Waves & Optics

FIZIKA SPhO Training

October 2025

Contents

1 Notes

1.1 Waves

Consider dropping a pebble into a large pond. When the pebble impacts the water surface, it sends energy travelling as ripples through the water. This is an example of **waves**, present in the form of ripples on the water surface.

Waves are a propagating disturbance from equilibrium. The most commonly studied waves are mechanical waves (for example, water ripples or sound in air) and electromagnetic waves (for example, light). Waves can also be longitudinal (for example, sound) or transverse (for example, ripples). In this section, we study some basic properties of waves.

1.1.1 Basic Quantities

The most basic form of a wave is a sinusoidal wave. The sinusoidally varying quantity ψ can be expressed as

$$\psi(x,t) = A\sin(kx \pm \omega t + \varphi)$$

where x is position and t is time. The coefficient of t is negative if the wave is travelling in the $+x$-direction, and positive if the wave is travelling in the $-x$-direction.

ψ can be different quantities depending on what is being considered: it could refer to a displacement for a wave travelling in a rope, or it could refer to a pressure value for a sound wave travelling in air.

Here, we recap the basic quantities related to a sinusoidal wave.

- A is the amplitude of the wave, defined as the maximum displacement from equilibrium.
- λ is the wavelength of the wave, defined as the distance between two points on the wave at the same oscillation state. For example, this could be between two crests or troughs. k is the angular wavenumber of the wave, related to the wavelength by

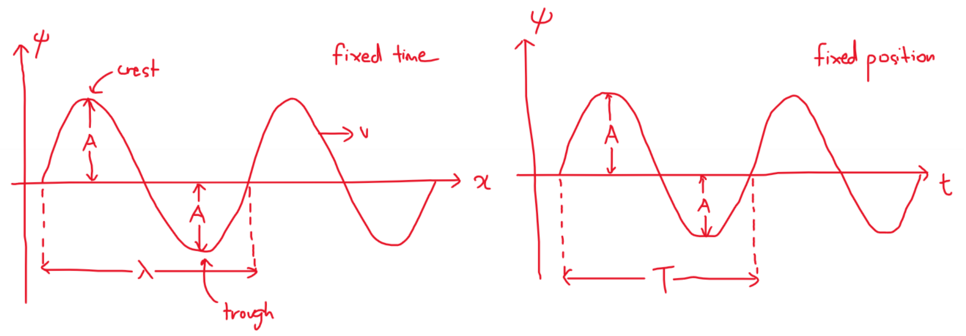
  $$k = \frac{2\pi}{\lambda} \tag{1}$$

  Sometimes, we will refer to k as the wavenumber of the wave.
- T is the period of the wave, defined as the time for any point at some fixed position on the wave to complete one oscillation cycle. For example, it could be the time it takes for a crest at some particular position to become a trough and back. f and ω are the frequency and angular frequency of the wave respectively, related to the period by

  $$f = \frac{1}{T} = \frac{\omega}{2\pi} \qquad \omega = \frac{2\pi}{T} = 2\pi f \tag{2}$$

  Sometimes, we will refer to ω as the frequency of the wave.
- v is the speed of the wave. Since the wave travels one wavelength in one period, we have

  $$v = \frac{\lambda}{T} = f\lambda \tag{3}$$



Notice that we can also derive a relation between k and ω,

$$v = \frac{2\pi}{k}\frac{\omega}{2\pi} = \frac{\omega}{k} \tag{4}$$

We can rearrange this equation to get

$$\omega = vk \tag{5}$$

which is called the *dispersion relation* for sinusoidal waves, essentially characterising ω as a function of k.

1.1.2 The Wave Equation

Waves are not always a travelling sinusoidal wave, although such waves are the simplest, although every wave can be represented as the superposition of such waves. All one-dimensional waves must satisfy the **wave equation**, a partial differential equation that relates the partial derivatives of the wave quantity ψ with position and time:

$$\frac{\partial^2\psi}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2\psi}{\partial t^2} \tag{6}$$

The general solution to this partial differential equation can be derived by the use of the substitution $x_- = x - vt$ and $x_+ = x + vt$ to simplify the differential equation, a method called D'Alembert's method. We will not go into the details here, but going through the derivation gives the general solution

$$\psi(x,t) = f(x - vt) + g(x + vt) \tag{7}$$

where f and g are arbitrary functions. This equation essentially means that a wave is comprised of a superposition of two functions of x travelling at speed v in opposite directions.

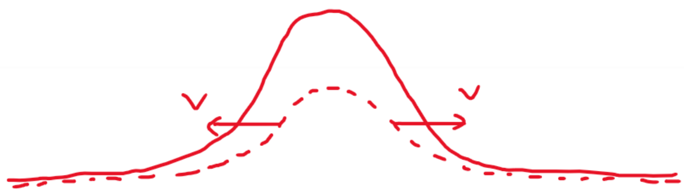
One thing of note is that the basic travelling sinusoidal wave is a special case of the wave equation. For example, such a wave travelling to the right can be represented with the functions $f(x) = A\sin(kx + \varphi)$ and $g(x) = 0$, in which case

$$\psi(x,t) = A\sin(k(x - vt) + \varphi) = A\sin(kx - \omega t + \varphi)\,.$$

1.1.3 Initial and Boundary Conditions

The solution to the wave equation allows us to determine the nature of waves subject to certain initial or boundary conditions. We shall mention two simple ones here. Here, we shall consider the wave to be a transverse wave on a string, where tension allows the wave to propagate.

Firstly, we consider an *infinite string released from rest*. Suppose the string initially has shape $y = \Psi(x)$.



We shall omit the derivation, but it can be shown that the time evolution of the wave will be

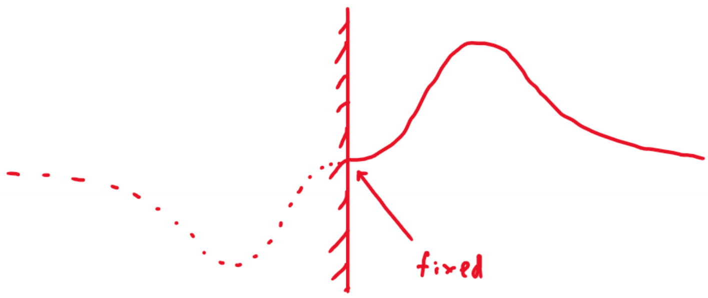
$$\psi(x,t) = \frac{1}{2}\left(\Psi(x-vt) + \Psi(x+vt)\right) \tag{8}$$

This can be done by considering the initial conditions (initial shape, starting at rest) and solving for the two functions f and g using some partial derivative manipulation.

In words, the initial shape is split into two halves moving in opposite directions at speed v. You can check that this satisfies the wave equation, so it is considered a wave.

Secondly, we can consider a *semi-infinite string with a fixed end released from rest*, with the fixed end at the origin, i.e. $\psi(0,t) = \Psi(0) = 0$. To derive the shape of the wave as a function of time, let us use the solution for an infinite string with some different shape Ψ_0 released from rest, imposing the condition for a fixed end:

$$\psi(0,t) = \frac{1}{2}(\Psi_0(-vt) + \Psi_0(vt)) = 0 \quad \Longrightarrow \quad \Psi_0(-vt) = -\Psi_0(vt)$$



Since this holds for all t, this implies Ψ_0 is odd. We must have $\Psi(x) = \Psi_0(x)$ for all $x \geq 0$, as that is the initial condition, so we can conclude that the time evolution of the wave can be described as

$$\psi(x,t) = \frac{1}{2}(\Psi_0(x-vt) + \Psi_0(x+vt)) \tag{9}$$

where

$$\Psi_0(x) = \begin{cases} \Psi(x) & x \geq 0 \\ -\Psi(-x) & x < 0 \end{cases} \tag{10}$$

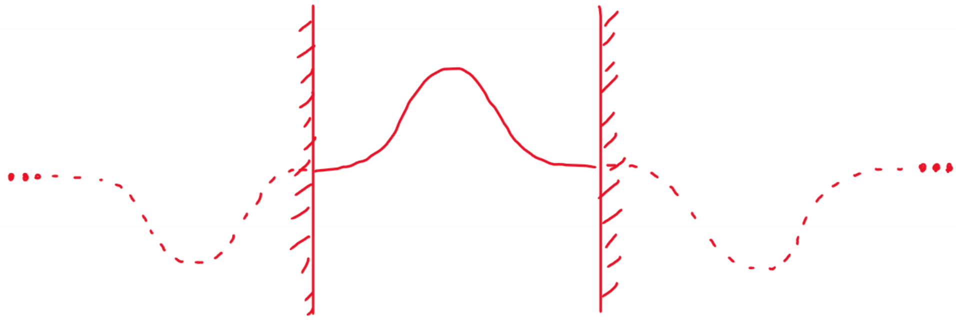
In words, we can consider there to be a "reflected image" behind the fixed end that is also released from rest.

1.1.4 Standing Waves

Now, we can consider a *finite string with two fixed ends released from rest*, with the fixed ends at $x = 0$ and $x = L$ respectively. It's easy to see from our analysis in the previous section that similar "reflections" will occur past each fixed end, so the time evolution of the wave can be described similarly with

$$\Psi_0(x) = \begin{cases} \Psi\left(L\left\{\frac{x}{L}\right\}\right) & \left\{\frac{x}{2L}\right\} \leq \frac{1}{2} \\ -\Psi\left(L\left\{\frac{x}{L}\right\}\right) & \left\{\frac{x}{2L}\right\} > \frac{1}{2} \end{cases} \tag{11}$$

where $\{\cdot\}$ denotes the fractional part.



It is clear that Ψ_0 is periodic, repeating every $2L$. This motivates us to return to the study of sinusoidal functions, leading us to the study of standing waves.

Instead of considering the solution to the wave equation, let's consider two identical (infinite) sinusoidal waves travelling in opposite directions. They will superimpose to form a new wave, that is valid due to the linearity of the wave equation.

$$\psi(x,t) = A_0 \cos(kx - \omega t + \varphi) + A_0 \cos(kx + \omega t + \varphi) = 2A_0 \cos\left(kx + \varphi\right) \cos \omega t$$

This is also a sinusoidal function, but instead of travelling at some constant amplitude, it is stationary with some oscillating amplitude. This is called a **standing wave**. If we now impose the boundary conditions by the fixed ends, being that $\psi(0,t) = \psi(L,t) = 0$,

$$\psi(0,t) = 2A_0 \cos\varphi \cos\omega t = 0 \quad \Longrightarrow \quad \varphi = -\frac{\pi}{2}$$

$$\psi(L,t) = 2A_0 \cos\left(kL - \frac{\pi}{2}\right) \cos\omega t = 2A_0 \sin kL \cos\omega t = 0 \quad \Longrightarrow \quad kL = n\pi, \quad n \in \mathbb{Z}^+$$

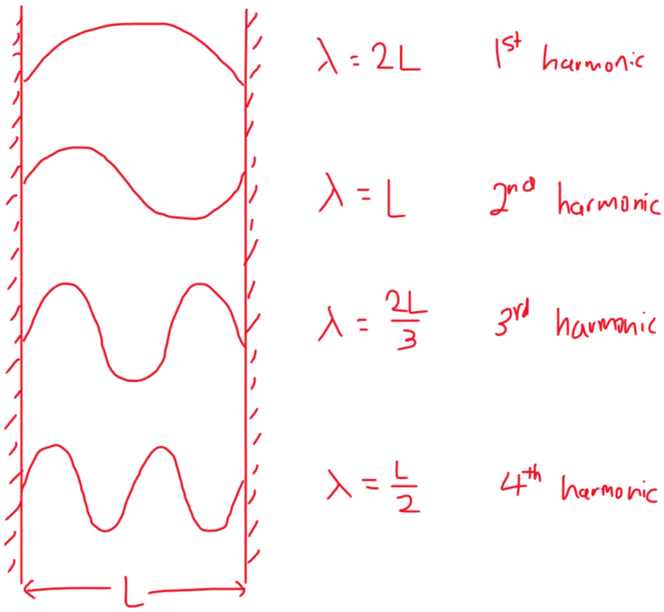
Therefore, we have that the standing wave can be represented as

$$\psi(x,t) = A \sin kx \cos \omega t \tag{12}$$

with a condition on the wavelength to ensure the boundary conditions are satisfied,

$$kL = n\pi \quad \Longrightarrow \quad \lambda_n = \frac{2L}{n} \tag{13}$$

where n is a positive integer. We term the wave corresponding to n as the n-th **harmonic**.



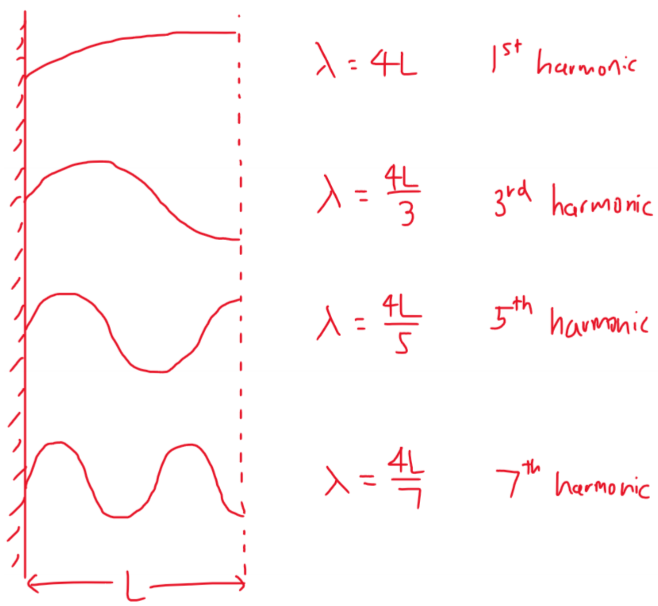
There are also standing waves corresponding to *one fixed end and one open end*. Here, an open end imposes a boundary condition that the string must be flat at that endpoint. In this case, we can solve for the new condition on the wavelength using this new boundary condition, being that $\frac{\partial \psi}{\partial x}(L,t) = 0$,

$$\frac{\partial \psi}{\partial x}(L,t) = 2kA_0 \cos kx \cos \omega t = 0 \quad \Longrightarrow \quad kL = \left(n - \frac{1}{2}\right)\pi, \quad n \in \mathbb{Z}^+$$

Therefore, we have that the standing wave with an open end has the condition on the wavelength to ensure that the boundary conditions are satisfied,

$$kL = \left(n - \frac{1}{2}\right)\pi \quad \Longrightarrow \quad \lambda_n = \frac{4L}{2n-1} \tag{14}$$

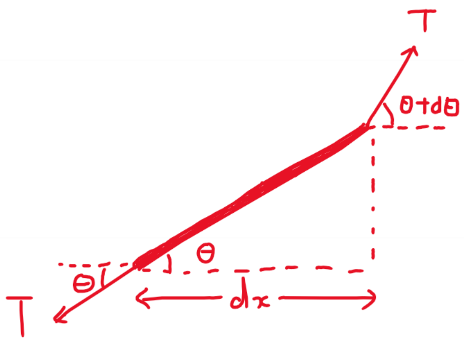
where n is a positive integer. We term the wave corresponding to n as the $(2n-1)$-th harmonic.



1.1.5 String Waves

The most common mechanical wave you will consider are **string waves**. This is a common example of a transverse wave.

Consider a string with linear mass density μ under tension. We neglect gravity and we assume that the amplitude of the wave is small, so that we can neglect any longitudinal motion of the string. Then, we can consider an infinitesimal segment of the string of horizontal length $\mathrm{d}x$ with mass $\mathrm{d}m = \mu\,\mathrm{d}x$.



Since the segment remains vertically constrained, the horizontal forces must balance, so $T\cos\theta$ is constant where T is the tension in the string. Since we've already assumed that the amplitude of the wave is small, θ is small so that to first order, $\cos\theta \approx 1$. Therefore, we can conclude that the tension T is constant everywhere in the string.

Now, considering the vertical forces on the string, using the approximation $\sin\theta \approx \tan\theta$ since θ is small,

$$\begin{aligned}
\mathrm{d}m\,a &= T\sin(\theta+\mathrm{d}\theta) - T\sin\theta \\
\mu\,\mathrm{d}x\,\frac{\partial^2\psi}{\partial t^2} &\approx T\left(\tan(\theta+\mathrm{d}\theta) - \tan\theta\right) \\
&= T\left(\frac{\partial\psi}{\partial x}(x+\mathrm{d}x\,,t) - \frac{\partial\psi}{\partial x}(x,t)\right) \\
&= T\,\mathrm{d}x\,\frac{\partial^2\psi}{\partial x^2} \\
\frac{\partial^2\psi}{\partial x^2} &= \frac{\mu}{T}\frac{\partial^2\psi}{\partial t^2}
\end{aligned}$$

where we employed the definition of the derivative to simplify the right hand side. As a result, we obtain the wave equation! We can solve for the speed of travelling waves in the string as

$$\frac{1}{v^2} = \frac{\mu}{T} \quad\implies\quad v = \sqrt{\frac{T}{\mu}}. \tag{15}$$

Example 1.1. A uniform string of length L hangs vertically from a ceiling. Find the time it takes for a wave to travel from the bottom of the string to the top.

At height x from the bottom of the string, the tension is $\mu x g$, since the tension in the string needs to support the weight below it. Therefore, the time it takes for the wave to travel a distance $\mathrm{d}x$ up the string is

$$\mathrm{d}t = \frac{\mathrm{d}x}{v} = \sqrt{\frac{\mu}{\mu x g}}\,\mathrm{d}x = \frac{\mathrm{d}x}{\sqrt{gx}}$$

Integrating over the length of the string,

$$t = \int_0^L \frac{\mathrm{d}x}{\sqrt{gx}} = 2\sqrt{\frac{L}{g}}.$$

1.1.6 Energy and Intensity

Finally, we shall go into some brief analysis of energy and intensity. Once again, suppose we have a wave in a string of linear mass density μ and tension T. The energy of the wave is comprised of kinetic and potential energy. If we consider an infinitesimal segment of length $\mathrm{d}x$ again, the kinetic energy is

$$\mathrm{d}K = \frac{1}{2}\mu\left(\frac{\partial\psi}{\partial t}\right)^2 \mathrm{d}x \tag{16}$$

On the other hand, the potential energy comes from the stretching of the string from its equilibrium length. The new length of the infinitesimal segment upon stretching is $\sqrt{1+\left(\frac{\partial\psi}{\partial x}\right)^2}\,\mathrm{d}x$, and the restoring force in the string is tension so the potential energy stored by that segment is simply the work done by tension to allow the string to stretch,

$$\mathrm{d}U = T\left(\sqrt{1+\left(\frac{\partial\psi}{\partial x}\right)^2}\,\mathrm{d}x - \mathrm{d}x\right) \approx \frac{1}{2}T\left(\frac{\partial\psi}{\partial x}\right)^2 \mathrm{d}x \tag{17}$$

where we again used the assumption that the amplitude of the wave is small so that we can use the first order approximation. Therefore, the energy stored in that infinitesimal segment is

$$\mathrm{d}E = \frac{1}{2}\left(\mu\left(\frac{\partial\psi}{\partial t}\right)^2 + T\left(\frac{\partial\psi}{\partial x}\right)^2\right)\mathrm{d}x \tag{18}$$

Example 1.2. Calculate the time-averaged power of a sinusoidal wave of amplitude A and frequency ω travelling at speed v in a string of linear mass density μ.

For a sinusoidal wave,

$$\psi(x,t) = A\cos(kx-\omega t)$$

where we have neglected the phase since we are considering the average. We also have

$$v = \frac{\omega}{k} = \sqrt{\frac{T}{\mu}} \qquad \Longrightarrow \qquad \mu\omega^2 = Tk^2$$

Then, the energy stored in an infinitesimal segment is

$$\begin{aligned}\mathrm{d}E &= \frac{1}{2}\left(\mu\left(-\omega A\sin(kx-\omega t)\right)^2 + T\left(-kA\sin(kx-\omega t)\right)^2\right)\mathrm{d}x \\ &= \frac{1}{2}\left(\mu\omega^2 + Tk^2\right)A^2\sin^2(kx-\omega t)\,\mathrm{d}x \\ &= \mu\omega^2A^2\sin^2(kx-\omega t)\,\mathrm{d}x\end{aligned}$$

Taking the time average, we have

$$\langle \mathrm{d}E\rangle = \mu\omega^2A^2\left\langle\sin^2(kx-\omega t)\right\rangle\mathrm{d}x = \frac{1}{2}\mu\omega^2A^2\,\mathrm{d}x$$

We can see that the time average of the energy in each infinitesimal segment is the same everywhere, so we have that the average total energy stored in one wavelength of the string wave is thus

$$E_\lambda = \frac{1}{2}\mu\omega^2A^2\lambda$$

The power is the amount of energy transferred per unit time. Since we are considering the energy over a wavelength, the duration we consider is the period. Therefore, the time-averaged power is

$$P_{\text{ave}} = \frac{1}{2}\mu\omega^2 A^2 \frac{\lambda}{T} = \frac{1}{2}\mu\omega^2 A^2 v.$$

In this example, we observe that the power of a 1-dimensional sinusoidal wave is proportional to the square of its amplitude. For propagating 3-dimensional waves, the analogous quantity that is proportional to the square of its amplitude is *intensity*:

$$I = \frac{P}{4\pi r^2} \propto A^2 \tag{19}$$

where P is the power of the source of waves and r is the distance from the source to the point of consideration. Here, the relationship between the power of the wave source and the intensity comes from the fact that energy is conserved and the energy is uniformly distributed across the surface area of a sphere centered on the power source. In turn, this allows us to deduce that the amplitude of a 3-dimensional propagating wave is inversely proportional to distance:

$$A \propto \frac{1}{r} \tag{20}$$

We can also use this line of reasoning to deduce that for *cylindrical waves* emitted by an infinitely long line source, the intensity I is inversely proportional to the distance r to the point source, and so the amplitude A is inversely proportional to the square root of the distance.

1.2 Physical Optics

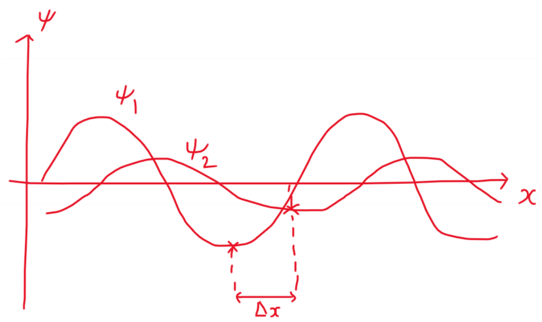
We begin our discussion on light. In physical optics, we shall describe light propagation by waves. Physical optics typically applies when the scale of consideration is comparable to the wavelength of light.

1.2.1 Interference

As we mentioned earlier, the wave equation is linear. This means that the sum of two solutions to the wave equation yields another solution to the wave equation. The idea of adding two waves together is called **superposition**, and it gives waves the ability to interfere with each other as they interact.

Consider two overlapping waves of equal wavenumber and frequency but not necessarily the same amplitude nor phase,

$$\psi_1(x,t) = A_1 \cos(kx - \omega t + \varphi_1)$$
$$\psi_2(x,t) = A_2 \cos(kx - \omega t + \varphi_2)$$



We can see that the crests and troughs of both waves do not occur at the same position. This is because both waves have a phase difference, $\Delta\varphi = \varphi_2 - \varphi_1$. We can then deduce that the spatial separation Δx between two points corresponding to the same phase on both waves are related to the phase difference by

$$\frac{\Delta\varphi}{2\pi} = \frac{\Delta x}{\lambda} \tag{21}$$

A wave source is termed to be *coherent* if the phase difference between the waves it emits is constant, and we shall only consider coherent sources here. These waves are also said to be in phase – waves which are not in phase are said to be out of phase.

There are two special cases of interest, where the phase difference between the two waves are $\Delta\varphi = 0$ and $\Delta\varphi = \pi$ respectively. In the first case, when both waves superimpose,

$$\psi(x,t) = A_1\cos(kx - \omega t + \varphi_1) + A_2\cos(kx - \omega t + \varphi_1) = (A_1 + A_2)\cos(kx - \omega t + \varphi_1)$$

so the amplitudes of both waves add up directly. This is called **constructive interference**. On the other hand, in the second case, when both waves superimpose,

$$\psi(x,t) = A_1\cos(kx - \omega t + \varphi_1) + A_2\cos(kx - \omega t + \varphi_1 + \pi) = (A_1 - A_2)\cos(kx - \omega t + \varphi_1)$$

so the amplitudes of both waves subtract. This is called **destructive interference**.

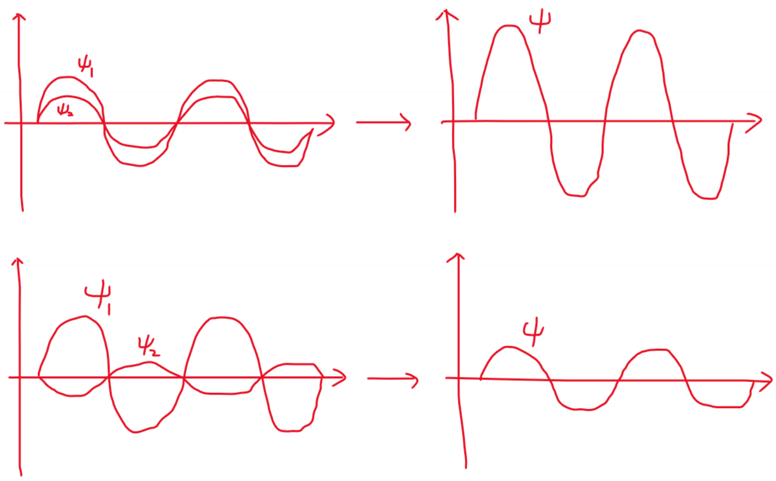
Equivalently, we can see that two waves will constructively interfere when they have a path difference of

$$\Delta x = m\lambda \tag{22}$$

while they will destructively interfere when they have a path difference of

$$\Delta x = \left(m - \frac{1}{2}\right)\lambda \tag{23}$$

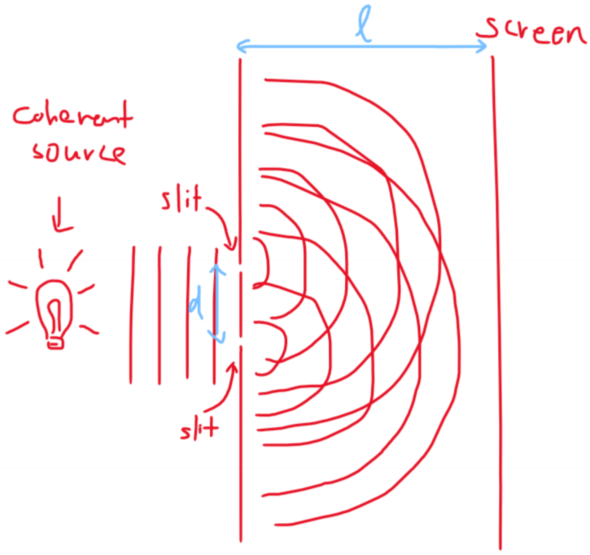
where m is an integer. This arises from the relation between Δx and $\Delta\varphi$, as well as the fact that waves are periodic with their phase repeating every 2π.



This may appear to violate the conservation of energy. For example, if two string waves of equal amplitude destructively interference, at some point in time, the string will be completely flat, having lost energy. However, this can be attributed to the fact that all the energy of the string waves have been converted to the kinetic energy of the string.

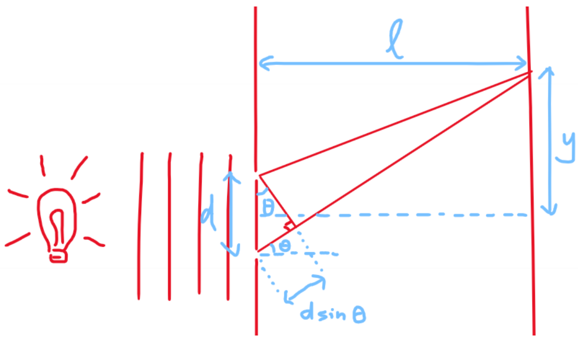
1.2.2 Double-slit Diffraction

In 1801, Thomas Young conducted an experiment to demonstrate the wave nature of light, called **Young's double-slit experiment**. In this experiment, a coherent light source of wavelength λ is placed in front of a long board with two thin slits separated by distance d. The resultant light illuminates a screen placed at distance ℓ to the slits.



The usual expectation is that there will only be two spots on the screen illuminated on the light, corresponding to the head-on positions of the slits on the screen. However, what was actually observed looked something like this:

The pattern actually shown exhibits repeating bands of bright and dark spots. This is called **diffraction**, and the pattern formed on the screen is said to be a diffraction pattern. To see why it occurs, consider two light rays, one from each slit, reaching the same point on the screen:



First, since the incoming light is coherent, the two light rays have the same phase when they just exit the slit. Then, due to the spatial separation between the two slits, the light rays have different path lengths once they reach the same point on the screen, and so they will have some phase difference. These light rays will interfere with each other on the screen depending on their phase difference, causing the bright and dark spots to appear (corresponding to constructive and destructive interference respectively).

We can calculate the positions of the bright and dark spots by analysing the geometry of the system. Here, we shall make our first assumption, which is that $d \ll l$. This allows us to make the simplification that both light rays are approximately parallel, and we can consider the path length difference to be the length shown in the diagram below,

$$\Delta x = d \sin\theta$$

where θ is the angle the rays make to the horizontal. This gives us the phase difference between the two rays as

$$\varphi = \frac{2\pi d \sin\theta}{\lambda}$$

Then, the bright spots occur when the path difference is a multiple of λ, so

$$d \sin\theta = m\lambda \tag{24}$$

where m is an integer. On the other hand, the dark spots occur when the path difference is a half-integer multiple of λ, so

$$d \sin\theta = \left(m - \frac{1}{2}\right)\lambda \tag{25}$$

We can also deduce the maximum number of bright spots by finding the maximum m allowable, using the fact that $\sin\theta < 1$:

$$m\lambda = d\sin\theta < d \quad \Longrightarrow \quad m < \frac{d}{\lambda}. \tag{26}$$

The maximum number of bright spots can then be calculated from the maximum integer m.

Finally, we can calculate the positions of the bright and dark spots on the screen. Here, we shall make our second assumption, which is that $\theta \ll 1$. This allows us to use the small angle approximation $\sin\theta \approx \tan\theta$, so we can calculate the position as

$$y = \ell \tan\theta \approx l \sin\theta$$

We can then deduce that the positions of the bright spots, which we also call the maximums, are

$$y_{\text{bright}} = \frac{m\lambda\ell}{d} \tag{27}$$

and the positions of the dark spots, which we also call the minimums, are

$$y_{\text{dark}} = \left(m - \frac{1}{2}\right)\frac{\lambda\ell}{d}. \tag{28}$$

Now, we wish to calculate the relative intensity of these bright and dark spots compared to the central bright spot. We can do this by considering the sinusoidal electric fields of each light ray and summing them up. However, summing sinusoidal functions can get challenging and tricky eventually. This is a great opportunity to reintroduce *phasors*, representing the electric fields of each light ray by a phasor.

Let consider the diagram shown earlier again. Suppose the electric field at the point on the screen by the light ray that passed through the top slit can be represented by the phasor

$$\widetilde{E}_1 = E_0$$

Then, due to the phase difference between the two light rays, the electric field by the light ray that passed through the bottom slit can be represented by the phasor

$$\widetilde{E}_2 = E_0 e^{i\varphi}$$

where φ is the phase difference between the two light rays. Due to superposition, these electric fields will add, so we will have

$$\widetilde{E} = E_0 + E_0 e^{i\varphi} = 2E_0 e^{\frac{i\varphi}{2}} \cos\frac{\varphi}{2}$$

obtained by factorisation and the complex definition of trigonometric functions. Finally, recall that the intensity is proportional to the square of the amplitude, so

$$I \propto |\widetilde{E}|^2 \propto \cos^2\frac{\varphi}{2}$$

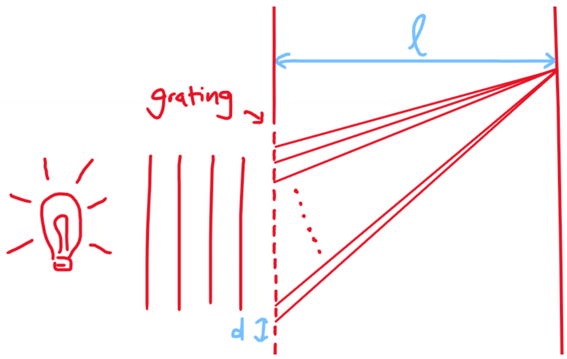
If we let the intensity at the central maximum be I_0, we have

$$I = I_0 \cos^2\left(\frac{\pi d \sin\theta}{\lambda}\right) \tag{29}$$

This is the intensity as a function of the angle of the ray. It is then easy to see, by considering its maximums and minimums, that we can obtain the same angular positions for the bright and dark spots in the diffraction pattern using this expression.

1.2.3 Diffraction Gratings

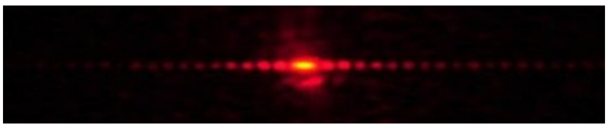
Diffraction gratings are essentially a large array of slits. Similar to the double-slit experiment, shining coherent light onto a diffraction grating will also cause a diffraction pattern (as implied by the name).



Diffraction gratings are characterised by their number of gratings per unit length. With that information, the slit distance d can be calculated as

$$d = \frac{1}{\text{number of gratings per unit length}} \tag{30}$$

The diffraction pattern by a diffraction grating is as shown:



Compared to a double-slit diffraction pattern, the maxima are much more accentuated.

The same reasoning can be applied to determine that for a diffraction grating, the bright spots occur when the path difference between two rays from adjacent slits are a multiple of λ, so

$$d \sin\theta = m\lambda \tag{31}$$

where m is an integer. When this is true, every adjacent pair of rays will constructively interfere, producing a very bright spot on the screen. We do not talk about the minima for a diffraction grating, as the regions between maxima are typically approximately equal in darkness as compared to the intensity of the maxima.

Similarly, we can then conclude that the position of the bright spots, under the second approximation, is

$$y_{\text{bright}} = \frac{m\lambda l}{d} \tag{32}$$

We can now proceed to determine the intensity of the diffraction pattern using phasors again. Suppose there are N slits in total, and that the electric field at the point of the screen by the light ray that passed through the very top slit of the grating can be represented by the phasor

$$\widetilde{E}_1 = E_0$$

Then, for the n-th slit from the top, the electric field by the light ray that passed through that slit can be represented by the phasor

$$\widetilde{E}_n = E_0 e^{(n-1)i\varphi}$$

where φ is the phase difference between two adjacent light rays. Adding all the electric fields,

$$\begin{aligned}\widetilde{E} &= \sum_{n=0}^{N} E_0 e^{(n-1)i\varphi} \\ &= E_0 \frac{1 - e^{Ni\varphi}}{1 - e^{i\varphi}} \\ &= E_0 e^{\frac{(N-1)i\varphi}{2}} \frac{\sin\frac{N\varphi}{2}}{\sin\frac{\varphi}{2}}\end{aligned}$$

obtained by a geometric sum. Finally, we can conclude that

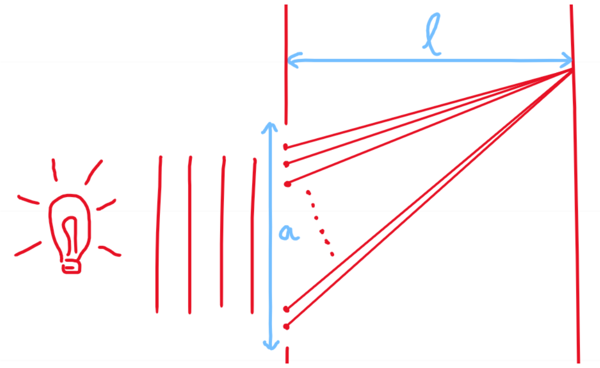
$$I = I_0 \frac{\sin^2\left(\frac{N\pi d \sin\theta}{\lambda}\right)}{N^2 \sin^2\left(\frac{\pi d \sin\theta}{\lambda}\right)} \tag{33}$$

where the factor of N^2 in the denominator is to ensure that the limit as $\theta \to 0$ is I_0.

1.2.4 Single-Slit Diffraction

The picture of the double-slit diffraction pattern shown in the double-slit interference section is a real picture taken. However, it is cropped to only show a portion of the pattern. If we actually expand to see the full picture, the following pattern is shown:

Mysteriously, the overall diffraction pattern *has its own bright and dark spots*! This is actually because the slits have finite width, which we have ignored in previous derivations. It turns out that slits of finite width also cause diffraction effects, known as single-slit diffraction, also known as Fraunhofer diffraction.



To analyse a single slit with finite width a, we can divide the region between the slit into N regions and consider a wave-source emitted from the middle of that region. Then, we can take the limit as $N \to \infty$ to get the full single-slit diffraction pattern. Splitting the single slit into N regions with one wave-source emitted from each is essentially the same as an N-slit diffraction grating with slit distance $\frac{a}{N}$, so the intensity pattern is

$$I = I_0 \frac{\sin^2\left(\frac{\pi a \sin\theta}{\lambda}\right)}{N^2 \sin^2\left(\frac{\pi a \sin\theta}{N\lambda}\right)}$$

If we take the limit as $N \to \infty$, we get the intensity pattern is

$$I = I_0 \frac{\sin^2\left(\frac{\pi a \sin\theta}{\lambda}\right)}{\left(\frac{\pi a \sin\theta}{\lambda}\right)^2} = I_0 \left[\operatorname{sinc}\left(\frac{\pi a \sin\theta}{\lambda}\right)\right]^2 \tag{34}$$

where $\operatorname{sinc} x = \frac{\sin x}{x}$ for $x \neq 0$ and $\operatorname{sinc} 0 = 1$. This is the single-slit diffraction pattern, and also the single-slit diffraction envelope to the double-slit diffraction pattern that we saw earlier.

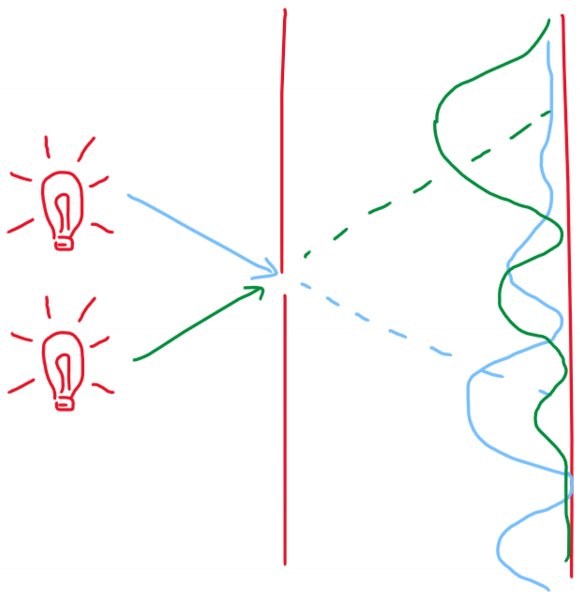
From this intensity pattern, we can determine that the minima of the single-slit diffraction occurs when

$$\frac{\pi a \sin\theta}{\lambda} = m\pi \qquad \Longrightarrow \qquad a\sin\theta = m\lambda \tag{35}$$

where m is an integer. We do not consider the maxima as they are not exactly at the middle of the bright spots and are challenging yet unenlightening to determine exactly.

1.2.5 Rayleigh Criterion

Let's suppose we have two sources viewed through a single-slit.



The diffraction patterns due to both sources will overlap. If we wanted to be able to *resolve* both sources (i.e. be able to visually tell them apart), we need the diffraction patterns to be some distance apart from each other in order to be able to visually separate them. The **Rayleigh criterion** states that the maximum angular separation for two sources to be just resolved occurs when the central maximum from one source coincides with the first-order minimum of the other.

For the rectangular slits that we have considered previously, we simply have that the maximum angular half-width θ between the sources should be

$$a \sin\theta \approx \lambda \quad \Longrightarrow \quad \theta \approx \frac{\lambda}{a} \tag{36}$$

Typically, we use circular apertures as opposed to the rectangular slits that we have considered so far. The treatment of the Fraunhofer diffraction for a circular aperture is much more mathematically involved, requiring the use of Bessel functions. However, the result is that the Rayleigh criterion for a circular slit is

$$\theta \approx \frac{1.22\lambda}{D} \tag{37}$$

where D is the diameter of the aperture.

1.3 Geometrical Optics

In geometrical optics, we shall describe light propagation by rays. Geometrical optics typically applies when the scale of consideration is much larger than the wavelength of light.

1.3.1 Reflection and Refraction

The trajectories taken by light are based off **Fermat's principle**. This principle states that the path taken by light between two points is the one that takes the least time. This principle can also be rephrased in terms of *optical path length*.

The optical path length between two points A and B is

$$\text{OPL} = \int_C n \, \mathrm{d}s \tag{38}$$

where n is the refractive index of the medium and $\mathrm{d}s$ is an infinitesimal displacement vector along the path C from A to B. Recall that the refractive index n of a medium is defined by

$$n = \frac{c}{v} \tag{39}$$

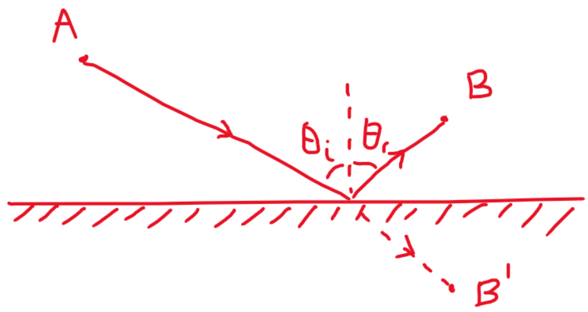
where v is the speed of light (the phase velocity) in the medium. Hence, Fermat's principle can then be rephrased to state that the path taken by light between two points is the one that has the shortest optical path length, since

$$\mathrm{d}t = \frac{\mathrm{d}s}{v} = \frac{n\,\mathrm{d}s}{c} = \frac{\mathrm{d(OPL)}}{c}$$

i.e. optical path length is proportional to time. This allows us to determine certain laws regarding the trajectories that light must follow.

The first is that light travels in a straight line through an unobstructed medium of uniform refractive index. This is clear both intuitively and by Fermat's principle. The medium has uniform refractive index, so the optical path length is proportional to the total path length. Since the shortest distance between two points is a straight line, light must travel in a straight line between two points.

The second is the **law of reflection**. Consider the path taken by light from point A to B reflecting over a reflective mirror near both points.

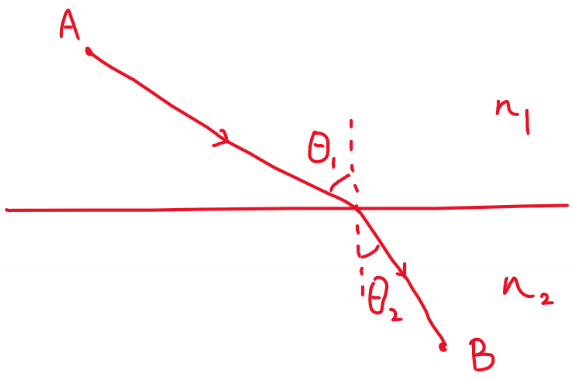


Since the light travels in the same medium, we essentially need to minimise the total length of the path. It is clear that by reflecting the path of the light after reflection across the mirror, the total path length remains the same, so we can deduce that in order the minimise the total path length, the light must travel from A directly towards the image of the endpoint B' across the mirror. Some simple geometry allows us to deduce the law of reflection, which is that the angle of incidence θ_i and angle of reflection θ_r of the light ray are equal:

$$\theta_\mathrm{i} = \theta_\mathrm{r} \tag{40}$$

Here, the angles are defined with respect to the normal of the surface. Here, we can then see that an object placed at point B will form a mirror image at point B'.

The third is the **law of refraction**, also termed Snell's law. Consider the path taken by light from point A to B refracting across the boundary between two mediums of refractive indices n_1 and n_2 respectively.



It can be easily shown via the use of some calculus that the path of least optical path length from points A to B must have the angle of incidence θ_1 and angle of refraction θ_2 satisfy the law of refraction

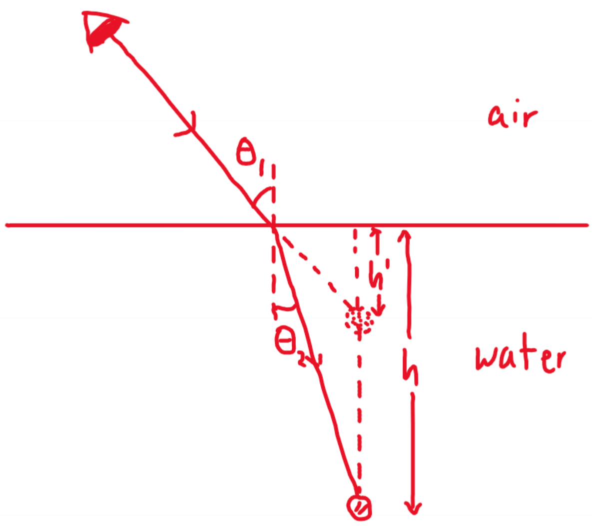
$$n_1 \sin\theta_1 = n_2 \sin\theta_2 \tag{41}$$

where the angles are also defined with respect to the normal of the surface.

In reality, light rays do not generally purely reflect off or refract through the boundary between two different optical mediums. Light will usually split into two beams of different intensities, a reflected beam and a refracted beam.

Example 1.3. Suppose an object is suspended underwater in neutral equilibrium at a distance d below the surface of the water. Let the refractive indices of air and water be 1 and n respectively. Find the apparent depth of the object to an observer above the surface of the water seeing the object directly from above.

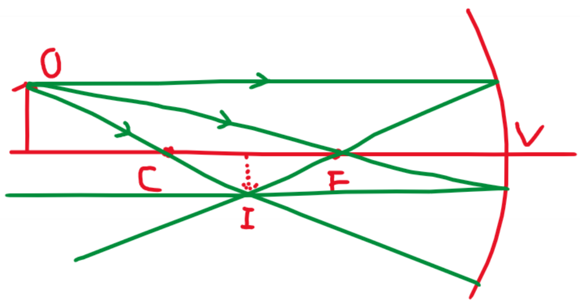
Apparent depth comes about because humans perceive light to travel in a straight line, and so do not account for the bending of light due to refraction.



From the diagram above, we can see that the apparent height h' and actual height h are related by

$$h' \tan\theta_1 = h \tan\theta_2$$

which, for small angles of incidence (and thus, angle of refraction) as the observer is looking at



the object from directly above, we can use the small angle approximation $\tan\theta \approx \sin\theta$ to obtain

$$h' \sin\theta_1 = h \sin\theta_2$$

Therefore, dividing by the equation from Snell's law, we have the apparent height is

$$\frac{h'}{n_1} = \frac{h}{n_2} \quad \Longrightarrow \quad h' = \frac{h}{n}. \tag{42}$$

1.3.2 Total Internal Reflection

Snell's law works for all angles of incidence when light enters an optically denser medium. However, when light enters a less optically dense medium, Snell's law sometimes can no longer apply. To see this, we can solve for the angle of refraction as

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

When $n_1 > n_2$, there are angles θ_1 for which the right hand side is more than one, so that there is no solution for θ_2. The value of θ_1 for which the right hand side is thus termed the **critical angle**, and its value can be determined as

$$\frac{n_1}{n_2}\sin\theta_c = 1 \quad \Longrightarrow \quad \theta_c = \sin^{-1}\frac{n_2}{n_1} \tag{43}$$

For angles of incidence larger than the critical angle, light rays no longer refract into the second medium, but rather purely reflect over the boundary obeying the law of reflection, This is called **total internal reflection**.

1.3.3 Reflection off Spherical Surfaces

Reflection and refraction at planar surfaces are simple to deal with. Sometimes, we need to consider reflection and refraction at curved surfaces, usually a spherical one. Something special about spherical surfaces is that they have light focusing properties, which we shall show later. We will also only be dealing with paraxial light rays (rays that are close to the axis of the mirror).

Suppose we have an object O placed in front of a spherical mirror. If the centre of the spherical mirror is on the same side as the object, it is called a *concave* mirror, else it is called a *convex* mirror.

We term the axis passing through the centre of the spherical mirror and the object the principal axis. We can show using some simple geometry that for rays parallel to the principal axis, no matter the distance to the axis, their reflections will be concurrent at some point on the principal axis. This point is called the **focal point** of the mirror. The focal length of the mirror is defined to be the distance of the focal point to the vertex of the mirror.

Consider the following rays emanating from the top of the object, termed the principal rays:

1. the ray parallel to the principal axis,
2. the ray passing through the centre of the spherical surface,
3. the ray passing through the focal point of the spherical surface.

We can use some simple geometry to deduce that the reflections of these three rays are concurrent. In fact, we can show that for any ray emanating from the top of the object, its reflection will pass through that same point. This point is the point of the top of the **image** of the object formed by the spherical mirror.

We define the following quantities as well as their sign conventions:

1. s and s' are the distances of the object and the image to the mirror respectively. They are positive if the object/image are in front of the mirror and negative if they are behind.
2. y and y' are the heights of the object and image respectively. They are positive if they are above the principal axis and negative if they are below.
3. R and f are the radius and focal length of the mirror respectively. They are positive if the centre and the focal point are in front of the mirror and negative if they are behind.
4. M is the magnification of the image compared to the object. It is positive if the image and object are on the same side of the principal axis and negative if they are on opposite sides.

We also use the following terminology to describe the image:

1. The image is real if the rays pass through the image, while the image is virtual if the rays do not pass through the image (but rather have to be projected backwards to pass through the image).
2. The image is magnified if its size is larger than the object (i.e. $|M| > 1$), while the image is diminished if its size is smaller than the object (i.e. $|M| < 1$).
3. The image is upright if it is on the same side as the object (i.e. $M > 0$), while the image is inverted if it is on the opposite side as the object (i.e. $M < 0$).

Using some simple geometry, we can also determine the following equations that relate the various quantities:

$$f = \frac{R}{2} \tag{44}$$

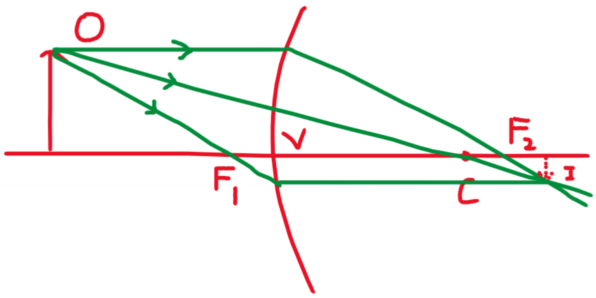
$$\frac{1}{s} + \frac{1}{s'} = \frac{1}{f} = \frac{2}{R} \tag{45}$$

$$M = \frac{y'}{y} = -\frac{s'}{s} \tag{46}$$

These equations hold for both concave and convex mirrors.

1.3.4 Refraction through Spherical Surfaces

When dealing with refraction, we have slightly different conventions for the definitions of quantities.



1. s and s' are the distances of the object and the image to the mirror respectively. They are positive if the object/image are in front of/behind the mirror respectively and negative otherwise.
2. y and y' are the heights of the object and image respectively. They are positive if they are above the principal axis and negative if they are below.
3. R is the radius of the mirror. It is positive if the centre is behind the mirror and negative if it is in front.
4. M is the magnification of the image compared to the object. It is positive if the image and object are on the same side of the principal axis and negative if they are on opposite sides.
5. n and n' are the refractive indices in front of and behind the mirror respectively.

The principal rays, defined similarly to that of the reflection off a spherical surface, are also shown.

Using some geometry combined with Snell's law, we can determine the following equations that relate the various quantities:

$$\frac{n}{s} + \frac{n'}{s'} = \frac{n' - n}{R} \tag{47}$$

$$M = \frac{y'}{y} = -\frac{ns'}{n's} \tag{48}$$

1.3.5 Lenses

Now, we get to lenses. Lenses are comprised of an optical material with two spherical surfaces. Lenses also have the properties that they are focusing.

To determine the properties of a lens, let's define the radii of the lens and their sign conventions:

1. R_1 is the radius of the front spherical surface. R_1 is positive if its centre is behind the mirror and negative if it is in front.
2. R_2 is the radius of the back spherical surface. R_2 is positive if its centre is behind the mirror and negative if it is in front.

Then, we can derive the focal length for a lens by considering the image formed from the first spherical surface as the object for the second spherical surface. We denote the quantities for the intermediate image as s'', y'', and we shall suppose the refractive indices of air and the lens material are 1 and n respectively. Then,

$$\frac{1}{s} + \frac{n}{s''} = \frac{n-1}{R_1} \qquad \frac{n}{-s''} + \frac{1}{s'} = \frac{1-n}{R_2}$$

where we assume that the distance between the vertices of both spherical surfaces are negligible, called the thin lens approximation. Adding both equations gives the relation between s and s' as

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

It is then easy to see by considering special cases for the object and image positions that there is one focal point, and the focal length satisfies the equation

$$\frac{1}{f} = (n-1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) \tag{49}$$

which is called the **lensmaker's equation**. Then, we can show that the following equations for a spherical mirror also hold for a lens:

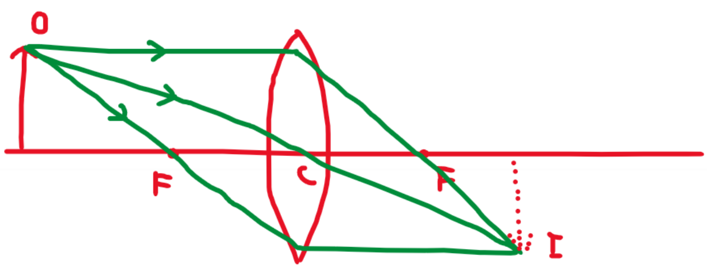
$$\frac{1}{s} + \frac{1}{s'} = \frac{1}{f} \tag{50}$$

$$M = \frac{y'}{y} = -\frac{s'}{s} \tag{51}$$

Lens can be categorised into two main categories: *converging* and *diverging* lens, having focal lengths $f > 0$ and $f < 0$ respectively.

When considering ray diagrams involving lens, the following principal rays are also important:

1. the ray parallel to the principal axis,
2. the ray passing through the centre of the lens (not the centres of either surface),
3. the ray passing through the focal point of the lens.



1.4 Ideas

1.4.1 Beats

Open two tone generators and play two sinusoidal tones with frequencies 440 Hz and 441 Hz at the same time. Do you expect what you hear?

Humans can barely even tell apart frequencies that are 1 Hz apart played alone, yet when they are played together, an obvious oscillating variation in amplitude can be heard. In fact, if you pay attention to the duration it takes to make one period, you find that the frequency of the oscillation is exactly 1 Hz. This is the concept of *beats*.

Let's quantify this. Suppose we have two sound waves of equal amplitudes, with difference in frequencies $\Delta\omega$ and difference in wavenumbers Δk, with $\Delta\omega \ll \omega$ and $\Delta k \ll k$.

$$\psi_1(x,t) = A\cos(kx - \omega t)$$
$$\psi_2(x,t) = A\cos\left((k+\Delta k)x - (\omega+\Delta\omega)t\right)$$

When they superimpose, we will get

$$\begin{aligned}\psi(x,t) &= A\cos(kx-\omega t) + A\cos\left((k+\Delta k)x - (\omega+\Delta\omega)t\right) \\ &= 2A\cos\left(\left(k+\frac{\Delta k}{2}\right)x - \left(\omega+\frac{\Delta\omega}{2}\right)t\right)\cos\left(\frac{\Delta k}{2}x - \frac{\Delta\omega}{2}t\right)\end{aligned}$$

The resultant wave is the product of two sinusoidal functions. The first is a sinusoidal function with the average frequency of its component waves and the average wavenumber of its component waves. Since the difference in frequencies and wavelengths are small, this sinusoidal wave has a phase velocity of $\frac{\omega+\frac{\Delta\omega}{2}}{k+\frac{\Delta k}{2}} \approx \frac{\omega}{k}$, essentially the same as the original two component waves.

The second is a sinusoidal wave with frequency $\frac{\Delta\omega}{2}$ and wavenumber $\frac{\Delta k}{2}$. This is essentially an envelope for the other sinusoidal wave. The velocity for this envelope is $\frac{\frac{\Delta\omega}{2}}{\frac{\Delta k}{2}} = \frac{\Delta\omega}{\Delta k}$.

Now, we see that the origin of beats comes from the envelope. Let's consider a particular position in space, $x = 0$, and consider only the envelope as its only wave. Then, the amplitude of the wave is

$$\psi_{\text{env}}(0,t) = 2A\cos\left(\frac{\Delta\omega}{2}t\right)$$

The intensity is proportional to the square of the amplitude, so

$$I \propto \cos^2\left(\frac{\Delta\omega}{2}t\right) = \frac{1+\cos\left(\Delta\omega\, t\right)}{2}$$

As we can see, if the frequencies of the original two waves were f_1 and f_2, then the frequency of the resulting envelope, called the beat frequency, is

$$\Delta f = |f_2 - f_1| \tag{52}$$

This explains the resultant 1 Hz beat heard in the example.

1.4.2 Phase and Group Velocity

In the previous example, we saw that if two waves similar in frequency and wavenumber are superimposed, there is a resulting envelope moving at a different velocity from its constituent parts. This allows us to introduce the idea of *phase and group velocity.*

Phase velocity is the velocity of sinusoidal waves that we have been talking about in the chapter. The definition for the phase velocity is

$$v_{\text{p}} = \frac{\omega}{k} \tag{53}$$

which is the exact same equation as that which we derived when we introduced the basic quantities. However, we now have a definition for the group velocity,

$$v_{\text{g}} = \frac{\mathrm{d}\omega}{\mathrm{d}k} \tag{54}$$

This is analogous to the velocity of the envelope. For sinusoidal waves, both of these quantities are equal. However, for general waves, these two quantities may differ.

1.4.3 Doppler Effect

If you've ever heard an ambulance pass by, you might have noticed that the pitch of its siren changes as it passes by you. The actual frequency of the siren doesn't actually change, so why do you hear a different frequency?

This phenomenon is called the *Doppler effect*. Even though the actual frequency of the siren does not change, if you or the source are moving towards each other, by the time the next crest of the wave is emitted, it has already moved closer to you, meaning it will arrive closer to the previous crest than usual. This increases your perceived frequency. A similar logic applies for when you or the source are moving away from each other.

Let the wave speed be v, and suppose v_s and v_o are the source and observer's velocities, defined to be positive when moving towards the other.

Suppose a crest of the wave is emitted at time $t = 0$, when the source and the observer are at distance d. The crest reaches the observer when

$$vt + v_\text{o}t = d \quad \implies \quad t = \frac{d}{v + v_\text{o}}$$

The next crest of the wave is emitted after one period, $t = T_\text{s}$, when the source and the observer are at distance $d' = d - (v_\text{s} + v_\text{o})\,T_\text{s}$. Using our previous derivation, the crest reaches the observer at

$$t = T_\text{s} + \frac{d'}{v + v_\text{o}} = T_\text{s} + \frac{d - (v_\text{s} + v_\text{o})T_\text{s}}{v + v_\text{o}} = \frac{v - v_\text{s}}{v + v_\text{o}}T_\text{s} + \frac{d}{v + v_0}$$

Therefore, the time interval between the two adjacent crests received by the observer, which is the perceived period, is

$$T_\text{o} = \frac{v - v_\text{s}}{v + v_\text{o}}T_\text{s}$$

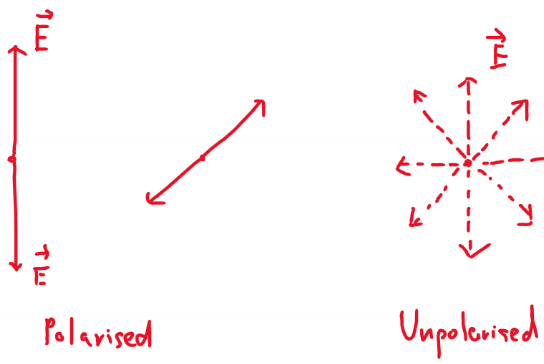
and so, taking the reciprocal, we arrive at the expression for the perceived frequency due to the Doppler effect,

$$f_\text{o} = \frac{v + v_\text{o}}{v - v_\text{s}}f_\text{s} \tag{55}$$

Realise that since you and the source are moving towards each other, the perceived frequency increases as expected. If instead, you and the source were moving away from each other, the signs of v_s and v_o would flip, resulting in a decreased frequency as expected.

1.4.4 Polarisation

Light is an electromagnetic wave that can exhibit *polarisation*, i.e. the ability to oscillate in more than one orientation. This is because, as a transverse wave, its electric field oscillates in one axis, which can be rotated.



Linearly polarised light consists of an electric field that only oscillates in one axis. In contrast, unpolarised light has a random, time-varying polarisation that can be treated as a uniform mixture of linearly polarised light in each direction.

Polarisers are optical filters that only allow one specific polarisation of light to pass through and blocks all other polarisations. If polarised light with electric field amplitude E_0 passes through a polariser that is at an angle θ to the oscillation of the electric field, only the component of the electric field that is parallel to the polariser remains, so

$$E = E_0 \cos\theta$$

Since intensity is proportional to the square of the amplitude, we obtain

$$I = I_0 \cos^2\theta \tag{56}$$

also called Malus's law.

If unpolarised light passes through a polariser, independent on the angle, the intensity of the polarised light out of the polariser will then be

$$I = I_0 \left\langle \cos^2\theta \right\rangle = \frac{1}{2} I_0 \tag{57}$$

since unpolarised light is a uniform mixture of linear polarisations at all angles.

Light can also be polarised when it passes between the boundaries of two optical media, in which both the reflected and refracted ray are polarised to some extent. It turns out that at some angle of incidence, when the reflected and refracted rays are perpendicular to each other, the reflected is completely polarised. Using Snell's law, we can deduce that this happens when

$$n_1 \sin\theta_B = n_2 \sin(90^\circ - \theta_B) \implies \tan\theta_B = \frac{n_2}{n_1} \tag{58}$$

which is called Brewster's angle.

1.4.5 Thin-film Interference

One subtle thing we have not discussed is how phase evolves in an optical medium. In an optical medium, since the phase velocity of light is decreased by a factor of n, the ratio between the phase difference of two points and their spatial separation actually increases by a factor of n, i.e.

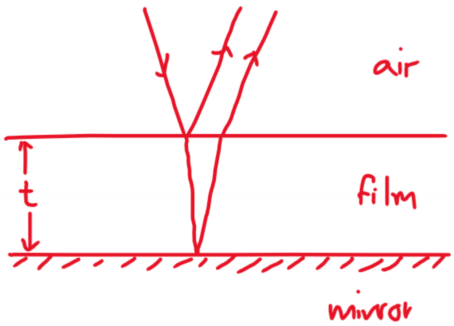
$$\frac{\Delta\varphi}{2\pi} = \frac{n\Delta x}{\lambda} = \frac{\Delta(\text{OPL})}{\lambda} \tag{59}$$

where OPL is the optical path length as we have introduced earlier.

It turns out there is also one more caveat regarding optical media and phases. If a ray is reflected off an optically denser medium, it turns out that there is a **π-radian phase shift** in the ray! On the other hand, the refracted ray is not phase shifted. Also, if a ray is reflected off a less optically dense medium, the ray is not phase shifted either. This has to do with the electromagnetic properties of light, which we will not discuss here.

These facts allow us to consider one specific example of interference, *thin-film interference.*

Example 1.4. Coherent light is shined perpendicularly onto a thin soap film of refractive index $n > 1$ and thickness t that rests on top of a mirror less optically dense than the film. Determine the conditions for the destructive interference of the ray reflected off the surface of the film and the ray reflected off the mirror and refracted back out of the film.



In the above figure, where the angle of incidence is exaggerated for clarity, we can observe the two intended rays that are to destructively interfere. They are able to interfere due to the difference in optical path length of the two rays, which is

$$\Delta(\text{OPL}) = nt$$

Furthermore, note that since the film is optically denser than air, the ray that reflects off the film experiences a π-radian phase shift. On the other hand, since the mirror is less optically dense than the film, the ray that refracts through the film and reflects off the mirror does not experience a phase shift at all, even after refracting into the air. Therefore, for the two rays to destructively interfere, we must have

$$\Delta(\text{OPL}) = nt = \left(m - \frac{1}{2}\right)\lambda + \frac{\lambda}{2} = m\lambda$$

where m is an integer. The extra $\frac{\lambda}{2}$ comes from the π-radian phase shift of one ray.

Notice that what this equation implies is that if the thin-film is sufficiently small (significantly smaller than $\frac{\lambda}{n}$), the film actually appears completely dark from viewed from above! This is because the path difference between the two rays are insignificant, so due to the phase shift, they destructively interfere.

1.4.6 Variable Refractive Index

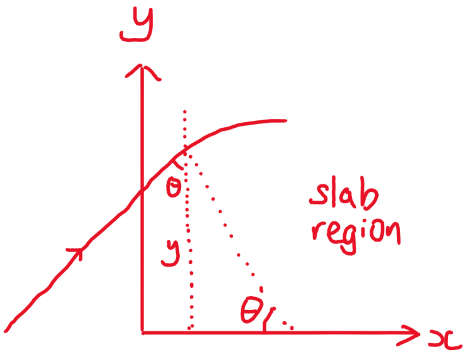
Snell's law tells us that for a ray entering another optical medium,

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

This actually can be rephrased in another way that is useful in situations where we have a varying refractive index:

$$n \sin\theta = \text{constant} \tag{60}$$

Example 1.5 (Ricardo)**.** Show that light travels in circular arcs in a slab where the speed of light varies linearly with height.



Since the speed of light varies linearly with height, let $v = \alpha y$ where y is the vertical position and let's suppose $\alpha > 0$. Then $n = \frac{c}{\alpha y}$, so We have

$$\frac{c}{\alpha y}\sin\theta = \frac{c}{\alpha y_0}\sin\theta_0 \qquad \Longrightarrow \qquad \sin\theta = \frac{y}{y_0\csc\theta_0}$$

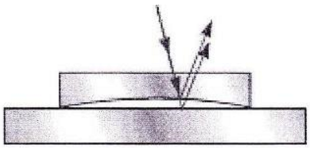
The key observation that allows us to deduce that light travels in circular arcs is that the equation obtained describes the angle from the vertical in a circle of radius $y_0 \csc\theta_0$ centered on the x-axis, as also illustrated in the diagram above. This gives us the conclusion that light will take that circular trajectory within the slab.

Of course, we can prove this rigorously by relating $\sin\theta$ to $\frac{\mathrm{d}y}{\mathrm{d}x}$ and solving for the equation of the curve taken. We will omit this derivation for brevity.

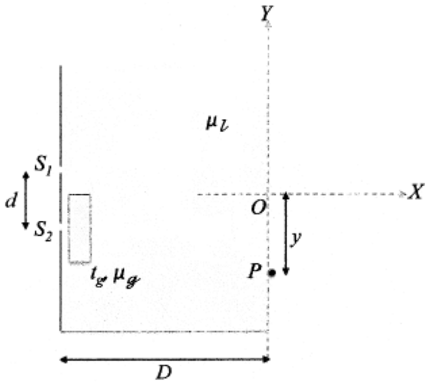
2 Problems

Problems are arranged in roughly increasing difficulty.

Problem 2.1 (SPhO 2010)**.** A plano-concave lens having index of refraction 1.50 is placed on a flat glass plate, as shown in the figure below. Its curved surface, with radius of curvature 8.00 m, is on the bottom. The lens is illuminated from above with yellow sodium light of wavelength 589 nm, and a series of concentric bright and dark rings is observed by reflection. The interference pattern has a dark spot at the centre, surrounded by 50 dark rings, of which the largest is at the outer edge of the lens. (i) What is the thickness of the air layer at the centre of the interference pattern? (ii) Calculate the radius of the outermost dark ring. (iii) Find the focal length of the lens.



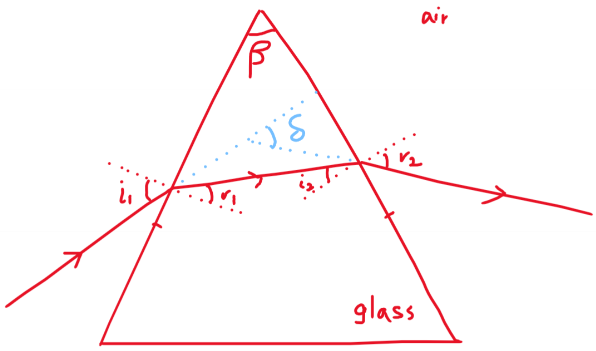
Problem 2.2 (SPhO 2012)**.** In a Young's double slit experiment, the region between screen and slits is immersed in a liquid whose refractive index varies with time t (in seconds) as $n_l = 2.50 - 0.25t$ until it reaches a steady state value of 1.25. The distance between the slits and the screen is $D = 1.00$ m and the distance between the slits S_1 and S_2 is $d = 2.00 \times 10^{-3}$ m. A glass plate of thickness $t_g = 3.60 \times 10^{-5}$ m and refractive index $n_g = 1.50$ is introduced in front of one of the slits. Note that the illuminations at S_1 and S_2 are from coherent sources with zero phase difference.



(i) Consider the point P on the screen at a distance y from O ($S_1O = S_2O; OP = y$). Obtain the expression for the optical path difference Δx in terms of the refractive indices. (ii) If P is the central maximum, obtain the expression for y as a function of time t. (iii) Obtain the time t_m when the central maximum is at point O, equidistant from S_1 and S_2 (i.e. $S_1O = S_2O$). (iv) Determine the speed v of the central maximum when it is at O.

Problem 2.3 (SPhO 2018)**.** A copper wire with mass m is stretched between two fixed points, distance l apart and a tension F_T is applied to it. When the copper wire is vibrating in the fundamental mode together with a 256 Hz tuning fork, a beat frequency of 5 Hz is observed. The copper wire is removed and a brass (which is an alloy made of copper and zinc) wire, with the same length and diameter, is stretched between the same two fixed points. The same tension is again applied to the brass wire. It is found that in this case, the brass wire, vibrating in the fundamental mode, resonates with the 256 Hz tuning fork when the two are vibrated together. Take the densities of copper and zinc to be 8940 kg m^{-3} and 7140 kg m^{-3} respectively. (i) State an equation for the speed of the wave on the string in terms of m, l and F_T only. (ii) Determine the percentage by mass of zinc in the brass wire. State any assumptions you make in your calculation.

Problem 2.4. Consider a light ray incident onto a glass prism.



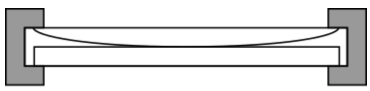
(a) Determine the deviation angle δ in terms of i_1 and n. (b) Prove that for small apex angles β and small angles of incidence i_1, $\delta \approx (n-1)\beta$. (c) Prove that the minimum $\delta_{\min}$ satisfies

$$\sin\left(\frac{\delta_{\min}+\beta}{2}\right) = n \sin\frac{\beta}{2}$$

Problem 2.5 (SPhO 2014)**.** The index of refraction of glass can be increased by diffusing in impurities. It is then possible to make a lens of constant thickness. Given a disk of radius a and thickness d, express the index of refraction $n(r)$ as a function of n_0 (refractive index at the centre of the lens), r (radius from the centre of the lens), d and f (focal length). You may assume $d \ll a$.

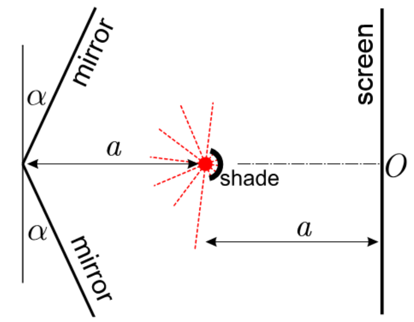
Problem 2.6 (NBPhO 2018)**.** In this problem, we analyse the working principle of a speed camera. The transmitter of the speed camera emits an electromagnetic wave of frequency $f_0 = 24\,\text{GHz}$ having waveform $\cos(2\pi f_0 t)$. The wave gets reflected from an approaching car moving at speed v. The reflected wave is recorded by the receiver of the speed camera. (i) Find the frequency f_1 of the reflected wave. (ii) In the speed camera, the received waveform is multiplied with the original emitted waveform. Express all frequency components present in the multiplied signal. (iii) Given the lowest frequency component present in the multiplied signal is $f_{\text{low}} = 4.8\,\text{kHz}$, calculate the speed of the car v. The trigonometric identity $\cos\alpha\cos\beta = \frac{1}{2}(\cos(\alpha+\beta)+\cos(\alpha+\beta))$ may be useful for this problem.

Problem 2.7. From above, coherent monochromatic light of wavelength λ illuminates a convex curved glass with radius of curvature R that lies on another flat piece of glass, forming a diffraction pattern on the flat glass. Find the radius of the n-th bright ring of the diffraction pattern.

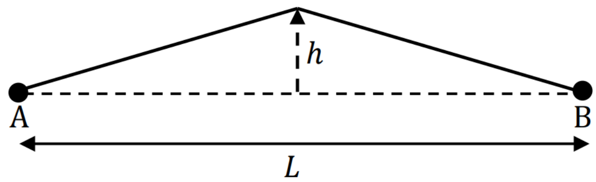


Problem 2.8 (SPhO 2006)**.** A man stands on a long plane concrete runway above which a uniform vertical temperature gradient results in a uniform gradient in the refractive index of the air, $n(x, y) = n_0(1+\alpha y)$, where $|\alpha| = 1.5\times 10^{-6}\,\text{m}^{-1}$. As a result, he cannot see the runway ($y = 0$) beyond a certain distance d. If his eyes are 1.7 m above the runway, find the value of d. Does the temperature rise or fall with increasing height?

Problem 2.9 (Est-PhO 2004)**.** A screen, two mirrors, and a source of monochromatic light are positioned as shown in the figure below. Due to a shade, only reflected light from the source can reach the screen. There will be a striped interference pattern on the screen, where the distance between consecutive stripes is d. Express the wavelength of the light λ in terms of d and the distance a (see the figure below). Assume that $a \gg d$.

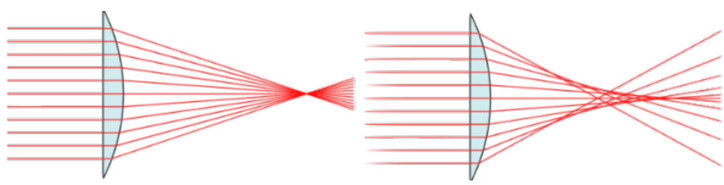


Problem 2.10 (APhO 2010 T1)**.** Consider an elastic string stretched between two fixed ends A and B, as shown in the figure below. The linear mass density of the string is μ. The speed of propagation for transverse waves in the string is c. Let the length $\overline{AB}$ be L. The string is plucked sideways and held in a triangular form with a maximum height $h \ll L$ at its middle point. At time $t = 0$, the plucked string is released from rest. All effects due to gravity may be neglected.



(a) Find the period of vibration T for the string and plot the shape of the string at $t = \frac{T}{8}$. In the plot, specify lengths and angles which serve to define the shape of the string. (b) Find the total mechanical energy of the vibrating string in terms of μ, c, h and L.

Problem 2.11 (Kevin Zhou)**.** Parallel light rays coming in along the $+\hat{\mathbf{x}}$ direction enter a lens of index of refraction n, whose left edge is at $x = 0$ and whose right edge is described by the function $x(y)$. If all the light beams are to be focused at $x = f$, as shown at left below, what kind of curve does $x(y)$ have to be?



You should find that $x(y)$ is not an arc of a circle, which implies that a spherical lens will fail to focus all incoming horizontal light to a point. Instead, we will get spherical aberration, as shown at right above. However, most lenses are spherical because it's easier to make them that way.

Problem 2.12. Monochromatic light of wavelength λ is incident normally on a single slit of width a. The lower half of the slit is covered with a thin membrane which causes a phase shift of π rad for light that passes through it. (a) Derive the intensity I of the diffracted light on a screen far away from the slit in terms of intensity I_0 at the centre of the diffraction pattern of an uncovered single slit and the phase difference 2α between the light waves from the top and middle of the slit. (b) Sketch a graph to show how intensity depends on the angular position θ when $a = 10\lambda$.

Problem 2.13. A linear source AB and its image through a thin lens $A'B'$ are shown below. Construct geometrically the lens (position and orientation) and its focal points.

3 Advanced Problems

These problems are way too difficult to be tested in a modern-day SPhO. If you have completed all the previous problems and are down for a challenge, try these!

Problem 3.1 (NBPhO 2013 Q9)**.** An excellent problem that tests your geometrical optics skills by requiring you to read off an image.

Problem 3.2 (IPhO 2021 T1B)**.** An excellent problem on the optics of earthquake waves.