$.

Consider points ABCD. Noting rays are normal to wavefronts, and time for wavefront crest to move distance BC = time to move distance AD:




$$t_{BC} = \frac{l_{BC}}{c\ /\ n_1} = t_{AD} = \frac{l_{AD}}{c\ /\ n_2}$$

$$n_1 l_{BC} = n_2 l_{AD}$$

[*We will see in Wave Optics that this condition of equal optical path length $OPL = nl$, corresponds to equal phase change, and how we move from input wavefront to the refracted wavefront*]. Then using geometry:

$$l_{BC} = l_{AC} sin\theta_1;\ l_{AD} = l_{AC} sin\theta_2$$

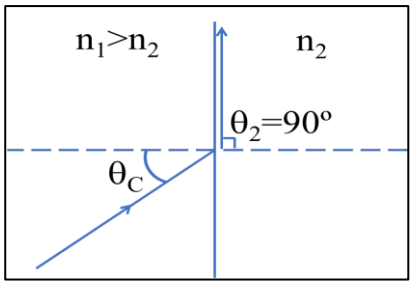
$$n_1 l_{AC} sin\theta_1 = n_2 l_{AC} sin\theta_2$$

$$\boldsymbol{n_1 sin\theta_1 = n_2 sin\theta_2}$$

We have derived Snell's Law. A similar derivation can be made for the reflection law.

## Further Consequences of Snell's Law

### Critical Angle $(\theta_c)$



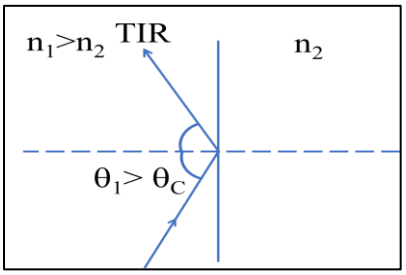

Consider $n_1 > n_2$ for rays approaching from medium 1.

Critical angle $\theta_1 = \theta_c$ occurs when $\theta_2 = 90°$ $(sin\theta_2 = 1)$

$$n_1 sin\theta_1 = n_2$$

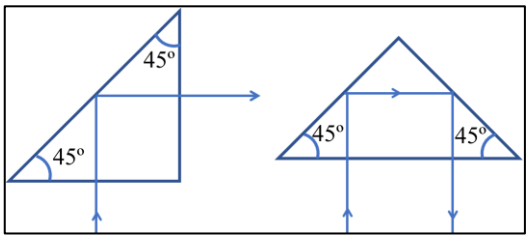
Formula for critical angle: $\boldsymbol{\theta_c = sin^{-1}\left(n_2/n_1\right)}$

### Total Internal Reflection (TIR)




At angles $\theta_1 > \theta_C$ there is no solution to Snell's Law as $sin\theta_2 > 1$ gives no permitted refracted ray angle. Light experiences **total internal reflection** (TIR) with 100% efficiency (for a perfect surface).

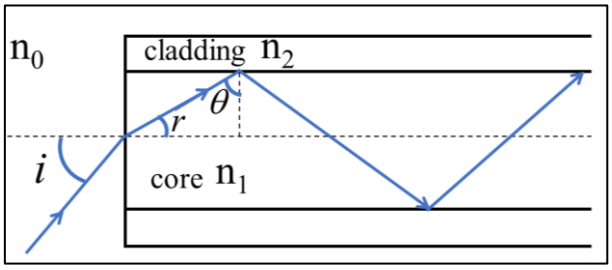
**Prism TIR** Consider glass ($n_g$=1.5) to air ($n_a$=1) interface: $\theta_c = sin^{-1}\left(1/n\right) = 42°$.




We can use 45° glass prisms for TIR and beam-steering since 45° angle of incidence at side faces exceeds critical angle $\theta_c = 42°$. Prisms using TIR are used in modern binoculars.

**Optical Fibre** ($n_{core} > n_{cladding}$).

TIR at interface between inner core $n_1$ and outer cladding $n_2$ in a flexible glass optical fibre provides "lossless" guiding of light for a range of input angle *i*, (the *acceptance angle* of the fibre).



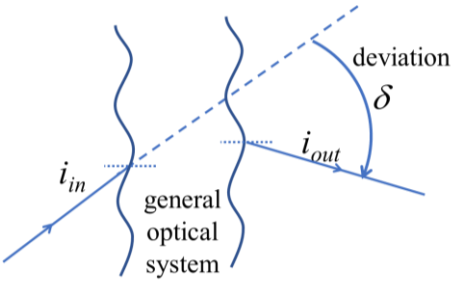

Optical fibres for light guiding and optical communications are the most important commercial application of TIR. Fibre optical communications underpins modern-day ultra-high data rate transport across global distances. Without optics, the Internet as we know it would not exist.

*We will cover Fibre Optics later* (Lecture 14) *in both a Ray and Wave Model picture*.

# DEVIATION, DISPLACEMENT and DISPERSION

## DEVIATION

In optical systems (e.g. lenses or mirrors) we are often not so much concerned with ray refraction/reflection at a surface but the net transformation of the ray. We can define the **angle of deviation** or just **deviation** (**δ**) as the difference between the input and output ray angles:

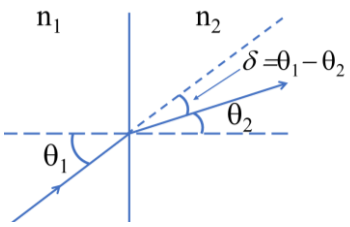



$$\boldsymbol{\delta = i_{in} - i_{out}}$$

**Simple examples of deviation include:**

**Single refracting interface**:

At refractive index boundary, the deviation $\delta = \theta_1 - \theta_2$ (for magnitude of deviation).




The deviation is a complicated function of input angle even in this simple case due to the sine transformation of Snell's Law. For example, taking the case: $n_1 = 1;\ n_2 = n$, then Snell's Law gives $sin\theta_2 = (sin\theta_1)/n$ and deviation:

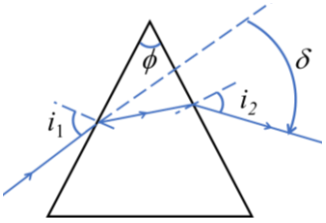
$$\boldsymbol{\delta = \theta_1 - sin^{-1}[(sin\theta_1)/n]}$$

Deviation is much simpler for the small angle (paraxial) case where Snell's Law become linear to input angle: $n_1\theta_1 = n_2\theta_2$ and deviation is also linear to input angle:

$$\boldsymbol{\delta = \theta_1 - \theta_1/n = \theta_1(1 - 1/n)}$$

**Prism**: is an optical refracting element with two non-parallel sides.

We can calculate deviation δ for prism for input angle $i_1$ by performing Snell's Law twice at the two interfaces. The total deviation δ is the sum of the two surface deviations:




$$\boldsymbol{\delta = \delta_1 + \delta_2}$$

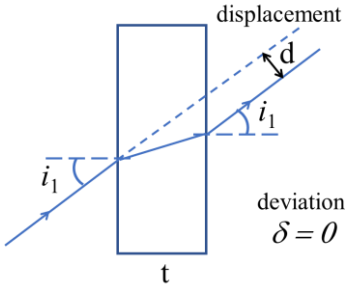
The deviation is in general a complex function of input angle. For small prism apex angle $\phi$, deviation is much simpler (Optics Problem Sheet 1, Qu. 3a) and independent of input angle:

$$\boldsymbol{\delta = (n - 1)\phi}$$

## DISPLACEMENT

**Parallel Plate**: For the case of a parallel plate with thickness t, the two surfaces of refraction produce *equal* and *opposite deviations*. This results in net zero deviation $\delta = 0$.

There is, however, a shift in the ray path called **displacement (d)**, which is the normal distance between the projection of input ray and the output ray. The displacement takes on a simple form for the paraxial (small angle) case (Optics Problem Sheet 1, Qu. 3b):



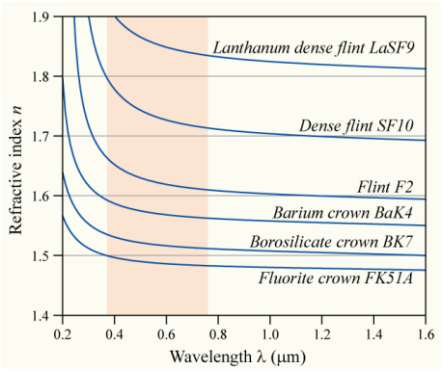

$$\boldsymbol{\delta = t(1 - 1/n)\theta_1}$$

## DISPERSION

All optical materials (e.g. glass) have refractive index that varies with wavelength $\boldsymbol{n \equiv n(\lambda)}$. A consequence of this is that the deviation also depends on wavelength

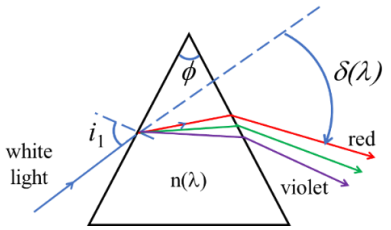
$$\boldsymbol{\delta = \delta(\lambda)}$$

This phenomenon is known as **dispersion**.




Examples of dispersion include:

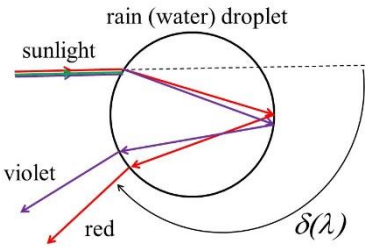
**Prism dispersion.** In glass, long wavelengths (red) have lower refractive index than short wavelengths (blue). The deviation is higher for the blue compared to red and different colours (wavelengths) are spread into difference output angles.




**Rainbow** is dispersion from a combination of refraction at input and output surfaces and an intermediate internal reflection* at back surface in water (rain) droplet. In rainbow spectrum, the deviation $\boldsymbol{\delta(\lambda)}$ is greater than 90° and it is common to use the complementary angle:

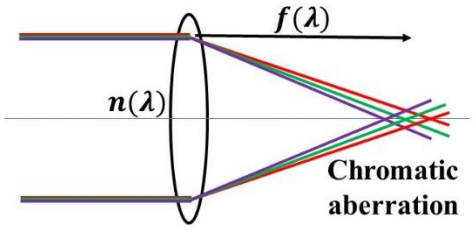
$$\boldsymbol{\theta(\lambda) = \pi - \delta(\lambda)}$$

[**Internal reflection is a weak surface reflection, not TIR*]




**Angular dispersion in (glass) lenses** $\boldsymbol{\delta(\lambda)}$ where blue light is deviated more than red light leads to wavelength dependent focal length $\boldsymbol{f(\lambda)}$.

This is an "error" in imaging leading to "blurring" of a colour image due to the wavelength dependent spread in image point position. In optical theory, imaging error is called aberration. This colour dispersion error is called **chromatic aberration**.



# Fermat's Principle & Lens Design

**Fermat's Principle** is a unifying basis for ray optics. Remarkably, it can derive key laws of ray optics (Snell's Law; reflection law), and it forms the basis for design of imaging elements (lenses; mirrors) and underlies principles for the design of optical imaging instruments.

The original version (1662) sometimes known as Fermat's "least-time" Principle states: *the path taken by a ray between two points is the one traversed in the* ***least time***.

A fuller modern version of the principle is the following:

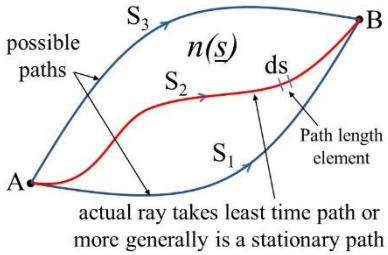
**Fermat's Principle**: *the path taken by a ray between two points* ***is stationary**** *with respect to small variations in the path.*

(**Least time is a special, but the most common, case*).

To understand this, consider a generalised medium with spatially dependent refractive index $n(s)$, where $s$ is a position coordinate. There are an infinite number of possible ray paths $S_n$ between 2 points A and B. Taking an infinitesimal path element with length *ds* at position *s* where the light speed is $c/n(s)$:

Time to travel length *ds*: $dt = \frac{ds}{c/n(s)} = \frac{n(s)ds}{c}$

Total path time = $T = \int_A^B dt = \int_A^B \frac{n(s)ds}{c}$




Fermat's Principle says of the possible paths, the actual ray path taken is the one with *least time* $\boldsymbol{T_{min}}$ or more generally is a stationary path where small variation from this path to first order has the same path time. We might like to write this as a turning point $\frac{dT}{ds} = 0$, but since path variable *s* is unspecified in this general case, we write Fermat Principle using a notation from *calculus of variations*

$$\boldsymbol{\delta T = 0}$$

where $\boldsymbol{\delta T}$ should be taken to mean the path time variation due to a small variation in path. In some specific optical systems that we encounter (see later) we can write $\boldsymbol{\delta T = \frac{dT}{d\varepsilon}}$ in a derivative form where $\boldsymbol{\varepsilon}$ might be a position *x*, radius *r*, or angle $\theta$ coordinate that fully specifies the path between two points for the geometry considered. We take that derivative to be zero to satisfy Fermat.

Since $T = \int_A^B \frac{n(s)ds}{c} = \frac{OPL}{c}$ where $OPL = \int_A^B n(s)ds$ is **optical path length** and noting c is constant, Fermat's Principle $\boldsymbol{\delta T = 0}$ can also be stated as:

$$\boldsymbol{\delta[OPL] = 0}$$

It is this latter definition that we will employ, noting that if OPL path is a **minimum**, it is equivalent to least-time, but further noting that Fermat's full modern version is satisfied if path OPL is also a **maximum** or a **point of inflexion** (saddle-point).

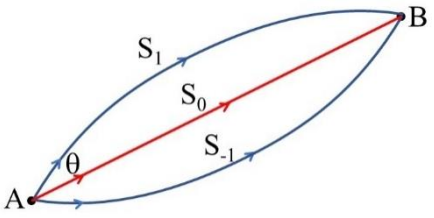
Some examples are given below:

### Case 1. Uniform medium: $n(s) = n = constant$

Central ray $S_0$ is a stationary (minimum time) path as OPL increases symmetrically for paths either side of it:

$OPL = n.l_{AB}$ (where $l_{AB}$ is physical length AB).

- Light travels in straight lines in homogeneous medium (least time path between AB)




### Case 2. Law of Refraction (Snell's Law)

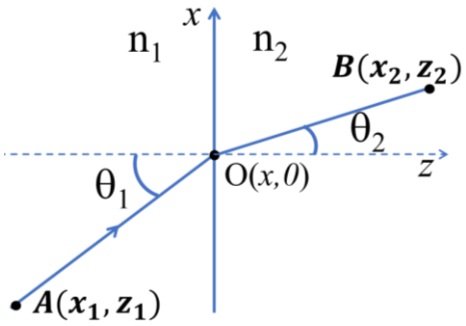
We can consider a path in the $x-z$ plane from a point $A(x_1, z_1)$ to a point $B(x_2, z_2)$ that crosses a refractive index boundary (taken as plane $z = 0$). We generalise the path by taken a crossing point $O(x, 0)$ at variable position *x* (and remembering from Case 1, least time for path sections AO and OB will be straight lines). We can uniquely define paths between *AB* as function of boundary crossing coordinate, *x*, and find $OPL(x)$ using Pythagoras, which we can then differentiate to find stationary path:

$$OPL(x) = n_1\sqrt{z_1^2 + (x - x_1)^2} + n_2\sqrt{z_2^2 + (x_2 - x)^2}$$

$$\frac{d(OPL)}{dx} = \frac{n_1(x - x_1)}{\sqrt{z_1^2 + (x - x_1)^2}} - \frac{n_2(x_2 - x)}{\sqrt{z_2^2 + (x_2 - x)^2}}$$

$$= n_1 sin\theta_1 - n_2 sin\theta_2 = 0$$

$$\boldsymbol{n_1 sin\theta_1 = n_2 sin\theta_2} \quad \textbf{(Snell's Law)}$$




Quite remarkably, the ray path followed by bending of light at the refractive index boundary as given by Snell's Law, is a path of least time! (It is easy to see counter-examples that it is not a maximum of time, nor is it a point of inflexion).

### Case 3. Law of Reflection

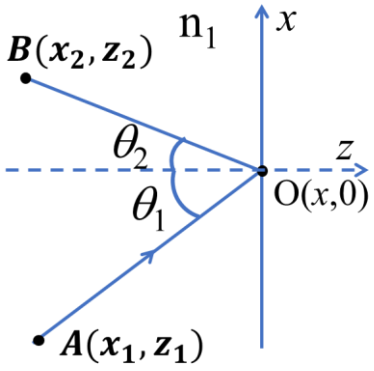
We can consider a path from A to B via plane mirror or reflecting surface (taken as plane $z = 0$). We generalise the path by taken a reflection point $O(x, 0)$ at variable position *x*. We find optical path length between AB as function *x*, $OPL(x)$ and then differentiate it to find stationary path:

$$OPL(x) = n_1\sqrt{z_1^2 + (x - x_1)^2} + n_1\sqrt{z_2^2 + (x_2 - x)^2}$$

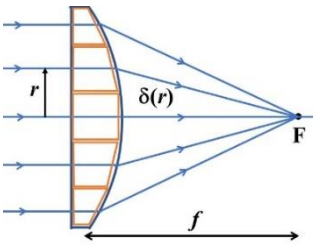
$$\frac{d(OPL)}{dx} = n_1\left[\frac{(x - x_1)}{\sqrt{z_1^2 + (x - x_1)^2}} - \frac{(x_2 - x)}{\sqrt{z_2^2 + (x_2 - x)^2}}\right]$$

$$= n_1[sin\theta_1 - sin\theta_2] = 0$$

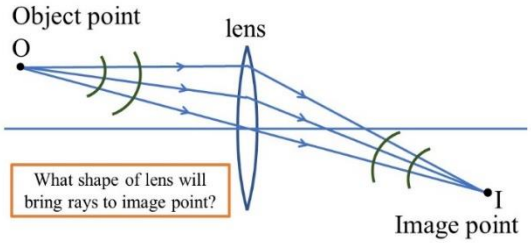
$$\boldsymbol{\theta_1 = \theta_2} \quad \textbf{(Reflection Law)}$$



## Fermat's Principle & Imaging




Lenses are important elements in imaging systems, but what is their best shape for imaging? We can consider a lens as a (continuum) set of prism elements providing varying deviation $\delta(r)$ so parallel incoming rays all converge at point F on axis, at a distance we define as the focal length $f$ of the lens.




This is an interesting way to thinking but a more powerful approach for lens design is to use Fermat's Principle. Consider an object O (point A) to be imaged at image I (point B) by a lens. Fermat says ray must take the least time (or stationary) path, but an object point O must be considered a spherical wave and *all ray paths from O* (over angular field of view of lens) must arrive at the same point I. *But Fermat says <u>all</u> these ray paths be least time (stationary) paths* - ***how is this possible?***

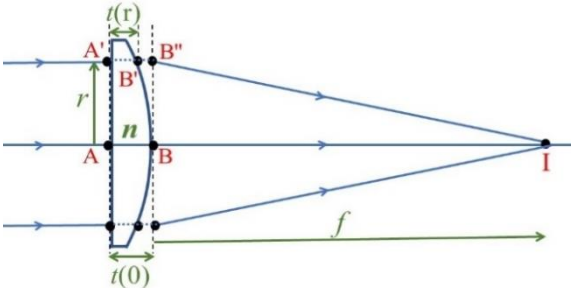
Fermat's Principle $\delta(OPL) = 0$, requires no variation of *OPL* with (small) change in path direction. This can be satisfied if all ray paths (S) from O to I have the **same optical path length**. We obtain **Fermat's Principle applied to imaging**:

$$\boldsymbol{OPL(s) = constant}$$

This criterion determines the shape of an imaging lens, mirror, lens combination or indeed any imaging system. This gives us a powerful tool for optical design.

## Design of a Thin Lens using Fermat's Imaging Principle

Consider an axial object point O (at ∞) producing parallel rays imaged to point *I* by a *plano-convex* lens with a flat surface and a bulged surface with thickness $A'B' = t(r)$ and central thickness $AB = t(0)$. To make all ray paths have the same OPL, we can add the most glass to direct axial ray OABI and less glass in ray path OA'B'I. Refraction occurs at point B', but for thin lens ($t(0) \ll f$) we can approximate the bend at point B" and consider a straight path A'B" with glass $OPL(A'B') = nt(r)$ and air $OPL(B'B') = \Delta t = t(0) - t(r)$ with $n = 1$.




Axial path: $OPL(0) = \text{ABI} = nt(0) + f$

Off-axis path: $OPL(r) = \text{A'B'B''I} = nt(r) + [t(0) - t(r)] + \sqrt{f^2 + r^2}$

For small angles $r \ll f$: $B''I = \sqrt{f^2 + r^2} = f(1 + r^2/f^2)^{1/2} \approx f + \frac{r^2}{2f}$

Fermat requires: $\boldsymbol{OPL(r) = OPL(0)}$

Hence $$nt(r) + [t(0) - t(r)] + f + \frac{r^2}{2f} = nt(0) + f$$

Rearranging: $\boldsymbol{t(r) = t(0) - \frac{r^2}{2(n-1)f}}$ **or** $\boldsymbol{\Delta t(r) = \frac{r^2}{2(n-1)f}}$

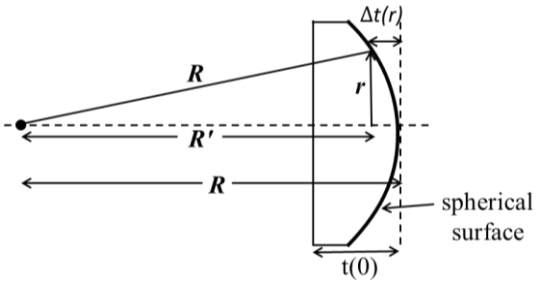
- We discover lens thickness $t(r)$ or surface shape $\Delta t(r) = t(0) - t(r)$ should be **parabolic, varying as $r^2$**.

Lens curvature (bulge) $C = \frac{1}{2(n-1)f}$ depends on the lens focal length (*f*) and its glass refractive index relative to air $(n - 1)$.

## Spherical Lenses

Although parabolic lens shape is good for small angle imaging, most lenses (and mirrors) have spherical surfaces as they are easier and cheaper to manufacture than parabolic surfaces.

For paraxial angles (small axial distances *r*) they approximate well to a parabola.




A binominal expansion for the spherical surface $\Delta t(r)$ for small axial *r*.

$$\Delta t(r) = R - R' = R - \sqrt{R^2 - r^2} = R - R(1 - (r/R)^2)^{\frac{1}{2}} \approx \frac{r^2}{2R} + \frac{r^4}{8R^3}$$

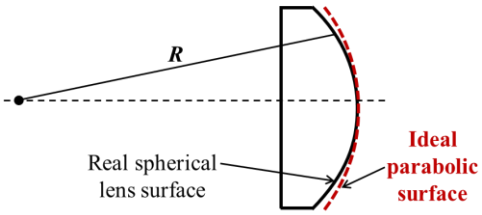
We find a spherical surface has a (paraxial) parabolic surface shape:

$$\Delta t(r) = \frac{r^2}{2R}.$$

Comparing with Fermat's derived parabolic shape: $\Delta t(r) = \frac{r^2}{2(n-1)f}$, we find:

$$f = \frac{R}{(n-1)}$$

Higher-order term in the expansion $(r^4/8R^3)$ is a deviation from the Fermat's parabolic shape. It is responsible for what is called **spherical aberration** with larger off-axis rays (*r*) more strongly focused than paraxial rays – they intercept the axis before the paraxial focal point F, leading to a "blurring" of the image point.



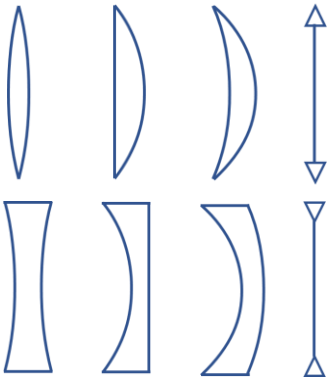

**Lens Maker's Equation**: More generally, lens has two spherical surfaces, with front and back curvatures $R_1$ and $R_2$, surface shapes $\Delta t_{1,2}(r) = \frac{r^2}{2R_{1,2}}$, and $t(r) = t(0) - \frac{r^2}{2R_1} - \frac{r^2}{2R_2}$.

Comparing to Fermat's lens shape $t(r) = t(0) - \frac{1}{2(n-1)f} r^2$ we get **Lens Maker's Equation**:

$$\frac{1}{f} = (n-1)\left(\frac{1}{R_1} + \frac{1}{R_2}\right)$$

where curvature *R* is positive for converging (convex facing into air) surface and negative for diverging (concave) surfaces. The Lens Maker's formula shows that a range of lens curvature combinations $\left(\frac{1}{R_1} + \frac{1}{R_2}\right)$ can achieve same focal length $f$ (e.g. if $R_1 = \infty$, we have a plano-convex lens case with $f = R_2 \,/\, (n-1)$, as previously).

A positive (converging lens) can be biconvex, plano-convex, or a concave-convex combination known as a positive meniscus lens. A similar set of negative lenses can also be considered (biconcave; plano-concave, negative meniscus lens).



A symbolic representation in ray tracing diagrams can be used of a positive lens (outward arrowheads) and negative lens (inward arrow-heads). This representation clarifies the thin lens location without worry about exact lens shape and ray deviations occurring in the one plane, rather than separately at the two surfaces.

# Imaging with Lenses & Mirrors

## Drawing Ray Diagrams to locate Images

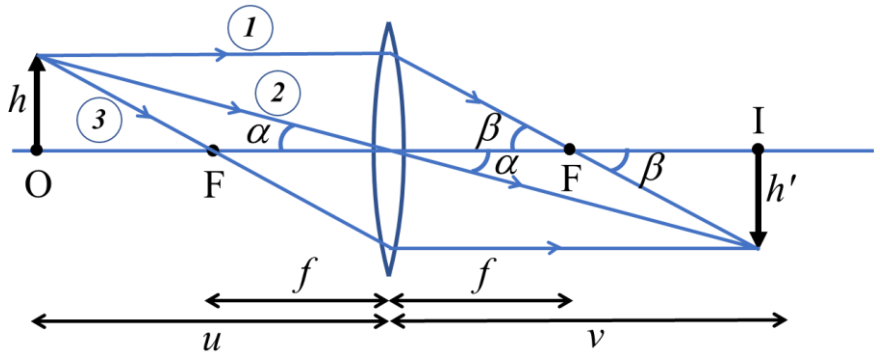
We consider a thin positive lens and define terms (later we will look at negative lens):

$u$ = object distance to lens; $v$ = image distance to lens

$f$ = focal length of lens [where incident parallel rays (object at infinity) converge to a point]

$F$ are axial front and back focal points at focal distance $f$ from lens.

Ray Diagram for a Positive (converging) Lens




There are **3 Principal Rays** (that follow rules of lenses) to draw and locate image. Drawn from the top of the object:

1. **Ray parallel to axis** passes through focal point F (or appears to diverge from *F*)
2. **Ray through centre of lens** is not deviated
3. **Ray through focal point *F*** **(**or directed to *F***)** travels parallel to axis

The terms in brackets (in red) apply to negative lens (see later)
Crossing of any two Principal Rays can locate image position I, *but 3rd Ray is a good check.*

## Thin Lens Formula

From principal ray diagram, assuming paraxial (small angle) approximation, $tan\theta \approx \theta$

$$\alpha = \frac{h}{u} = \frac{h'}{v} \;\; ; \;\; \beta = \frac{h}{f} = \frac{h'}{v-f}$$

$$\frac{h\prime}{h} = \frac{v}{u} \;\; ; \;\; \frac{h\prime}{h} = \frac{v-f}{f} = \frac{v}{f} - 1$$

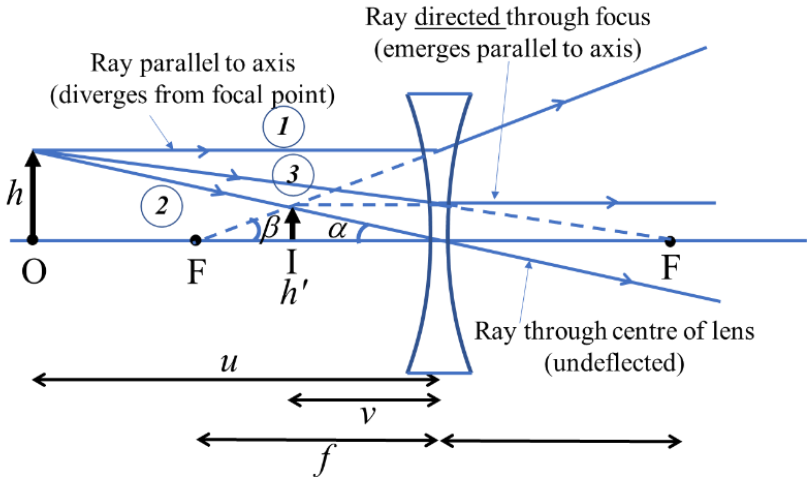
$$\boxed{\frac{\mathbf{1}}{\boldsymbol{u}} + \frac{\mathbf{1}}{\boldsymbol{v}} = \frac{\mathbf{1}}{\boldsymbol{f}}}$$

## Transverse Magnification Formula $m_T$

From ray diagram: $\boxed{\boldsymbol{m_T} = \frac{\boldsymbol{h\prime}}{\boldsymbol{h}} = \frac{\boldsymbol{-v}}{\boldsymbol{u}}}$

where negative sign is introduced to indicate image is inverted.

Ray Diagram for a Negative (diverging) Lens:




In terms of distances, you can show: $\frac{1}{u} - \frac{1}{v} = -\frac{1}{f}$

And transverse magnification: $m_T = \frac{h'}{h} = \frac{v}{u}$

**Instead of having different formula for positive and negative lenses, we will use one thin lens formula (and magnification formula) with sign convention:**

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f}$$

$$m_T = \frac{-v}{u}$$

- *u* positive (real) on left; negative (virtual) on right
- *v* positive (real) on right; negative (virtual) on left
- *f* positive for converging lens; negative for diverging lens.

## Real and Virtual Images

In the first Positive Lens diagram: Image is **real** (*v* = positive; right of lens). All 3 rays physically pass through image. Image would be observed on screen placed at I.

In the second Negative Lens diagram: Image is **virtual** (*v* = negative; left of lens). Apart from central ray, principal rays diverge from image but not physically through it. Image **not** observable on screen placed at I.

**But note:**

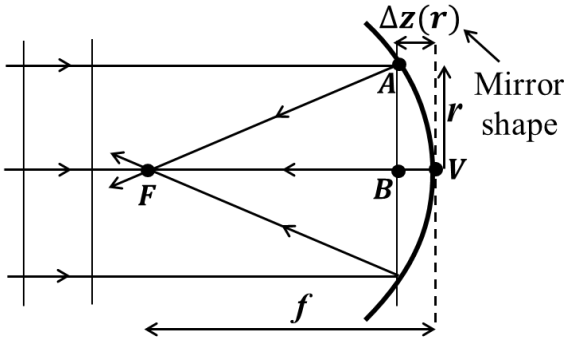
**NB1** Converging lens also produce a virtual image when object distance $u < f$. [Prove this]

**NB2** Although virtual image is not observable at image point I itself, you (or a camera) can see it because the lens of the eye (or camera) can convert a virtual image into a real image onto eye's retina (or camera detector array).

## Mirrors as Imaging Elements

### Fermat's Principle & Shape of Mirror for Imaging

Consider rays parallel to the optical axis (from a point object at infinity) incident on a curved mirror. The central vertex V of the mirror we take as origin $(0,0)$ of our coordinate system $(r, \Delta z)$ that specifies the shape of the mirror $\Delta z(r)$. To form an image all rays must be intersect at a common point which for object at infinity is the focal point F of the mirror. The distance VF we take as the focal length *f*.




Fermat's principle for imaging requires all optical paths from object to image to be equal, and since it is all in the same medium the physical path lengths must all be equal. The parallel central ray ($r$=0) and upper ray ($r$) have equal optical paths (plane wavefront AB) until point A meets the mirror. Fermat imaging requirement is that path lengths **AF** = **BVF** where A is at general mirror coordinate $(r, \Delta z)$ and B is at $(0, \Delta z)$. Using Pythagoras:

$$AF = \sqrt{(f - \Delta z(r))^2 + r^2}$$

and $BVF = f + \Delta z(r)$. If we consider $(AF)^2 = (BVF)^2$, we get

$$(f - \Delta z(r))^2 + r^2 = (f + \Delta z(r))^2$$

$$f^2 - 2f\Delta z + \Delta z^2 + r^2 = f^2 + 2f\Delta z + \Delta z^2$$

where many terms cancel to give:

$$\boxed{\Delta z(r) = \frac{r^2}{4f}}$$

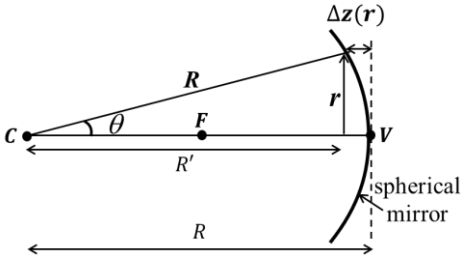
The "ideal" mirror shape for axial imaging is therefore **parabolic** (with a quadratic $r^2$ curvature dependence) with imaging in reflection rather than in transmission through a lens. **It is important to note that unlike the paraxial approximation we took for the lens shape, the parabolic mirror profile works for all (small & large) transverse distance $r$.**

- *The results we have got don't only relate to visible light. Parabolic radio telescope mirrors and satellite dishes are used to image and concentrate other EM radiation.*
- *If an object source (e.g. light bulb) is placed at focal point F of parabolic mirror, than following rays in reverse to above diagram, the emergent rays are parallel. This is the basis of a car headlamp to project a parallel beam.*

## Imaging with a Spherical Mirror

From our knowledge of a spherical lens surface, we know (for small paraxial angles) spherical mirror shape:

$$\Delta z(x) = \frac{r^2}{2R}$$




Comparing to parabolic mirror shape: $\Delta z(r) = \frac{r^2}{4f}$

we find (paraxial) focal length $f$ (distance VF) of mirror with curvature $R$ given simply by:

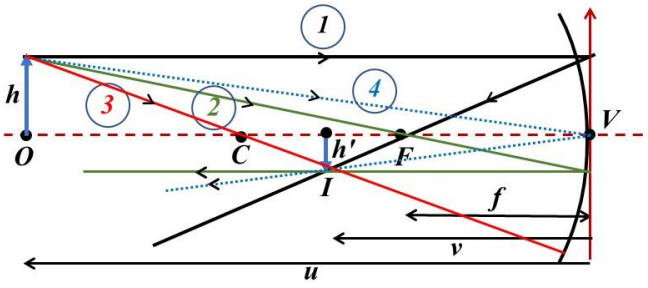
$$\boldsymbol{f = \frac{R}{2}}$$

The focal point F is midway between mirror vertex V and its centre of curvature C.

Spherical (or parabolic) mirror has equivalent focusing and imaging properties as a lens with focal length $f = R/2$, except, of course, the imaging is in reflection. We can also perform mirror ray tracing, with principal rays to locate image I, they follow similar rules as for the thin lens (and we consider a "thin" mirror $\Delta z \ll f, R$, with reflections occurring in plane $z = 0$ containing the mirror vertex V):

There are **4 Principal Rays** (that follow rules of mirrors):

1. Ray parallel to axis is reflected through focal point F (or for diverging mirror, appear to diverge from focal point F behind mirror).
2. Ray passing through a focal point F is reflected to travel parallel to axis.
3. Ray passing through centre of mirror curvature C (hits mirror surface at normal incidence) is reflected along same path.
4. Ray incident at vertex V of mirror is reflected at equal angle (law of reflection)




We get an equivalent mirror formula and magnification formula as with a lens, when using the object and image distance convention (positive on incidence side of mirror):

$$\boldsymbol{\frac{1}{u} + \frac{1}{v} = \frac{2}{R} = \frac{1}{f}}$$

$$\boldsymbol{m_T = \frac{h'}{h} = \frac{-v}{u}}$$

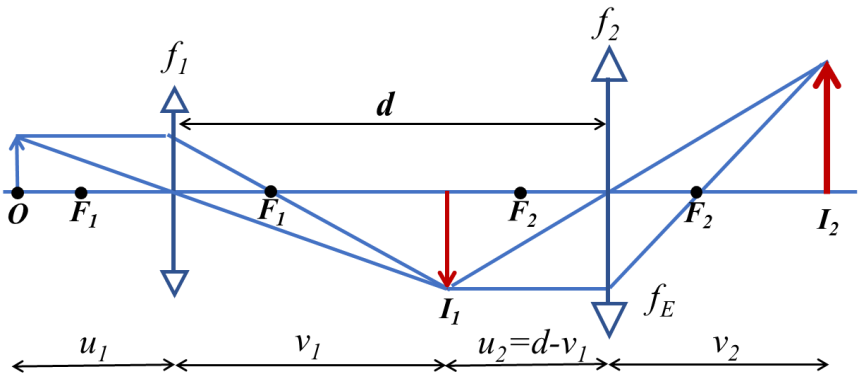
For light incident on a convex (diverging) mirror: we put negative focal length in mirror formula $f = -|R|/2$ with focal point F behind the mirror. For real object distances ($u$ positive), we get divergent reflected rays producing a *virtual image* point behind mirror ($v$ negative).

## Two-Lens (or multiple-lens) Imaging Systems

Optical instruments will often consist of more than a single lens (e.g. microscope, telescope). Multi-lens systems are also commonly used for aberration correction (e.g. smartphone camera). The simplest case is a two-lens system that we will consider here.

### Two-Lens Optical Systems

To find image of 2-lens combination: e.g. if you know $u_1$; and $f_1$; $f_2$; and $d$ (lens separation)




A) **Use lens formula twice to calculate final image**:

i) Use lens formula to find $v_1$ (1st image $I_1$)

ii) Use lens formula to find $v_2$ (2nd image $I_2$) with 2nd lens object distance $u_2 = d - v_1$

Combined transverse magnification of 2-lens system is product of each lens magnification:

$$M_T = m_1.m_2 = \left(\frac{-v_1}{u_1}\right)\left(\frac{-v_2}{u_2}\right)$$

B) (alternatively) **Use two successive Ray Diagrams:**

i) Draw 2 principal rays from O to $I_1$

ii) Then draw 2 principal rays from $I_1$ (new object) to $I_2$.

(Hint: use results of Method A to get scale of diagram before you start).

**Two Closely Spaced Lenses**. A special and important case is when the two lenses are placed so closely together that $\boldsymbol{d = 0}$. In that case, $u_2 = -v_1$.

For first lens: $$\frac{1}{v_1} = \frac{1}{f_1} - \frac{1}{u_1}$$

And for second lens ($u_2 = -v_1$): $$\frac{1}{v_2} = \frac{1}{f_2} - \frac{1}{u_2} = \frac{1}{f_2} + \left(\frac{1}{f_1} - \frac{1}{u_1}\right) = \frac{1}{f_T} - \frac{1}{u_1}$$

where the input object distance $u_1$ and final image distance $v_2$ form the thin lens formula. We find the two-lens system has a "single-lens" formula with an effective lens power $\frac{1}{f_T}$ the sum of the individual lens powers when $d = 0$:

$$\boxed{\frac{\mathbf{1}}{\boldsymbol{f_T}} = \frac{\mathbf{1}}{\boldsymbol{f_1}} + \frac{\mathbf{1}}{\boldsymbol{f_2}}}$$

## Optical Power D of Lens

We can define the optical power of lens (**D**) as the reciprocal of the focal length: $D = \frac{1}{f}$.

D is measured in SI unit as dioptres ($m^{-1}$)

e.g. $f = 200mm$ has an optical power $D = \frac{1}{f} = 5\ dioptres$.

(Note: *The unit Dioptres is used by opticians to classify your eye and prescribing lenses for eye correction glasses or contact lenses*).

We can add the lens powers of two (or more) closely spaced thin lenses (equivalent to: $\frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2} + \cdots$)

$$\boldsymbol{D = D_1 + D_2 + \cdots + D_n}$$

It is the simplicity of adding powers that makes the D parameter sometimes more useful to employ than quoting focal length *f*.

This lens power summation is used in the Lens Makers Formula can also be considered as the sum of focal power of each lens surface:

$$\boldsymbol{D = \frac{1}{f} = (n-1)\left(\frac{1}{R_1} + \frac{1}{R_2}\right) = D_1 + D_2}$$

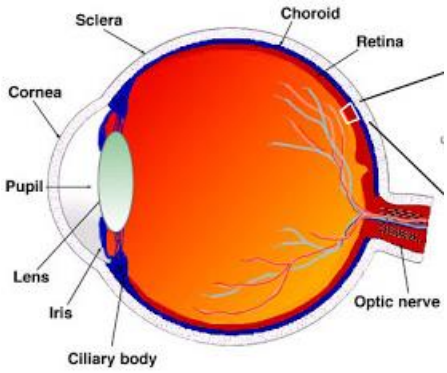
where $D_{1,2} = (n-1)\left(\frac{1}{R_{1,2}}\right)$.

# Optical Imaging Instruments – Part 1

Optical instruments augment the human eye to see smaller details (magnifier; microscope) and distant objects (telescope). We therefore start the topic of optical imaging instruments with the human eye: its physiology and its operation as an imaging system.

## THE HUMAN EYE

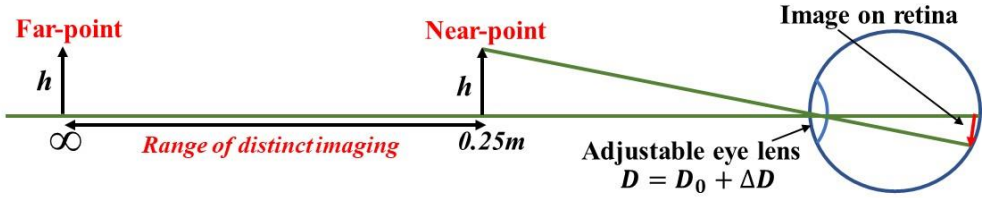
### The Human Eye: Physiology




- The eye can image light by the combined lensing provided by the curvature of the eye's outer **cornea** and an **internal lens**. **Ciliary muscles** can increase bulge of internal lens to increase its lensing power $\Delta D$ giving the eye a variable lensing power $D = D_0 + \Delta D$ to adjust its focus for objects at different distances.
- **Pupil** near internal lens is an adjustable aperture (diameter ~2-8mm) controlling amount of light entering the eye (less in bright light; more in the dark)
- **Retina** at back of eye has light sensitive cells (rods and cones) detecting the light and sending the image to the brain via the **optic nerve**.
- **Rods** detect black and white. They are most sensitive providing our night-time or dark vision (and can even be capable of single photon detection).
- **Cones** detect colour. There are 3 cone types with peak response notionally in the **red**, **green**, and **blue**. This is the basis of the 3 primary colours for vision, with relative excitation of cones providing our rich visualisation of colours.

### The Human Eye: Imaging

The lensing powers of the curved **cornea** $D_C \sim 40\ dp$ and the **internal lens** $D_L \sim 20\ dp$ produce a combined dioptric power $D_0 = D_C + D_L \sim 60\ dp\ (m^{-1})$. Changing the bulge of internal lens gives lensing adjustment $D = D_0 + \Delta D$ to image objects over a range of distances. By convention, for the "normal" eye this range is taken between a **far-point** $u_{max} = \infty$ and a **near-point** $u_{min} = 250mm\ (= 0.25m)$.

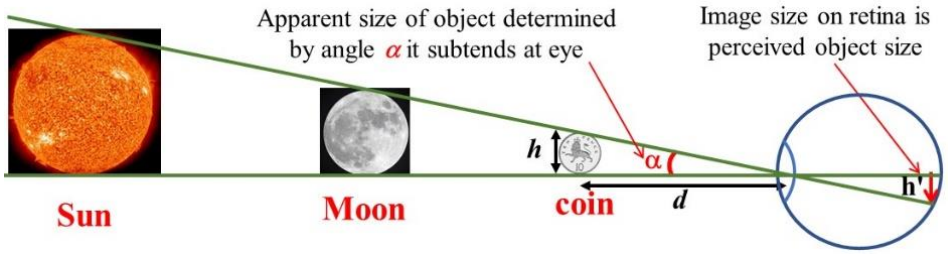



- Using the fixed eye lens to retina distance $v_e$ as the image distance in the lens formula gives relationship between eye lensing $\boldsymbol{D}$ and object distance $\boldsymbol{u}$: $\boldsymbol{D = \frac{1}{f} = \frac{1}{u} + \frac{1}{v_e}}$.
- At far-point ($u_{max} = \infty$), lens formula gives relaxed eye lensing $\boldsymbol{D_0 = \frac{1}{v_e} \sim 60\ m^{-1}}$, with (relaxed) focal length of eye $f_0 \sim 17\ mm$ equal to eye lens to retina distance $v_e$.
- At near-point $u_{min} = 0.25\ m$, $\boldsymbol{D = D_0 + \frac{1}{0.25}}$ needing increase in dioptric power $\boldsymbol{\Delta D = D - D_0 = 4\ m^{-1}}$ to maintain image on retina.
- A young person may have eye adjustment range $\Delta D \sim 10\ m^{-1}$ but this decreases with age. When $\boldsymbol{\Delta D < 4\ m^{-1}}$ you may need corrective glasses (or contact lenses) (e.g. for reading) with an additional positive lens power $D_G$ to augment the eye: $D = D_{eye} + D_G$.

- Some people have relaxed eye $D_0$ too strong to image long distances (a condition called myopia) and image is formed in front of retina. In this case the corrective lens will have a negative lensing power (diverging lens) to decrease *D*.
- The above model the eye is very similar to the way a camera and adjustment of its lens system forms an image on a digital detector array.

## Apparent size and Angular size

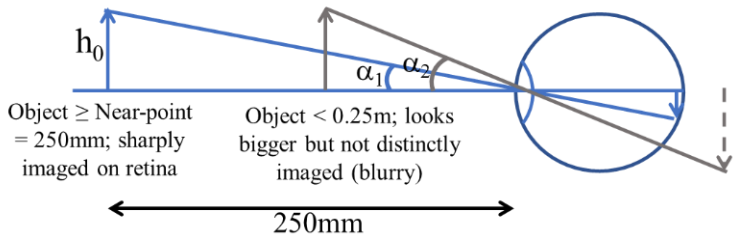
**Apparent size** of an object is solely determined by the size of the image (**h'**) on the retina. In a camera the apparent size would be the width of detector pixels illuminated.




The 'perceived' size is not the physical object size (**d**) itself (e.g. Sun, Moon, coin) but depends solely on the **angular size** (α) that the object subtends at the pupil of the eye, through which the light must enter. [*Note in above diagram, image is inverted, but the brain re-processes to interpret object as upright*].

**Angular size** of an object (for small angles) is given by: $$\alpha = \frac{h_0}{d}$$

The angular size increases as object distance d decreases but minimum distance the eye can image (for a "normal" eye) is 250 mm (0.25 m), the **near point**, sometimes called the **least distance of distinct vision**.




We define a new quantity for the human eye:

**Maximum distinct object angular size**: $$\alpha_{250} = \frac{h_0(mm)}{250}$$

This angular size is used to compare the improvement that some optical instruments provide compare to the limitations of the human eye.

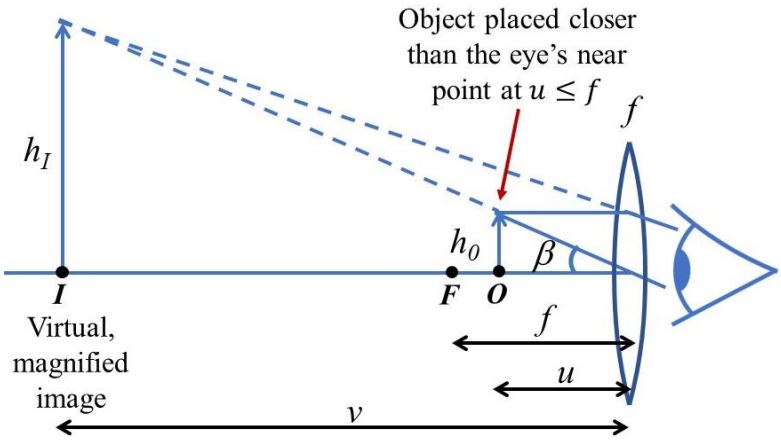
## SIMPLE MAGNIFIER

The simple magnifier is a single positive (converging) lens that when correctly placed with respect to the object provides a magnified image for the eye.

Consider a converging lens *f* with a close object: $u \leq f$.

Lens formula gives image location: $\frac{1}{v} = \frac{1}{f} - \frac{1}{u} < 0$

Since $|v| > u$, image is magnified: $m_T = \frac{-v}{u} > +1$ and since sign of *v* is negative, image is non-inverted (upright) and virtual, located behind lens on same side as object.




By adjusting lens to object distance, to making $|v| \geq 250\ mm$ (eye near-point) the magnified virtual image can be seen *distinctly* by the eye.

Image angular size: $\beta = \frac{h_1}{|v|} = \frac{h_0}{u}$ = object angular size (we assume eye next to thin lens *f*).

We define a new magnification quantity for eye imaging:

**Angular Magnification**: $$\boldsymbol{M_\alpha = \frac{angle\ subtended\ by\ lens\ image}{angle\ subtended\ by\ object\ at\ 250mm} = \frac{\beta}{\alpha_{250}}}$$

$\alpha_{250} = \frac{h_0(mm)}{250}$; $\beta = \frac{h_1}{v} = \frac{h_0}{u}$

$$M_\alpha = \frac{\beta}{\alpha_{250}} = \frac{250}{u(mm)}$$

Normal eye is more relaxed if image at ∞ rather than 250mm. This occurs when $u = f$.

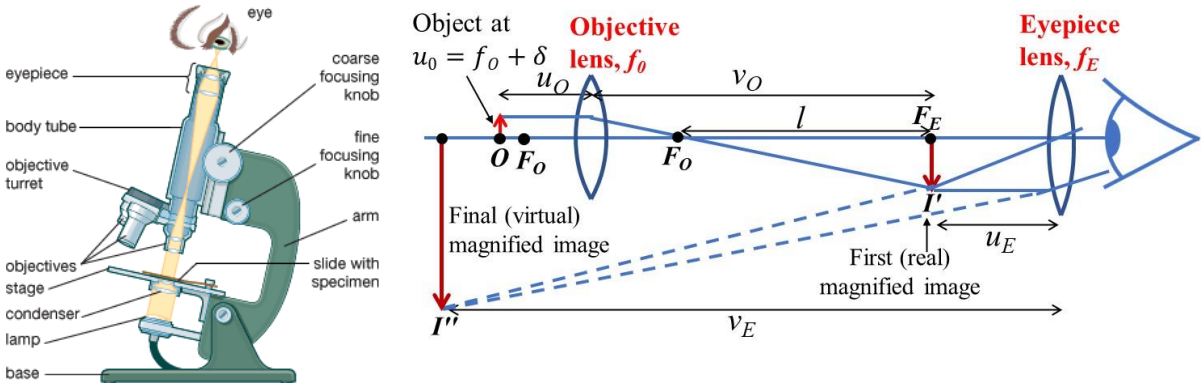
$$\boldsymbol{M_\infty = \frac{250}{f(mm)}}$$

This shows angular magnification is large for small *f*: e.g. $f = 25\ mm$; $M_\infty = 10$.

[*An alternative way to think of the simple magnifier is to consider when it is in contact with eye it provides the eye with a combined dioptric power* $D = D_{eye} + D_M$, *where* $D_M$ *is magnifying lens dioptric power. This strongly increased D reduces the effective near-point of the eye to allow the object to be closer to eye and have a larger angular size*].

## COMPOUND MICROSCOPE

The magnification of simple magnifying lens is limited as impractical to place eye sufficiently close to very short focal length (highly bulged) lenses. In practice, it is better to use a two-step magnification with a 2-lens system: the compound microscope, with an **objective lens** (placed near the object) and **eyepiece lens** (placed near the eye or viewer).




i) The objective and eyepiece lenses have (short) focal lengths $f_0, f_E$ and separated by distance $d = (f_0 + f_E + l)$, where $l$ is often much larger than the focal lengths. The objective lens creates a real image (magnification, $m_O$) that is the object for the eyepiece lens acting as a simple magnifier to giving a further stage of magnification ($m_E$). The compound magnification $M = m_O.m_E$ is the product of the two magnifications. To simplify the analysis, **in the following we take the intermediate image I' to be located at focal point of eyepiece lens $F_E$ and final image I'' at infinity**.

ii) To get a large objective magnification, object O (on microscope stage) has its distance $u_0 = f_0 + \delta$ (*with small* $\delta$) adjusted to create a real image at $v_0 = f_0 + l$ (situated at $F_E$) and magnification $m_O = -v_0/u_0$. Using lens formula: $\frac{1}{u_0} = \frac{1}{f_0} - \frac{1}{v_0}$

$$m_O = \frac{-v_0}{u_0} = \frac{-v_0}{f_0} + 1 = \frac{-(f_0 + l)}{f_0} + 1 = \frac{-l}{f_0}$$

iii) With first image $I'$ at eyepiece lens focal point, then $u_E = f_E$ and eyepiece acts as a simple magnifying lens with final image $I''$ at $v_E = \infty$ (for relaxed viewing).

$$M = m_O.m_E = \left(\frac{-v_0}{u_0}\right)\left(\frac{250}{f_E(mm)}\right)$$

$$\frac{1}{u_0} = \frac{1}{f_0} - \frac{1}{v_0}$$

$$\boxed{M = -\left(\frac{l}{f_0}\right)\left(\frac{250}{f_E(mm)}\right)}$$

Example: $f_0 = 4\,mm;\ f_E = 25mm; l = 160mm$

$$M = (-40)(10) = -400$$

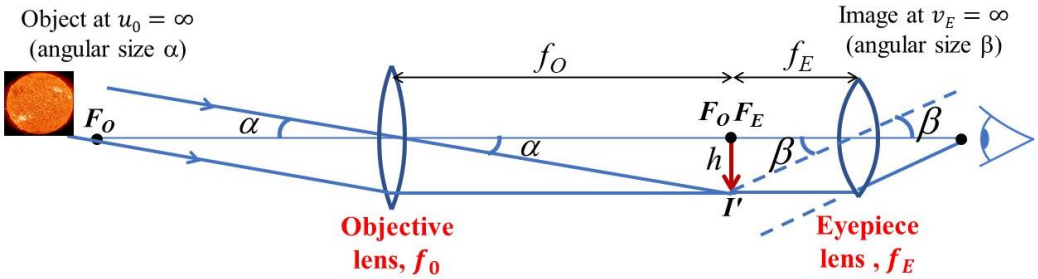
# Optical Imaging Instruments – Part 2

## (ASTRONOMICAL) TELESCOPE

Telescopes are used for viewing distant objects, especially for astronomical viewing. For the telescope the object is usually at infinity and the object's angular size is the key feature to be magnified. Telescopes can be constructed with lenses (refracting telescopes) or mirrors (reflecting telescope), we consider both.

## REFRACTING TELESCOPES (using lenses)

**Keplerian Telescope** design uses positive objective lens $f_o$ and positive eyepiece $f_E$ lens.




We take the case that the final object is at infinity and separate the two lenses by the sum of their focal lengths: $\boldsymbol{d = f_O + f_E}$ .

With object (angular size $\alpha$) at $u_O = \infty$, the objective lens produces a (real) image I' at its focal point, $F_O$, ($v_O = f_O$) which for $d = f_O + f_E$ is coincident with the eyepiece focal point $F_E$, ($u_E = f_E$) providing a final image at infinity with angular size $\beta$. From telescope diagram, angular sizes are: $\alpha = \frac{h}{f_o}$ ; $\beta = \frac{h}{f_E}$

Telescope angular magnification: $\boldsymbol{M_\alpha = \frac{-\beta}{\alpha} = -\frac{f_O}{f_E}}$

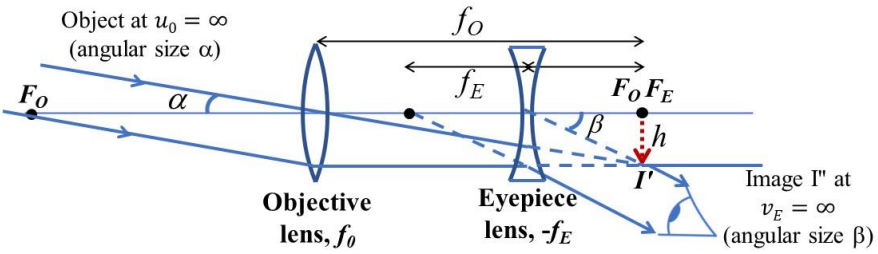
using negative sign as image is inverted. High magnification requires large focal length ratio using a long objective focal length $f_o$ and hence a long telescope tube length $d$.

**Galilean Telescope** design uses positive objective $f_o$ and negative eyepiece $-|f_E|$. By co-locating focal points $F_0$ and $F_E$, the first lens intermediate image I' at $F_O$ provides a *virtual* object at $F_E$ for the negative eyepiece lens producing final image at infinity. A more compact telescope is produced with shorter lens separation: $\boldsymbol{d = f_0 - |f_E|}$.

From ray diagram:
$\alpha = \frac{h}{f_o}$ ; $\beta = \frac{h}{|f_E|}$

Angular magnification:

$\boldsymbol{M_\alpha = \frac{\beta}{\alpha} = +\frac{f_O}{|f_E|}}$




This is the same magnification as Keplerian telescope, in terms of magnitude of focal lengths, but in a more compact length. The image is upright (not inverted).

[*In high-power laser applications, the Galilean telescope may be preferred to Keplerian due to its compact size but also no intermediate focus avoiding a high intensity focal point where air breakdown or optical damage can occur*].

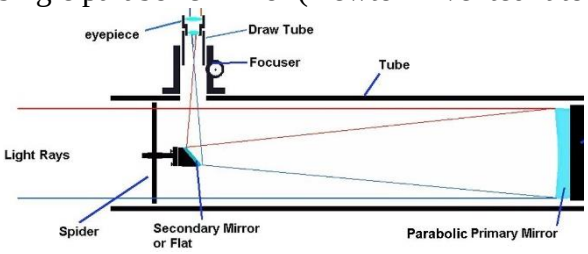
## REFLECTING TELESCOPES (using mirrors)

For good astronomical viewing, a large aperture objective lens is required for light collection (to visualise weak objects) and for high resolution. Increasing glass lens diameter becomes impractical due to its increasing thickness and weight (that would cause it to sag) so mirrors are preferred and used in all large telescopes and have other advantages:

- *they can be made thin and lightweight compared to a lens*
- *they can be supported by a rigid framework behind mirrored surface*
- *they don't suffer from chromatic aberration*
- *there are good designs for correction of other aberrations*
- *mirrors can reflect visible, IR, microwave, radio-wave EM parts of spectrum, whereas lenses are limited to glass transmission band in UV, visible and near-IR.*

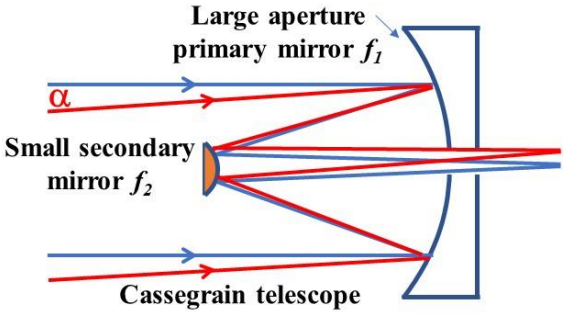
### Newtonian Telescope

This was the original reflecting telescope with a single parabolic mirror (Newton invented it to solve the problem of chromatic aberration produced by lens-based telescopes). The reflected beam is in the path of the incoming light, so a small 45° plane mirror is used to send the image to a side viewing port.



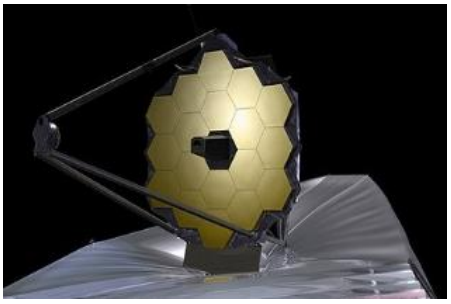

### Cassegrain Telescope

The Newtonian telescope has been largely superseded by the Cassegrain-style of telescope using a large diameter primary concave mirror (positive focal length $f_1$) and a small diameter secondary convex mirror (negative focal length $f_2$). The secondary mirror lies in the path of the incoming light and a hole is made in the centre of the primary to access the image. (*The hole doesn't diminish image as the secondary mirror has already blocked rays getting directly to that part of the mirror*). The Cassegrain can be thought of as equivalent to the Galilean refracting telescope.




Angular Magnification: $\boldsymbol{M_\alpha = -\frac{f_1}{f_2} = +\frac{f_1}{|f_2|}}$.

The folded Cassegrain is more compact than the Newtonian reflecting telescope. The two-mirror design allows aberration to be corrected better than a single mirror design. Most high-end astronomical telescopes including the Hubble and James Webb Space Telescopes use variants of the Cassegrain design.



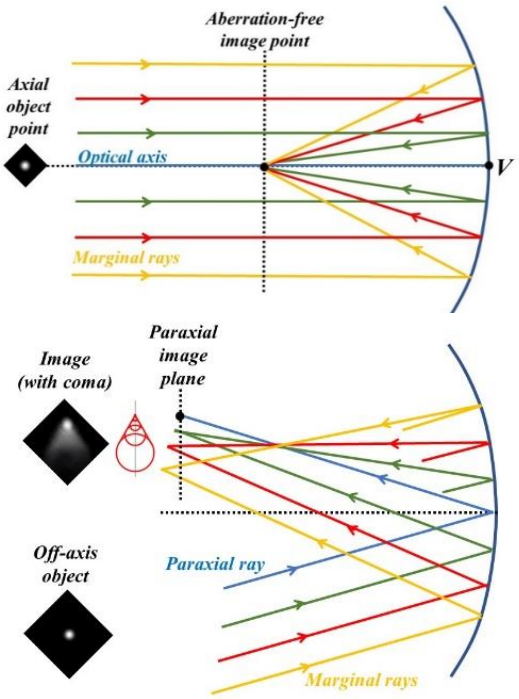
James Webb Space Telescope is designed to operate from red to infrared part of EM spectrum (well beyond the transmission band of glass lenses). Its 6.5m diameter primary mirror giving a large light collecting area. It operates at ultra-low temperature to allow sensitive detection into the long wavelength IR spectral region.

## ABERRATIONS and their CORRECTION

Fermat's Principle for perfect lensing requires equal optical path length of all ray paths from each object point (A) to its corresponding image point (B). In practice, there are deviations from this, that we call **aberrations**. Considerable effort in advanced optical design aims to minimise these in commercial optical devices such as telescopes, microscopes, and cameras.

## MIRROR ABERRATIONS

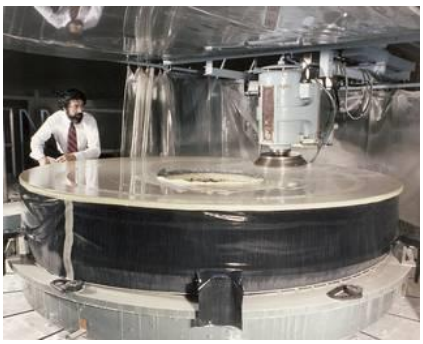
The law of reflection is independent of wavelength, so **mirror imaging does not suffer from chromatic aberration**. Fermat's Principle also showed that for an axial object perfect imaging can be achieved with a parabolic mirror with no spherical aberration.




**Coma Aberration**. However, imaging quality (and deviation from Fermat) becomes increasing worse as the object point moves off axis. Rays near the mirror vertex V form a paraxial image point but marginal rays (further from vertex V) have asymmetric path lengths to mirror surface and intersect at increasingly displaced position from the paraxial focus. The image of an off-axis point object displays a comet-like appearance giving the name **coma** to this off-axis aberration. This aberration leads to increasing image degradation for object points in an image that are further off-axis in the field of view.

**Mirror Aberration Compensation** is achieved in a Cassegrain reflecting telescope using two mirrors, giving a greater range of mirror surface curvature parameters and separation of mirrors for elimination of both on-axis and off-axial aberration compared to a single mirror. The classical Cassegrain telescope uses a parabolic (concave) primary mirror and a hyperbolic (convex) secondary mirror which in combination correct much of the off-axial aberration. The Hubble Space Telescope is a Cassegrain-type telescopes but both mirrors are hyperbolic profiles (known as the Ritchey-Chretien design) that further eliminates off-axial aberration by even wider choice of hyperbolic mirror parameters. It is noted that each mirror individually need not be aberration-free but in combination help to fulfil Fermat's Principle for equalising *OPL* ("*two wrongs can make a right*"). The main issue with hyperbolic mirrors is their difficulty of manufacture as they must be made with exquisite precision for large mirrors to small fraction of a wavelength surface accuracy.

*The Hubble Space Telescope had an error in manufacture of its 2.4m diameter primary mirror (see photo of mirror during manufacture). This led to the first images relayed to Earth in 1990 being blurry. A subsequent NASA mission in 1993 was required to retrofit a compensating optic to undo the mirror aberrations and produce the spectacular images that Hubble subsequently produced.*



## LENS ABERRATIONS

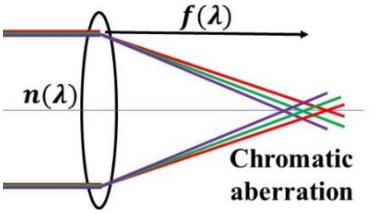
Lens aberration theory is much more complicated than mirrors when paraxial and thin lens approximations cannot be made:

- *refraction occurs through a "thick" volume of glass (rather than at a mirror surface)*
- *unlike the simple reflection law, refraction depends on the sine of angles*
- *wavelength-dependent refractive index of glass leads to chromatic aberrations.*

However, refracting systems working in transmission of transparent optics also provide a broader range of strategies for compensation. Aberration theory is very advanced, but this section provides some qualitative information to illustrate some key lens aberrations and compensation strategies and give an understanding why "the lens" in many optical instruments (e.g. cameras, microscopes) is composed of multiple lens elements.

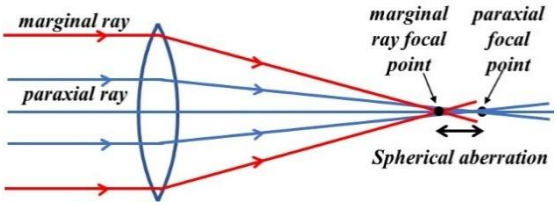
**Chromatic Aberration.** We previously noted that wavelength-dependent refractive index $n \equiv n(\lambda)$ leads to chromatic aberrations in lenses with wavelength dependent focal length $f(\lambda)$. For thin lenses, this can be quantified by the Lens Maker's formula:

$$\frac{1}{f(\lambda)} = [n(\lambda) - 1]\left(\frac{1}{R_1} + \frac{1}{R_2}\right)$$

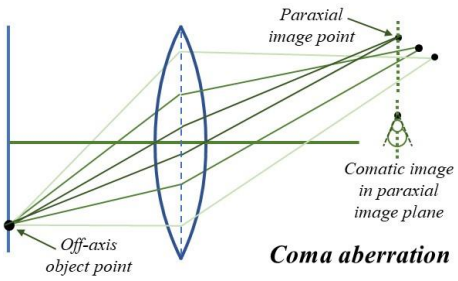



**Monochromatic Aberrations.** Even with single wavelength light, there are 5 primary aberrations: ***spherical aberration***; ***coma***; ***astigmatism***; ***field curvature***; and ***distortion***. We will consider (qualitatively) three of these:

**Spherical Aberration** occurs for on-axis object points when marginal object rays near the edge of the lens see excessive deviation and are imaged before the paraxial rays. This spreading of image point leads to image blurring.

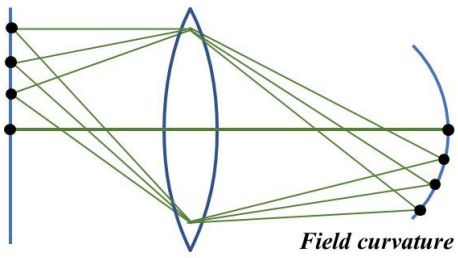



**Coma (off-axis) Aberration** as we saw for mirrors occurs for off-axis object points. Rays passing near the lens centre provide a "paraxial" image point but marginal rays near the edge of the lens are skew to the optical axis and imaged with increasing longitudinal and transverse displacement from the paraxial image point. The image in the paraxial plane has a comet-like appearance giving the aberration name **coma**.




(*A related aberration is **astigmatism** that gives different paraxial focal points in off-axis direction and orthogonal to this direction*).

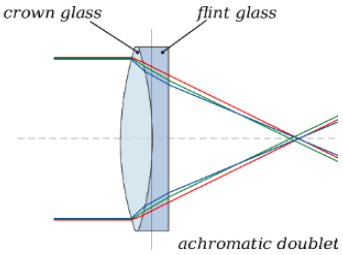
**Field Curvature Aberration** occurs when object points in a plane are mapped to image points on a curved surface. This is not surprising if you consider that off-axis object points are further from the centre of the lens and according to the thin lens formula their image distance should be closer to the lens. When using a planar detector (e.g.in camera), field curvature can lead to the edges of the image being out of focus.




(NB *Field curvature is an aberration associated with relative position of points, whereas spherical aberration and coma occur for individual points and known as point aberrations. A related aberration to field curvature is called **distortion***).

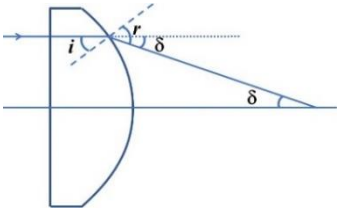
**Lens Aberration Correction.** Monochromatic and chromatic aberrations occur in optical systems (camera, microscope, telescope) and need to be minimised for good imaging.

**Chromatic Aberration** due to refractive index dispersion $n(\lambda)$ cannot be corrected with a single glass material. However, joining two glass materials, one with low index dispersion as a positive lens and another with high index dispersion as a negative lens, can create what is known as an achromatic doublet where $f(\lambda)$ is near constant over visible wavelengths. [*See Problem Sheet 2 for strategy and maths for this compensation*].




**Spherical Aberration** A strong spherical aberration occurs when paraxial approximation $sin\theta = \theta$ is not valid due to nonlinear deviation caused by Snell's Law based on the sine of angles $sin\theta = \theta - \theta^3/3$, due to the third-order $O(\theta^3)$ term even for parabolic lens surface.

Consider a plano-convex lens (in 1-D) with parabolic thickness: $t(y) = t(0) - \frac{y^2}{2(n-1)f}$, and convex surface angle: $\phi(y) = \frac{dt(y)}{dy} = -\frac{y}{(n-1)f}$. A ray parallel to optical axis at height y is incident at internal angle $i = -\phi(y)$ to convex surface and Snell's Law $nsini = sinr$ expanded to third-order: $n(i - i^3/3) = (r - r^3/3)$

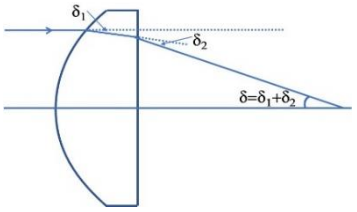



Deviation: $\delta = r - i = (n-1)i + \frac{1}{3}(r^3 - ni^3) \approx (n-1)i + \frac{1}{3}(n^3 - n)i^3 = \delta_0 + \Delta\delta$

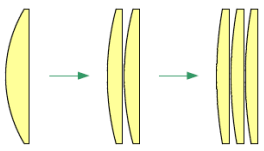
where paraxial deviation $\boldsymbol{\delta_0 = (n-1)i}$ and deviation error (aberration) $\boldsymbol{\Delta\delta = \frac{1}{3}(n^3 - n)i^3}$.

The deviation error: $\Delta\delta \propto i^3 \propto \delta_0^3$ leads to **spherical aberration** as excess deviation $\Delta\delta$ increases with marginal ray height *y* on lens.

One approach to reduce this spherical aberration is to minimise the ray angles $i$ on lens surfaces by distributing the ray deviation between the refracting lens surfaces. In our plano-convex case, by placing convex face first, incoming parallel rays are deviated at both lens surfaces. If we approximate equal split of deviation at each surface: $\delta_1 \approx \delta_2 \approx \frac{\delta_0}{2}$, then 2 surface aberration error: $\Delta\delta \propto 2.\left(\frac{\delta_0}{2}\right)^3 = \delta_0^3/4$ is 4 times less than single surface error. The following rules apply:



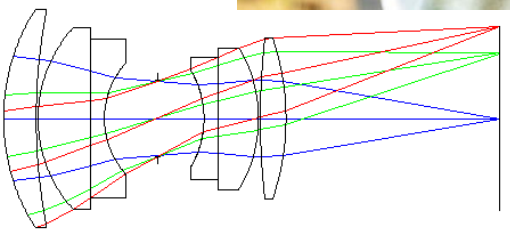

- For distant object (parallel incoming rays), it is best to place plano-convex with curved surface towards object
- However, if the object is near the focal point and image at infinity (parallel output rays) then placing lens flat face towards object will split the deviation best.
- For near equal distances object/image distances ($u \sim v$) a bi-convex lens is best split of deviations, and a meniscus lens is optimal for more general case.
- Even better than a single lens is to "bend" rays over a multi-lens system where more surfaces *n*, reduce error $\Delta\delta \propto n.(\delta_0/n)^3 = \delta_0^3/n^2$. This approach is used in high-end imaging systems .



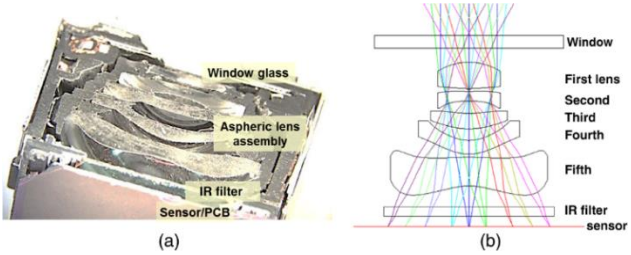
**Coma and field curvature aberration correction**: For wide angle imaging systems these need a multi-lens system for correction whilst also allowing "slow-ray bending". Field curvature can also be eliminated by placing a correction optics near to the detector with a refractive index profile to transform the curved image plane onto the flat detector array image plane (this is employed in smartphone cameras).

## REAL-WORLD OPTICAL SYSTEMS

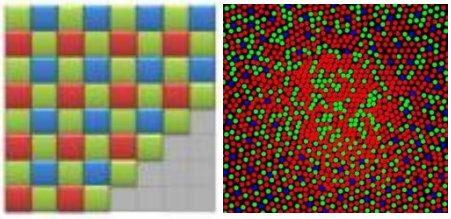
**CAMERAS.** A real-world **photographic camera** "lens" consists of many lens elements. This is needed for monochromatic aberration correction (spherical, coma, field curvature) and using different dispersion lens elements for chromatic aberration correction. In the ray diagram shown of a 6-element system, spherical aberration is removed by using "slow ray bending" over multiple surfaces. To minimise coma, there is an angular symmetry to the multi-lens design. This can be seen by the three sets of ray bundles from different angles (shown in different colours) that pass through a common centre point between the two sets of 3-lens elements. An aperture stop (pupil) is placed in this crossing point plane to pass a limited ray bundle centred on this common crossing point and block "unwanted" marginal rays that would degrade image quality. It is interesting to follow the path of the central ray of each of these three ray bundles – each of these go through the centre point and to first-order "see" a near-to straight path as if it is following a principal axis (although there is a small final bend noticeable in the widest-angle case to bring image point correctly to detector plane and eliminate field curvature, in this design). In effect, all object points experience a near "axial" path (there are no off-axis points) and coma aberration is eliminated.



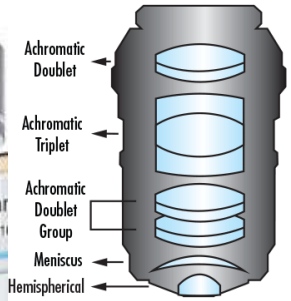
**Smartphone camera** also has a large number (e.g. 5 or 6) of lens elements. Amazingly, this whole multi-lens system and CMOS sensor needs to occupy a length of only ~ 5mm to sit within the thickness of the phone. A ray diagram of a 5-element design is shown (and a cut-through photo). You might notice the first 4 elements look and function a lot like the previous photographic lens example. The final (strange looking) refractive element next to the sensor provides field curvature correction to bring the image to match the flat plane of the sensor array. In smartphone, all lens elements are aspheric (non-spherical) for extra flexibility to achieve optical aberration correction over a wide angular range in such a short (few mm) distance. The lens elements are plastic materials to allow them to be moulded rather than using traditional glasses and using two plastic materials to provide chromatic aberration correction.

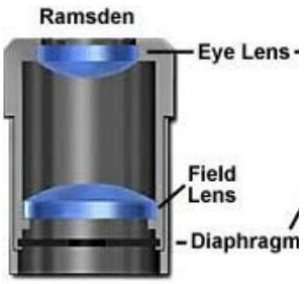



**CMOS Detector**. In both camera systems, there is focus adjustment to near and far points by relative movement of lens elements onto a light sensor, usually a CMOS detector array. The CMOS detector can see visible and near-IR light, so an IR filter is used to block the IR part of the spectrum. CMOS pixel size is now ~ 1μm (or even less) giving the high megapixel specification of modern cameras. To "see" colour, pixels are overlaid with a patterned array of red, green, blue colour filters to provide a colour sensing analogous to the way the eye's retina uses its three colour cones. Twice as many green pixels are used to imitate the eyes increased sensitivity to this spectral region.



**COMPOUND MICROSCOPE.** The microscope must resolve the smallest (micron scale) object details possible, limited only by wave-theory diffraction requiring wide-angle collection of the object light. To accomplish this, the objective "lens" is usually a very elaborate arrangement of lenses (see photo and diagram of high-end objective lens), providing full chromatic and monochromatic aberration correction. For highest magnification systems, the first element is a hemispherical lens that sits in near contact with the specimen slide (object) to increase the acceptance angle even towards the maximum possible (~ ± 90°). It can also operate with a liquid interface between lens and slide to increase image resolution that needs wave theory to explain. This objective lens provides the first real image for the compound microscope.


The eyepiece lens is also usually 2 or more lens elements (e.g. the Ramsden eyepiece) providing more comfortable viewing than the simple magnifier lens and viewer not needing to place eye too close to lens.

## DIFFRACTION-LIMITED IMAGING SYSTEMS and WAVE OPTICS

The aim of this somewhat lengthy section on aberration and correction strategies of imaging systems is to explain why real-world lenses are more complicated (multi-element systems) than a single thin lens that we learn about in basic optics within a paraxial approximation.

It also brings us to a key point of physics: geometrical optics shows that with paraxial operation or with excellent aberration correction in wider angle (non-paraxial) optical systems, we can, at least in principle, bring a point object to a perfect point image. But this is not correct! We have limited ourselves to a model of light based on rays. We know light is an **electromagnetic wave**, and waves experience diffraction and interference. In the **WAVE OPTICS** part of this course, we will re-examine imaging in a wave model and see that diffraction means that light can never be brought to a single point but will be a "point spread function" (PSF).

If an optical imaging system is corrected to eliminate all aberration, so that in a geometrical ray picture perfect imaging is achieved, the system is said to be **"diffraction-limited"**. In this case, the image point is entirely determined by diffraction to a fundamental limit of size determined by wave theory, for a given "aperture" of the imaging system. We will discover that larger aperture systems can make the image point smaller and give higher resolution. However, if the optical system has aberrations, we have already seen that there is image spreading. Such an aberrated system is not diffraction-limited, and in wave optics the point spread function is larger than its so-called diffraction-limited size. We will deal with wave imaging using Fourier Optics, a sub-topic of diffraction theory that uses the mathematics of Fourier Transforms.

# RAY OPTICS – SUMMARY FORMULA SHEET

| | |
|---|---|
| **Snell's Law:** | $n_1 sin\theta_1 = n_2 sin\theta_2$ |
| **Reflection Law** | $\theta_1 = \theta_r$ |
| **Critical Angle:** $n_1 > n_2$ | $\theta_c = sin^{-1}\left({}^{n_2}/_{n_1}\right)$ |
| **TIR** | $\theta_1 > \theta_c$ (45° prism; apparent depth; optical fibre) |
| **Deviation** ($\delta$) | dispersion $\delta(\lambda)$ in prism; rainbow; lens |
| **Fermat's Principle:** | $\delta[OPL] = 0$; stationary ray paths |
| **Lens shape:** | Parabolic: $\Delta t(r) = r^2/2(n-1)f$ (small angles) |
| **Spherical lens shape:** | $\Delta z(r) = r^2/2R$ (small angles) |
| **Lens Maker's Formula** | $\frac{1}{f} = (n-1)\left(\frac{1}{R_1} + \frac{1}{R_2}\right)$ (Spherical surfaces) |
| **Mirror surface:** | Parabolic: $\Delta z(r) = r^2/4f$ (all angles – axial point) |
| **Spherical mirror:** | $f = \frac{R}{2}$ |
| **Thin lens formula** | $\frac{1}{u} + \frac{1}{v} = \frac{1}{f}$ (& lens sign convention) |
| **Mirror formula** | $\frac{1}{u} + \frac{1}{v} = \frac{1}{f} = \frac{2}{R}$ (& mirror sign convention) |
| **Magnification formula** | $m_T = \frac{-v}{u}$ |
| **Ray diagrams** | Principal rays to locate image (lens/mirror cases) |
| **Optical power of lens** | $D = \frac{1}{f}$ dioptres ($m^{-1}$) |
| **Combining lens powers** | $\frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2}$; $D = D_1 + D_2$ |
| **Imaging Systems:** | |
| **Eye lensing power** | $D_{eye} = D_0 + \Delta D$; $D_0(60dpt) \approx D_C(40dpt) + D_L(20dpt)$ |
| **Far & near point** (eye) | Infinity to 250mm ($\Delta D = 0 - 4dpt$) |
| **Angular size** | $\alpha = \frac{h_0}{d}$ |
| **Max distinct angular size** | $\alpha_{250} = \frac{h_0(mm)}{250}$ |
| **Simple Magnifier:** | $M_\infty = \frac{250}{f(mm)}$ |
| **Compound microscope** | $M_\infty = -\left(\frac{l}{f_O}\right)\left(\frac{250}{f_E(mm)}\right)$ |
| **Refracting telescope** | $M_\alpha = -\frac{f_O}{f_E}$; $d = f_O + f_E$ (Kepler); $d = f_O - \lvert f_E \rvert$ (Galilean) |
| **Reflecting telescope** | $M_\alpha = \frac{f_1}{\lvert f_2 \rvert}$; $d = f_1 - \lvert f_2 \rvert$ (Cassegrain) |
| **Aberration** | Chromatic; spherical; coma; field curvature; Correction strategies. |

# WAVE OPTICS

In **Ray Optics**, we considered light rays travelling in straight or stationary paths in a relatively unapertured region of the optical system. In **Wave Optics**, we consider the wave nature of light, and in particular the phenomena of diffraction and interference.

**Diffraction** occurs when there are limiting apertures and light spreads to new directions. We will use Huygens-Fresnel Principle (secondary spherical waves) to mathematically model diffraction and explore the fundamental wave limits of **imaging** that go beyond the ray theory we previously discussed. **Interference** is an over-arching feature of (light) waves based on the Principle of Superposition requiring knowledge of the **amplitude** and **phase** of the waves. **Interferometers** can provide highly accurate measurement of phase which we will overview. For non-monochromatic light sources with erratic phases in time (and space) we will explore the more complex topic of **coherence** and its impact on interference.

## Representation of Light as EM waves

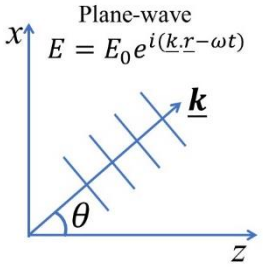
We describe light waves in terms of electric field *E,* and in this course use complex notation.

**Plane-wave:** $$E(\underline{\boldsymbol{r}},t) = A_0 e^{i(\underline{\boldsymbol{k}}.\underline{\boldsymbol{r}}-\omega t+\emptyset)} = E_0 e^{i(\underline{\boldsymbol{k}}.\underline{\boldsymbol{r}}-\omega t)}$$

where $E_0 = A_0 e^{i\emptyset}$ is a *complex* field amplitude.

(*Strictly, the real part* $\mathbb{R}\{E_0 e^{i(kz-\omega t)}\} = A_0 cos(kz-\omega t+\phi)$ *gives the real physical field, but it is more convenient to use the complex notation and drop the Real Part symbol*)

**Wavevector:** $$\underline{\boldsymbol{k}} = (k_x, k_y, k_z)$$




describes a general wave direction with wavevector magnitude $|\underline{\boldsymbol{k}}| = k = 2\pi/\lambda$ with $\lambda$ free-space wavelength and $k_i$ are the component in *x, y, z* directions. The wavevector is an important quantity when we consider diffraction with light travelling in different directions. For example, a wave travelling at angle $\theta$ to *z*-axis in *x*-*z* plane has $\underline{\boldsymbol{k}} = (k_x = ksin\theta, 0, k_z = kcos\theta)$

For a plane wave travelling in z-direction, $\underline{\boldsymbol{k}} = (0,0,k_z = k)$, we write:

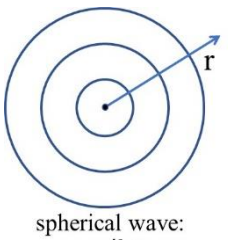
$$E(z,t) = E_0 e^{i(kz-\omega t)}$$

**Intensity:** $I = \frac{1}{2} c\varepsilon_0 |E_0|^2$ (in SI units) is time-averaged over an optical cycle. In this course, we are interested in light distribution rather than absolute value, so for simplicity just take:

$$I = |E_0|^2$$

**Spherical Waves**: are important waveforms in optics and have mathematical form:

$$E(r,t) = \frac{A_0}{r} e^{i(kr-\omega t)} = E(r) e^{i(kr-\omega t)}$$



spherical wave:

where $r = \sqrt{x^2+y^2+z^2}$ is the radial distance from the source. The intensity $I = |E(r)|^2 = \frac{A_0^2}{r^2} = P/4\pi r^2$ is the "inverse square law" to maintain constant power $(P)$ as the surface area $4\pi r^2$ of spherical wave expands, and hence the $1/r$ term in the field amplitude.

Note how using complex notation we can easily separate the space and time parts: $E(r,t) = \frac{A_0}{r} e^{ikr} e^{-i\omega t}$ and drop time part if that is convenient.

**Interference** is based on the **Principle of Superposition (PoS)**:

> "*when two or more (light) waves overlap, the resultant displacement (electric field **E** for light) is equal to the (vector) sum of the displacements due to each individual wave.*"

Mathematically, the resultant field given by PoS at a point *P* in space: $E_P = \sum_i E_i$ is the sum of individual fields $E_i(r,t) = E_{0i}e^{i(kr_i - \omega t)}$ where $r_i$ is a distance coordinate and assuming monochromatic light with all waves at same angular frequency $\omega$.

The phase change of a wave $\Delta\phi = k(r_2 - r_1) = k\Delta r$, or more generally, $\Delta\phi = kn\Delta r = k(OPL)$ in medium with refractive index *n*. Two waves, initially in phase, will constructive interference at other points in space when their path length difference satisfies: $\Delta\phi = k\Delta r = \frac{2\pi}{\lambda}\Delta r = m2\pi$, where *m* is integer, and hence when:

$$\Delta r = m\lambda$$

## DIFFRACTION

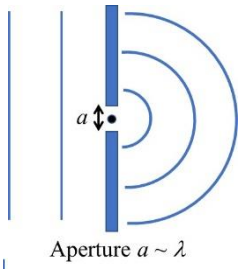
Diffraction occurs when light meets a limiting aperture and spreads to new directions.

**Huygens-Fresnel Principle** (HFP) is the key basis for describing diffraction using a modern version of Huygens geometrical use of secondary spherical waves but including the maths of their amplitude and phase (as done by Fresnel) and using **Principle of Superposition** (PoS).

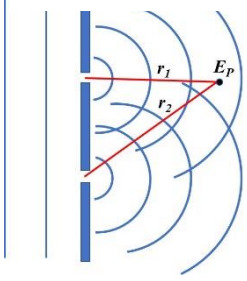
**HUYGENS-FRESNEL PRINCIPLE (HFP)** can be stated as:
"*Every unobstructed point of a wavefront serves as a source of spherical secondary waves. The amplitude of the optical field at any point beyond is the superposition of these waves taking account of their amplitudes and phases.*"

Let's illustrate a few cases qualitatively, then proceed to the maths:

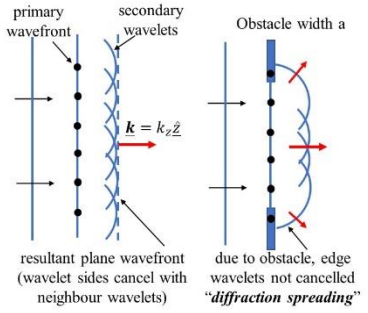
- A single narrow aperture (width $a \sim \lambda$) is considered as a single elemental point in the HFP and diffraction pattern is a single spherical wave spreading beyond aperture: $E_P = \frac{A}{r}e^{i(kr-\omega t)}$.
  [*Note: we only consider forward (hemi-)spherical wave – the neglect of backward component is explained by fuller EM theory*]




- For 2-narrow apertures, the diffraction pattern is the sum of 2 spherical waves: $E_P = E_1 + E_2$. *Constructive* interference occurs at points P where their relative path difference $\Delta r = r_2 - r_1 = m\lambda$ (as they have the same starting phase for normal-incidence plane-wave illumination of the apertures).
  [*Two-slit diffraction is sometimes termed two-slit interference*]

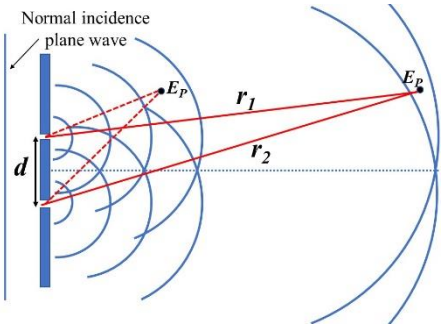



- An unapertured plane wave can be considered an infinite number of spherical waves. Neighbouring waves cancel sideways, giving a net new plane wavefront in advance of the first, maintaining a single wavevector direction $\underline{\boldsymbol{k}} = k_z\underline{\hat{z}}$.
- An extended aperture produces a diffraction pattern that is a continuum of waves over unapertured region *A* at obstacle. Its sum is an integral: $E_P = \int_A E'(x)dx$. "Edge" secondary wavelets don't cancel sideways due to missing neighbours leading to wave spreading into the "shadow" of the aperture creating new wavevector directions.



## HFP applied to Diffraction from Two Narrow-Slits

Two narrow slits (width $a \sim \lambda$), separated by distance $d$, are illuminated by normal incidence plane wave light (angular frequency $\omega$). According to HFP, each slit acts as a spherical wave (point source). Let's consider the maths to find the diffraction pattern for this case:




1. The diffracted field a point P after the 2-slits $E_P$ is given by sum of spherical waves:

$$E_P = E_1 + E_2 = \frac{A}{r_1}e^{i(kr_1-\omega t)} + \frac{A}{r_2}e^{i(A/r_i-\omega t)}$$
$$= \left(\frac{A}{r_1}e^{ikr_1} + \frac{A}{r_2}e^{ikr_2}\right)e^{-i\omega t}$$

   where $r_i$ are the distances from each slit to point *P*. A few comments can be made:

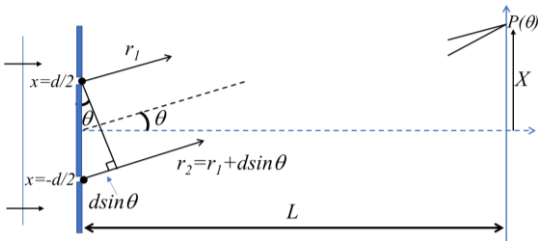
   i) We draw straight lines paths – this is not ray optics – but show distances $r_i$ that we need to calculate spherical wave amplitude $A/r_i$ and phase $kr_i$.
   ii) Both waves start with a common phase due to plane-wave illumination.
   iii) We can drop $e^{-i\omega t}$ term - it will disappear when taking intensity $I_P = |E_P|^2$.

2. Near aperture plane ($L \sim d$), spherical wave amplitudes $A/r_i$ can be very different, and the phase difference $k(r_2 - r_1) = k\Delta r$ varies in a complex way. Further away ($L \gg d$) we can take $r_2 \approx r_1$ and assume equal spherical wave amplitude: $\frac{A}{r_1} \approx \frac{A}{r_2} = E_0$. and write:

$$E_P = E_0 e^{ikr_1}\left(1 + e^{ik\Delta r}\right)$$

   where $\Delta r = r_2 - r_1$ is the path length difference, which while small compared to other distances ($\Delta r \ll L, d$) must be considered for interference on wavelength scales.

3. At sufficiently far distance (*we will analyse what this is later*), ray paths to point *P* at angular coordinate $\theta$ can be approximated as parallel. The triangle in the figure, shows path length difference is simply: $\boldsymbol{\Delta r = dsin\theta}$




4. The diffraction pattern is given by:

$$E_P = E_0 e^{ikr_1}\left(1 + e^{ikdsin\theta}\right)$$

5. Noting: $\left(1 + e^{ikdsin\theta}\right) = 2e^{ikdsin\theta/2}cos(kdsin\theta/2)$ [since $cosx = \frac{1}{2}\left(e^{ix} + e^{-ix}\right)$], we can write diffraction pattern as:

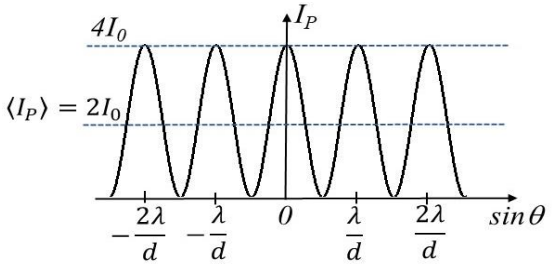
$$E_P(\theta) = 2E_0 e^{ikr_0}cos(kdsin\theta/2)$$

   where $r_0 = r_1 + dsin\theta/2$ is the path distance measured from the centre of the 2 slits.

6. Intensity far-field diffraction pattern $I_p = \left|E_P{}^2\right|$ is given by

$$I_P = 4I_0 \cos^2\left(k\frac{d}{2}sin\theta\right)$$

   where $I_0 = |E_0|^2$ is intensity of a single slit.




   i) *Note: $I_{max} = 4I_0$; $I_{min} = 0$; but average $\langle I \rangle = 2I_0$. Power is conserved on average, but redistributed across diffraction pattern.*

7. Maxima occur at angles when: $k\frac{d}{2}sin\theta = \frac{2\pi}{\lambda}\frac{d}{2}sin\theta = m\pi$

$$\boxed{dsin\theta_m = m\lambda}$$

and maxima positions at $sin\theta_m = m\lambda/d$ have equal spacing $\lambda/d$ in terms of $sin\theta$, the $\theta_m$ positions are equal only for small angles.

8. Minima occur at intermediate angles: $dsin\theta_m = \left(m + \frac{1}{2}\right)\lambda$

9. The distribution is an angular one, but for small angles $sin\theta \approx \theta \approx tan\theta \approx X/L$, we get pattern in screen $X$ coordinates at axial screen distance $L$:

$$\boxed{I_P = 4I_0 \cos^2\left(\frac{\pi dX}{\lambda L}\right)}$$

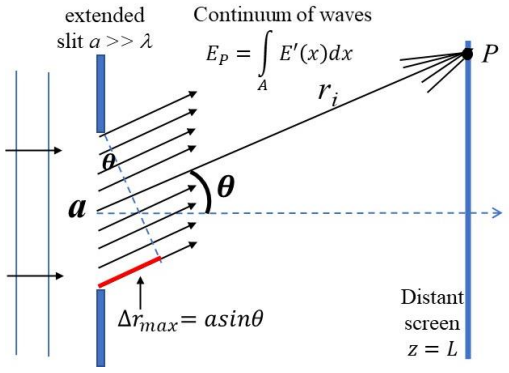
with maxima at: $X_m = m\frac{\lambda L}{d}$ $(m = 0, \pm1, \pm2, \ldots)$; and equal spacing: $\Delta X = \frac{\lambda L}{d}$

As an example: $\lambda$ = 500 nm; $L$ = 1m; $d$=0.1mm: $\Delta X = \frac{(5\ 10^{-7})(1)}{(10^{-4})} = 5\ mm.$

## Near & Far-field Diffraction

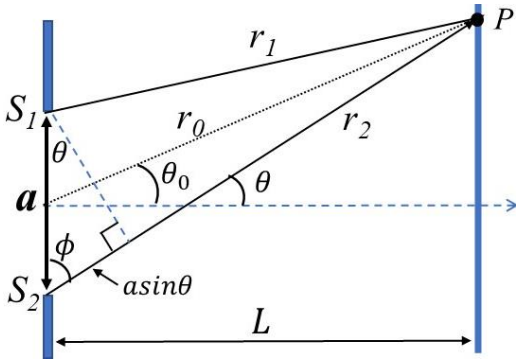
We can apply HFP to more complicated apertures such as an extended slit with width $a$. In this case, we must consider a continuum of elemental points $dx$ at position $x$ extending across slit $|x| \leq \frac{a}{2}$ each contributing a spherical wave. The resultant diffraction field $E_P$ (dropping common term $e^{-i\omega t}$) is:

$$E_p = \int_{-a/2}^{a/2} \frac{e^{ikr}}{r} dx$$




We can consider a distant screen at $z = L$ and again approximate parallel rays from all elemental aperture points to screen point $P$ at angular coordinate $\theta$. We will do this in the next Hand-Out to get the extended slit "far-field" diffraction pattern. But first we want to ask a more general question: How far away is the "far-field" when we can consider a parallel ray approximation from aperture points to the observation screen?

The diagram shows the two extreme ray paths from the edge points $S_1$ and $S_2$ of the aperture (width $a$) to point $P$ in the observation plane. A perpendicular line dropped down from $S_1$ to ray path 2 creating a triangle with base length $asin\theta$. Previously, we noted that when the two ray paths are parallel then $\Delta r = r_2 - r_1 = asin\theta$, but this is not true at finite distance $L$ to observation screen. We can get the more general relationship for the 2 ray paths by using the Law of Cosines for the triangle $S_1S_2P$:




$$r_1^2 = r_2^2 + a^2 - 2ar_2 cos\phi$$

Since $cos\phi = sin\theta$, $r_2^2 - r_1^2 = 2ar_2 sin\theta - a^2$, and noting $r_2^2 - r_1^2 = (r_2 - r_1)(r_2 + r_1)$:

$$\Delta r = \frac{2r_2}{(r_2 + r_1)} asin\theta - \frac{a^2}{(r_2 + r_1)}$$

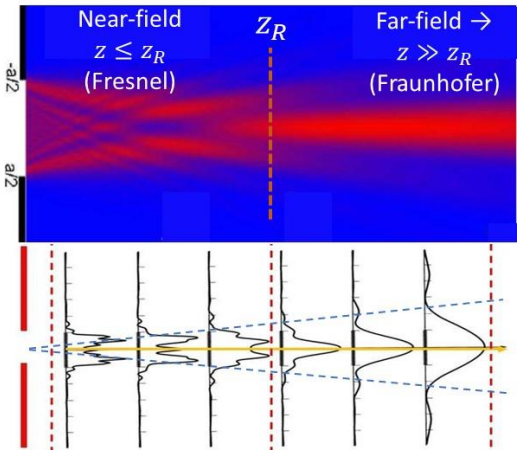
This expression is exact so far, but since we are normally considering small aperture size compared to screen distance $a \ll L$, then $r_1 \approx r_2 \approx r_0$, giving

$$\Delta r = asin\theta - a^2/2r_0$$

For parallel ray case $\Delta r = asin\theta$ we need to neglect the final term, and its contribution to net interference, meaning it must be smaller than wavelength scale: $a^2/2r_0 \ll \lambda$. If we take $r_0 = L$ (axial screen distance) as characteristic distance of importance, we find far-field condition:

$$\boxed{L \gg z_R = a^2/2\lambda}$$

where the characteristic distance $z_R = a^2/2\lambda$ is known as the **Rayleigh distance**.




- When $L \gg z_R$ we have **far-field diffraction** (also known as Fraunhofer diffraction)
- Conversely, when $L \leq z_R$, we have **near-field diffraction** (also known as Fresnel diffraction).
- Note that Rayleigh distance $z_R$ varies as the square of aperture size, and large apertures can have a very distant far-field. As an example: if $a = 1mm$, $\lambda = 500nm$, then $z_R = 1m$; (*so for aperture few times larger, it would be hard to achieve reach the far-field inside a lab/room*)
- The near-field pattern can be quite complex, changing rapidly in form with distance $L$ from aperture due to the extra term $-a^2/2r_0$.
- At long distances $L \gg z_R$, the far-field angular pattern $E_p(\theta)$ does not change form as relative ray path lengths $r_2 - r_1 = asin\theta$ are independent of distance $L$.

## Generalised Far-field Diffraction Integral

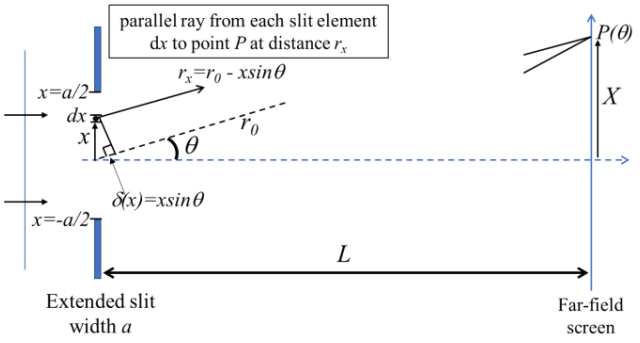
Let's return to the integral for the single extended slit diffraction pattern:

$$E_p = \int_{-a/2}^{a/2} \frac{e^{ikr_x}}{r_x} dx.$$

We can define a parallel ray distance to point P:

$$\boxed{r_x = r_o - xsin\theta}$$

from aperture element at position $x$ with respect to a ray distance $r_o$ from the centre of the slit ($x = 0$) to obtain the far-field diffraction integral:




$$E_p(\theta) = \frac{e^{ikr_o}}{r_0} \int_{-a/2}^{a/2} e^{-ikxsin\theta} dx$$

The integral gives the angular form of the diffraction pattern $E_p(\theta)$. The pre-multiplier $e^{ikr_o}/r_0$ corresponds to an expanding spherical wave, showing that the angular form of the diffraction pattern should be considered as expanding on a spherical (wave) surface – rather than on a flat observation plane at axial distance $L = r_0$.

We can extend this analysis to a more general aperture with a transmission function $A(x)$, which in the most general case may not just be binary (1 or 0, as the case of a slit) but may also vary continuously in amplitude (amplitude mask) or even in phase (phase mask) or both. By using transmission aperture function $A(x)$, we can now extend the HFP integral limits to $\pm\infty$:

$$E_p(\theta) = C(L)\int_{-\infty}^{\infty} A(x)e^{-ikxsin\theta}dx$$

where $C(L) = \frac{e^{ikL}}{L}$ is the pre-factor of the expanding pattern at axial distance $L$.

**For small angles of diffraction**, we can take the same paraxial approximation that we made in Ray Optics: $sin\theta \approx \theta \approx tan\theta = X/L$, where $X$ is the coordinate at the screen at distance $L$, and write the *paraxial* far-field diffraction integral in screen spatial coordinate $X$:

$$E(X) = C(L)\int_{-\infty}^{\infty} A(x)e^{-ikXx/L}dx$$

## Far-field Diffraction Integral is a Fourier Transform of Aperture $A(x)$

As we are often interested just in the form of the diffraction pattern, we can often choose to drop the pre-factor and write:

$$E(\theta) = \int_{-\infty}^{\infty} A(x)e^{-ikxsin\theta}dx = \int_{-\infty}^{\infty} A(x)e^{-ik_x x}dx$$

where we simplify the exponential argument by introducing the term $k_x = ksin\theta$, and we note that this quantity is the transverse (x-component) of the wavevector of the light field.

We arrive at a key point: the far-field diffraction pattern has the form of a Fourier Transform of the aperture transmission function $A(x)$:

$$\boldsymbol{E(k_x) = \int_{-\infty}^{\infty} A(x)e^{-ik_x x}dx = \mathcal{F}[A(x)]}$$

In equivalence to the Fourier Transform $\boldsymbol{E(\omega) = \int_{-\infty}^{\infty} E(t)e^{-i\omega t}dx = \mathcal{F}[E(t)]}$ between a function in time ($\boldsymbol{t}$) and its decomposition in terms of angular frequency ($\boldsymbol{\omega}$), the far-field diffraction integral is a spatial Fourier Transform between space ($\boldsymbol{x}$) into its angular spatial frequency ($\boldsymbol{k_x}$) components. We can understand the physical meaning of angular spatial frequency $k_x = ksin\theta$ as being the transverse wavevector of a light wave with wavevector magnitude $k$, and propagating at a diffraction angle $\theta$. This relationship between the spatial form of the light and its Fourier Transform is one we will return to later – and a branch of optics known as Fourier Optics.

# Far-field diffraction patterns

We shall start from the far-field diffraction integral we generalised with aperture transmission function $A(x)$:

$$E_p(\theta) = C\int_{-\infty}^{\infty} A(x)e^{-ikxsin\theta}dx$$

and noting:

i) $A(x)$ weights the strength of spherical waves (from HFP) across points on incident plane-wave with integral limits to $\pm\infty$;
ii) observing in far-field $L, r_0 \gg z_R = a^2/2\lambda$ ($a$ is characteristic aperture size)
iii) and where $C = E_0 e^{ikr_0}$ is an expanding spherical wave with amplitude $E_0 = A/r_o$

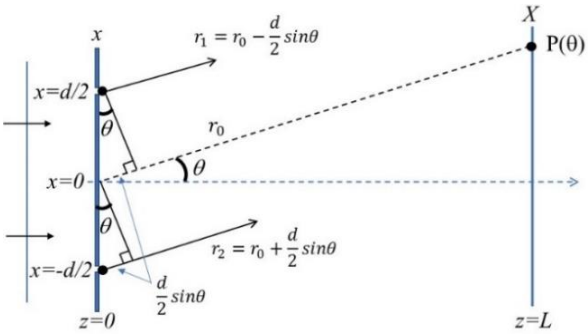
## 2-narrow slit diffraction pattern

Let's consider again the case of two narrow slits separated by distance *d*.

What is the aperture function, $A(x)$?

The pair of slits can be created by an aperture function composed from a pair of Dirac delta functions. For the slits positioned symmetrically at $x = \pm d/2$:

$$\boldsymbol{A(x) = \delta(x - d/2) + \delta(x + d/2)}$$




The diffraction integral becomes:

$$E_p(\theta) = C\int_{-\infty}^{\infty} [\delta(x - d/2) + \delta(x + d/2)]e^{-ikxsin\theta}dx$$

The Dirac delta functions when inside an integral have the property of selecting discrete values of the integrand when arguments of delta function go to zero at $x = \pm d/2$. Hence

$$E_p(\theta) = C\left[e^{-ik\frac{d}{2}sin\theta} + e^{+ik\frac{d}{2}sin\theta}\right]$$

and far-field 2-slit (field) diffraction pattern can be written as:

$$E_P(\theta) = 2Ccos(kdsin\theta/2)$$

Intensity far-field diffraction pattern $I_p = |E_P{}^2|$ is given by

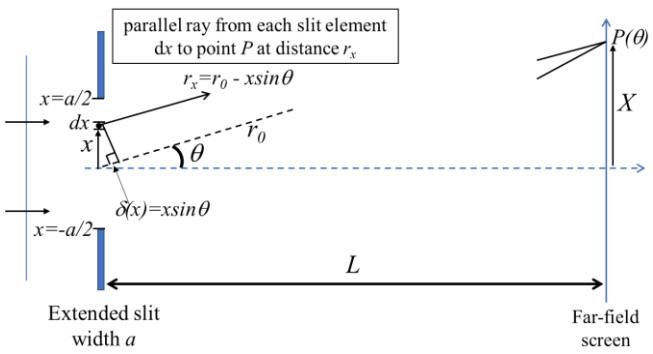
$$I_P(\theta) = 4I_0\cos^2(kdsin\theta/2)$$

where $I_0 = |C|^2 = |E_0|^2$ is intensity due to a single slit (spherical wave) as we showed in last lecture. The 2-slit diffraction pattern is a cosine function in (field) amplitude and cosine squared function in intensity with maxima at: $dsin\theta_m = m\lambda.$

At small angles $sin\theta \approx \theta \approx tan\theta \approx X/L$: maxima are at $X_m = m\frac{\lambda L}{d}.$

## Extended Single Slit Diffraction

An extended slit of width, $a$, extending from $-a/2 \leq x \leq a/2$, has a stepped transmission function we will denote as:
$\boldsymbol{A(x) = rect(x/a)}$




where

$$A(x) = rect(x/a) = \begin{cases} 1, & |x| \leq a/2 \\ 0, & x > a/2 \end{cases}$$

Our generalised far-field diffraction pattern is:

$$E_p(\theta) = C \int_{-\infty}^{\infty} rect(x/a).e^{-ikxsin\theta}\,dx$$

which becomes:

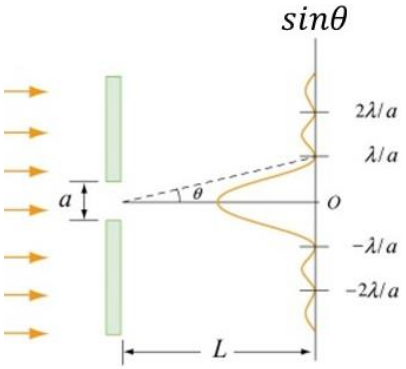
$$E_p(\theta) = C \int_{-a/2}^{a/2} e^{-ikxsin\theta}\,dx$$

Solving:

$$E_P(\theta) = \left[\frac{e^{-iksin\theta x}}{-iksin\theta}\right]_{-a/2}^{a/2} = \frac{e^{-\frac{iksin\theta a}{2}} - e^{+\frac{iksin\theta a}{2}}}{-iksin\theta}$$

$$E_P(\theta) = a\frac{\sin(kasin\theta/2)}{(kasin\theta/2)} = a\,sinc(kasin\theta/2)$$

where we used $sinx = \left(e^{ix} - e^{-ix}\right)/2i$; and $sinc(x) = (sinx)/x$ (and noting $\lim_{x\to 0}(sinx)/x = 1$).




- Single slit intensity pattern: $I_P(\theta) = a^2 sinc^2(\frac{kasin\theta}{2})$
- Central peak field $E_P(0) = a$ increases as slit width $a$ increases (more input power)
- **Minima** at *sine* zeroes: $kasin\theta/2 = \pi asin\theta/\lambda = m\pi$; $m = \pm 1, \pm 2$ ... (but not $m = 0$, as denominator also zero, giving central maximum).
- First zero is at: $sin\theta_1 = \frac{\lambda}{a}$. It is scale of pattern and inversely proportional to slit width $a$. (*Remembering far-field diffraction integral is a Fourier Transform, this reciprocal relationship is a general feature of function $A(x)$ and its Fourier Transform*)
- As central peak scales as $a$ and its width scales as $1/a$, area of field pattern is conserved. As $a \to \infty$, the diffraction pattern $E_P(\theta) \to \delta(\theta)$ (Dirac delta function). This does not mean a small point in space, but light is a unidirectional (plane-wave) in angle $\theta = 0$, travelling in $z$-direction.
- For small angles (and finite slit widths), we can convert to screen coordinates: $sin\theta \approx \theta \approx X/L$: and first minimum is at $X_1 = \lambda L/a$.

## Cosine Aperture Function: $A(x) = \cos(2\pi x/d)$

The aperture transmission function $A(x) = |A(x)|e^{i\phi(x)}$ can also in general vary continuously (amplitude intermediate to 0 or 1) and include phase (e.g. different glass thickness).

An important example is a cosine aperture function $A(x) = \cos(2\pi x/d)$. Our generalised diffraction integral (and neglecting pre-factor C) becomes:

$$E_P(\theta) = \int_{-\infty}^{\infty} \cos(2\pi x/d) e^{-iksin\theta x} dx$$

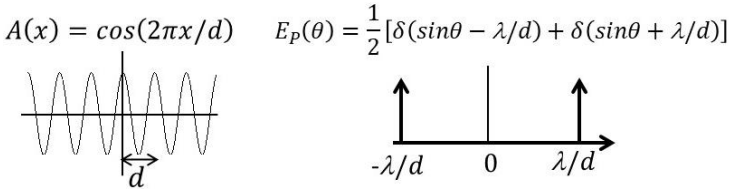
We can solve this by writing: $\cos(2\pi x/d) = \frac{1}{2}\left(e^{i2\pi x/d} + e^{-i2\pi x/d}\right)$

$$E_P(\theta) = \frac{1}{2}\int_{-\infty}^{\infty} \left(e^{i2\pi x/d} + e^{-i2\pi x/d}\right) e^{-iksin\theta x} dx$$

$$E_P(\theta) = \frac{1}{2}\int_{-\infty}^{\infty} \left[e^{-i\left(ksin\theta - \frac{2\pi}{d}\right)x} + e^{-i\left(ksin\theta + \frac{2\pi x}{d}\right)x}\right] dx = \frac{1}{2}[\delta(a_+x) + \delta(a_-x)]$$

where $\delta(a_\pm x)$ are Dirac delta functions with $a_\pm = ksin\theta \mp \frac{2\pi}{d}$. The diffraction pattern is a discrete pair of diffracted plane waves (when delta function arguments go to zero) at angles given by:

$$\boxed{sin\theta_\pm = \pm\frac{2\pi}{kd} = \pm\frac{\lambda}{d}}$$




And, again for small angles $sin\theta \approx \theta \approx X/L$: $\boxed{X_m = \pm(\lambda L/d)}$.

The cosine "aperture" may seem artificial, but it turns out to be very important for analysing periodic diffraction gratings. For instance, an even periodic function (with period *d*) can be Fourier decomposed as a harmonic series of cosines:

$$A(x) = a_0 + \sum_{m=1}^{\infty} a_m cos\left(m\frac{2\pi x}{d}\right)$$

Optical elements with periodic transmission (or reflection) are known as diffraction gratings. From Fourier decomposition, we see we don't need to manufacture a cosine element itself, but mathematically we can analyse diffraction from periodic aperture structures from what we have learnt about the cosine function.

By extension, all functions, even when not period can be analysed in terms of cosine Fourier decomposition – and this is the basis of wave imaging theory.

# FOURIER OPTICS

## Far-field Diffraction as a Fourier Transform

A branch of Optics called **Fourier Optics** allows the deployment of Fourier mathematics to reveal insight and the form of diffraction patterns. More importantly, it provides the basis for formulating the propagation of light and the wave theory of imaging. In far-field, the diffraction integral has the form of the Fourier transform of the aperture function $A(x)$:

$$E(k_x) = \int_{-\infty}^{\infty} A(x)e^{-ik_x x}dx = \mathcal{F}[A(x)]$$

*[We will neglect any pre-multipliers ($1/\sqrt{2\pi}$) on Fourier Transform or Inverse Transform Integrals as we are mainly interested only in the form of the diffraction].*

**Space** coordinate $\boldsymbol{x}$ (m) of aperture $A(x)$ and **angular spatial frequency $\boldsymbol{k_x = ksin\theta}$** ($m^{-1}$) in diffraction pattern are Fourier Transform variables. The space Fourier Transform is analogous to the time Fourier Transform: $E(\omega) = \int_{-\infty}^{\infty} E(t)e^{-i\omega t}dt = \mathcal{F}[E(t)]$ where **time $\boldsymbol{t}$** (s) and **angular frequency $\boldsymbol{\omega}$** ($s^{-1}$) are the Fourier Transform variables.
[NOTE: *The choice of sign convention of the exponential doesn't change the form of the Fourier Transform only inverting the coordinate system*].

In diffraction, the term angular spatial frequency $k_x = ksin\theta = \frac{2\pi}{\lambda}sin\theta$ is used and is analogous to the term angular frequency ω. It is convenient to use angular frequency ($\omega$) rather than frequency $\nu = \omega/2\pi$, to not need to write the $2\pi$ factor and similarly, it is also convenient in diffraction to use angular spatial frequency $k_x$. In imaging theory, optical device response is sometimes also expressed in terms of **spatial frequency $\boldsymbol{u_x}$** where $u_x = \frac{k_x}{2\pi} = \frac{sin\theta}{\lambda}$ and $E(u_x) = \int_{-\infty}^{\infty} A(x)e^{-i2\pi u_x x/\lambda}dx$.

To interpret from Fourier maths back to the physical diffraction pattern, the angular coordinate $\theta$ is embodied in the $sin\theta$ part of $k_x$. For small angles, we can also use spatial coordinate $X$ on far-field screen at distance $L$, noting $k_x = ksin\theta \approx k\theta = kX/L = 2\pi X/\lambda L$.

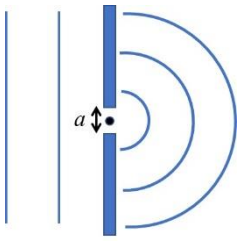
The table below shows some Fourier Transforms of importance in Optics:

### Table of some important 1-D functions $A(x)$ and their Fourier Transforms in Optics:

| Diffraction system | Aperture function: $A(x)$ | Diffraction pattern: $\mathcal{F}[A(x)]$ |
|---|---|---|
| Narrow slit (at $x = 0$) | $\delta(x)$ | 1 |
| Narrow slit (at $x = d/2$) | $\delta(x - d/2)$ | $e^{-ik_x d/2}$ |
| Two-narrow slits (separation, $d$) | $\delta(x + d/2) + \delta(x - d/2)$ | $2\cos(k_x d/2)$ |
| Cosine aperture | $cos(2\pi x/d)$ | $\frac{1}{2}\left[\delta\left(k_x + \frac{2\pi}{d}\right) + \delta\left(k_x - \frac{2\pi}{d}\right)\right]$ |
| Extended single slit (width $a$): $rect(x/a)$ | $A(x) = \begin{cases} 1, & \lvert x\rvert \le a/2 \\ 0, & x > a/2 \end{cases}$ | $a\, sinc(k_x a/2)$ |
| Diffraction grating (Dirac $\delta$-comb) | $\sum_{n=-\infty}^{\infty} \delta(x - nd)$ | $\sum_{m=-\infty}^{\infty} \delta(k_x - \frac{2\pi m}{d})$ |
| Gaussian (laser beam) | $e^{-(x/w)^2}$ | $\sqrt{\pi}w\, e^{-k_x{}^2 w^2/4}$ |

We have met some of these apertures, but let's explore them in Fourier transform maths.

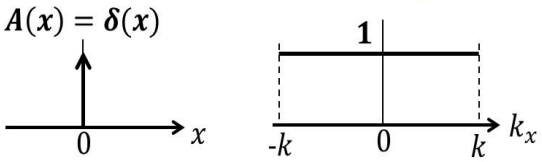
**Single narrow slit** (positioned at $x = 0$) acts as a source for a single spherical wave with equal amount of light diffracted in all directions ($\theta$) and hence equally across all angular spatial frequencies $k_x = ksin\theta$.




Its aperture function $A(x) = \delta(x)$ and its Fourier Transform

$$E(k_x) = \int_{-\infty}^{\infty} \delta(x) e^{-ik_x x} dx = 1$$

$E(k_x)$ in maths is unity at all $k_x$ to infinity but is a physically limited in optics: $k_x = \pm k$, at limits in diffraction angles $\theta = \pm 90°$.

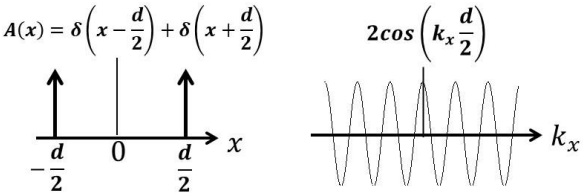



Fourier Transform theorems can explain more complex apertures. Some examples are below:

- **Shift Theorem**: $\mathcal{F}[f(x - x_0)] = \mathcal{F}[f(x)]e^{-ik_x x_0}$

  Shifting slit position: $\mathcal{F}\left[\delta\left(x - \frac{d}{2}\right)\right] = \mathcal{F}[\delta(x)]e^{-ik_x d/2} = e^{-ik_x d/2}$

  **2-slit diffraction**: $\mathcal{F}\left[\delta\left(x + \frac{d}{2}\right) + \delta\left(x - \frac{d}{2}\right)\right] = e^{-ik_x d/2} + e^{ik_x d/2} = 2\cos(k_x d/2)$




- **Fourier Inversion Theorem (Fourier Pairs)**:

  Knowing FT of pair of Dirac delta functions (2 narrow slits) is a cosine:

  $$\mathcal{F}\left[\delta\left(x + \frac{d}{2}\right) + \delta\left(x - \frac{d}{2}\right)\right] = 2\cos(k_x d/2)$$

  and Fourier theory says the converse that FT of a cosine is a pair of Dirac delta function, which have previously also shown occur at $k_x = \pm 2\pi/d$:

  $$\mathcal{F}\left[cos\left(\frac{2\pi x}{d}\right)\right] = \frac{1}{2}\left[\delta\left(k_x + \frac{2\pi}{d}\right) + \delta\left(k_x - \frac{2\pi}{d}\right)\right]$$

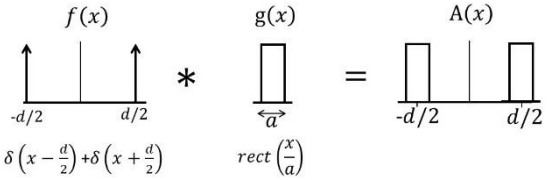
- **Convolution Theorem**: $\mathcal{F}[f(x) * g(x)] = \mathcal{F}[f(x)].\mathcal{F}[g(x)]$

  The convolution theorem is important for diffraction and imaging theory (see below).

**2-Extended Slits** (slit width $a$; separation $d$)

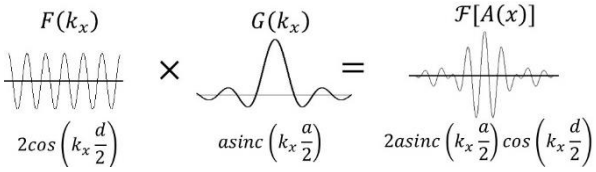
This aperture can be expressed as a convolution:

$$A(x) = \left[\delta\left(x + \frac{d}{2}\right) + \delta\left(x - \frac{d}{2}\right)\right] * rect\left(\frac{x}{a}\right)$$




Using Convolution Theorem:

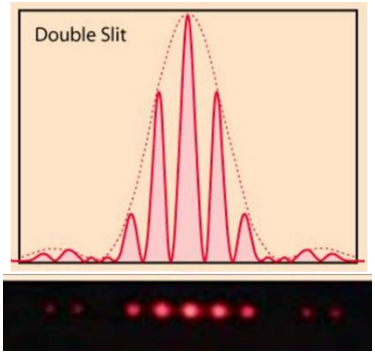
$$\mathcal{F}[A(x)] = \mathcal{F}\left[\delta\left(x + \frac{d}{2}\right) + \delta\left(x - \frac{d}{2}\right)\right].\mathcal{F}\left[rect\left(\frac{x}{a}\right)\right]$$




Hence:

$$E_P(k_x) = \mathcal{F}[A(x)] = 2a\ sinc(k_x a/2)\cos(k_x d/2).$$

- 2-extended slit pattern = 2-narrow slit pattern $2\cos(k_x d/2)$ <u>modulated</u> by single slit pattern $a\ sinc(k_x a/2)$
- Maxima: $dsin\theta_m = m\lambda$ [ $\frac{k_x d}{2} = m\pi$ in cosine argument]
- Single slit envelope has minima at $asin\theta_{m'} = m'\lambda$
- **<u>Note missing 2-slit maxima</u>** when $a$ is multiple of $d$ e.g. if $d = 3a$, every third 2-slit maximum is situated at a single-slit minima ($m = 3m'$).




## N infinitely-narrow Slits (with separation *d*)

This is equivalent to N equally-spaced $\delta$-functions:

$$A(x) = \sum_{n=0}^{N-1} \delta(x - nd)$$

We can get the Fourier Transform by using the shift theorem on the delta functions:

$$\mathcal{F}[A(x)] = \sum_{n=0}^{N-1} e^{-ik_x nd}$$

(*A physical way of understanding this is we have here summed the slit contributions due to the parallel path lengths ($ndsin\theta$ for nth slit) in the far-field at angle $\theta$*).

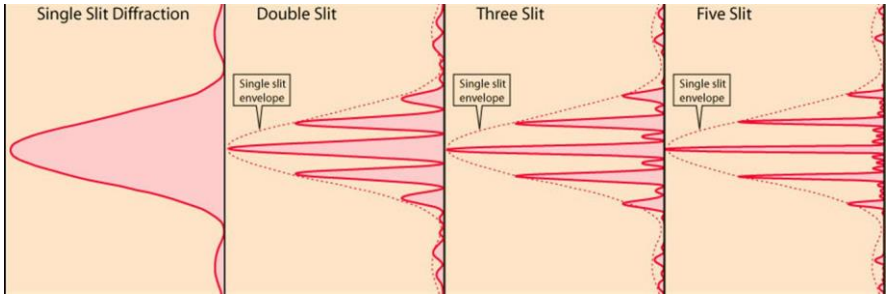
We can solve this by noting summation is a geometric series to get solution:

$$\mathcal{F}[A(x)] = \sum_{n=0}^{N-1} \left(e^{-ik_x d}\right)^n = \frac{1 - e^{-iNk_x d}}{1 - e^{-ik_x d}} = e^{-i(N-1)k_x d/2} \frac{\sin(Nk_x d/2)}{\sin(k_x d/2)}$$

Maxima occur when sin numerator and denominator both go to zero $\frac{k_x d}{2} = m\pi$ corresponding to $dsin\theta_m = m\lambda$ (same condition as with 2 slits) with peak value $|\mathcal{F}[A(x)]| = N$ linearly increasing with number of slits $N$.

The first zero is at $\frac{Nk_x d}{2} = \pi$ corresponding to $sin\theta_1 = \lambda/Nd$ decreasing with number of slits $1/N$.

**<u>Diffraction from N-extended slits (width *a*)</u>** is a simple extension that can be solved by convolving the N-slit $\delta$-functions by the slit function $rect(x/a)$ and using the convolution theorem for the Fourier Transform. Examples $N = 1,2,3,5$ in diagram below show that the N-narrow slit diffraction patterns are all modulated (multiplied) by the extended single-slit pattern. The peak positions are independent of *N*, but their width narrows with *N*.
(*The central peak strength also increase with N, but is not visible in diagram as central peak has been normalised to unity*).

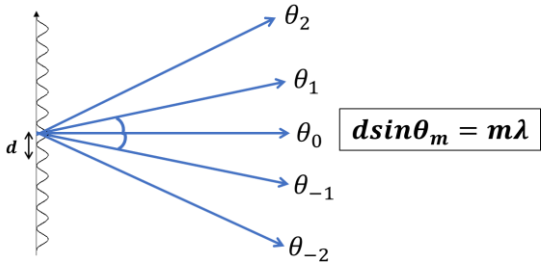


**DIFFRACTION GRATINGS** For the infinite slit case ($\boldsymbol{N \to \infty}$), we have what is called a diffraction grating, an infinite comb of $\delta$-functions – the limit of the $N$-slit case. Its Fourier transform is another infinite comb of $\delta$-functions in $k_x$ space:

$$\frac{2\pi}{d}\sum_{m=-\infty}^{\infty}\delta(k_x - m\frac{2\pi}{d})$$

with discrete **diffraction orders** $k_{x,m} = m\frac{2\pi}{d}$.
In terms of angles, this is the **grating formula**:

$$\boldsymbol{d sin\theta_m = m\lambda}$$




- The grating formula is true for any **periodic aperture function** (diffraction grating) e.g. a **square wave grating** or a **cosine grating**.
- A generalised grating (with period $d$) can be Fourier decomposed as a harmonic series of cosines: $A(x) = a_0 + \sum_{m=1}^{\infty} a_m cos\left(m\frac{2\pi x}{d}\right)$ and as we know each cosine gives rise to a discrete pair of angular spatial frequencies $k_{x,m} = \pm m\frac{2\pi}{d}$ or diffraction angles (known as diffraction orders). The relative strengths of the diffracted orders are given by the coefficients of the cosine terms $a_m$.
- It should be noted that the diffracted orders must be limited to angles $\theta_m < \pm 90°$.
- Main application: A grating spectrometer uses the fact that the diffracted angle in nay diffraction order depends on wavelength $\boldsymbol{\lambda}$ to provide spectral map of incident light on a 1-D detector array placed after the diffraction grating.

## Lens as a Fourier Transforming Element

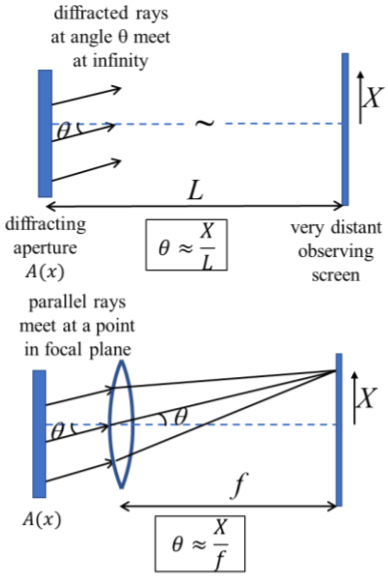
Whilst far-field diffraction pattern is Fourier Transform of aperture $A(x)$, far-field $z \gg z_R = a^2/2\lambda$ may be very distant and change strongly with aperture dimension $a$.

A lens can be used to perform a Fourier Transform operation by noting that all diffracted rays from aperture at angle $\theta$ will be brought to a point in the ***focal plane of the lens*** at position $X = f\theta$ (for small angles). The focal plane is therefore a Fourier map of angles (angular spatial frequencies).

The lens Fourier transform has scaled units: $\boldsymbol{k_x = ksin\theta \approx k\theta = kX/f}$ [$f$ replaces far-field screen distance $L$ in small angle formula for diffraction]

$$E(X) = \int_{-\infty}^{\infty} A(x)e^{-ikXx/f}dx$$

You can choose your lens focal length to get pattern size you want (e.g. to fit onto a camera detector) – and you can do it in a compact space on a small bench or in an instrument.



# 2-D Apertures

**2-D Diffraction Apertures** We can easily extrapolate diffraction patterns to a 2-D Fourier Transform of a 2-D aperture $A(x,y)$ with area element $dA = dxdy$ and transform variables $k_x = ksin\theta_x; k_y = ksin\theta_y$ for the two transverse components of the wavevector:

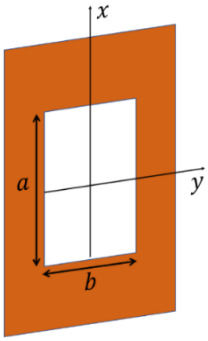
$$E(k_x,k_y) = \iint A(x,y)e^{-i(k_x x+k_y y)}dxdy = \mathcal{F}[A(x,y)]$$

**Rectangular Aperture**: For separable aperture function $A(x,y) = A(x)A(y)$, integral is product of 1-D Transforms:

$$E(k_x,k_y) = \int_{-\infty}^{\infty} A(x)e^{-ik_x x}dx \int_{-\infty}^{\infty} A(y)e^{-ik_y y}dy = \mathcal{F}[A(x)].\mathcal{F}[A(y)]$$

An example is the 2-D rectangular aperture: $A(x,y) = rect\left(\frac{x}{a}\right)rect\left(\frac{y}{b}\right) = A(x)A(y)$

$$A(x,y) = \begin{cases} 1, & |x| \le a/2; |y| \le b/2 \\ 0, & otherwise \end{cases}$$




It has 2-D Fourier Transform:

$$\boldsymbol{E(k_x,k_y) = ab\ sinc\left(\frac{k_x a}{2}\right)sinc\left(\frac{k_y b}{2}\right)}$$

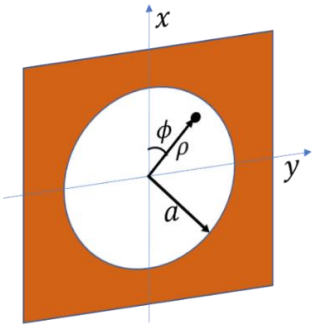
with central peak amplitude $ab$ increasing in proportion to the area of the 2-D aperture.

**Diffraction from a Circular Aperture – the Airy Pattern**

This is a very important 2-D aperture in optical systems, as most lenses and limiting apertures are circular. To get far-field diffraction pattern of circular aperture we must convert from Cartesian to cylindrical coordinates. The Cartesian 2-D Fourier Transform is:

$$E(k_x,k_y) = \iint A(x,y)e^{-i(k_x x+k_y y)}dxdy = \iint A(x,y)e^{-ik(xX/L+yY/L)}dxdy$$

and where we have explicitly introduced the (X, Y) coordinates of the observation screen using $k_x = ksin\theta_x = \frac{kX}{L}; k_y = ksin\theta_y = \frac{kY}{L}$. Aperture $A(x,y)$ must be changed to radial and azimuthal angle coordinates $A(\rho,\phi)$ and diffraction pattern screen coordinates change from $(X,Y)\ to\ (\mathrm{R},\Phi)$




$$x = \rho cos\phi\ ; y = \rho sin\phi$$

$$X = Rcos\Phi\ ; Y = Rsin\Phi$$

The circular aperture has unity transmission within radius $a$ (zero transmission otherwise):

$$A(\rho,\phi) = circ(\rho/a)$$

Area element $dA = dxdy$ becomes $dA = \rho d\rho d\phi$ and with coordinate change:

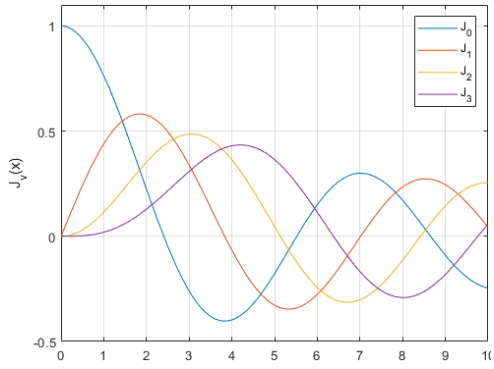
$$(xX/L + yY/L) = (\rho R/L)(cos\phi cos\Phi + sin\phi sin\Phi) = (\rho R/L)cos(\phi - \Phi)$$

$$E(R,\Phi) = \int_{\rho=0}^{a}\int_{\phi=0}^{2\pi} e^{-i(k\rho R/L)\cos\phi}\,\rho d\rho d\phi$$

where we set $\Phi = 0$ since diffraction must be axially symmetric and independent of $\Phi$.

The inner integral is a **Bessel function of zero order** $J_0(u)$**,** a function commonly met in Physics:

$$J_0(u) = \frac{1}{2\pi}\int_{\phi=0}^{2\pi} e^{-iu\cos\phi}\, d\phi$$




Hence: $E(\mathrm{R}) = 2\pi\int_{\rho=0}^{a} J_0(k\rho R/L)\rho d\rho$

A Bessel recurrence relationship allows the integral of $J_0$ Bessel function to be expressed as a Bessel function of first order $J_1$: $\int_0^u J_0(u')u'du' = uJ_1(u)$.

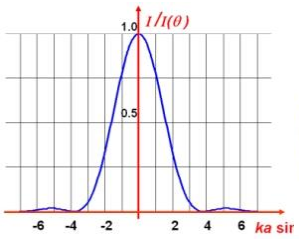
With substitution of variables, we find the solution:

$$\boldsymbol{E(\mathrm{R}) = \pi a^2 \frac{2J_1(kaR/L)}{kaR/L} = \pi a^2\left[2.jinc\left(\frac{kaR}{L}\right)\right]}$$

Noting that $jinc(u) = \frac{J_1(u)}{u} = \frac{1}{2}$ at $u = 0$, it is seen that the central (peak) amplitude of diffraction pattern $E(0) = \pi a^2$ scales with area of circular aperture $\pi a^2$. In terms of intensity and angular coordinate of diffraction pattern: $sin\theta = R/L$ we can write:

$$\boldsymbol{I(\theta) = I(0)\left[\frac{2J_1(kasin\theta)}{kasin\theta}\right]^2 = I(0)[2.jinc(kasin\theta)]^2}$$

- The intensity diffraction pattern of the circular aperture is called the **Airy pattern**. It is an important function in circular imaging systems.
- It has a bright central region called the **Airy disc** (measured to first zero).

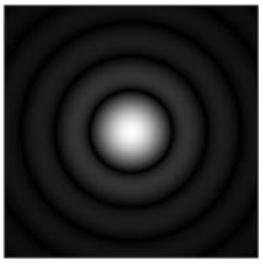




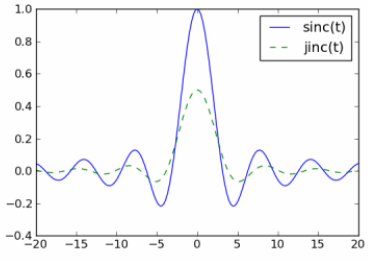
- A set of rings correspond to the zeroes in the Bessel function $J_1(u)/u$ occurring at $u = kasin\theta \approx 3.83, 7.02, 10.17$ ....
- Airy disc has angular size to first zero at $u_1 = kasin\theta_1 \approx 3.83$

$$\boldsymbol{sin\theta_1 = \frac{3.83\lambda}{2\pi a} = 1.22\frac{\lambda}{D}}$$

where $D = 2a$ is the diameter of the circular aperture.



- It is interesting to compare diffraction patterns from the 1-D slit $rect(x/a)$ and 2-D circular aperture $circ(\rho/a)$. They have very similar forms $sinc(u)$ and $jinc(u)$, and first zeroes at $sin\theta_1 = \lambda/a$ for the slit and $sin\theta_1 = 1.22\lambda/D$ for the circle (diameter $D = 2a$ is equivalent dimension to slit width $a$). The difference in geometrical factor $\frac{3.83}{\pi} = 1.22$ due to where the sine and $J_1$ Bessel functions have first zeroes can be consider arising from HFP summation over a circular aperture $circ(\rho/a)$ compared to the 1-D slit $rect(x/a)$ or 2-D square aperture $rect(x/a).rect(y/a)$.

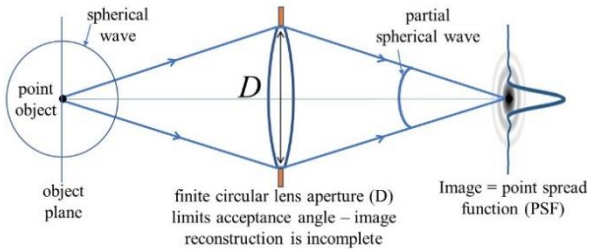


# Angular Resolution of Optical Systems

## Point Spread Function (PSF)

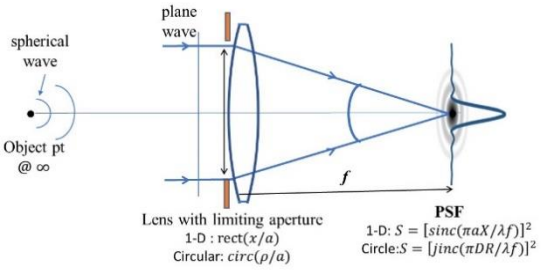
**Idealised Imaging System:** Image of a point object is a point image is a "fantasy" of ray optics that is predicted when all aberrations are removed from imaging system.

**Real Imaging system**: Even aberration-free systems are "diffraction-limited" in wave optics, since imaging systems having a limiting (usually circular) aperture e.g. camera lens, eye, telescope. Only a part of the spherical wave from an object point is collected by the lens representing a loss of information. An object point is not imaged to a point, but to a so-called **point spread function (PSF).**




We will use notation: **amplitude PSF = $s$** and **Intensity PSF = $S = |s|^2$**. For imaging incoherent objects, the use of intensity PSF is appropriate, and for now we assume no aberrations.

The PSF theory is complicated for the general imaging case, but the key result is obtained by considering distant point object (at $u = \infty$). Light field at the lens is a plane-wave that meets a limiting aperture and image of point at infinity is in the focal plane of the lens that we know will be the Fourier Transform of the aperture i.e. a *sinc* function for a 1-D slit or *jinc* (Airy) diffraction distribution for a circular aperture.



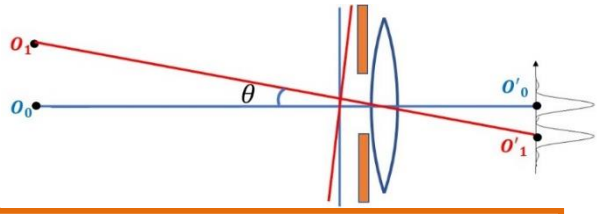

**Cylindrical lens** (with 1-D limiting width $a$) **intensity PSF: $S = sinc^2(\pi aX/\lambda f)$**. It has 1st minimum (characterising its spread size) and assuming small angles ($\theta_1 = X/f$) at:

$$\theta_1 = \lambda / a$$

**Circular lens** (limiting diameter D) **intensity PSF (Airy pattern) $S = jinc^2(\pi DX/\lambda f)$** and 1st minimum at:

$$\theta_1 = 1.22\lambda / D$$

**Resolution Criterion**: Finite image point size, limits ability to resolve two closely spaced points in object. We can quantify this resolution by the Rayleigh Criterion:




**Rayleigh Criterion:**

*Two image points are just resolved when the central maximum of one image pattern coincides with the first minimum of the other.*

For optical system with circular limiting aperture diameter $D$, 2 incoherent objects are resolved when their angular separation $\theta$:

$$\theta \geq 1.22\frac{\lambda}{D}$$

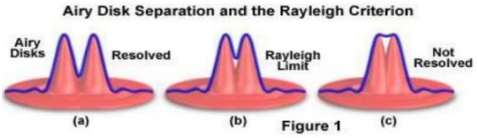
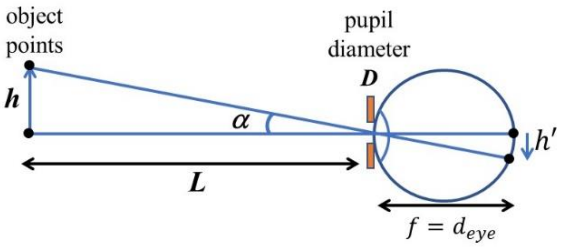



Figure 1

Note, PSF is an angular distribution and resolution is of the angular separation of the object points.

**Actual separation** $h$ of the object points at distance $L$ from the lens is given (for small angles) is given by: $\boldsymbol{h = L\theta}$. **PSF size**: If object at infinity, the image is in focal plane of the lens and Airy disc radius at: $\boldsymbol{X_1 = f\theta_1 = 1.22\lambda f \,/\, D.}$

## Resolution of Eye




Typical pupil diameter D = 2 mm; visible wavelength λ = 500 nm

Minimum resolved angular separation of 2 objects when: $\alpha = \theta_{eye}$

$$\theta_{eye} = 1.22\frac{\lambda}{D} = 1.22\frac{5\ 10^{-7}}{2\ 10^{-3}} = 3\ 10^{-4} rad$$

Effective eye-lens to retina distance: $d_{eye}$ ~ 15 mm;

Separation size on retina $h' = d_{eye}.\theta_{eye} = (15).(3\ 10^{-4})\ mm = 4.5\ \mu m$

Interestingly, typical separation of light-receptor cone cells in eye ~ 2.5 $\mu m$ is small enough to detect (resolve) this detail at retina.

## Resolution of Space Telescopes:


**Hubble Space Telescope (HST)** primary mirror diameter D = 2.4m and angular resolution:

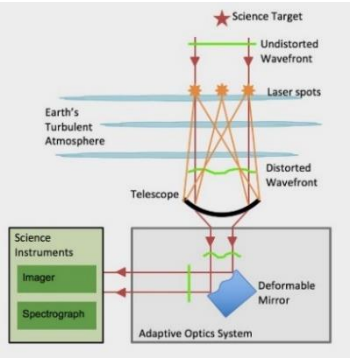
$$\theta_{Hubble} = 1.22\frac{\lambda}{D} = 1.22\frac{5\ 10^{-7}}{2.4} \approx 2.5\ 10^{-7} rad$$

**James Webb Space Telescope (JWST)** (launched 25 Dec 2021) has a primary mirror with diameter $D$ = 6.5 m; and images both visible and infra-red: λ = 600 nm – 28.5 μm. In visible it has higher resolution and much more collection area. In IR, resolution is not as good due to longer wavelength, but can see early Universe soon after Big Bang and observe through dust that absorbs visible light. To detect weak IR object, its detectors operate at very low temperature (T = 50 K) to reduce thermal noise.

**Space Telescopes:** Despite high cost, telescopes put in space can achieve their limit of resolution, and pointed in any direction and not dependent on geographical Earth position.

**Ground-Based Telescopes:** are cheaper to make and to larger sizes but viewing is disturbed by **atmospheric turbulence** from refractive index fluctuations in the air distorting the light wavefront from stars and limiting angular resolution to $\boldsymbol{\theta_{min} \sim 5.10^{-6}\ rads}$. Larger diameter telescopes do not have better resolution but only more light collection to see weaker objects.

**Laser Guides Stars:** However, new generations of ground-based telescopes are being built with increasing aperture size using **adaptive optics** to correct for atmospheric distortion using **lasers** that excites a "point-like" region of sodium in a layer in the upper atmosphere at 90km whose fluorescence creates an artificial "star". This **guide-star** is a known point source used for adaptive correction of atmospheric wavefront distortion $\varphi(r)$ using a "*deformable mirror*" to create a reverse distortion $-\varphi(r)$ to bring wavefront back to its undistorted form – a case of "two wrongs make a right".

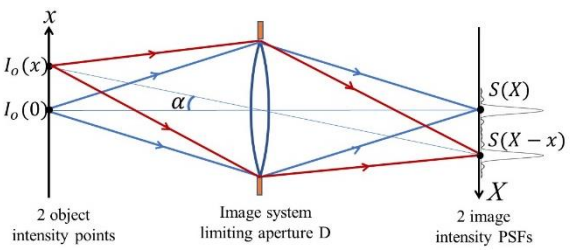


# Imaging Quality of Optical Systems

## Wave Theory of Imaging

The purpose of imaging systems rather than dealing with one or a few object points is to map an entire continuous 2-D object space $(x, y)$ onto its image space $(X, Y)$. Much imaging is incoherent (e.g. eye; camera; most microscopy) and object points are uncorrelated in phase to each other and it is appropriate to add image point intensities rather than fields.

**Incoherent object/images**: We will consider an incoherent object with intensity distribution $I_0(x, y)$ and use the intensity point spread function $S(X, Y) = |s(X, Y)|^2$ for the image. The intensity distribution $I_0(x, y)$ can be considered composed of point elements $I_0(x, y)dxdy$ (Dirac delta functions) that after passing through the imaging system are mapped to the image space as intensity PSFs $S(X, Y)$ weighted by the strength of the object point $I_0(x, y)dxdy$. For mathematical convenience, we consider imaging with unity magnification and no inversion.




Central object point: $I_0(0,0)dxdy$ maps to central image spread $dI_i = [I_0(0,0)dxdy]S(X, Y)$

but more generally $I_0(x, y)dxdy$ maps to displaced $dI_i = [I_0(x, y)dxdy]S(X - x, Y - y)$

Image $I_i(X, Y)$ is sum over all imaged intensity object points:

$$I_i(X, Y) = \iint I_0(x, y)S(X - x, Y - y)dxdy$$

This integral shows that **image is convolution of object and intensity PSF**:

$$I_i = I_0 * S$$

Image is blurred by the point spread function at each image point. We can analyse this further by taking the Fourier Transform of the image and using the convolution theorem:

$$\mathcal{F}[I_i] = \mathcal{F}[I_0] \times \mathcal{F}(S)$$

$\mathcal{F}[I_i]$ is **spatial frequency spectrum of image**

$\mathcal{F}[I_0]$ is **spatial frequency spectrum of object**

$\mathcal{F}(S)$ is the **optical transfer function (OTF)**

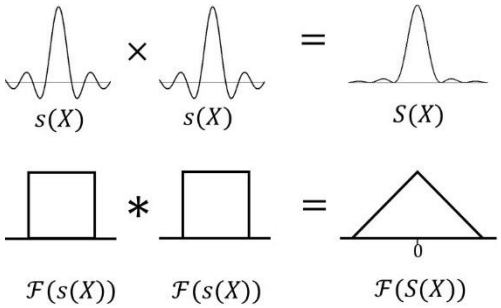
**Optical Transfer Function (OTF)** is the spatial frequency response of the optical system. It multiplies the object spatial frequency spectrum to give an attenuated image spatial frequency spectrum. In general, *OTF* will tend to attenuate high spatial frequencies most severely and has a cut-off frequency (due to limiting image system aperture). This leads to **blurring** and **contrast reduction** in image, notably affecting small or sharp features in object.

**OTF=** $\mathcal{F}(S)$ is the Fourier Transform of the intensity point spread function (*PSF*). For a circular lens, the intensity *PSF* is the Airy intensity pattern. For mathematical simplicity, we approximate the 2-D circular aperture $circ(\rho/a)$ by the 1-D slit function $rect(x/a)$ whose amplitude PSF in image has form: $s(X) = sinc((\pi aX)/\lambda f) = sinc(\alpha X)$ and intensity PSF is

$$S(X) = |s(X)|^2 = sinc^2(\alpha X)$$

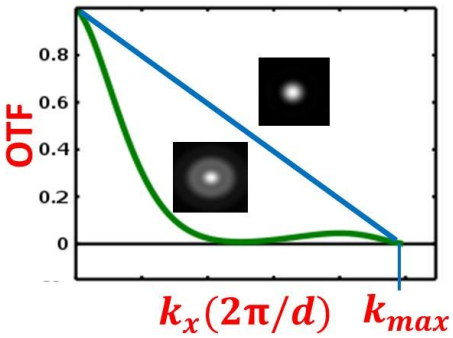
We know *PSF* $s(X) = sinc(\alpha X)$ has a Fourier Transform in the form $\mathcal{F}(s) = rect(\beta k_x)$ [as $sinc$ and $rect$ functions are Fourier pairs]. Therefore, using the convolution theorem:

$$\boldsymbol{OTF = \mathcal{F}(S) = \mathcal{F}(s.s) = rect(\beta k_x) * rect(\beta k_x) = triangle(\beta k_x)}$$




The triangle function is the overlap (convolution) of the $rect$ function with itself.

**OTF Plot** is shown in the graph for an imaging system with no aberrations (in blue) and with aberrations (in green). OTF is plotted against angular spatial frequencies $k_x = 2\pi/d$, where $d$ is the period of a cosine component in the object (in its Fourier decomposition). There is a cut-off spatial frequency above which OTF is zero, due to the sharp aperture edges of the imaging system, defining a maximum diffraction angle (spatial frequency in object) that it can collect and reproduce in image.




For 1-D lens (slit aperture), OTF is the triangular function. Inset in the graph is the 2-D Airy pattern which is intensity PSF for a 2-D circular aperture $circ(\rho/a)$ [note its OTF is like the triangular function – but "tent" shaped with a slight curve from the convolution of a circular aperture with itself]. For an aberrating system, OTF (green plot) is more attenuating for spatial-frequencies due to a larger PSF (shown in inset) leading to greater image blurring and contrast loss even though cut-off frequency is the same.

## How to use the OTF graph for optical imaging systems?

We know $\mathcal{F}[I_i] = \mathcal{F}[I_0]$ x $OTF$ reduces contrast in the image, but how do we apply it?

A spatial pattern can be Fourier decomposed into its cosine components. Consider a cosine intensity object (where 1 is added so intensity is positive everywhere):

$$I_0(x) = 1 + cos(2\pi x/d)$$

It has Fourier Transform (angular spatial frequency spectrum):

$$\mathcal{F}[I_0(x)] = \delta(k_x) + \tfrac{1}{2}\delta[(k_x - 2\pi/d) + \delta(k_x + 2\pi/d)]$$

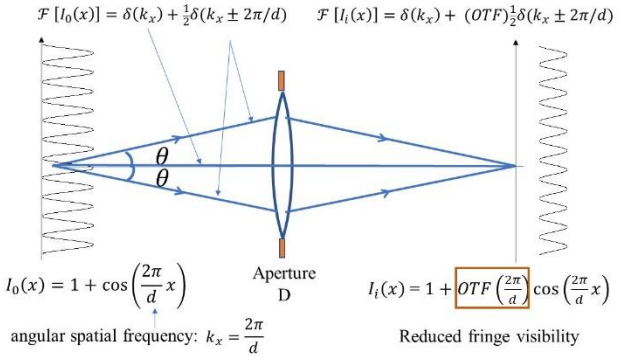
The 1 gives $k_x = 0$ term, and cosine has discrete pair of angular spatial frequency $k_x = \pm 2\pi/d$. Image spatial frequency spectrum $\mathcal{F}[I_i] = \mathcal{F}[I_0]$ x $OTF$ is given by:

$$\mathcal{F}[I_i(x)] = OTF(0)\delta(u_x) + OTF(2\pi/d)\tfrac{1}{2}\delta[(k_x - 2\pi/d) + \delta(k_x + 2\pi/d)]$$

The image (by re-transforming):

$$I_i(x) = 1 + OTF(2\pi/d).cos(2\pi x/d)$$

where $OTF(2\pi/d)$ is at spatial frequency $2\pi/d$, and it is customary to normalise $OTF(0) = 1$.




**Fringe Visibility $V$**. The reduced cosine modulation of the image can be quantified by a fringe visibility: $V = \frac{I_{max} - I_{min}}{I_{max} + I_{min}}$.

- For object: $I_{min} = 0$ and $\boldsymbol{V_0 = 1}$
- For image: $I_{max} = 1 + MTF$ and $I_{min} = 1 - MTF$, and $\boldsymbol{V_i = MTF}$. From the above *OTF* graphs, higher spatial frequencies with reduced *MTF* have decreasing visibility.

**Cut-off Frequency**. Due to the limiting aperture *D* of the imaging system, there is also an absolute maximum spatial frequency $k_{max} = (2\pi/d)_{max}$, due to the maximum diffraction angle $\theta_{max}$ collected, given by diffraction formula: $d_{min} sin\theta_{max} = \lambda$. We can therefore define a minimum resolvable spatial (cosinusoidal) object feature:

$$\boldsymbol{d_{min} = \frac{\lambda}{n_0 sin\theta_{max}} = \frac{\lambda}{NA}}$$

- $NA = n_0 sin\theta_{max}$ is called the numerical aperture of optical system. If refractive index $n_0$ is included between object and lens (e.g. used in oil-immersion microscope objective lens), the shorter medium wavelength ($\lambda/n_0$) increases resolution.
- This formula is a limit for imaging of linear optics. It is a statement that the minimum possible resolvable feature $d_{min} \sim \lambda$, at maximum possible angle ($sin\theta_{max} \sim 1$).
- *Our analysis was based on normal-incidence plane wave illumination, for wide-angle illumination the resolving power can be increased by a factor of two*: $d_{min} = \lambda/(2.NA)$.
- Nonlinear optics allowed super-resolution beyond the diffraction-limit and won the Nobel Prize in 2014 – but we'll not discuss here as it is beyond the course content.
- In many imaging systems, the numerical aperture is (much) less than this maximum, but is quite acceptable e.g. in telescopes the aim is to resolve small angular size $\alpha_{min}$ of distant objects (e.g. stars), and not physical size $d_{min}$.

## Abbé's Theory of Image Formation

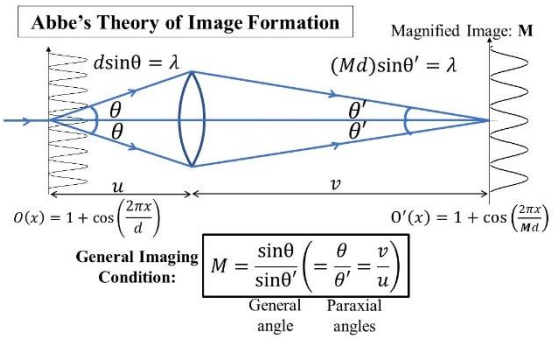
In paraxial ray optics, we made *small angle* geometrical approximations ($sin\theta \approx tan\theta \approx \theta$) for imaging to get the thin lens formula and magnification formula: $\frac{1}{u} + \frac{1}{v} = \frac{1}{f}$ ; $M = \frac{-v}{u}$.

Diffraction theory tells us requirements for *large angle* imaging. A cosine object with spatial period $d$: $I_0(x) = 1 + cos(2\pi x/d)$ has a pair of symmetric diffraction angles given by:

$$dsin\theta_{\pm} = \pm\lambda$$

If its image is magnified by a factor *M*, then, in the image space (*X*), the image has the form: $I_i(X) = 1 + cos(2\pi X/Md)$ with period $d' = Md$. But diffraction theory tells us this magnified cosine image must be formed by a pair of waves at (diffraction) angles given by:

$$Mdsin\theta'_{\pm} = \pm\lambda$$




Imaging system "bend" the pair of (diffracted) object waves at diverging angles $sin\theta_{\pm}$ to form converging waves at angles $sin\theta'_{\pm} = sin\theta_{\pm}/M$, leading to **Abbé's Sine Rule** for imaging:

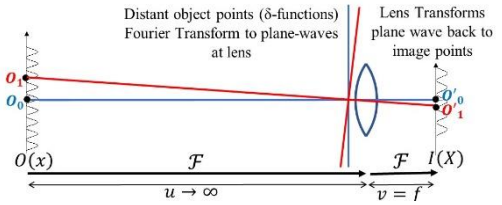
$$M = \frac{d'}{d} = \frac{sin\theta}{sin\theta'}$$

- From paraxial ray optics, we also know magnification in terms of image/object distances: $M = v/u = \theta/\theta'$ and is consistent with Abbé's Sine Rule (at small angles).
- The geometry in the diagram shows that bending the ray at a single plane as approximated for a "thin-lens", the triangles formed at the lens show the large angle relationship for imaging is $M = \frac{tan\theta}{tan\theta'}$, which is not the Sine Rule. It can therefore be deduced that large angle imaging cannot be achieved with a thin lens, explaining the need for more complex multi-lens element systems we looked at end of Ray Optics part of course.
- When Sine Rule is satisfied in an imaging system, the system is free of all aberrations.

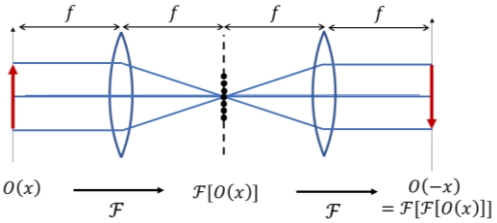
## Understanding imaging as a double-Fourier Transform (double diffraction) process.

We've learnt that a lens forms a Fourier Transform of the light distribution before the lens in the focal plane (at least in magnitude) - but the image is not the Fourier Transform of the object – it is the same as the object (to within a magnification factor). This means there must be a further Fourier Transform. Imaging must be considered as a double-Fourier Transform (or double diffraction) process. How can we think of this? Let's take some special cases.

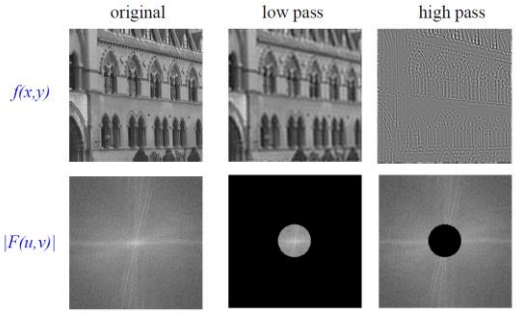
1. **Distant Object $u = \infty$** (e.g. astronomical object). The light field distribution is fully in the far-field and a Fourier Transform of the object (in angular $k_x$ space) when it reaches the lens (points become plane-waves). From the thin lens formula, the lens then forms the image at the focal plane of the lens ($v = f$) for object is at infinity ($u = \infty$). But this is also a Fourier Transform of the light entering the lens. So, we see two Fourier Transforms have been performed to get from object to image: the first Transform is from object (space) to angular spatial frequency space (at lens), where lens Transforms to (image) space. The finite lens aperture truncates plane-wave before the second transformation occurs - the origin of the point spread function (*PSF*).




2. **4-f Imaging System.** The 4-f system shown in diagram explicitly demonstrates the double Fourier Transform nature of image formation in this 2-lens imaging system. It is so-called because the distances between all 4 relevant planes in the system are *f*. The first lens produces a Fourier Transform of the object in its focal plane, the second lens performs a further Fourier Transform in its focal plane to reconstruct the image of the object (*it is inverted – as second transform is not the Inverse Fourier Transform*).




The 4-f imaging system is a good experimental test-bed for inserting a "spatial-filter" in the first Fourier plane. For example, by placing a small pinhole at centre of the first Fourier plane (low pass filter), high spatial frequencies are lost leading to blurring of sharp edge/small features in the final image; or with a central disc (high pass filter), low spatial frequencies are blocked leading to picking out (enhancing) edges/small features in the image. In microscopy, this latter case is used to visualise phase objects, that are revealed when the zero order (low-frequency term) is removed – a technique known as phase contrast microscopy.



# INTERFEROMETERS

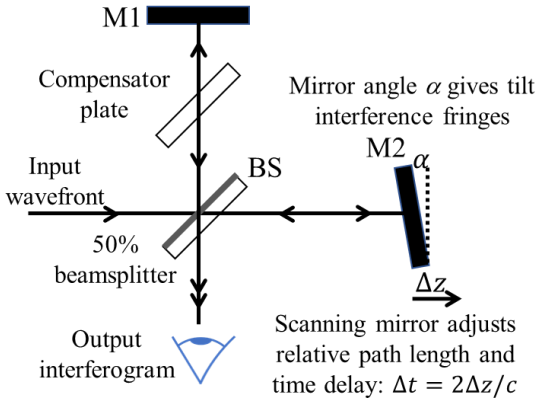
Two-slit and diffraction grating systems can be considered as interferometers operating by **division of wavefront**, as they provide interference of spatially separated regions from the incident wavefront to create a diffraction pattern.

Most interferometers operate without recourse to diffraction by **division of amplitude**, often using a beamsplitter (or partially reflecting surface). Examples of these are given below.

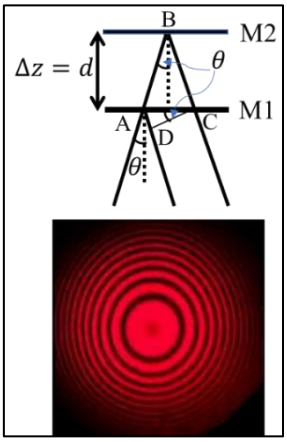
## Interferometers operating by Division of Amplitude

### Michelson Interferometer (MI)

Michelson interferometer consists of a 50% beamsplitter (BS) and two arms with mirrors M1 and M2. The BS splits the incident wavefront equally and the two mirrors return the beams to interfere at the beamsplitter to give an output interference pattern (interferogram). One arm of the interferometer has an adjustable mirror ($\Delta z$) to control the relative optical path between the two arms. In principle, this mirror can make the optical path length of the two arms arbitrarily equal. However, due to dispersion in the beamsplitter glass substrate it is common to add a compensator glass plate of equal thickness to the beamsplitter, placed in the arm that is split before passing through the BS substrate. The two arms now pass through equal amounts of glass and optical path length of all wavelengths can be equalised (allowing even white light interferograms).




The Michelson can be considered conceptually as in Figure with M2 placed at its effective distance d from mirror M1, which is the view of an observer at the output port. Two special cases can be considered:

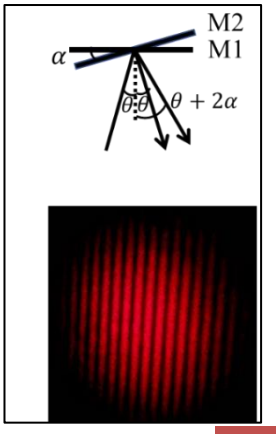



i) **Mirrors are parallel and separated by distance d**.

Consider path length difference $\Delta = ABC - AD$ between rays from back mirror M2 and front mirror M1 separated by distance d.

$$\Delta = \frac{2d}{cos\theta} - 2dtan\theta sin\theta = \frac{2d}{cos\theta}(1 - sin^2\theta) = 2dcos\theta$$

There is constructive interference and maxima when: $\Delta = 2dcos\theta = m\lambda$. Due to the symmetry, the maxima are circular fringes.




**ii) Mirrors have a relative tilt angle $\alpha$ but no separation ($d$=0).**

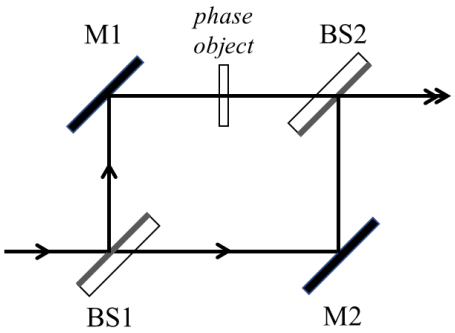
There will be linear tilt ($\cos^2$) fringes in the direction of the tilt due to interference of two waves at reflected angles $\theta$ and $\theta + 2\alpha$, or relative angle $2\alpha$. Comparing interference of these two tilted wavefronts, maxima of interference pattern occur at

$$x_m sin2\alpha \approx x_m 2\alpha = m\lambda$$

giving a fringe spacing: $\Delta x = \lambda/2\alpha$

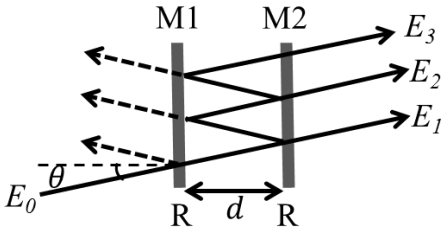
More generally, pattern is a combination of circular and tilt fringes.

**Mach-Zehnder Interferometer (MZI)** is similar to the Michelson having an input beamsplitter (BS1) and two beam arms, but instead of being redirected back to input beamsplitter, the two beams are steered to a second output beamsplitter (BS2) via two mirrors (M1/M2). No compensator plate is required as both arms see identical beamsplitter substrates. A test object can be placed in one arm and the other arm acts as a reference. The MZI works well with thin phase objects (e.g. biological tissue; air turbulence), where phase in the object can be visualised in the intensity contours of the output interferogram.




**Fabry-Perot Interferometer (FP)** consists of two parallel partially reflecting mirrors (M1/M2) with intensity reflectivity R. An input ray at angle $\theta$ is reflected multiple times between the mirrors, and remarkably even with high mirror reflectivity R, unity transmission through the pair of mirrors can be obtained if all transmitted beams are constructive.

With input wave amplitude, $E_0$, and mirror amplitude reflection and transmission $r$ $and$ $t$. Then $E_1 = E_0 t^2$; $E_2 = E_1 r^2 e^{i\phi}$; and more generally $E_n = E_1 \left(r^2 e^{i\phi}\right)^n$ where $\phi = k(2dcos\theta)$ is the round-trip phase difference (as we saw with Michelson). The resultant output is the sum of an infinite set of these fields:




$$E_t = \sum_{n=0}^{\infty} E_i = E_0 t^2 \sum_{n=0}^{\infty} \left[r^2 e^{i\phi}\right]^n = \frac{E_0 t^2}{1 - r^2 e^{i\phi}}$$

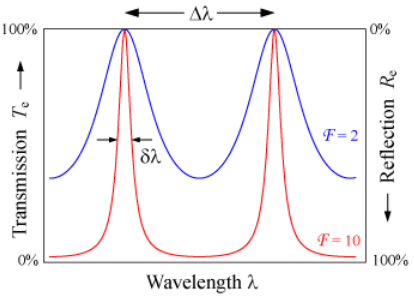
when $\phi = m2\pi$; $e^{i\phi} = 1$ and since $1 - r^2 = t^2$, $E_t = E_0$ and transmission $\frac{E_t}{E_0} = 1$!

More generally, transmitted intensity: $T = \frac{|E_t|^2}{|E_0|^2} = \frac{t^4}{|1 - r^2 e^{i\phi}|^2} = \frac{t^4}{1 + r^4 - 2r^2 cos\phi}$ and using trigonometric identity: $cos\phi = 1 - 2sin^2(\phi/2)$; and $1 - r^2 = t^2$

$$T = \frac{(1 - r^2)^2}{(1 - r^2)^2 + 4r^2 sin^2(\phi/2)} = \frac{1}{1 + F sin^2(\phi/2)}$$

$F = 4R/(1 - R)^2$ is known as the coefficient of finesse.

- A key FP application is high-resolution spectroscopy of a light source (or signal) so one considers FP response as a function of wavelength using $\phi = k(2dcos\theta) = 4\pi dcos\theta/\lambda$. Transmission peaks $T = 1$ occur when $\phi = m2\pi$ at wavelength separation: $\Delta\lambda = \lambda_0^2/2dcos\theta$, and transmission peak width (FWHM): $\delta\lambda \approx 2\Delta\lambda/\pi\sqrt{F}$, that determines ability to resolve two closely spaced wavelengths which improves with higher mirror reflectivity *R*.

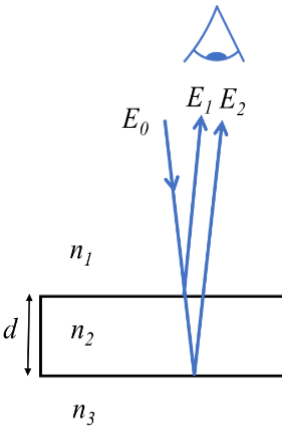



- The Fabry-Perot is the basis for most laser cavities, with a high reflectivity mirror and a partially reflecting mirror acting as the output coupler. The laser can only oscillate at a discrete set of frequencies that are those matched to the Fabry-Perot spectral peaks.

## Thin-Film Interference

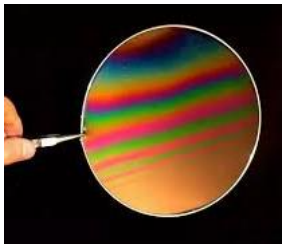
When light meets a boundary between materials with different refractive index, as well as refraction, there is a finite surface reflection. This can be derived from EM theory, and for normal incidence, amplitude reflectivity is given by the simple relationship:

$$r_{12} = \frac{E_1}{E_0} = \frac{n_1 - n_2}{n_1 + n_2}$$




where light is incident from medium with index $n_1$ and onto medium with index $n_2$. For air ($n_1 = 1$) to glass ($n_2 = 1.5$), $r_{12} = -0.2$ and intensity reflectivity $R_{12} = |r_{12}|^2 = 0.04 = 4\%$. The negative sign in $r_{12}$ indicates a π phase shift on reflection from a low to high index boundary. The reflectivity from high to low index has the same magnitude but zero phase shift.

With more than one surface, interference occurs from reflections at each interface (see diagram) and if layer(s) are thin with thickness on sub-wavelength scale, it is called **thin-film interference** e.g. colours appearing on an oil film on water; and colour bands in a soap bubble or soap film (see picture) due to constructive or destructive depending on relative phase $\delta = k.OPD = k(2n_2d)$ due to the double-pass optical path difference where $d$ is film thickness of second medium with refractive index $n_2$ and accounting for the phase shifts on reflection from the surfaces. For oil films on water: medium 1 is air ($n_1 = 1$), medium 2 is oil ($n_2 = 1.58$) and medium 3 and water ($n_3 = 1.33$). There is a π phase shift at first interface but not second. For soap bubble or soap film: medium 1 is air ($n_1 = 1$); medium 2 is soap ($n_2 = 1.34$) and medium 3 is air ($n_1 = 1$) with a π phase shift at first interface but not second.



### i) Anti-reflection coating

One important thin-film application is applying an anti-reflection coating to an optical glass surface to eliminate unwanted reflection e.g. lens surfaces in a camera or spectacles. In this case: medium 1 is air ($n_1 = 1$), medium 2 is coating layer ($n_2$), and medium 3 is glass ($n_3 = 1.5$). To minimise reflection, you require destructive interference, and ideally if the two reflections are equal $r_{12} = r_{23}$, reflection can be eliminated.

We treat the case of normal incidence light with reflection coefficient from media 1 to 2: $r_{12} = \frac{E_1}{E_0} = \frac{n_1 - n_2}{n_1 + n_2}$ and second reflectivity from media 2 to 3: $r_{23} = \frac{E_2}{E_0} = \frac{n_2 - n_3}{n_2 + n_3}$.

Destructive interference for a given colour $\lambda$ (usually mid-visible wavelength e.g. green) occurs when the double-pass film thickness is such that:

$$OPD = 2n_2d = \lambda/2$$

To cancel reflection with $r_{12} = r_{23}$: $\frac{n_1 - n_2}{n_1 + n_2} = \frac{n_2 - n_3}{n_2 + n_3}$ which leads to $(n_1 - n_2)(n_2 + n_3) = (n_2 - n_3)(n_1 + n_2)$ and a simple relationship between the refractive indices:

$$n_2 = \sqrt{n_1 n_3}$$

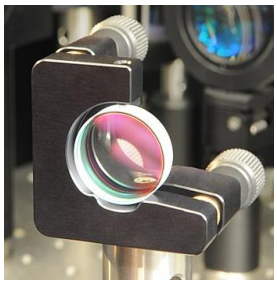
For normal glass $n_3 \approx 1.5$, we find optimum $n_2 = \sqrt{1.5} = 1.22$, is lower than real dielectric materials available for coating. A common coating choice is magnesium fluoride glass $MgF_2$ with $n_2 = 1.38$. Although a single layer is not perfect it significantly reduces the reflectivity

(from 4% to ~1%) for wavelength band in region of the destructive interference wavelength. Better anti-reflection with $R \to 0$ can be achieved by using multiple layers.

[*In the analysis above, for simplicity, we approximated that the incident field* $E_0$ *is negligibly reduced at the first surface but (surprisingly) full theory considering multiple reflections between the two surface gives the same answer even when reflectivity at first surface is high*!]

ii) **High reflectivity coating**:

Mirrors with extremely high reflectivity: $R > 99.9\%$ can be produced by coating a stack of alternate high and low refractive index films on a glass substrate. With constructive interference at each boundary, small individual reflections can add up to an arbitrarily high overall reflectivity with sufficient layers. These mirrors are used in lasers as they are lossless (non-absorbing), unlike metal mirrors that have poorer reflectivity $R \to 0.98 - 0.99$ due to their finite absorption that can also heat and damage a metal mirror surface at high laser powers. Laser mirrors made with multi-layer thin films are also called dielectric mirrors (as opposed to metallic mirrors). They can be made to reflect strongly at a particular laser transition or with broadband reflection spectrum.



Many other thin-film devices can be engineered using the properties of interference e.g. interference as filters (e.g. high-pass and low-pass filters; band-pass filters; notch filters) for spectral control or analysis of radiation. With large numbers of dielectric layers, using a computer-based transfer-matrix approach is convenient (computing surface reflection at layer boundaries and phase change due to propagation inside layers) and allows for the design and response of many useful thin-film interference devices.

# Coherence and Interferometry

## Coherence

Up to now, we've assumed perfect monochromatic plane-wave illumination when considering diffraction and interference. Real light sources are not perfect and can have complex spatial wavefronts and spread of frequencies (polychromatic) and this will clearly have an impact on interference.

**Definition:** Coherence relates to the phase relationship of a wave with itself (**self-coherence**) or between two waves (**mutual coherence**) and the ability to produce interference effects.

The concept of coherence and its treatment in textbooks can be complicated and abstract. We will take a more pragmatic approach and start with a familiar example of a 2-slit interference to develop two key concepts in interference: **fringe visibility** and a **coherence function**.

**Example: 2-slit diffraction (slit separation *d*).** Consider the 2-slit case but with a more general incident field $E(\underline{r},t)$ producing two unequal field amplitudes with different phases at the two slits: $E_1 = |E_1|e^{+i\phi_1}$ and $E_2 = |E_2|e^{+i\phi_2}$ with phase difference $\Delta\phi = \phi_2 - \phi_1$ at the slits.

The diffracted field at angle $\theta$ is given by: $E_p = E_1 + E_2 e^{i\delta}$ where $\delta = kdsin\theta$ is the relative phase delay between the two beam paths. The intensity pattern is given by:

$$I_p(\delta) = \langle |E_p|^2 \rangle = \langle (E_1 + E_2 e^{i\delta})(E_1^* + E_2^* e^{-i\delta}) \rangle = \langle |E_1|^2 \rangle + \langle |E_2|^2 \rangle + \langle E_1 E_2^* e^{-i\delta} + E_2 E_1^* e^{i\delta} \rangle$$

where $\langle - \rangle$ denotes time-average.

If the phase difference in the two beams **$\Delta\phi$ is fixed**, then:

$$I_p(\delta) = I_1 + I_2 + 2\sqrt{I_1 I_2} cos(\delta + \Delta\phi).$$

With equal slit intensity $I_1 = I_2 = I_0$, (and using $1 + cosx = 2cos^2(x/2)$) we get back our original 2-slit expression, $I_p(\delta) = 4I_0 cos^2(\delta/2 + \Delta\phi/2)$, consisting of $cos^2$ *fringes* but with a shift in the pattern by $\Delta\phi/2$.

**Fringe visibility**: A key feature of interference is *fringe visibility V* that we define as:

$$V = \frac{I_{max} - I_{min}}{I_{max} + I_{min}}$$

where $I_{max}$ and $I_{min}$ are fringe maximum and minimum intensity.

For our 2-slit example with intensity distribution: $I_p(\delta) = I_1 + I_2 + 2\sqrt{I_1 I_2} cos(\delta + \Delta\phi)$:

$$I_{max} = I_1 + I_2 + 2\sqrt{I_1 I_2}$$

$$I_{min} = I_1 + I_2 - 2\sqrt{I_1 I_2}$$

$$V = \frac{2\sqrt{I_1 I_2}}{I_1 + I_2}$$

Equal slit intensity $I_1 = I_2$ yields a fringe visibility $V = 1$ (100% fringe visibility). This defines a **perfectly coherent** incident light source with fixed relative phase relationship $\Delta\phi$ between the two field components, and if true regardless of separation *d* between the slits.

**But what if $\Delta\phi$ is not fixed in time?**

**Spatial Coherence**: When we have a spatially extended light source (e.g. a star) with relative phase of light emission spatially varying at different points on the source, the wavefront produced can change in time. We can define a space correlation function between the field at one point on the wavefront and the field displaced by a transverse distance d

$$\gamma(d) = \frac{\langle E_1(x)E_2^*(x+d)\rangle}{\sqrt{I_1 I_2}}$$

where denominator $\sqrt{I_1 I_2}$ provides normalisation. In optics, this correlation function is known as a coherence function.

We can measure this function using two-slit interference for varying slit separation $d$ and looking at the fringe visibility of the interference expression:

$$I_p(\delta) = I_1 + I_2 + 2\sqrt{I_1 I_2}\gamma(d)cos(\delta)$$

For coherent light (wavefront relative phase constant in time): $\gamma(d) = 1$ we get 100% fringe visibility for equal intensities: $V = 1$.

For incoherent light $\gamma(d) = 0$, intensity $I_{max} = I_{min}$ and our visibility function $V = 0$ i.e. **we lose interference**.

For generally "partially-coherent" light (real light sources), $\gamma(d)$ is a decreasing function of slit separation $d$.

**Temporal Coherence Function**: We can also define a time correlation function:

$$\gamma(\tau) = \frac{\langle E_1(t)E_2^*(t+\tau)\rangle}{\sqrt{I_1 I_2}}$$

This is an important parameter for treating **non-monochromatic** light. We can consider a plane-wave with central frequency $\omega_0$: $E(t) = E_0(t)e^{i(k_0 z - \omega_0 t)}$ but with a field amplitude $E_0(t) = |E_0(t)|e^{i\phi(t)}$ that can be varying in time in magnitude and phase.

A Michelson interferometer is appropriate for characterising temporal coherence as it splits a beam into two paths and provides a relative time-delay $\tau$ by varying path length of one of its two arms. When the beams recombine, we obtain an output interference signal as a function of delay time τ:

$$I_p(\tau) = I_1 + I_2 + 2\sqrt{I_1 I_2}\gamma(\tau)cos(\omega_0\tau)$$

where $\gamma(\tau)$ is the temporal self-coherence function:

$$\gamma(\tau) = \frac{\langle E(t)E^*(t+\tau)\rangle}{I}$$

We can find this function by variation of intensity on a photodetector as delay is scanned. Another visual approach to find the temporal coherence function $\gamma(\tau)$ is by measuring visibility of tilt fringes as function of time-delay $\tau$:

$$I_P(\tau) = I_1 + I_2 + |\gamma(\tau)|2\sqrt{I_1 I_2}cos(k'x)$$

For monochromatic light with $E(t) = E_0 e^{-i\omega_0 t}$, we obtain $\gamma(\tau) = e^{i\omega_0\tau}$, and temporal coherence function has unity magnitude $|\gamma(\tau)| = 1$ for all time delays.

For non-monochromatic light, at zero time-delay $|\gamma(0)| = 1$ but at longer time delays there is imperfect correlation (Δϕ is not fixed) leading to a reduced coherence $|\gamma(\tau)| < 1$. We can

define a **coherence time, $\tau_C$** when the coherence function has fallen from unity at $\tau = 0$ to a value $|\gamma(\tau_C)| = |\gamma(0)|e^{-1}$.

There is a Fourier Transform relationship between time and frequency and for light (waves) they are related by the **bandwidth theorem** that states:

$$\Delta t \Delta \nu \sim 1$$

In the context of interferometry, the effective time in this theorem $\Delta t \approx \tau_C$ is the coherence time. Hence, coherence time depends on the bandwidth of the light source:

$$\tau_C \approx \frac{1}{\Delta \nu}$$

Using $\nu = \frac{c}{\lambda}$ we can also express coherence time in terms of a wavelength bandwidth $\Delta\lambda$:

$$\Delta \nu = -\frac{c}{\lambda^2}\Delta\lambda$$

e.g. for a "super-fluorescent" laser diode source: $\Delta\lambda \sim 100\ nm$; $\lambda \sim 0.8\ \mu m$ **:**

$$\Delta \nu = -\frac{3\ 10^8}{(8\ 10^{-7})^2}\ 10^{-7} \approx 5\ 10^{13}\ Hz$$

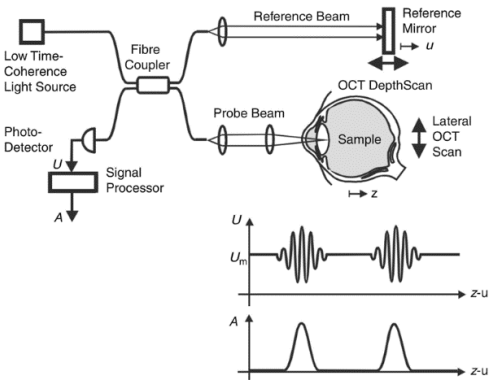
$$\tau_C \approx \frac{1}{\Delta \nu} = 2\ 10^{-14} s$$

We can also define a coherence length (distance travel by light in a coherence time):

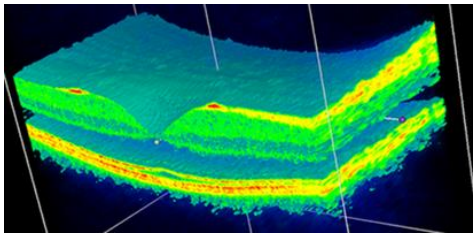
$$l_C = c\tau_C = 6\ 10^{-6} m = 6\ \mu m$$

## Optical Coherence Tomography (OCT)

The very short coherence length of broadband light provides a useful tool for coherence interferometry for mapping depth (z) information via reflections from object of interest. This technique is now commonly used for medical imaging and known as **optical coherence tomography (OCT)** e.g. see 3-D retina mapping image (see diagrams)




This can be done by and using reflection from the object structure of interest (e.g. retina) in one arm of a Michelson interferometer while the reference mirror of the other arm is scanned (see diagram with a fibre-optic implementation of Michelson). When the time delay to any reflecting structure is matched to reference arm ($\tau = 0$), then interference fringe visibility is high. By using low coherence length light this interference decays with small displacement resolving double-pass depth $\Delta l = \frac{l_C}{2} = 3\ \mu m$ using the super-fluorescent light source example above. By also x-y scanning light over sample a 3-D mapping can be built up, a technique known as tomography (see retina OCT scan figure).



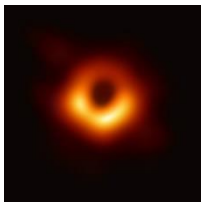
## Astronomical Interferometry (this topic is non-examinable)

We've noted that varying the two slit distance d measures the spatial coherence function of the light $\gamma(d) = \langle E_1(x)E_2^*(x+d)\rangle$. This method has been used to find the angular size of the astronomical objects. In 1920, the first measurement of the angular size of a star (Betelgeuse) was made using a modified 2-pinhole interferometer (Michelson stellar interferometer) with variable pinhole separation.

## Spatial Array Telescopes

- We can go one step further, instead of using 2 pinholes we can use an array of pinholes to gain more information about the object – not only assess angular size but to image ("resolve") sub-angular size object information. This has become the basis of astronomical interferometry where the "pinholes" are arrays of telescopes. In principle, the resolving power of an individual telescope of aperture $D$ $(1.22\lambda/D)$ is increased to that of the maximum separation of the telescopes $D_{max} \gg D$.

- **Radio telescope arrays**. Most astronomical interferometry has been done with radio telescopes as the long period radio wavelength signals are easier to measure and signals are not affected by atmosphere. The picture shown is of the Very Large Array (VLA) radio telescope array in New Mexico. It comprises 27 independent antennas each with a 25-meter dish diameter distributed in a Y-configuration, and the length of the 3 arms can be changed by physically moving the antennae to arm lengths up to 21km. The antennae perform what is termed **aperture synthesis** interferometry and the act as an effective single antenna of variable diameter, with angular resolution ~ 0.1 arcseconds ($5.10^{-7}$rad). [compare our formula for resolution $\theta = 1.22\lambda/D$ with radio wavelength 1cm and diameter ~ 20km: $\theta \sim 510^{-7} rad$].



  Radio telescope arrays even spanning the planet have been linked to provide the first image of a black-hole in 2019 using the "Event Horizon Telescope" – a planet-scale array of eight ground-based radio telescopes. Other new planetary scale radio telescopes arrays are in construction e.g. Square Kilometer Array (SKA) with two main centres in South Africa and Australia. Such telescopes will have ability for massive mapping of star/galaxies in high resolution, increasing catalogue by many times currently known.

- **Optical telescope arrays** have been built and operated (Very Large Telescope VLT) and ever bigger new ones are under construction (e.g. Extremely Large Telescope ELT, see artist picture). The optical wavelength gives a big resolution advantage over radio-waves. A major challenge is you must use adaptive optics (with laser guide stars to provide reference point sources in the object field) to cancel the atmospheric turbulence and allow ground-based operation to achieve its full resolution potential. The ELT is scheduled to resolve exoplanets amongst other mission goals.

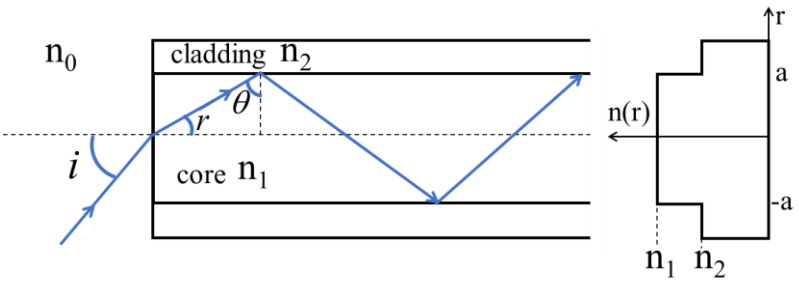
# FIBRE OPTICS

**Optical Fibre** is a long, thin flexible cylindrical glass waveguide that transports light using total internal reflection (TIR). It can transport light over long distance with low loss for illumination and imaging (e.g. in medical endoscopes) but its single most important application is for communication of information. Indeed, optical fibres underpin ultra-high data rates over global distances that make the internet possible. We will consider the optical fibre in both a Ray Model and a Wave Model and consider its application in Optical Communication.

## RAY PICTURE

### Step-Index Fibre

The diagram shows the cross-section of input end of a step-index cylindrical optical fibre with an inner **core** glass $n_1$ with higher refractive index ($n_1 > n_2$) than an outer **cladding** glass $n_2$.




To use a ray model and neglect wave (and interference) effects we consider the core radius a to be large ($a \gg \lambda$). Such large core fibre is called multimode fibre (the term mode coming from the wave model that we will discuss later).

Over a range of input angles $i$ (from $i = 0\ to\ i_{max}$), light will experience TIR at interface between core and cladding if internal interface angle $\theta \geq \theta_C$.

**Acceptance Angle**: $i_{max}$ when $\theta = \theta_C$.

Noting $r + \theta = \pi/2$, we use Snell's Law at the input face of the fibre:

$$n_0 sini_{max} = n_1 sinr_{max} = n_1 cos\theta_c = n_1(1 - sin^2\theta_c)^{1/2} = n_1\left(1 - \left({}^{n_2}/_{n_1}\right)^2\right)^{1/2}$$

$$\boxed{NA = n_0 sini_{max} = (n_1^2 - n_2^2)^{1/2}}$$

where from imaging theory, the quantity $n_0 sini_{max}$ is known as the numerical aperture (NA) of the optical system.

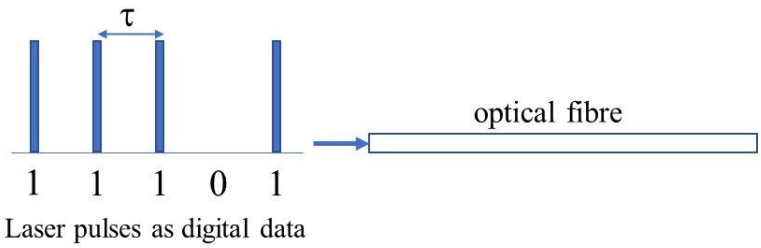
e.g. for $n_1 = 1.55;\ n_2 = 1.50;\ n_0 = 1\ (air) \Rightarrow NA = 0.39;\ i_{max} = 23°$.

Within the acceptance angle, light will see TIR at the core/cladding interface and experience near lossless guiding. This is basis of optical waveguiding in an optical fibre.

## Optical Communications

One of the most important applications of optical fibres is for optical communications, where laser pulses are used as the binary bits for data transfer (e.g. pulse = 1; no pulse = 0). Let's consider data transfer of a highly multimode fibre optics in the ray picture.

Laser pulses come in with time spacing $\tau$, corresponding to a date rate $R = 1/\tau$.



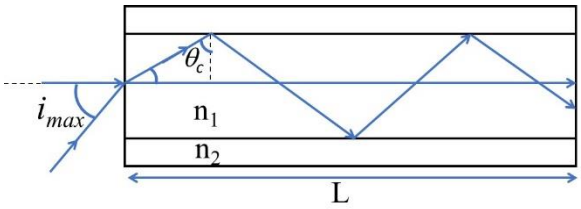
Laser pulses as digital data

Multimode fibre is cheap, but it has major problems for high-speed data transfer due to the different paths (angles) the light pulses can take leading to pulse spreading due to what is known in wave theory as **modal dispersion**.

Consider a multimode fibre with length L and the multiple ray paths (angles) allowed in the fibre between an axial path and the maximum off-axis ray path:

Axial path length = L;

Maximum off-axial path length (ray angle $\theta_c$):
$L' = \frac{L}{sin\theta_c} = \frac{n_1}{n_2}L$




Path length difference: $\Delta L = L' - L = \left(\frac{n_1}{n_2} - 1\right)L$

Spread in pulse duration: $\Delta t_s = \frac{n_1}{c}\Delta L = \frac{n_1}{n_2}(n_1 - n_2)\frac{L}{c} \approx \Delta nL/c$

where $\Delta n = (n_1 - n_2)$ is usually very small and we can take $n_1/n_2 \approx 1$

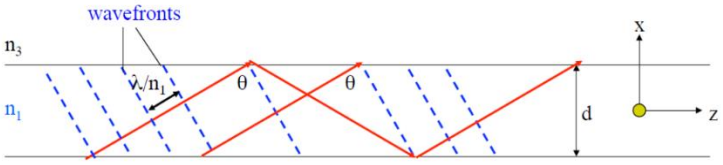
e.g. for $n_1 = 1.51;\ n_2 = 1.50; L = 3km; \Delta t_s = (0.01)\frac{3\ 10^3}{3\ 10^8} = 10^{-7}s$

Since pulse separation $\tau$ but be greater than this (so 1s and 0s don't overlap), bit rate must be limited to $< \frac{1}{\Delta t_s} = 10\ MHz$ in this example (not exactly super-fast broadband!). Hence data rate is very low or very short fibre cable length $L$ must be used.

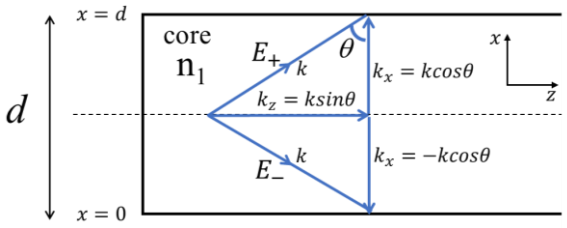
In practice, all long distance high-data rate optical communication is done in what is known as single mode fibre which has only one effective path and no modal dispersion.

## Wave Picture

The Ray picture is interesting to see how light can undergo TIR in step-index fibre over long fibre lengths. The Ray Model is suitable for very large core diameters, but the Wave Model is essential when considering small core fibres as used in optical telecommunications. Instead of rays we must consider the extended wavefronts of light in fibre light guide and the interference of the light with itself occurring after it reflects from the light guide boundaries.




To simplify analysis, consider a 1-D planar waveguide with core diameter d and boundary walls at $x = 0\ and\ x = d$. Light must be considered as two waves bouncing at angles $\pm\theta$ to the surface normal of boundary walls with wavevector $\boldsymbol{k}_{\pm} = (k_x, k_y, k_z) = (\pm kcos\theta, 0, ksin\theta)$. The two waves at angles $\pm\theta$ will interfere:




$$E = E_1 + E_2 = E_0 e^{i(\boldsymbol{k}_+ .\boldsymbol{r} - \omega t)} + E_0 e^{i(\boldsymbol{k}_- .\boldsymbol{r} - \omega t)} e^{i\Delta\phi} = E_0 e^{i(k_z z - \omega t)} \left[ e^{ikxcos\theta} + e^{-ikxcos\theta} e^{i\Delta\phi} \right]$$

where $\Delta\phi$ allows for a phase shift on reflection from the boundary walls of the waveguide.

### 1-D Metal Waveguide (with 100% reflecting walls)

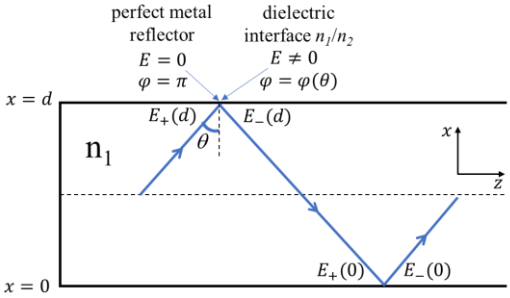
If we consider the extreme case of the boundary wall being a perfect metal reflector, the E-field at wall boundaries is zero $E = 0$ and phase change on reflection is $\Delta\phi = \pi$.

$$E(x,z) = E_0 e^{i(k_z z - \omega t)} \left[ e^{ikxcos\theta} - e^{-ikxcos\theta} \right] = 2iE_0 sin[kxcos\theta] e^{i(k_z z - \omega t)}$$

If $E = 0$ at $x = 0$ and $x = d$, then only discrete angles θ can satisfy boundary conditions: $kdcos\theta_m = m\pi$.

The boundary conditions lead to a discrete set of **mode** patterns given by

$$E_m(x,z) = E_{m0} sin\left[\frac{m\pi x}{d}\right] e^{i(k_z z - \omega t)}$$




These **waveguide modes** with amplitude distribution $sin[m\pi x/d]$ in the transverse direction ($x$) have the form of standing waves but unlike standing waves on a guitar string (for example) these modes are propagating along the z-axis of the waveguide.

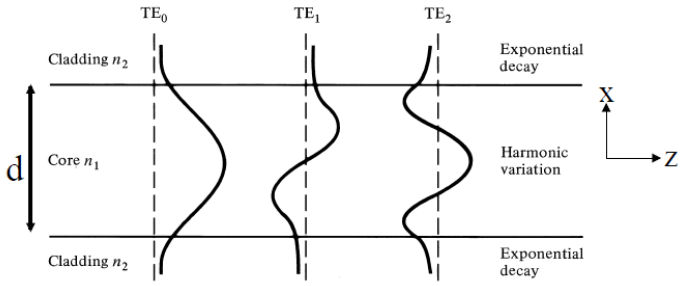
**For optical fibre.** Maxwell's equations lead to boundary conditions that do not allow the E-field to be zero at the refractive index interface $n_1/n_2$ and means E-field is non-zero in cladding $n_2$. But how is this possible if TIR means all the light is reflected at the boundary?

Consider E-field at boundary and Snell's Law: $n_1 sin\theta_1 = n_2 sin\theta_2$. Then, for core angles greater than critical angle, when TIR occurs, wavevectors of light are given by:

$$n_1 k_0 sin\theta_1 = n_2 k_0 sin\theta_2 > n_2 k_0$$

where $k_0 = 2\pi/\lambda$. Whilst there is no physical wave that can propagate in the cladding, we can get a solution $sin\theta_2 > 1$ if $\theta_2$ is imaginary. The field solution in the cladding takes on a form: $e^{ik_2x} \rightarrow e^{-\kappa x}$ with a decaying field amplitude. It is not a propagating wave, and it has no associated power flow which means TIR can still occur with 100% efficiency. This field is known as an "evanescent wave" in EM theory. It is a classical analogue to quantum tunnelling that you will meet in quantum physics.

The solution for the glass waveguide is more complicated than the sine solutions for the 1-D metal waveguide since the boundary is not a node, and in a cylindrical optical fibre, have the form of Bessel functions – but still similar form to the sine functions but extending into the cladding as shown in diagram.



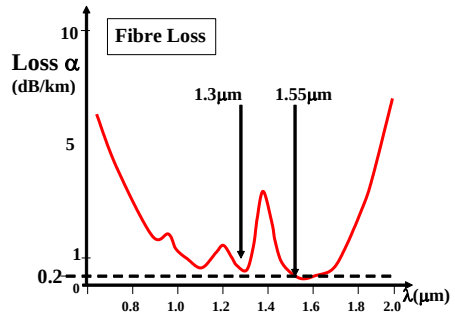

Fibre mode theory is beyond the scope of this course, but a key parameter in the theory is a quantity known as the normalised frequency $V = \frac{2\pi a}{\lambda} NA$ where $NA = n_0 sini_{max}$. When $V < 2.405$, only a single (lowest-order) mode can exist, and this fibre is called a "single-mode" fibre. [*The number 2.405 is the value of the argument where the $J_0$ Bessel function has its first zero*]. Single-mode fibre operation usually means small core size $a$ ~ few microns (approaching wavelength scales).

## Optical Communications

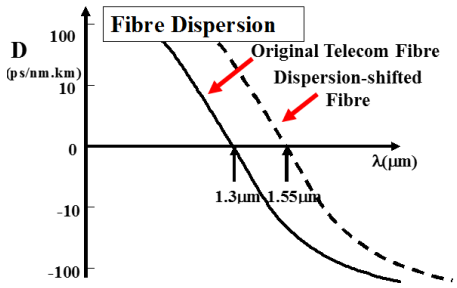
- **Modal dispersion** we saw previously due to different modes taking different paths in multi-mode fibre is eliminated by using single mode fibre.
- **Fibre attenuation** ($\alpha$) limits distance for data transfer when signal becomes too weak to distinguish binary 0s and 1s due to noise (*signal-to-noise limit*).
- **Fibre dispersion** (D) temporally broaden pulse duration $t_p$ and limits data rate due to (short) laser pulses have an (large) spectral bandwidth with different frequencies travelling at speeds due to material dispersion $n(\lambda)$.

### Telecom (silica) Fibre Characteristics

- Older telecom silica fibre operated at 1.3 $\mu m$ where attenuation $\alpha$ is small, and silica has a material dispersion minimum (D=0).
  [NB. *Telecom fibre attenuation is expressed in dB/km units, a logarithmic scale useful for adding system losses*].




- Today, fibre communication is mainly at 1.55 $\mu m$, which has lowest attenuation ($\alpha$ ~0.15dB/km) and using dispersion-shifted fibre to get D=0 at this wavelength region. [This is also the wavelength region for the erbium-doped fibre amplifier (EDFA) that can periodically boost the signal to offset the fibre attenuation and allowing data transfer at multi-1000 km (trans-Atlantic/Pacific) global distances.

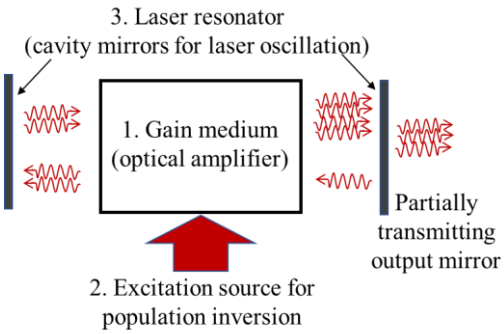



**Single Mode Optical Fibre** is essential to get long distance high data rates. Then limit is only due to material dispersion n(λ) which we can set D~0, to get data rates ~ 100GHz. This data rate is more a limit of switching speed of electronics.

Optical bandwidth of fibre is large and Terabit data rates can be achieved in fibre by using wavelength division multiplexing (WDM) where several wavelength bands (channels) can be simultaneously transmitted in fibre. The multiplexing (channel combining) and demultiplexing (where channels are separated) can be done by passive optics and no electronic switching needed.

# LASERS

LASER = "light amplification by stimulated emission of radiation". The first laser device, a flashlamp-pumped Ruby Laser, was demonstrated in 1960. Lasers now enable communication, manufacture, medicine, sensing and the advancement of cutting-edge fundamental science.

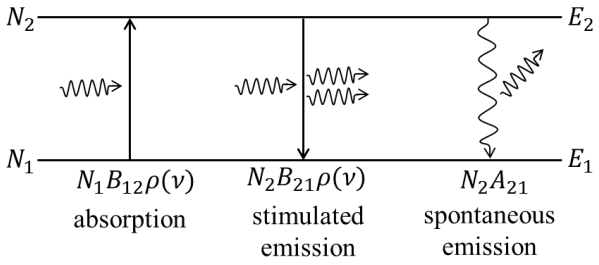



**A laser device has 3 key elements: gain medium + excitation source + resonator.**

We overview first the basic physics of the gain medium.

## Laser Amplification Process: Stimulated Emission

There are 3 main processes for light-matter interaction when light has a frequency ($\nu$) with photon energy ($h\nu$) matched to a transition between two energy levels in a material $h\nu = E_2 - E_1$. These are: absorption; stimulated emission; and spontaneous emission. In a quantum description, absorption involves annihilation of a photon and transfer of electron from lower energy to high energy state. Stimulated emission is the reverse process where an incident photon stimulates the release of another (identical) photon and transfer of electron from high to low energy state. It is the coherence of this stimulated emission process that provides the unique quality of laser light. Spontaneous emission also emits a photon from a high energy to low energy state but the photon direction and phasing is random, and its timing is statistical but occurring within a characteristic decay time of the transition ($\tau = A_{12}^{-1}$).




**Rate Equations**. Generally, light flux is high and a photon description unnecessary. Then these processes are described by average rates governed by the Einstein coefficients: $A_{21}; B_{12};\ B_{21}$, the populations densities in lower and upper energy levels, $N_1\ and\ N_2$, respectively, and photon energy density $\rho$. We can write a rate equation for the upper-level population:

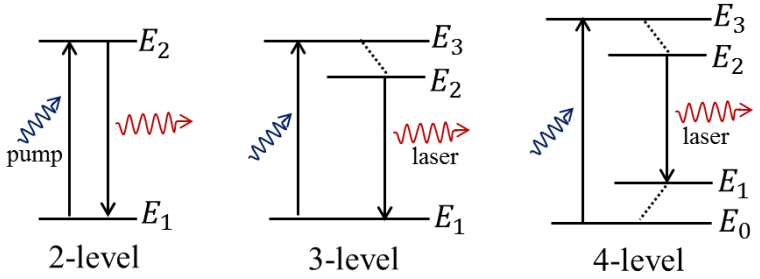
$$\frac{dN_2}{dt} = N_1B_{12}\rho - N_2B_{21}\rho - N_2A_{21}$$

and $\frac{dN_1}{dt} = -\frac{dN_2}{dt}$, in a two-level system with conserved total population $N_T = N_1 + N_2$. We can also describe the growth (or loss) of light intensity $I = \rho c$ as a function of propagation distance $z$ in the medium (by using $d/dt \rightarrow cd/dz$) and taking common case ($B = B_{12} = B_{21}$) and neglecting spontaneous emission (good approx. when light flux in cavity is high):

$$\frac{dI}{dz} = \frac{Bh\nu}{c}(N_2 - N_1)I = \sigma\Delta NI$$

where term $\Delta N = N_2 - N_1$ is the population difference and $\sigma$ is a cross-section (units of area). The solution for fixed $\Delta N$ is $I(z) = I(0)e^{\sigma\Delta Nz}$. Under normal conditions $N_1 \gg N_2$, $\Delta N < 0$ and light sees exponential absorption (Beer's Law). But if we can find a way to make $N_2 > N_1$, then we have condition known as **population inversion** $\Delta N = N_2 - N_1 > 0$ and light will experience exponential amplification. This is the basis of laser action.

**Energy-level systems for population inversion**. Population can be transferred from the ground state by external excitation (or pump) source that can be optical, electrical, or chemical depending on the physical material system. There are 3 main energy level systems that can be considered for laser action.




**2-level system**: population inversion is not possible as in thermal equilibrium the Boltzmann distribution demands: $\frac{N_2}{N_1} = e^{-h\nu/k_BT} < 1$, so you cannot make a laser from a 2-level system.

**3-level system** (e.g. ruby laser): here pump mechanism is from ground state to a third (higher-lying) level $E_3$ which rapidly and non-radiatively decays to the upper lasing level $E_2$. Population inversion is possible now, but over half of initial ground state population ($N_T$) must be transferred to upper laser level: $N_2 > N_T/2$ to exceed the ground state population. This requires high pump rate which is possible (e.g. ruby laser) but difficult and inefficient.

**4-level system** (e.g. nearly all lasers): In this system, pumping is again to a higher lying level $E_3$ that rapidly non-radiatively decays to the upper lasing level $E_2$, but the lower laser level $E_1$ is not the ground-state which is at a lower energy level $E_0$. Now production of inversion is easy, as lower lasing level $E_1$ is empty and any transfer to $N_2$ creates a population inversion. Good 4-level laser systems have a lower laser level $E_1$ with a fast decay rate to ground-state $E_0$, so it remains empty $N_1 \approx 0$ even when lasing occurs. Then inversion $\Delta N = N_2 - N_1 \approx N_2 > 0$ under all pumping, and laser needs only small pumping rate.

## Laser Material and Gain Equations

For a laser gain medium with a pumping rate R for producing upper-state population $N_2$ which has a spontaneous decay lifetime $\tau = {A_{21}}^{-1}$, we can write:

$$\frac{dN_2}{dt} = R - \frac{N_2}{\tau}$$

assuming no intracavity flux (no lasing). In steady-state (setting the equation to zero), the upper-state population $N_2 = R\tau$, is directly proportional to pumping rate $R$.

The growth rate of light intensity $I(z)$ at the lasing transition $h\nu = E_2 - E_1$ for an ideal 4-level system (with $N_1 \approx 0$) is given by

$$\frac{dI}{dz} = \sigma N_2 I$$

This equation has a simple exponential solution (assuming $N_2$ constant) and for a length of gain medium $L$: $I(L) = I(0)e^{g_0 L}$, where gain coefficient $g_0 = \sigma N_2$. All laser transitions have some bandwidth, so gain has a line-spread function: $G(\nu) = e^{g_0(\nu)L}$ where $g_0(\nu) = \sigma(\nu)\Delta N$ and the cross-section $\sigma(\nu)$ embodies the transition lineshape.

**Threshold condition for laser oscillation.** By placing the gain medium in a laser cavity to provide feedback the cumulative gain can be made to increase by multiple passes. For a cavity

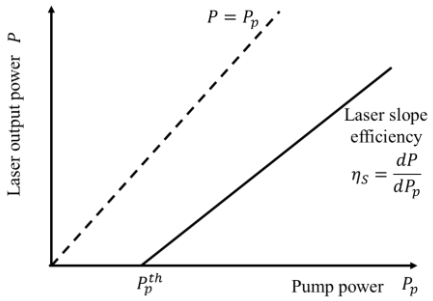
with two mirrors with reflectivity $R_1$ *and* $R_2$, the **round-trip condition for laser oscillation** (assuming no losses other than mirror transmission) is:

$$R_1 R_2 e^{2g_0 L} \geq 1$$

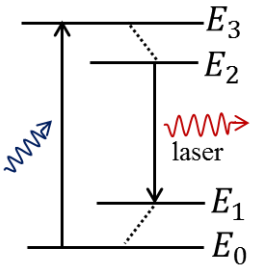
where $e^{2g_0 L}$ is the double pass (round-trip) gain. Usually, $R_1 = 1$ and the other mirror $R_2$ is partially reflecting to act as the output coupling for the laser output power. When the above equation is an equality this is the condition for **laser threshold** with threshold gain coefficient ${g_0}^{th}$ and with threshold inversion ${N_2}^{th}$. When pumping above threshold, the lasing causes the light to grow to a level that will "clamp" the inversion $N_2$ to this threshold value. The resultant laser output power has an equation of the form:

$$P = \eta_S \left(P_p - P_p^{th}\right)$$

valid for $P_p \geq P_p^{th}$, with a threshold pump $P_p^{th}$ required to overcome cavity losses to see nett gain. Above threshold pumping, there is clamping of the inversion to its threshold value and there is a linearity of output power against pumping above threshold characterised by a laser slope efficiency $\eta_s = dP/dP_p$ (the gradient of the power curve graph).




The laser slope efficiency $\eta_S$ will allows be less than unity. The most fundamental limit to efficiency is the photon energy efficiency. This is most easily seen with optical pumping, if every absorbed pump photon with energy $h\nu_p$ leads to an output laser photon with energy $h\nu_L$ this is the maximum possible laser slope efficiency given by:



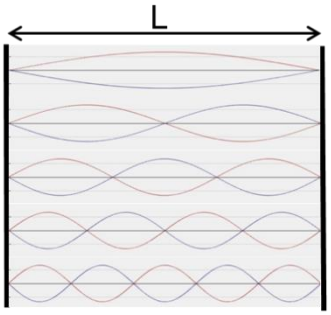

$$\eta_{ph} = \frac{h\nu_L}{h\nu_P} = \frac{hc/\lambda_L}{hc/\lambda_P} = \frac{\lambda_P}{\lambda_L};$$

There are normally several other factors that decrease the slope efficiency. The choice of lasers is based on finding systems where the efficiency is high.

## Laser Modes

The laser cavity confines light waves with defined end boundary conditions, and this leads to laser modes that are discrete (quantised) self-producing solutions after each round-trip.

**Spectral Modes**. In terms of spectral modes these are a set of discrete standing wave solution with laser cavity length equal to an integer number of half wavelengths (i.e. round-trip = integer number of wavelengths or integer number of $2\pi$ in phase).



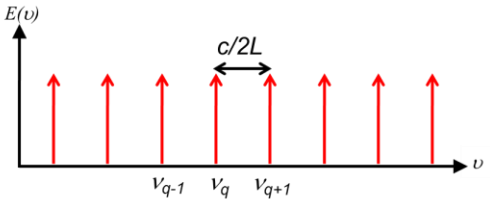

$$L = q\frac{\lambda_q}{2} = q\frac{c}{2\nu_q}$$

where $q$ is an integer with resonant standing wave discrete wavelengths $\lambda_q$ that must also be within the gain bandwidth of the laser transition $\sigma(\nu)$. Hence, since $\lambda_q = c/\nu_q$

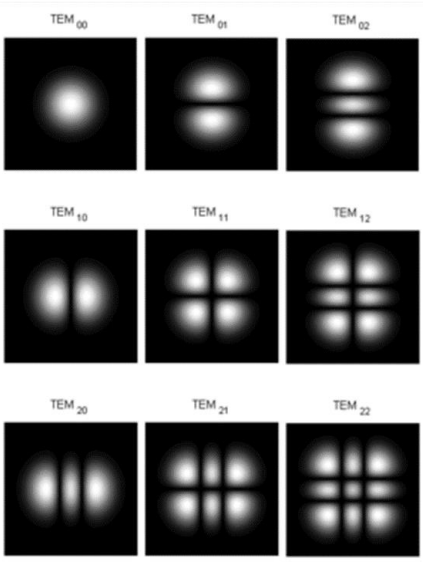
$$\nu_q = q\frac{c}{2L}$$

with equal mode frequency spacing

$$\Delta\nu = \nu_{q+1} - \nu_q = \frac{c}{2L}$$




**Spatial Modes**. The laser spatial distribution must also self-reproduce after a cavity round-trip resulting in spatial modes. The form of spatial modes is more complex to analyse theoretically, as you must invoke diffraction theory. However, we can gain insight by considering that light propagating in the cavity undergoes diffraction and that in the far-field light has the form of a Fourier Transform of the initial E-field distribution. This insight provides the clue that the **self-reproducing spatial modes are self-Fourier Transforms** i.e. after diffraction the spatial form of the light is unchanged.




The lowest order such self-Fourier Transform is a **Gaussian Beam**. Higher-Order laser modes that form a fuller set of self-Fourier Transforms are **Hermite-Gaussian beams** (in Cartesian coordinates – see diagram of a set of low order HG modes of which lowest order is Gaussian).

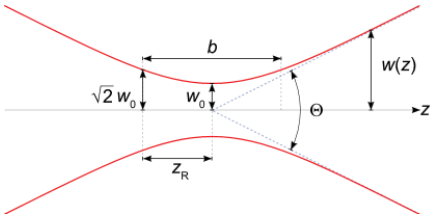
## Gaussian Laser Beams

The exact form for the Gaussian beam (fundamental mode of a laser) is derived from Maxwell's wave equation under a paraxial approximation, and can be written in the form:

$$E(r,z,t) = E_0\frac{w_0}{w(z)}e^{-r^2/w^2(z)}e^{-ikr^2/2R(z)}e^{i\phi(z)}e^{i(kz-\omega t)}$$

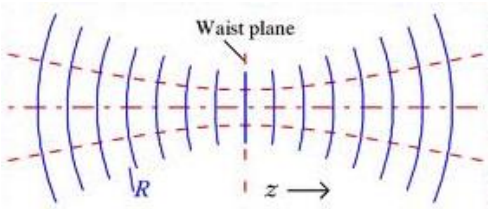
with waist size *w(z)*, radius of curvature *R(z)*, Guoy phase $\phi(z)$, and Rayleigh length $z_R$.

In amplitude, it has Gaussian transverse distribution given by: $|E(z)| = E_0(z)e^{-r^2/w^2(z)}$ with (1/e) waist size $w(z)$ that varies with propagation distance $z$ from a minimum waist size $w_0$ according to:




$$w(z) = w_0\left[1 + \frac{z^2}{z_R^2}\right]^{1/2}$$

The beam has spherical wavefronts $e^{-ikr^2/2R(z)}$ with radius of curvature: $R(z) = z + \frac{z_R^2}{z}$ and where Rayleigh distance in $w(z)$ *and* $R(z)$ equations is given by: $z_R = \pi w_0^2/\lambda$.




Rayleigh distance is of the form we discussed as the characteristic distance to be in the far-field. For Gaussian beam it more specifically defines where $w(z = z_R) = \sqrt{2}w_0$ and when $z \gg z_R$ (far-field): $w(z) \approx w_0\, {}^{z}/_{z_R} = {}^{\lambda z}/_{\pi w_0}$ allowing to define a far-field divergence angle: $\theta = \lim_{z\to\infty} \frac{w(z)}{z} = \frac{\lambda}{\pi w_0}$.

- Guoy phase term $\phi(z) = tan^{-1}\left({}^{z}/_{z_R}\right)$ is due to the "strange" form of the wavefronts R(z) near the minimum waist size (focal region around *z=0*) and is the difference in axial phase of the Gaussian beam compared to a geometrical plane (or spherical) wave.
- The pre-multiplier term $\frac{w_0}{w(z)}$ is for conservation of power. There is decreasing field amplitude as the beam expands (inverse square law).

## Laser Properties:

- **High spatial coherence**: low divergence and focusable to minimum spot size (required in many applications e.g. coupling to optical fibres; laser cutting)
- **Monochromatic**: can have very narrow spectral linewidth - near single frequency
- **Pulsed**: laser dynamics can be modulated to achieve very short and high peak power pulses with nanosecond, picosecond, and femtosecond pulse durations.
- **Broad Tunability**: Some lasers have large gain bandwidth allowing it to be widely tuned in wavelength (e.g. Ti:sapphire laser 650 – 1050 nm)

## Main laser types (in common operation today):

- **semiconductor (or diode) lasers** with conduction and valence band structure as the upper and lower "levels".
- **solid-state lasers** formed by a solid host doped with an active lasing species (e.g. Nd:YAG or ruby). Solid-state gain medium can be bulk form or as optical fibre.
- **gas lasers**. Two important examples are infra-red $CO_2$ lasers and ultraviolet excimer lasers.
- There are many other laser types: liquid; chemical; free-electron; X-ray.

| DIODE | SOLID-STATE | $CO_2$ (IR) | EXCIMER (UV) |
|---|---|---|---|
|  |  |  |  |

## Some key applications of lasers include:

- Industrial laser material processing
- Optical communications
- Medical surgery and diagnostics
- Optical sensing (e.g lidar for driverless cars)
- Defence and security
- Fundamental science