Fundamental Concepts

Course content introduction

The science of light is a discipline that appears to have found the secret of eternal youth. Every generation seems to come upon the subject and leave it with fresh discoveries, fresh ideas, and fresh inventions. As if to underline this statement, the Nobel Prize for Physics in 2005 was awarded to Glauber, Hall and Hänsch for work in the field of optics. The past 25 years have seen another revolution in optical developments. Lasers, a solution without a problem in my youth, are opening up a public communications revolution, new imaging techniques based in confocal microscopy, colour holography, extremely sensitive fluorescence time-resolved molecular probes and extraordinary advances in basic physics such as single atom traps or the study of light travelling at speeds of only metres per second. Lasers are intimately linked with quantum optics developments and optoelectronics, a subject closely tied in with the huge growth in applications of fibre-optics and fibre-optic based sensing probes. *Photonics* is the modern word that covers many of these developments and is now widely used where the word *Optics* was used before.

Below all this, the optics that has been discovered during the previous 300 years is a rich and diverse set of phenomena that goes to the very roots of our understanding of nature, on the one hand, and has produced superb tools that are used the world over in laboratories of all disciplines. Optical science isn’t just for laboratories, it’s a very public science. How many are fascinated by optical phenomena we see around us, from rainbows to colour mixing, or enjoy the chance to look through a microscope or telescope, or take pictures with a still or video camera? We can’t cover all the fascinating optics there is in a set of 24 lectures but I hope you’ll agree that the subject is well worth studying. It will make you think, it will show you the beauty in nature and it will equip you with knowledge that is extremely practical and widely useful.

Any well-stocked science library will have metres of books on ‘optics’ already on the shelves. Why should you keep reading these notes? Perhaps you shouldn’t. If you’re looking to solve highly technical problems in optics, then look at some of the many other books on the shelves. This course is about context. It is intended as a broad introduction to the subject that will give you insight into the historical, social and scientific context of the science of light. It is aimed both at those who will be going on to a more specialist course in physics in the future and also at those who will specialise in quite different areas of the physical, biological and medically related sciences. I’d like to think that the numerate liberal arts student will also find the course intelligible. By the end of the course, the conscientious student will be able to ‘do’ quite a lot of optical science. That is a useful skill to acquire but the central objective of the course is that you will be able to talk sense about light science. Everyone is interested in light. Optical ideas should be part of the gift of the scientifically literate to today’s culture. Optical ideas are the culmination of centuries of inquisitive minds probing the secrets of nature. The results are unexpected, thought provoking and powerful. They are well worth the time invested in them.

The numerical skill and mathematical knowledge expected in this course is pretty much that which any student of the sciences should have these days. It is ‘basic maths’. An appendix assembles all the maths results used at various points throughout the course. A few sections will be flagged as advanced. These are intended to set the scene for those going on to subsequent study in the field or who have already acquired more than the average mathematical background. There is one difficulty with optics that the older generation don't appreciate. The most natural way to present many optical arguments is through the use of diagrams. For the older generation, reading the relationship between different parts of a diagram is simple to the point of being intuitive. School children used to be drilled for years in Euclid's books on geometry and one of the pay-offs was the skill acquired in following geometrical arguments. Euclid's books have fallen out of the school syllabus, in Britain at least, and basic geometry isn't much to be found in Universities either. Unfortunately for this development, geometry and optics are intimately linked. It's surely no coincidence that Euclid himself wrote a treatise on optics, too. At least it is usually assumed that the same Euclid was responsible for both topics. The visual content of diagrams is so strong that this course will continue to present arguments via diagrams. If this experience makes you more confident in using diagrams, so much the better.

Finally, a few words on how to study this course. These notes accompany lectures that will use related PowerPoint presentations. Come to the lectures, listen and think about the subject during the lectures. Participate in lecture demonstrations. Make brief notes during the lecture if you like but the text here is intended to leave you free to engage your brain, not your pen. One popular technique is to read the relevant section of the notes in advance so that during the lecture you can reinforce what seemed clear and have a second bite at what's left you puzzled. Don't skip lectures and simply read the notes. You not only miss out on the personal explanation but you reduce the quality of the learning experience, making life unnecessarily difficult. It should go without saying that you shouldn't miss the accompanying computer class-room sessions.

Fundamentals

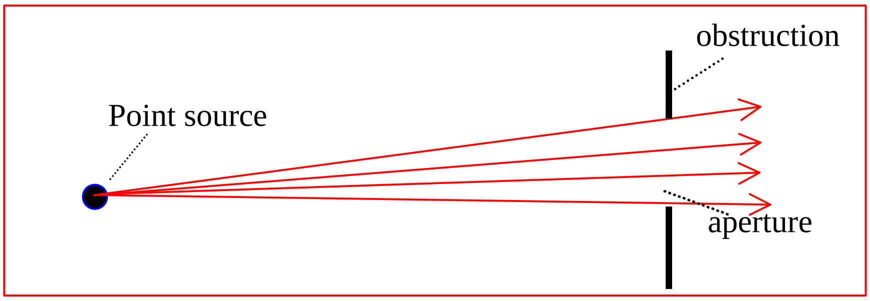
The *fundamentals* of this section are **straight-line propagation**, **reflection**, and **refraction** of light. "I know all about these", you may say, "we did them at school". Maybe so. I'll try not to dwell on the obvious but to concentrate on the new. The lectures, though, are intended to be self-contained and you will find here clear statements of the fundamental laws of light and how they relate to each other.

Rays of light

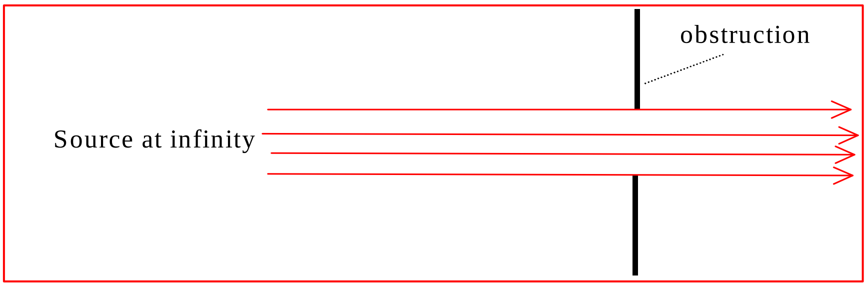
What 'everyday' phenomena convince you that light travels in straight lines? You should be able to come up with several suggestions. One piece of indirect evidence that I like to think is pretty convincing is that the perceived shape of objects and their relationship to each other appears the same as we alter our distance from them. This can only happen if light travels in straight lines. Many diagrams in this book include rays of light. A **ray** of light is the direction of propagation of light energy. You will not be surprised to find in most diagrams that the rays of light are straight lines.

For future use, I'll introduce the concepts of a pencil of light and a beam of light. These words describe propagating light in different circumstances. A **pencil** of light is a bundle of rays coming from a single point source of light and limited by some aperture. If the source is a very long way away, it is said to be **at infinity** and the pencil of light from a point on the source is a **parallel pencil**. It should be clear in the following sketch that all the rays in such a pencil are indeed parallel. Finally, a **beam of light** is a collection of pencils from an extended source, as illustrated in the accompanying figure.

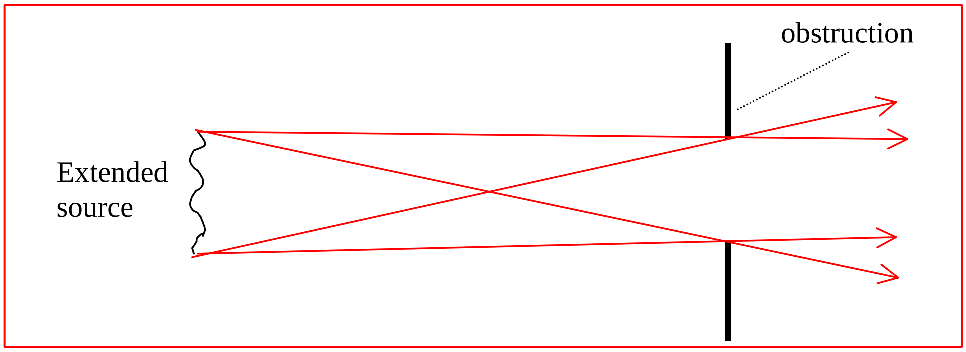
A **pencil** of rays



A **parallel pencil**



A **beam** of rays



I think it was Richard Feynman who said that optics was either simple or very complicated. In my experience there is more to everything you see than meets the eye. Even apparently simple optical ideas introduced at school have great depth. However, if we want to make progress, we mustn't stop at every opportunity.

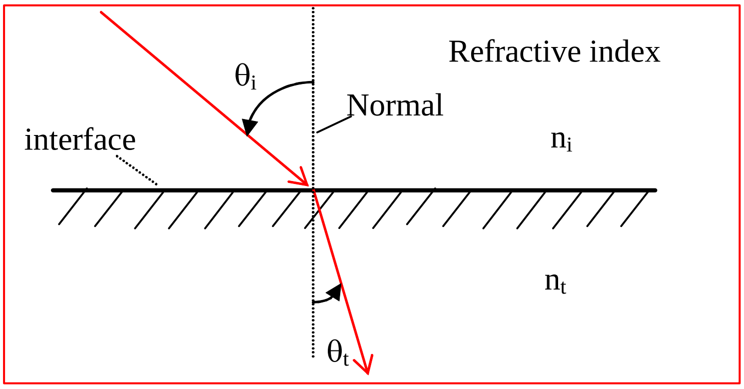
The relationship between waves and rays in optics is fascinating, and one that is especially relevant in today's world where lasers and coherence properties of light are very important. What I want to do in these opening lectures is to show you that sometimes a wave description is the most natural way to understand what is happening, and sometimes a ray description. The really fundamental properties like propagation, reflection and refraction can be described in both terms. Light is neither one nor the other. Light is light; the rest is analogy.

Refraction

Refraction is the bending of light at an interface between two media. What everyday phenomena illustrate refraction? Again, you shouldn't be stuck for a few ideas, from simple demonstrations like the apparent bending of a stick placed at an angle into water to the more complex imaging of lenses. You have probably met refraction in our first year physics lab, if not a long time before.

Snell's Law

A change in the direction of propagation of light at an interface is described by the bending of a ray of light at the surface. This alters the angle that a ray makes to the perpendicular to the surface, a direction called the surface **normal**. In the diagram below, θ_i is known as the **angle of incidence**. After transmission through the surface, the light ray makes an angle θ_t, known as the **angle of transmission** or, more commonly, as the **angle of refraction**. The relation between these two angles is known as Snell's law, named after its discoverer Willebrord Snell (1591 – 1626) of Leyden.



Snell's law is a key result in optical science. It tells you what is going on in the phenomenon of refraction. With Snell's law you can explain how our eye lenses form images, how any lens in a microscope, telescope, camera, spectacles or optical device of your choice forms an image. You can explain why a swimming pool appears shallower than it really is or such natural phenomena as rainbows and mirages. It's scarcely an exaggeration to say that a succession of scholars at least had been looking for the explanation of refraction for some 2000 years. How ironic, then, that Willebrord Snell did not appreciate the importance of his discovery and died a young man without realising that he had carved his name in stone in the history of physics.

What is Snell's law? In modern notation, Snell's law says that "$n \sin\theta$" is conserved as the light ray crosses the interface, where n is a constant associated with a transparent medium called its refractive index. θ is the angle the ray makes with the surface normal, as above, and $\sin\theta_i$ is the trigonometrical function "sine θ_i ". In other words, **Snell's law** can be written:

$$n_i \sin\theta_i = n_t \sin\theta_t$$

where n_i is the refractive index in the medium that the light comes from (the medium with the **incident ray**) and n_t is the refractive index of the medium containing the **transmitted ray**.

Snell's law is a summary of experimental fact. It can be deduced from more than one underlying model of the behaviour of light. Indeed, Descartes, who publicised Snell's law quite soon after its discovery in the 17th century, gave several mutually conflicting deductions, all of which are now ignored except by historians.

The **refractive index** of a medium is a measure of the **speed of propagation**, v, of light in the medium. This speed is related to the speed in vacuum by the factor of n, the refractive index. Thus:

$$v_x = c/n_x \quad .$$

'x' can label any medium. For example, you might talk about n_{air}, n_{water} or n_{glass}. The bigger the refractive index, the slower the speed of propagation. Most transparent media you are likely to come across have a refractive index of between 1 and 2, implying that the speed of propagation is up to two times slower in these media than it is in vacuum. If you look back at the earlier statement of Snell's law, you will see that what really matters is just the ratio of the speed of light on either side of the surface, i.e.

$$n = n_t/n_i \ .$$

n is sometimes called the refractive index of the interface. In terms of n, Snell's law is written as

$$\sin\theta_i = n \sin\theta_t \ .$$

It is in this form that you often see Snell's law in textbooks. Snell himself didn't discover the concept of refractive index but he did realise that the ratio of the sines of the angles in either media was a constant for different angles of incidence, as the equation above says.

Refraction phenomena will recur throughout this course and I shan't labour the point with worked examples here.

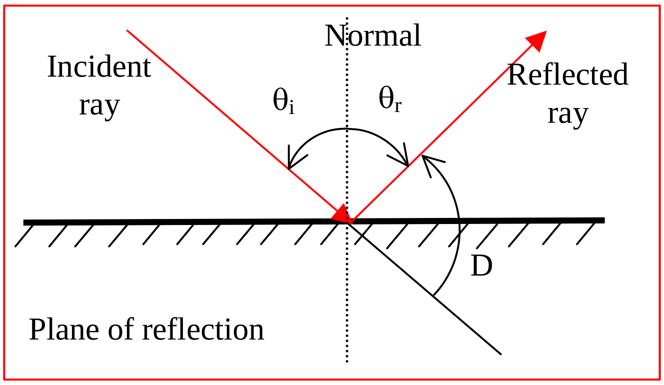
Reflection

Johannes Kepler (1571 - 1630) was one of the founders of modern astronomy. He also wrote two seminal books on optical science in which he laid out the foundations of the subject in a recognisably modern way. Earlier writers and thinkers had concentrated on trying to unravel the nature of light and explain optical phenomena as a whole. They asked questions like: "how does vision work?"; "why do we see small images in a clear glass ball?"; "what relation is the image in a mirror to reality?". Kepler realised that it is more productive to begin by asking simpler questions, such as what happens to a ray of light in different circumstances. Images are not described as entities that behave as a whole but as a compound production of rays of light. Describe how the rays behave and the properties of images will follow. This is essentially the modern method.

We begin by following Kepler's sound approach by looking at the reflection of a ray of light. See the diagram below. A reflected light ray obeys two laws:

1. $\theta_r = -\theta_i$, noting the direction of measurement. θ_r is the angle of reflection.

2. The incident ray, normal and reflected ray are in the same plane, called the **plane of reflection**.

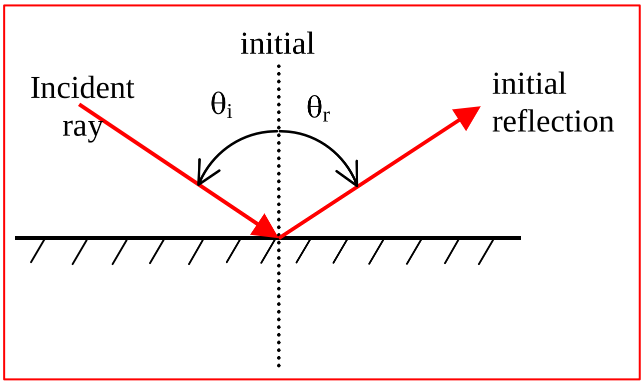


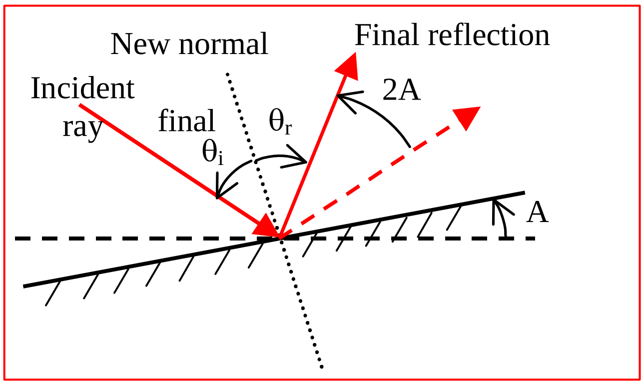
Various consequences follow.

- The deviation, D, of a reflected ray caused by the reflection is: $D = 180° - 2\theta_i$

- The optical lever:

Tilt a mirror about an axis perpendicular to the plane of reflection, through an angle A in the same direction as θ_i.





The change in θ_i is written $\delta\theta_i$. From the diagram, $\delta\theta_i$ = -A, because the normal has twisted through angle A in such a direction as to reduce θ_i . Hence,

$$\delta D = -2 \times \delta\theta_i = 2A.$$

i.e. the reflected beam twists through twice the angle of tilt of the mirror, which is **the optical lever effect**.

What happens when the axis of rotation is not perpendicular to the plane of reflection? The optical lever effect is diminished by an amount that depends on the orientation of the rotation axis. You can easily see that in the special case when the rotation axis lies parallel to the normal, then the angle of deviation is unaltered by rotation, because the reflecting plane is simply rotated unchanged.

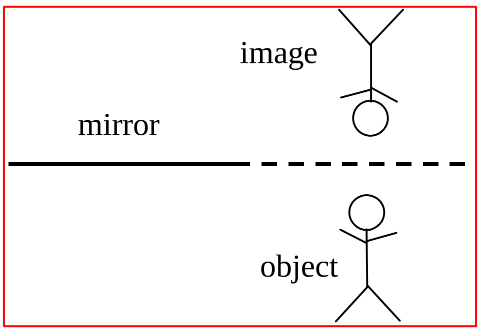
Some applications of the optical lever are discussed in the lecture. They include the historically important optical galvanometer, the laser printer and one form of modern high-brightness data projector.

- [Aside for mathematicians: The law of reflection can be considered mathematically as Snell's law, with

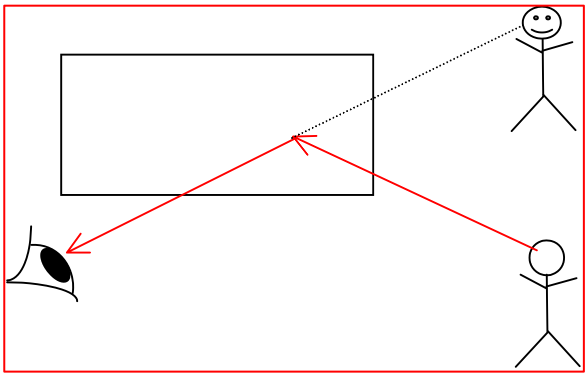
  $n_t = -n_i$; $\theta_t \rightarrow \theta_r$].

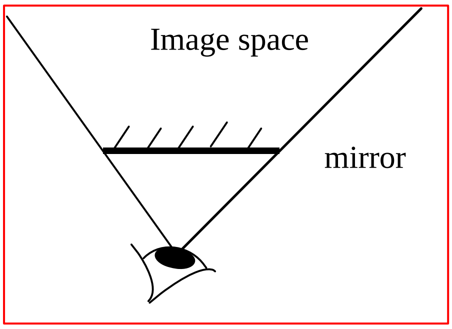
Applications and occurrences of reflection

Plane mirrors produce images with particular properties:



- Location of image. The image is as far behind the plane of the mirror as the object is in front.

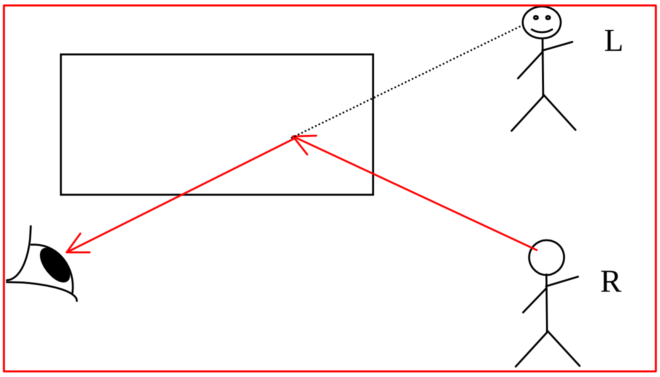




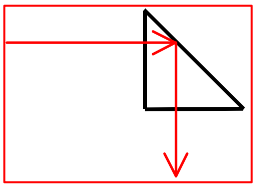
- Field of view. To decide how much you can see in a mirror, it is very helpful to understand the useful idea of **image space**. Image space is the set of coordinates where an image might be located. The mirror acts as a window into image space. This is shown in the illustration.

  An alternative view for deciding how much you can see in a given mirror is to imagine that your eye is in the position of its mirror image (in image space) and is looking from image space through the mirror into your real surroundings (called object space in light science). Again, the frame of the mirror defines how much can be seen, namely how much the image of your eye can see in object space. In other words, you get a view of the scene as if you were located where your image is and are looking at the scene through the 'window' of the mirror.

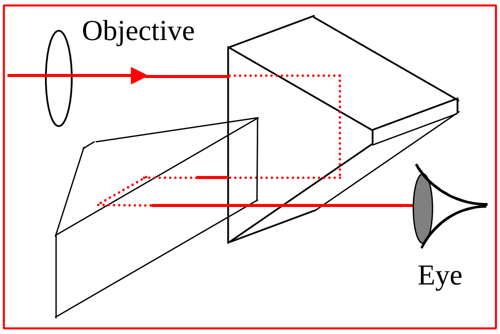
- Every reflection changes the handedness of the image. An even number of reflections leaves the handedness unchanged; an odd number changes it.



- Image forming light passes through a 90°prism, as shown in the diagram below. *Is the handedness of the image changed*?



- *How many reflections are there in traditional prism binoculars, whose components are illustrated on the right*?



- An overhead projector has only one mirror. *Why do written overheads not appear as mirror reflections on the screen*?

Light as a wave phenomenon

So many everyday optical phenomena can be described and explained using light rays that you may be forgiven for thinking that you only need to introduce waves when talking about a few uncommon phenomena, like the interference and diffraction of light. This is not true, on two counts. First, interference and diffraction phenomena are all around us, if we know where to look (for example in the colours of soap bubbles, starling’s wings, beetles, some butterflies, supernumerary rainbows and much more). Secondly, the fundamental phenomena of propagation, refraction and reflection can be equally well accounted for as wave phenomena.

I’m assuming that most in the class will have met the key concepts used to describe waves but, even so, a refresher wouldn’t go amiss.

A digression on waves

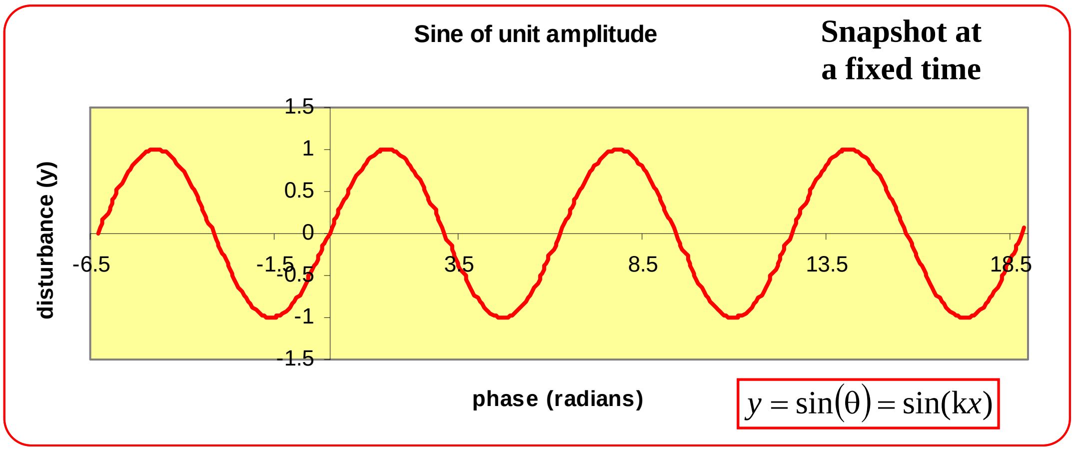
Waves are a dynamic, moving phenomena. They occur in space and time. They are difficult to represent on a diagram. Look at the waves on the sea and you’ll see a pattern that never repeats in detail. What are the essential features of a wave? They would seem to be hard to pin down. It was the great Frenchman Jean Baptiste Joseph Fourier (1768 – 1830) who provided the key idea that makes discussing wave phenomena much simpler than it would otherwise be. Fourier realised that arbitrarily complex wave patterns can be described in terms of a sum of simple waves, known as sine waves. (In his own words of 1811: *Tout mouvement périodique peut être regardé comme la superposition d’un certain nombre de mouvements sinusoïdaux partiels*).

Sine waves had already been investigated by the mathematicians. They are easy to understand and can be described by only a few ideas. So we only need to understand sine waves then? Right? Yes. Complicated periodic waves are treated as a Fourier series of sine waves, namely a sum of harmonically related waves. 'Harmonic' just means that each wave has a frequency that is an integral multiple of the lowest frequency present. The most general aperiodic waves are treated as Fourier transforms. Neither Fourier series nor Fourier transforms are part of this course but both techniques are relevant to several phenomena in optics.

Sine waves come into Standard Grade Physics, so they must be quite simple to describe. They are. You can describe basic sine waves with just three numbers: their amplitude (how big they are), their wavelength (how stretched out they are) and their frequency (how rapidly they oscillate). That’s only one more number than you need to describe a straight line. Here is a very brief reminder of the language and symbols used when talking about sine waves.

Sine waves have no beginning or end, in space or time. The disturbance is represented by a sine function.

- $y = \sin(\theta)$ varies periodically between ± 1, repeating as θ increases by 360°, or 2π radians as it is usually described in physics and mathematics;
- $y = A \sin(\theta)$ varies between $\pm A$; A is called the **amplitude** of the disturbance, θ is called the **phase** of the sine term

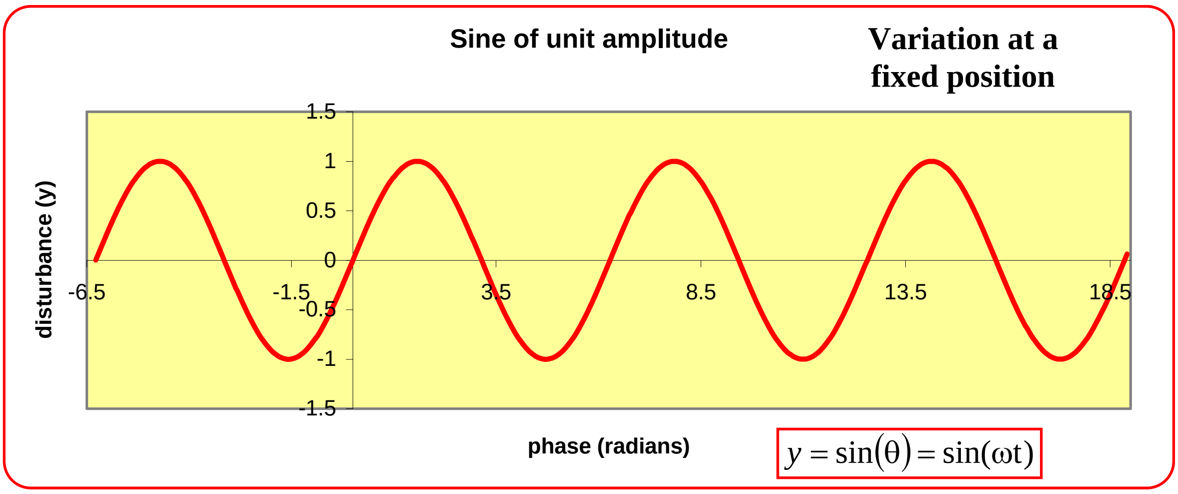


- You can 'freeze' what a wave looks like by plotting a section of it at a fixed time, like taking a photograph with a fast shutter speed. The picture above shows a sine wave:

The distance after which a wave repeats is called its **wavelength**, as you know, and is usually represented by the letter λ (Greek "lambda"). Wavelengths are measured in metres (**m**) or a related unit like nm. What's special about sine waves is that they each have a distinct wavelength.

What we have plotted above is represented by y = sin(θ), where θ is in radians. To make that represent a snapshot of a wave, the horizontal axis must also be in radians in which case the above is written y = sin(kx). Since the wave repeats in space every wavelength λ and in phase every 2π, then $k\lambda = 2\pi$ and hence the constant $k = 2\pi/\lambda$. This is an important relationship and defines the **angular wavenumber** k of a wave. Angular wavenumbers so defined are measured in units such as rad m^{-1} or rad mm^{-1}.

- A wave isn't just a wiggle in space, it's a wiggle in time too. Likewise, we could plot what happens if we measure the disturbance at a fixed position, some definite x, with time. With a sine wave passing a fixed position, we get just the same sort of curve:



The time after which a wave repeats is called its **period**, as you know, and is usually represented by the letter T and measured in seconds (**s**) or a related unit like μs. What's special about sine waves is that they each have a distinct period. The number of repetitions per second is known as the **frequency**, measured in Hertz (**Hz**) and usually represented by the letter f, or sometimes the Greek letter ν ("nu"). If you think about the basic concepts of period and frequency, you will see that $f \equiv \nu = 1/T$. One of Fourier's great insights was that the very concepts of wavelength and frequency apply only to single sine waves. Waves of other shapes that repeat regularly contain more than one frequency and more than one wavelength.

What we have plotted above is represented by $y = \sin(\theta)$, where θ is in radians. To make that represent the disturbance produced by a wave at a point, the horizontal axis must also be in radians, in which case the above is written $y = \sin(\omega t)$. Since the wave repeats in time every time interval T and in phase every 2π, then $\omega T = 2\pi$ and hence the constant $\omega = 2\pi/T$. This is an important relationship and defines the **angular frequency** ω (Greek "omega") of a wave. Angular frequency is measured in **rad s^{-1}**. You can see that $\omega = 2\pi\nu$.

- Putting all the above together, you'll not be surprised that a general way of writing a sine wave is $y = A \sin(kx - \omega t)$, representing a sine wave travelling in the positive x direction. A is the amplitude of the wave; $(kx - \omega t)$ is the **phase of the wave**. This expression will help you see what's going on in several optical phenomena that we'll introduce later on.

- When you watch a wave travelling, you probably follow a peak, or a trough, or the middle of the wave. What you follow is a point of constant phase. Since all the parts of a sine wave travel together, you can focus on any bit to decide the speed (or velocity) of the wave.

For example, look at the succession of points that keeps the phase at 0. These are connected by the relationship $(kx - \omega t) = 0$. One point where the phase is zero is clearly the point $x = 0$ at $t = 0$. At a time t later, the point of zero phase has travelled to position x, where $(kx - \omega t) = 0$, or $x = \omega t/k$. Hence the velocity, v, of a point of constant phase is:

$$\boxed{\mathrm{v} = \frac{\text{distance travelled}}{\text{time taken}} = \frac{x}{t} = \frac{\omega}{k} = \lambda \mathrm{f}}$$

- One immediate consequence of this very general relationship for waves is that the wavelength in vacuum $\lambda_{vac} = c/f$ is given simply in terms of the reciprocal of the frequency. The higher the frequency, the shorter the wavelength.

- Another consequence is that the wavelength in a medium, λ_{medium} is less than the wavelength in vacuum by a factor equal to the refractive index of the medium. This follows because:

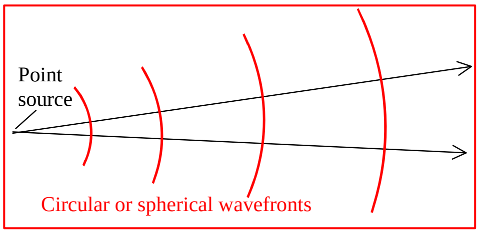
$$\boxed{\lambda_{\text{medium}} = \frac{\mathrm{v}}{\mathrm{f}} = \frac{c}{n\mathrm{f}} = \lambda_{\text{vac}}/n}$$

- Coming back briefly to Fourier's ideas, remember that a general wave can be expressed as a sum of sine waves. If all the component sine waves travel at the same speed, then the general wave will travel unchanged in shape. If the component sine waves travel at different speeds, then the shape of the wave will change with time.

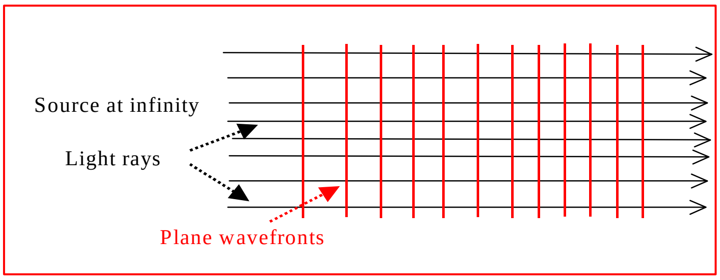
- Finally, in two or three dimensions you can watch waves travelling outward. The lines of maxima (or any other well defined phase) define the **wavefronts** of the wave. They typically have a characteristic shape, like expanding circles in 2D, spheres in 3D or, for waves travelling uniformly in one direction they are simply planes. Most of the waves of interest to us will be **plane waves**.

- We can now make the link between waves and rays:

  Rays are perpendicular to wavefronts in uniform isotropic media.



Huygens' Principle



Christiaan Huygens was a man of great ability and insight. His principle, **Huygens' Principle**, is one of the oldest pieces of 'modern physics' and one of the few advanced ideas from the 17th century that has survived still intact today. It is applicable far more widely than to the problem of optical refraction, though this was the phenomenon that he formulated his principle to explain.

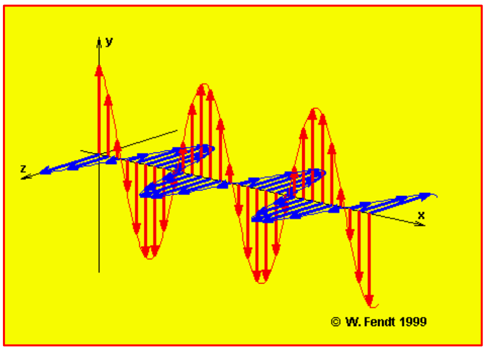
The principle is a *geometrical recipe* for deducing the propagation path of a wave (it doesn't have to be light), when the wavelength is not important. The principle is *inductive*. Here are the steps (which aren't spelt out this clearly in Hecht's *Optics*).

1) Take the wavefront at some time.
2) Treat each point on the wavefront as the origin of the subsequent disturbance.
3) Construct a sphere (circle) centred on each point to represent possible propagation of the disturbance in all directions in a little time.
4) Where the confusion of spheres (circles) overlap, the possible disturbances all come to nought.
5) The common tangent of the system of spheres (circles) defines the new wavefront a little time later.
6) Starting with the new wavefront, the construction goes back to step 2 to see where the wavefront reaches a little later on; and so on… .

When Huygens' principle is applied to simple circumstances like refraction and reflection, it predicts the correct law for both. [See s*lides and lecture discussion with Java applets*].

Equally importantly, it implies that for a transparent material both reflection and refraction necessarily occur together. It also embodies the modern quantum concept that the direction in which radiation travels through a medium is dictated by the sum total of the responses of all the electrons in the system to the presence of the incident radiation.

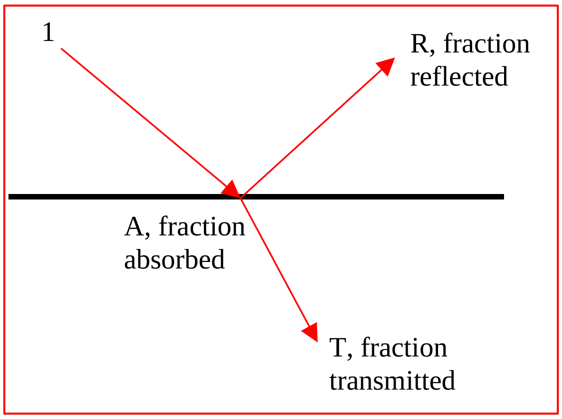
If Huygens’ wave description explains so many optical phenomena, then you might think it is conclusive proof that light must be a wave phenomenon. The one thing wrong with this line of reasoning is that we shall soon see that the same phenomena can also be described from a particle and ray viewpoint. However, the investigation of diffraction, interference and polarisation phenomena in the first half of the 19th century provided convincing evidence that light behaved as waves and James Clerk Maxwell, professor here at Marischal College, Aberdeen, in the 1850s was the first to realise and point out that light was an electromagnetic wave.



The term 'electromagnetic wave' means that when the wave is the present there is of both an electric field and a magnetic field disturbance. For a sine wave in space, both these fields vary in phase sinusoidally in magnitude and are oriented at right angles to the direction of propagation. For this reason, light is said to be a *transverse wave*. In addition, the magnetic field is at right-angles to the electric field. All this can be seen in the simulation shown in the course presentation, which is available on the web from Walter Fendt's page of Java applets.

The fraction reflected and the fraction transmitted are not given by Huygens' Principle. What can be said very simply but very importantly is that conservation of energy implies that the three possible outcomes for light landing on a surface must account for all the incident energy. The three outcomes are *reflection*, *transmission* and *absorption*. In symbols:

$$R + A + T = 1 .$$

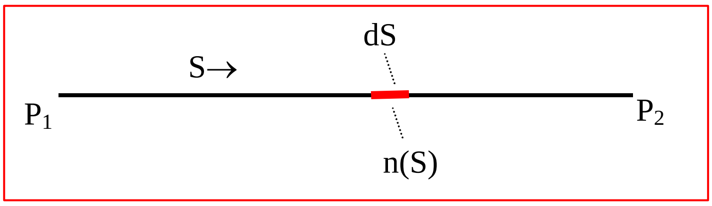


It was the brilliant Frenchman Augustin Fresnel, in the early 19th century, who worked out what fractions are reflected and transmitted, making the analogy between light waves and elastic waves. We call these fractions the *Fresnel Coefficients*. These coefficients depend on the polarisation of the light, as we shall see much later in the course.

Modern optics treats reflection and refraction as a boundary value problem involving electromagnetic waves, following the work of James Clerk Maxwell. As experimental physics extends to a very wide range of wavelengths (e.g. at synchrotron radiation facilities, NASA's Chandra X-ray space telescope or ESA's X-ray multi-mirror version, etc.) there is an increasing demand for 'optics' over the range of wavelengths from microwaves to x-ray regions of the spectrum. It is important we know the physics behind reflection and refraction so that we can predict what will take place in spectral regions where we ourselves can't see.

Huygens also looked at anisotropic refraction and produced the modern explanation of a complex phenomenon of great relevance to mineralogists and modern optical device technologists. We shall meet some of his ideas on this subject later in the course.

Optical path length



The concept of *optical path length* is one that is not only central to a different way of looking at light propagation but will also be useful in practice in several later chapters. The scenario is that we shall look at the path taken by light getting from one point to another. These points are labelled P_1 and P_2 in nearby diagrams. That path may simply be a straight line or it may have kinks in it due to reflection and refraction. The difference between optical path length and a straightforward measure of the actual distance along a light path is that the optical path length takes into account the slowing of light by the refractive index of the medium through which the light is passing.

The refractive index of a medium is just a measure of the slowing factor for light travelling in that medium. If the refractive index is as large as 2, then this means that light travels at half the speed it does in vacuum; if n = 1.5, then light travels at two-thirds the speed it does in vacuum, and so on.

The optical path length in any small region is the physical path length multiplied by the refractive index. If S is a measure of how far along the path you look, in symbols:

$$d\,(OPL) = n(s)\,dS$$

$$\therefore OPL = \int_{P_1}^{P_2} n(s)dS\ .$$

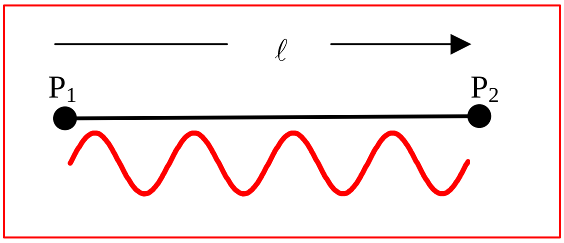
By way of analogy, consider how such a slowing factor might be useful in describing a journey that covers different kinds of roads. Say that on a fast road you can travel at full speed. On a bypass road you can travel at two-thirds full-speed and for the last part of the journey you can travel at only one-quarter full-speed. The corresponding slowing factors are 1.5 and 4. The *journey length*, by analogy with the optical path length, is defined for any part of your journey as the actual distance you are going multiplied by the slowing factor for the road you are travelling on. If your journey is 10 km by fast road, 6 km on the bypass road and 2 km on the slow road then the *journey length* by our definition will be 10 + 6*1.5 + 2*4 = 27 km. This means that the total journey time is just what it would have taken to travel 27 km on fast roads only. If an alternative route has only 3.5 km on the bypass instead of 6 km but an extra one km on the slow road, is it faster? For this alternative route, the *journey length* would be 10 + 5.25 + 12 = 27.25, a little bit longer and hence the journey will be slower.

The *optical path length* is the very same concept applied to optical paths. The next sections show how it works.

- *Propagation time*

The time, *t*, for a wave to propagate along a path P_1P_2 through a medium is determined by the optical path length and the speed of light in vacuum.

$$\boxed{dt = \frac{dS}{v(s)} = \frac{n(S)dS}{c} = \frac{d(OPL)}{c} \qquad \therefore t = \frac{OPL}{c}}$$



- *Number of wavelengths in path $P_1 \rightarrow P_2$*

The number of wavelengths in the path P_1P_2 is determined by the optical path length and the vacuum wavelength.

Remember that $\lambda(S) = \lambda_{vac}/n(S)$. For the small path length d(S)

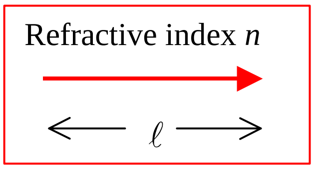
$$no.\ of\ \ wavelengths\ \text{is} = \frac{d(S)}{\lambda(S)} = \frac{n(S)dS}{\lambda_{vac}} = \frac{d(OPL)}{\lambda_{vac}}$$

$$\therefore the\ total\ \ no.\ of\ \ wavelengths\ \text{in}\ \mathrm{P}_1 \rightarrow \mathrm{P}_2\ = \frac{OPL}{\lambda_{vac}}$$

[Alternatively: no. of wavelengths = no. of cycles = (time of travel) × v = OPL × v/c = OPL/λ_{vac}].

- *Optical path length and the phase change along a path*

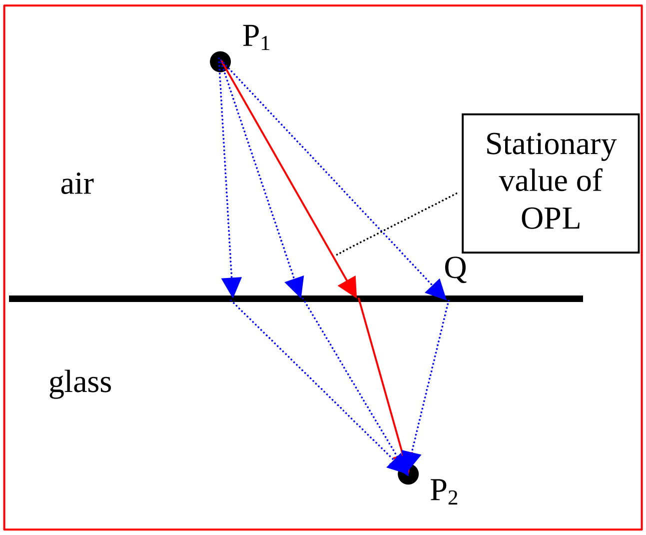
The phase change along path P_1P_2 is determined by the optical path length and the vacuum angular wavenumber.



In vacuum, a light wave changes its phase by 2π radians (≡ 360°) for each wavelength, λ_{vac}, in its distance travelled.

Hence in a path length ℓ, the phase change is $2\pi\ell/\lambda_{vac} = k_{vac}\,\ell$.

In a medium of refractive index n, the phase change in a path length ℓ is "$nk_{vac}\,\ell$" or OPL × k_{vac} (*Hecht* uses "$k_o n\,\ell$").



We shall need this result later on.

Fermat's Principle

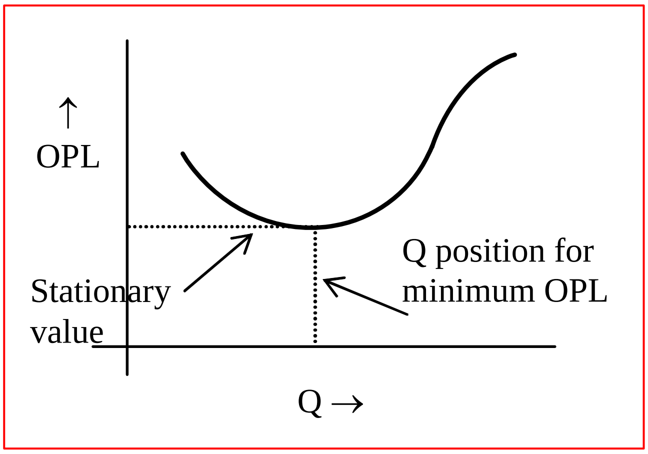
Fermat's Principle is a variational principle. Such principles play a fundamental role in several branches of modern physics. In optics, Fermat's principle provides a fresh outlook. See Hecht's *Optics*.

The principle:

Of all the geometrically possible paths that light could take between point P_1

and P_2, the actual path has a stationary value of the OPL.

A stationary value means that the rate of change of the value is zero. E.g. there is a minimum (usual case) in the OPL, or a maximum or, occasionally, a saddle point in 2 dimensions. If the OPL is stationary, then so is the time of travel for light. Hence in the accompanying picture, the actual path taken by light is the one that takes the minimum time. This is the case for Snell's law and for the circumstance of equal angles of incidence and reflection required by the law of reflection.



See Hecht's *Optics*, or other texts, for an evaluation of the geometry. For variation of the point Q, the position giving a stationary value gives Snell's law. Similarly for reflection and the laws of reflection. See the 'Albert' suite of software in the computer classrooms for a useful simulation illustrating Fermat's principle, and also the Java simulation in the lecture.

Digression

To solve the problem of the burning tepee on the slide, you need to think in the same way as for Fermat's principle. What is the shortest path if the brave can run at the same speed both before and after he reaches the river? The red lines show that it is just the path taken by a light ray coming from a lamp held by the brave if it were to be reflected at the bank onto the tepee. The problem is geometrically the same as if the brave were on the other side of the bank, starting at the dot, and now it is obvious that the straight line from the dot to the tepee is the shortest path.

However, if the brave laden by the bucket of water can travel only at half the speed with it, then he is better to travel further downstream without the bucket and take a shorter route with the bucket. How much further downstream? The geometry is just the same as for a light ray being refracted into a medium where it can only travel at half speed, i.e. a medium with refractive index 2. Snell's law tells you the relationship between the angles θ_1 and θ_2 in the diagram. The brave would be wise not to try to work out exactly where he has to go but to make a rough guess, for finding the precise point on the bank that makes Snell's law true isn't simple. Light, of course, does it without thinking!

Implications of Fermat's principle

- Note immediately the *principle of reversibility* of light rays is implicit, namely that if a ray propagates from $P_1 \rightarrow P_2$ along a known path, then the reverse path is the path of propagation from $P_2 \rightarrow P_1$.

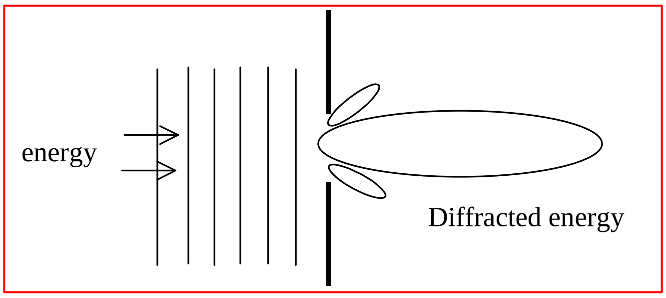
- The paths of rays through a lens from one object point to one image point must have the same optical path. You can see why a lens has a tapering shape. This result has important implications in understanding lens imaging; in particular it is implicit in our use of a lens to show Fraunhofer diffraction, as we'll meet later in the course, and also in the lab in PX2505 for those who take it.

Departures from geometrical optics

- *Diffraction*: the propagation of light waves around obstacles.

Diffraction will be a topic in the latter part of this course.

- Likewise the wave phenomenon of *interference* will come later.



- *Quantisation of illumination*

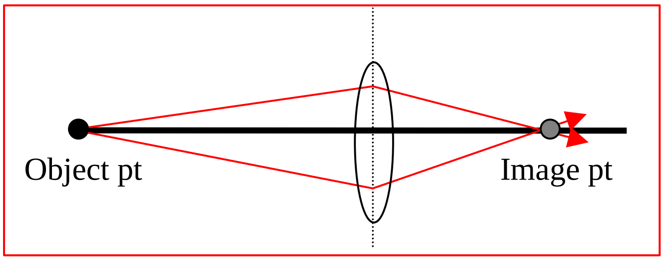
The geometrical view is that if a uniform pencil of light is restricted to a fraction of its original size or amplitude, then the illumination will be reduced by the same fraction. This takes no account of the granularity of light energy. This granularity says that light energy is always bundled up as *photons*, each with energy E that depends on the frequency, ν, of the light.

$$\boxed{E = h\nu}$$

where *h* is Planck's constant. Hecht's figure 1.8 shows evidence for the existence of photons, illustrating a picture being built up slowly in very poor illumination. In some circumstances light sources are intrinsically weak, such as light coming from a distant galaxy, and the arrival of light in photons is conspicuous.

Planck's constant *h* is one of the 'fundamental constants of physics'; in fact it is the most recent fundamental constant to be recognised and that was over a century ago now. In SI units Planck's constant is very small, 6.626×10^{-34} Js, but it is nonetheless very important. Max Planck was awarded the 1918 Nobel Prize in Physics *in recognition of the services he rendered to the advancement of Physics by his discovery of energy quanta*. It was thoroughly deserved.

The *photoelectric effect* provided the clinching evidence for photons, as Einstein realised. Photons are part of the world of quantum optics. They are a particle concept, whose main attributes are **energy** and **momentum**. Photons are the natural language to use when discussing why solids emit light, how lasers work, how photomultipliers, image intensifiers, CCD cameras and many other devices of quantum optics work. All these phenomena will be discussed much later in the course.

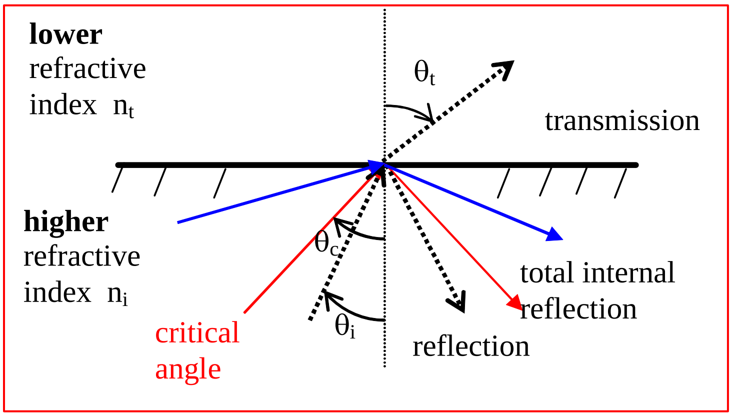


It was another Frenchman, Louis de

Broglie, who in 1924 realised how the photon description and the wave description were related. The connection is very simple. The *momentum*, *p*, of a photon is related to its wavelength, λ, also through Planck's constant:

$$\boxed{p = h/\lambda} \quad .$$

Total Internal Reflection



Total internal reflection is a phenomenon that takes place when light travels from one medium to another **of lower refractive index**. For example, the light must be coming **from glass** out **into air**, or **from** water **into** air, or from glass into water and so on. In addition, not all such light is totally internally reflected, only light that strikes the interface between the two media at a sufficiently large angle. How big an angle? We'll find out now.

- Internal reflection isn't a phenomenon that suddenly happens. In the reflection scenario shown in the figure above, there is a progressive rise in internal reflection with increasing angle of incidence, θ_i. When θ_i is small, there is both transmission and internal reflection (see the diagram). As θ_i increases, the transmission becomes nearly parallel to the surface and the fraction internally reflected becomes bigger and bigger. At the incident angle that mathematically gives transmission parallel to the surface (θ_t = 90°), the transmitted ray has no energy and all the energy goes into the internal reflection. At this point, the angle θ_i is said to be **the critical angle**, θ_c, shown in red on the diagram. Total internal reflection takes place at this angle and all larger angles of incidence.

The critical angle is easily calculated using the facts above:

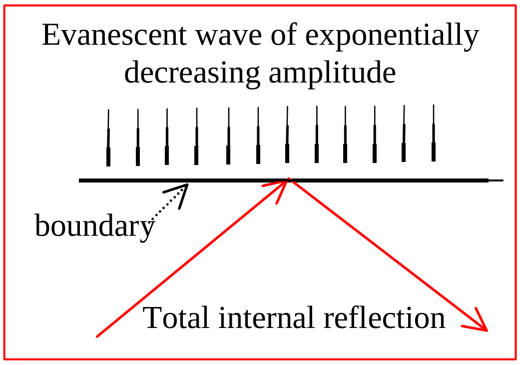
$$\boxed{\begin{aligned}
&n_i \sin\theta_c = n_t \sin 90^o \quad \text{Snell's law} \\
&\therefore \sin\theta_c = \frac{n_t}{n_i} = \frac{1}{n} \quad if \;\; n_t = 1 \\
&\therefore \quad \theta_c = \sin^{-1}(1/n) \quad n \text{ is the refractive index of the incident light medium}
\end{aligned}}$$

e.g. if n = 1.6, θ_c = ? Answer 38.7°.

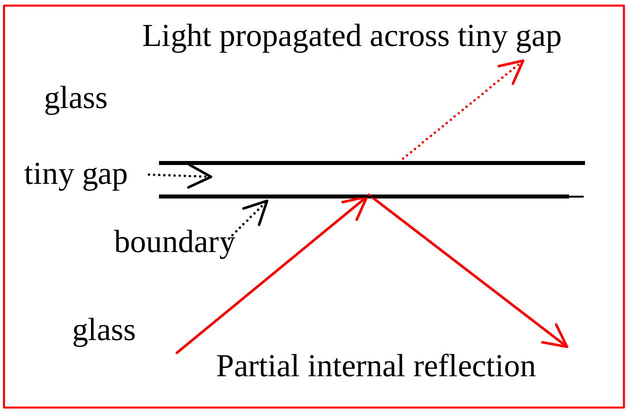
- Notice that for all $\theta_i > \theta_c$, total internal reflection occurs.

- Total internal reflection occurs both naturally and by design in optical equipment. Can you think of any examples? A few areas are: reflecting prisms; fibre optics; light guides (illuminated fountains, motorway signs, etc.).

- *The evanescent wave*



The evanescent wave is a piece of physics that used to be considered slightly exotic but which has now come in to the realm of the useful and important. There is a difference between reflection at a metallic surface, like aluminium, and an insulating surface like glass. In the metallic surface the electric field has to be zero inside the metal; at the glass/air interface there is no reason why the electric field has to be zero beyond the reflecting surface. In fact it is not.



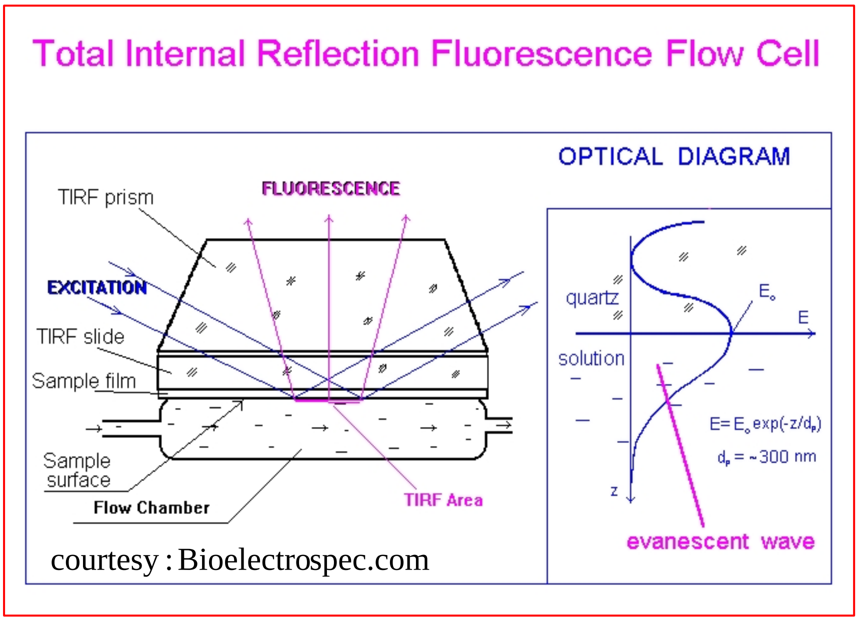
When total internal reflection is taking place, a wavefield extends into the low refractive index medium. This wavefield doesn't propagate perpendicular to the surface but decreases exponentially, and rather quickly. It is called the **evanescent wave**. It is associated with an electric field moving parallel to the surface. Hence in a fibre optic cable such as will be discussed shortly, the cladding around the central core must be thick enough to accommodate this evanescent wave and also have a low absorption.

If a second surface is brought close to the totally internally reflecting surface, into the evanescent wavefield, then the total internal reflection is *frustrated*. Some energy crosses the narrow gap and propagation resumes in the neighbouring medium. The fraction of light that crosses is very dependent on the gap size and can therefore be controlled by very careful positioning of the second surface.

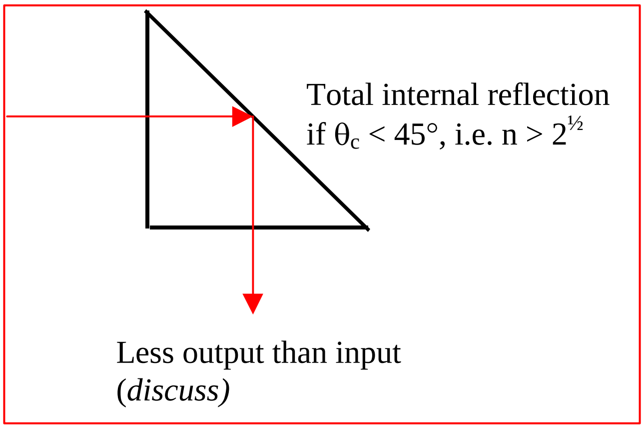
Use is made of **frustrated total internal reflection** in such devices as variable beam-splitters and the deliberate feed of optical energy into fibre-optic lines. The next two paragraphs describe another application of the evanescent wave.

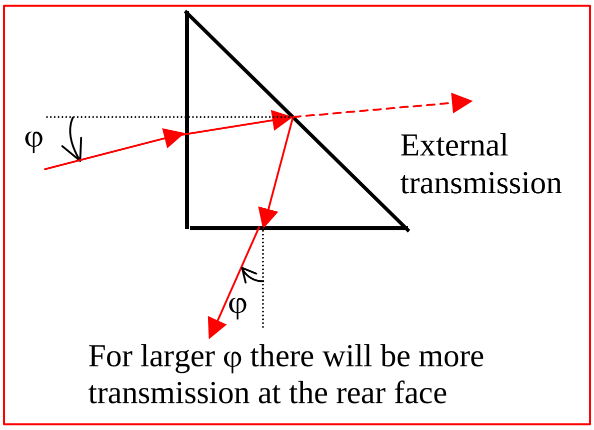
Total internal reflection fluorescence is a technique used to detect very small concentrations of biological molecules, which may be proteins, single-strand DNA, drugs and so on. The question the technique tries to answer is "what is the effectiveness of binding of pairs of molecules, such as antigens to antibodies, drugs to proteins and hybridisation of DNA?". Don't trust any biology I offer in this course but here is the optics.

A thin layer of the specific sensor is coated on the inside of an optical slide past which a solution with the interesting molecules flows. The accompanying diagram is courtesy of one of the companies in the field. The molecules, call them proteins, have a fluorescent label attached, like fluorescene. Alternatively, they may fluoresce naturally. The coupling between antigen and antibody takes place best when the antibodies are coated onto a surface and not when they are rolling about in the solution. The slide is illuminated by totally internally reflected light from the outside with a wavelength that excites the fluorescence. The evanescent wave, which extends for a few hundred nm over the sample surface that is coated with antibodies, excites the fluorescence of the bound molecules but not the fluorescence of proteins loose in the solution. By measuring the fluorescence signal, the effectiveness of the binding can be studied and the concentration of molecules in the solution inferred. It is the evanescent wave that makes the technique surface sensitive.



Reflecting prisms

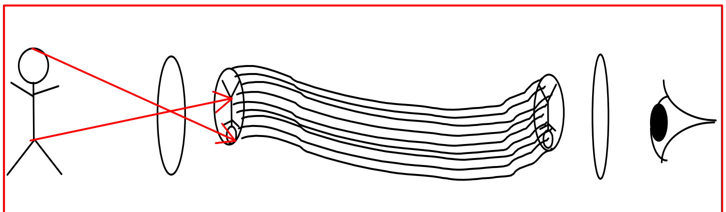


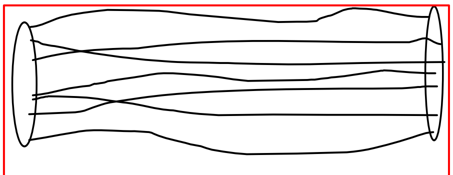


Although the normally incident ray is internally reflected for a glass prism, not all the incident rays need be reflected.

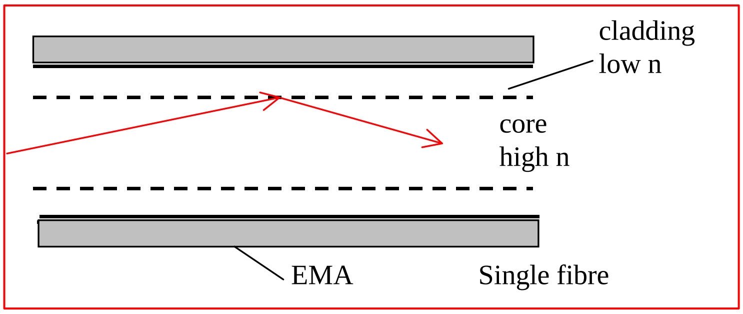
Fibre optics

- Original patents on fibre optic transmission were granted to John Logie Baird in the 1930's.


Coherent bundle


Incoherent bundle

- Optical fibres can be **coherent,** for transmitting images, or **incoherent**, for transmitting a light signal. Communications links may contain just a single fibre.



- The diagram to the right shows the constituents of a typical fibre optic: core, cladding, EMA (extra mural absorption to suppress any residual evanescent wave)

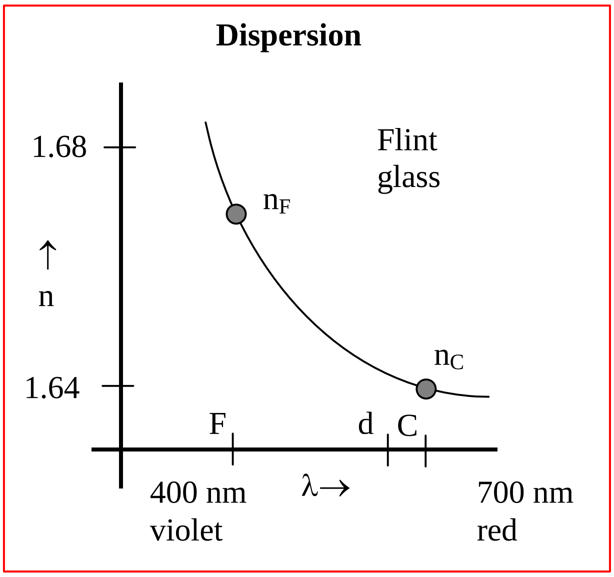
Advantages:

Optical transmission of images
- Flexible
- Long
- Little loss
- Straightforward construction

Communications
- Not subject to electrical interference
- Closed circuit
- Very high bandwidth (subject to refractive dispersion and propagation dispersion)
- Long-life
- Disadvantage: repeaters, when needed, involve conversion to electrical signals

Dispersion

Dispersion is the phenomenon that causes prismatic colours. Dispersion gives the brilliance to cut diamonds, the gemstones of choice for the wealthy and the married alike. Diamonds have made nations rich, men greedy and women envious. Dispersion gives the rainbow its colours and the spectroscope its power. At the heart of a phenomenon of awe is a property that is simply described, though less simply understood. Optical dispersion is just the variation of refractive index with wavelength. The simple quantity that was labelled *n* depends on wavelength. To make this clear, *n* is written as n_λ.



- Dispersion increases more rapidly at shorter wavelengths, resulting in the plot of n_λ versus wavelength λ to be curved.

- Cauchy's **empirical** formula:

$$n_\lambda = n_0 + \frac{A}{\lambda^2}\left(+\frac{B}{\lambda^4}+....\right)$$

- n_0 and A are parameters which, when specified, determine n_λ at all wavelengths. The formula has validity around the wavelengths that have been used to determine the parameters, usually in the optical region.

- Dispersion curves for different glasses do not simply scale. i.e. there is not one 'universal' curve that can be fitted by a simple change of scale. The Cauchy formula is the best reasonably simple scaling that can be done.

- Glass catalogues quote n_λ at certain standard wavelengths of selected Fraunhofer spectral lines. The following are particularly relevant here:

Fraunhofer letter	Origin	Wavelength nm
C	Red hydrogen	656.27
D	Na yellow	589.4
d	He yellow	587.56
F	Blue hydrogen	486.13

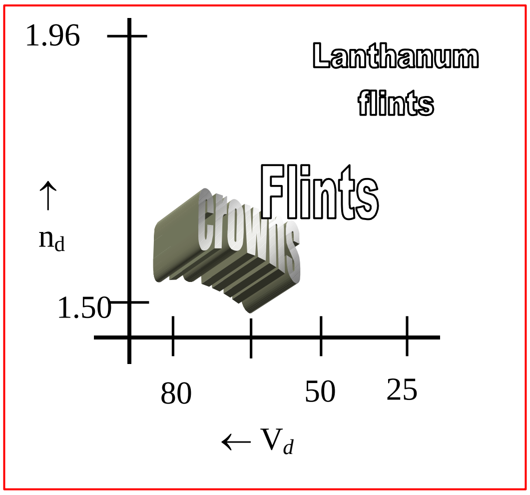
- A single parameter measure of dispersion is the *Abbe number* V_d.

$$V_d = \frac{n_d - 1}{n_F - n_C}$$

(Some texts use the Na D wavelength to define the Abbe number, calling it V_D).

- The larger the dispersion, the smaller is the Abbe number.

Optical glasses are displayed on a n_d/V_d graph with different types of glass occupying their own regions, as shown here.



- Note the naming of glasses, where the 6-digit number includes elements from n_d and V_d: e.g. BK7 517642 implies that

$$n_d = 1.517; \quad V_d = 64.2$$

- Manufacturers quote n_d and V_d, from which it is possible to calculate n_λ for all wavelengths. See homework examples to come. The gist of the idea is to use Cauchy's formula at the wavelengths F and C, giving

$$n_F - n_C = \frac{A}{\lambda_F^2} - \frac{A}{\lambda_C^2} = \frac{n_d - 1}{V_d}$$

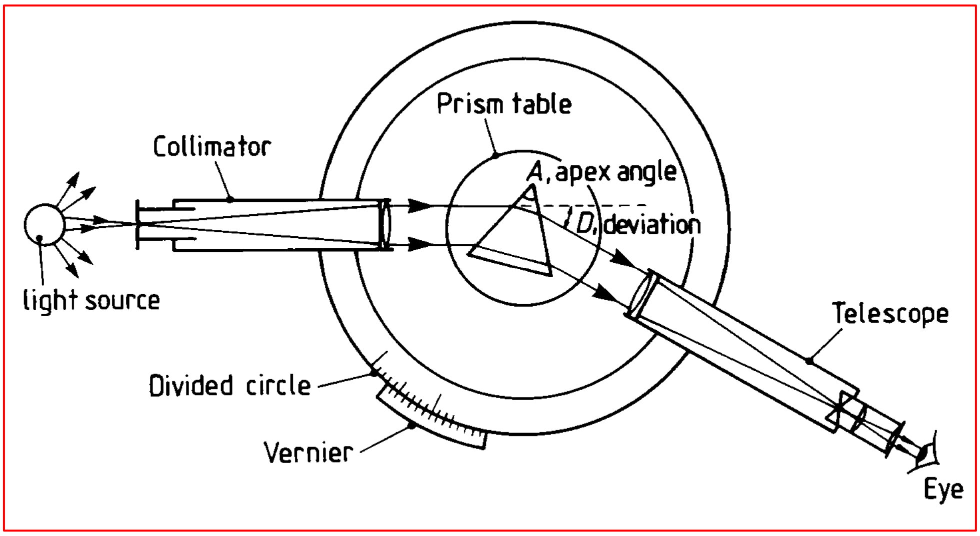
$$\therefore\ A = \frac{n_d - 1}{V_d} \times \left(\frac{1}{\lambda_F^2} - \frac{1}{\lambda_C^2} \right)$$

$$\textit{and}\ \ n_0 = n_d - \frac{A}{\lambda_d^2}$$

- Can you think of any more phenomena that depend on dispersion? Two more meteorological examples are ice-particle halos and the green flash.

The spectrometer

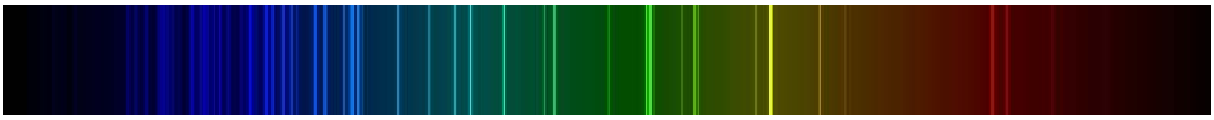
Because of dispersion, any single refraction of light will tend to split the light into its spectrum of colours. The formation of a spectrum by a prism is the example that many think of first, but the refracting shape doesn't have to be obviously prismatic. Think of the spectrum of sunlight created on a huge scale by spherical water droplets making a rainbow. Think of the spurious colouring produced by any single component lens. In this last case, the formation of a spectrum is a humbug that degrades the performance of the optical system in which the lens is used. I'll show shortly how to reduce the extent of the spectrum produced by dispersion in lens systems. Sometimes, though, it is the spectrum itself that is really of interest. I'll say a lot more about the spectrum of light in the next chapter of the course.



Various devices are made to display and measure light spectra. Many in the class will have met such a device in the first-year lab. It's called a **spectrometer**. The spectrometer in the first year lab is a **prism spectrometer**, though it can also be used with a diffraction grating to produce a spectrum. The illustration is taken from our first year lab manual. The components of the prism spectrometer are:

- A narrow **slit** to admit the light whose spectrum you want to show or measure.
- A **collimator** whose purpose is to create a parallel pencil out of the light passing through each part of the slit.
- The **prism**, mounted on a turntable that can be twisted around by hand.

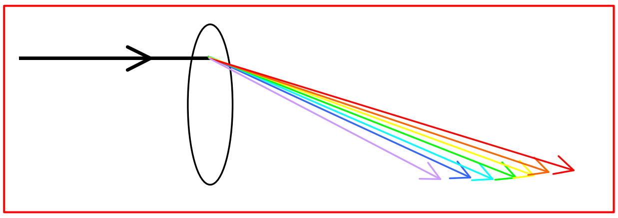
A **telescope**, which takes the deviated parallel pencil of light coming out of the prism and refocuses it to a point located at 'infinity', thereby making an image of the entrance slit.



- This image is in a different place for each colour of the incident light and hence a line is created for each colour present in the incident light. We talk about a **spectral line**. If the spectrum is continuous, then you will see the spectrum spread out into a band. If the spectrum is discontinuous, then you'll see the individual lines of colour that make it up. The accompanying illustration shows a spectrum (of Krypton) with a continuous background and superimposed bright spectral lines.

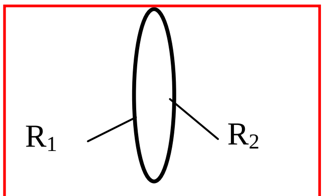
Achromatic Doublet

If dispersion is left unchecked, it will ruin the performance of any lens system, introducing what is known as *chromatic aberration*. The solution to this great difficulty is to keep the red and blue rays close together as much as possible as they travel through the optical system. How to do this was discovered by John Dollond and patented by him in 1758. He replaced a single lens by a combination of two lenses, called an *achromatic doublet*. His own interest was in making refracting telescopes and by doing this for a telescope's objective lens he more or less secured for himself the world market in good refractors. The same solution is still applied nowadays to a wide range of optics. Hence the achromatic doublet is a central concept in optical instrument design.



The following paragraphs give an idea of how the achromatic doublet works.

- The location of the image formed by a lens is determined by the lens's power D. D depends on the two radii of curvature of the lens and the refractive index of the glass. Power is just the reciprocal of the focal length f of the lens. D and f are determined by the lensmakers' equation. There is no need to remember this.



Radii of curvature

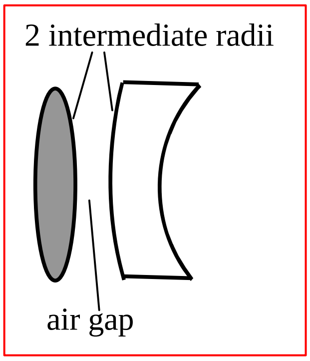
The lensmakers' equation for a **thin lens** is:

$$\text{lens power: } \mathsf{D} = \frac{1}{f_\lambda} = (n_\lambda - 1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right)$$
$$= (n_\lambda - 1)\rho$$

where ρ is a geometrical factor.

- Notice that D and f depend on the colour of light, giving rise to axial and lateral chromatic aberration of an image.

- There is no way of merging back into a single image point all the separated rays of different colours originating from a single object point. By merging just the red and the blue, the total spread is reduced by around a factor 10. This is the basis of the correction of chromatic aberration.

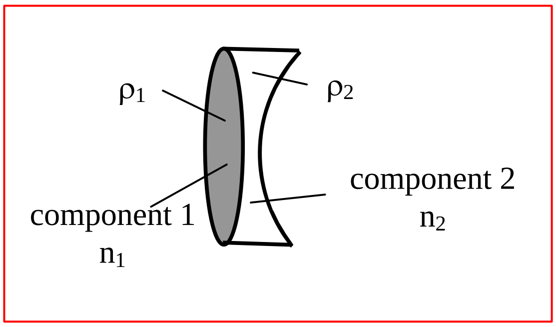


- Hecht's *Optics* discusses how to do this, in the context of a general discussion of image aberrations. The gist of it is that every lens is made from a combination of two different components. Such a two-component lens is called a **doublet**. The two components are made from different glass.

The ideas behind the doublet are these:

1) The power of a combination of lenses is the sum of the powers of the components.

2) To make a combination achromatic, the combined power in the red is set equal to the power in the blue. In symbols:



$$D_{Red} = D_{Blue}$$

This condition can be realised in practice provided the Abbe numbers (V_d) of the two different components are not in proportion to their refractive indices. The more different they are, the easier it is to make a doublet.

Typically, the front component is **crown** glass with $V_1 \sim 65$,
the rear " **flint** " $V_2 \sim 45$.

3) What is done is to cancel out the spread of colours produced by the first component using a second (negative) component of higher dispersion. By this means the focusing effect of the first component is not cancelled out too, leaving a residual positive focusing.

Although the focal length in the red and blue are the same, those of intermediate colours are slightly different. The correction is not perfect across the spectrum.

4) The achromatic doublet is used throughout modern optical instrumentation. The idea is that it is essential to keep the red and blue rays together during their passage through the microscope, camera or whatever rather than try to combine them at the end of the instrument.

5) The achromatic doublet has at least one further parameter that can be used to control the imaging of the combination. Depending on how the doublet is constructed, the extra parameters may be:

- ♦ the intermediate radius of curvature
- ♦ 2 intermediate radii of curvature
- ♦ presence or absence of a definite air gap (the gap may be filled with glue such as ‘Canada Balsam’)

The adjustable parameter(s) can be used to control spherical aberration or other potential imaging defects.

End of Fundamentals