9. Wave phenomena

Contents

Essential Ideas

The solution of the harmonic oscillator can be framed around the variation of kinetic and potential energy in the system.

Single-slit diffraction occurs when a wave is incident upon a slit of approximately the same size as the wavelength.

Interference patterns from multiple slits and thin films produce accurately repeatable patterns.

9.1 Simple harmonic motion

NATURE OF SCIENCE:

Insights: The equation for simple harmonic motion (SHM) can be solved analytically and numerically. Physicists use such solutions to help them to visualize the behaviour of the oscillator. The use of the equations is very powerful as any oscillation can be described in terms of a combination of harmonic oscillators. Numerical modelling of oscillators is important in the design of electrical circuits. (1.11)

AHL

Essential idea: The solution of the harmonic oscillator can be framed around the variation of kinetic and potential energy in the system.

Understandings:

- The defining equation of SHM
- Energy changes

Recall from chapter 4 that if the **acceleration** $\boldsymbol{a}$ of a system is:

- **directly proportional to its displacement** $\boldsymbol{x}$ **from its equilibrium position**

  and

- **directed towards the equilibrium position**

  then the system will execute SHM.

This is the formal definition of SHM.

We expressed this definition mathematically as

$\boldsymbol{a} = -\text{const} \times \boldsymbol{x}$

The negative sign indicated that the acceleration was directed towards the equilibrium position. Mathematical analysis shows that the constant is in fact equal to ω2 where ω is the angular frequency (defined above) of the system. Hence equation 4.5 becomes

$\boldsymbol{a} = -\omega^2 \boldsymbol{x}$

If a system is performing SHM, then to produce the acceleration, a force must be acting on the system in the direction of the acceleration. From our definition of SHM, the magnitude of the force F is given by

$\boldsymbol{F} = -\text{k}\boldsymbol{x}$

where k is a constant and the negative sign indicates that the force is directed towards the equilibrium position of the system. (Do not confuse this constant k with the spring constant. However, when dealing with the oscillations of a mass on the end of a spring, k will be the spring constant.)

To understand the solutions of the SHM equation, let us consider the oscillations of a mass suspended from a vertically supported spring. We shall consider the mass of the spring to be negligible and for the extension x to obey the rule $F = kx$. F $(= mg)$ is the force causing the extension. Figure 901(a) shows the spring and a suspended weight of mass m in equilibrium. In Figure 901(b), the weight has been pulled down a further extension x_0.

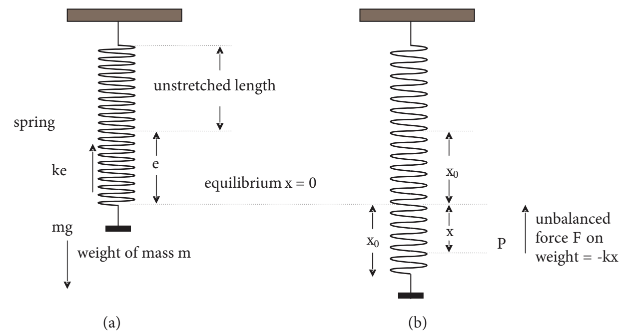


Figure 901 SHM of a mass suspended by a spring

In Figure 901 (a), the equilibrium extension of the spring is e and the net force on the weight is $mg - ke = 0$.

In Figure 408 (b), if the weight is held in position at $x = x_0$ and then released, when the weight moves to position P, a distance x from the equilibrium position $x = 0$, the net force on the weight is $mg - ke - kx$. Clearly, then the unbalanced force on the weight is $-kx$. When the weight reaches a point distance x above the equilibrium position, the compression force in the spring provides the unbalanced force towards the equilibrium position of the weight.

The acceleration of the weight is given by Newton's second law;

$F = -kx = ma$

i.e. $a = -\frac{k}{m}x$

This is of the form $a = -\omega^2 x$ where $\omega = \sqrt{\frac{k}{m}}$, that is, the weight will execute SHM with a frequency

$T = 2\pi\sqrt{\frac{m}{k}}$

The displacement of the weight x is given by

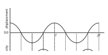
$x = x_0 \cos \omega t = x_0 \cos \sqrt{\frac{k}{m}} t$

This is the particular solution of the SHM equation for the oscillation of a weight on the end of a spring. This system is often referred to as *a harmonic oscillator*.

Alternatively, if the weight was at its equilibrium position when t = 0, then the trigonometric expression would be

$$x = x_0 \sin\omega t$$

The velocity v of the weight at any instant can be found by finding the **gradient** of the displacement-time graph at that instant. The displacement time graph is a cosine function and the gradient of a cosine function is a negative sine function. The gradient of

$x = x_0\cos\omega t$ is in fact  so

$$v = -\omega x_0 \sin\omega t = -v_0 \sin\omega t$$

where v_0 is the maximum and minimum velocity equal in magnitude to ωx_0.

Students familiar with calculus will recognise the velocity v as

$$v = \frac{dx}{dt} = \frac{d}{dt}\left(x_0 \cos\omega t\right) = -\omega x_o \sin\omega t$$. Similarly,

$$a = \frac{dv}{dt} = \frac{d}{dt}\left(-\omega x_0 \sin\omega t\right) = -\omega^2 x_0 \cos\omega t = -\omega^2 x$$

which of course is just the defining equation of SHM.

However, we have to bear in mind that ωt varies between 0 and 2π where $\cos\omega t$ is negative for ωt for $\frac{\pi}{2} \rightarrow \frac{3\pi}{2}$ and $\sin\omega t$ is negative for ωt in the range π to 2π. This effectively means that when the displacement from equilibrium is positive, the velocity is negative and so directed towards equilibrium. When the displacement from equilibrium is negative, the velocity is positive and so directed away from equilibrium

The sketch graph in Figure 902 shows the variation with time t of the displacement x and the corresponding variation with time t of the velocity v. This clearly demonstrates the relation between the sign of the velocity and sign of the displacement.

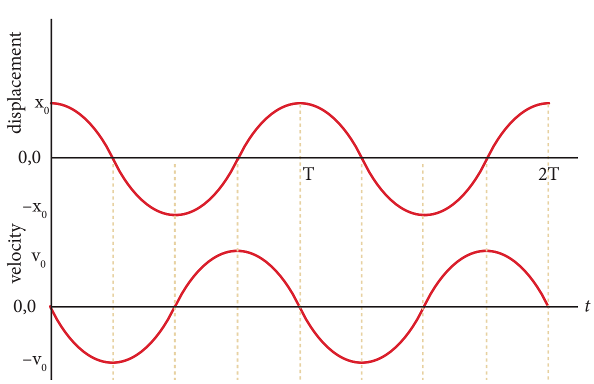


Figure 902 Displacement-time and velocity-time graphs

We can also see how the velocity v changes with displacement x.

We can express $\sin\omega t$ in terms of $\cos\omega t$ using the trigonometric relation

$$\sin^2\theta + \cos^2\theta = 1$$

From which it can be seen that

$$\sin\theta = \sqrt{1-\cos^2\theta}$$

Replacing θ with ωt we have

$$v = -\omega x_0\sqrt{1-\cos^2\omega t}$$

Remembering that $x = x_0\cos\omega t$ and putting x_0 inside the square root gives

$$v = -\omega\sqrt{\left(x_0^2 - x^2\right)}$$

Bearing in mind that v can be positive or negative, we must write

$$v = \pm\omega\sqrt{\left(x_0^2 - x^2\right)}$$

The velocity is zero when the displacement is a maximum and is a maximum when the displacement is zero.

The graph in Figure 903 shows the variation with x of the velocity v for a system oscillating with a period of 1 second and with an amplitude of 5 cm. The graph shows the variation over a time of any one period of oscillation.

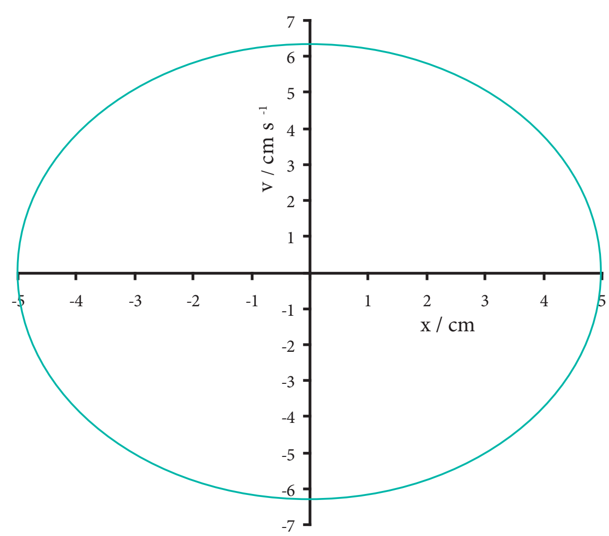


Figure 903 Velocity-displacement graph

Boundary Conditions

The two solutions to the general SHM equation are $x = x_0 \cos\omega t$ and $x = x_0 \sin\omega t$. Which solution applies to a particular system depends, as mentioned above, on the boundary conditions for that system. For systems such as the harmonic oscillator and the simple pendulum, the boundary condition that gives the solution $x = x_0 \cos\omega t$ is that the displacement $x = x_0$ when $t = 0$. For some other systems it might turn out that $x = 0$ when $t = 0$. This will lead to the solution $x = x_0 \sin\omega t$. From a practical point of view, the two solutions are essentially the same; for example when timing the oscillations of a simple pendulum, you might decide to start the timing when the pendulum bob passes through the equilibrium position. In effect, the two solutions differ in phase by $\frac{\pi}{2}$.

The table in Figure 904 summarises the solutions we have for SHM.

$x = x_0\cos\omega t$	$x = x_0\sin\omega t$
$v = -v_0\sin\omega t$	$v = v_0\cos\omega t$
$v = -\omega x_0\sin\omega t$	$v = -\omega x_0\cos\omega t$
$v = \pm\omega\sqrt{x_0^2 - x^2}$	$v = \pm\omega\sqrt{x_0^2 - x^2}$

Figure 904 Common equations

We should mention that since the general solution to the SHM equation is $x = x_0 \sin\omega t + x_0 \cos\omega t$ there are in fact three solutions to the equation. This demonstrates a fundamental property of second order differential equations; that one of the solutions to the equation is the sum of all the other solutions. This is the mathematical basis of the so-called **principle of superposition**.

SHM is a very good example in which to apply the Newtonian method discussed in Chapter 2; i.e. if the forces that act on a system are known, then the future behaviour of the system can be predicted. Here we have a situation in which the force is given by $-kx$. From Newton's second law therefore $-kx = ma$, where m is the mass of the system and a is the acceleration of the system. However, the acceleration is not constant. For those of you who have a mathematical bent, the relation between the force and the acceleration is written as $-kx = m\frac{d^2x}{dt^2}$. This is what is called a "second order differential equation". The solution of the equation gives x as a function of t. The actual solution is of the SHM equation is

$x = P\cos\omega t + Q\sin\omega t$ where P and Q are constants and ω is the angular frequency of the system and is equal to $\sqrt{\frac{k}{m}}$.

Whether a particular solution involves the sine function or the cosine function, depends on the so-called 'boundary conditions'. If for example $x = x_0$ (the amplitude) when $t = 0$, then the solution is $x = x_0\cos\omega t$.

The beauty of this mathematical approach is that, once the general equation has been solved, the solution for all systems executing SHM is known. All that has to be shown to know if a system will execute SHM, is that the acceleration of the system is given by Equation 4.5 or the force is given by equation 4.7. The physical quantities that ω will depend on is determined by the particular system. For example, for a weight of mass m oscillating at the end of a vertically supported spring whose spring constant is k, then $\omega = \sqrt{\frac{k}{m}}$ or , from equation 4.4 $T = 2\pi\sqrt{\frac{m}{k}}$ For a simple pendulum, $\omega = \sqrt{\frac{g}{l}}$ where l is the length of the pendulum and g is the acceleration of free fall such that $T = 2\pi\sqrt{\frac{l}{g}}$.

Examples

The graph in Figure 905 shows the variation with time t of the displacement x of a system executing SHM. (Some question parts were already done in the SL chapter 4).

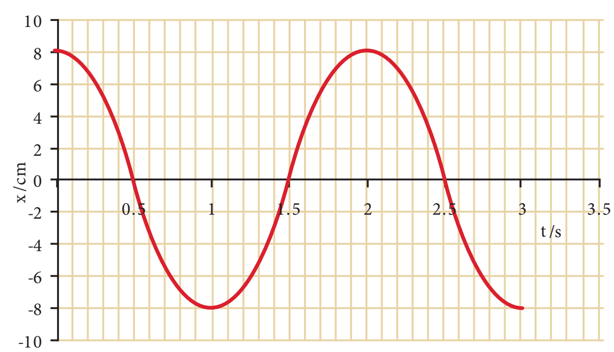


Figure 905 Displacement – time graph for SHM

Use the graph to determine the

(i) period of oscillation
(ii) amplitude of oscillation
(iii) maximum speed
(iv) speed at $t = 1.3$ s
(v) maximum acceleration

Solutions

(i) *2.0 s (time for one cycle)*
(ii) *8.0 cm*
(iii) *Using* $v_0 = \omega x_0$ *gives* $\frac{2\pi}{2} \times 8.0$
(remember that $\omega = \frac{2\pi}{T}$*) = 25 cm s*$^{-1}$

(iv) $v = -v_0 \sin\omega t = -25\sin(1.3\pi)$. *To find the value of the sine function, we have to convert the 1.3π into degrees (remember ω and hence ωt, is measured in radians)*

$1 \text{ radian} = \frac{180}{\pi}$ *deg therefore* $1.3\pi = 1.3 \times 180 = 234°$

therefore $v_1 = -25\sin(234°) = +20 \text{ cm s}^{-1}$.

Or we can solve using $\omega\sqrt{(x_0^2 - x^2)}$

from the graph at t = 1.3 s, x = – 4.8 cm

therefore $v = \pi \times \sqrt{(8.0)^2 - (4.8)^2} = 20 \text{ cm s}^{-1}$

(v) *Using* $a_{max} = \omega^2 x_0 = \pi^2 \times 8.0 = 79 \text{ m s}^{-2}$

Energy changes

We must now look at the energy changes involved in SHM. To do so, we will again concentrate on the harmonic oscillator. The mass is stationary at $x = +x_0$ (maximum extension) and also at $x = -x_0$ (maximum compression). At these two positions the energy of the system is all potential energy and is in fact the elastic potential energy stored in the spring. This is the total energy of the system E_T and clearly

$$E_T = \frac{1}{2}kx_0^2 \qquad \text{Equation 4.12}$$

That is, that for any system performing SHM, the energy of the system is proportional to the square of the amplitude. This is an important result and one that we shall return to when we discuss wave motion.

At $x = 0$ the spring is at its equilibrium extension and the magnitude of the velocity v of the oscillating mass is a maximum v_0. The energy is all kinetic and again is equal to E_T. We can see that this is indeed the case as the expression for the maximum kinetic energy E_{max} in terms of v_0 is

$$E_{max} = \frac{1}{2}mv_0^2$$

Clearly E_T and E_{max} are equal such that

$$E_T = \frac{1}{2}kx_0^2 = E_{max} = \frac{1}{2}mv_0^2$$

From which

$$v_0^2 = \frac{k}{m}x_0^2 \quad (\text{as } v_0 \geq 0)$$

Therefore

$$v_0 = \sqrt{\frac{k}{m}}\,x_0 = \omega x_0$$

which ties in with the velocity being equal to the gradient of the displacement-time graph (see 4.1.5). As the system oscillates there is a continual interchange between kinetic energy and potential energy such that the loss in kinetic energy equals the gain in potential energy and $E_T = E_K + E_P$.

Remembering that $\omega^2 = \frac{k}{m}$, we have that

$$E_T = \frac{1}{2}m\omega^2 x_0$$

Clearly, the potential energy E_P at any displacement x is given by

$$E_P = \frac{1}{2}m\omega^2 x$$

At any displacement x, the kinetic energy E_K is $E_K = \frac{1}{2}mv^2$ Hence remembering that

$v = -\omega\sqrt{(x_0^2 - x^2)}$, we have

$$E_K = \frac{1}{2}m\omega^2(x_0^2 - x^2)$$

Although we have derived these equations for a harmonic oscillator, they are valid for any system oscillating with SHM. The sketch graph in Figure 906 shows the variation with displacement x of E_K and E_P for one period.

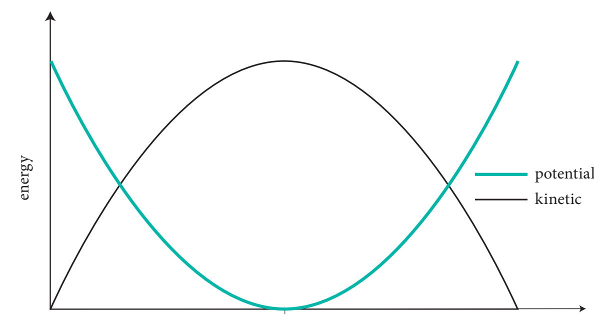


Figure 906 Energy and displacement

Example

The amplitude of oscillation of a mass suspended by a vertical spring is 8.0 cm. The spring constant of the spring is 74 N m–1. Determine

(a) the total energy of the oscillator

(b) the potential and the kinetic energy of the oscillator at a displacement of 4.8 cm from equilibrium.

Solution

(a) *Spring constant* $k = 74\ N\ m^{-1}$ *and*

$x_0 = 8.0 \times 10^{-2}\ m$

$E_T = ½\ kx_0^{\ 2}$

$= ½ \times 74 \times 64 \times 10^{-4} = 0.24\ J$

(b) *At* $x = 4.8\ cm$

Therefore $E_P = ½ \times 74 \times (4.8 \times 10^{-2})^2$

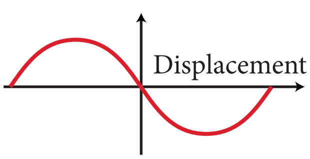
$= 0.085 \times 10^{-4}\ J$

$E_K = E_T - E_P = 0.24 - 0.085 = 0.16\ J$

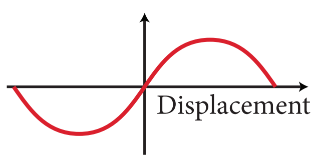
EXERCISE 9.1

1. Which graph shows the relationship between the acceleration a and the displacement *x* from the equilibrium position of an object undergoing simple harmonic motion?

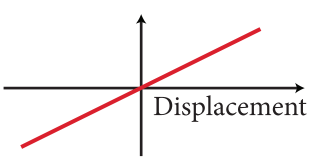
A. Acceleration



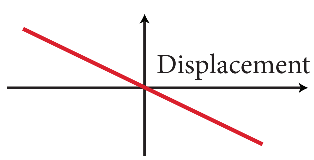
B. Acceleration



C. Acceleration



D. Acceleration



2. An object is undergoing simple harmonic motion about a fixed point P, and the magnitude of its displacement from P is *x*. Which one of the following statements is correct?

	Magnitude of the resultant force	Direction of the resultant force
A.	Proportional to *x*	Away from point P
B.	Inversely proportional to *x*	Away from point P
C.	Proportional to *x*	Towards point P
D.	Inversely proportional to *x*	Towards point P

3. The angular speed of the "minute" hand of an analogue watch is

A. π / 1800 rad s^{-1}
B. π / 60 rad s^{-1}
C. π / 30 rad s^{-1}
D. 120 rad s^{-1}

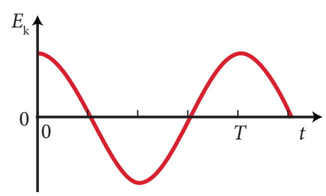
4. Which of the following is true of the magnitude of the acceleration of the object that is undergoing simple harmonic motion?

A. It is greatest at the midpoint of the motion.
B. It is greatest at the end points of the motion.
C. It is uniform throughout the motion.
D. It is greatest at the midpoints and the endpoints.

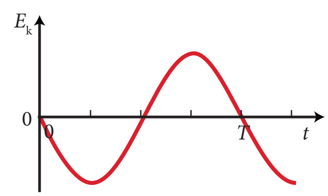
5. A particle oscillates with simple harmonic motion with a period *T*.

At time $t = 0$, the particle has its maximum displacement. Which graph The variation with time t of the kinetic energy of the particle is shown in which diagram below?

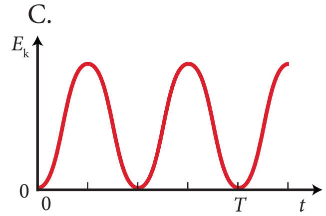
A.



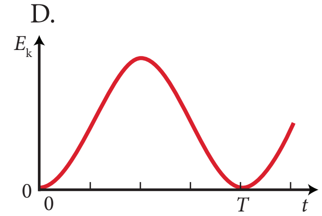
B.



C.



D.



6. Figure 907 below shows how the displacement of the magnet of mass 0.3 kg hanging from a spring varies with time for two oscillations.

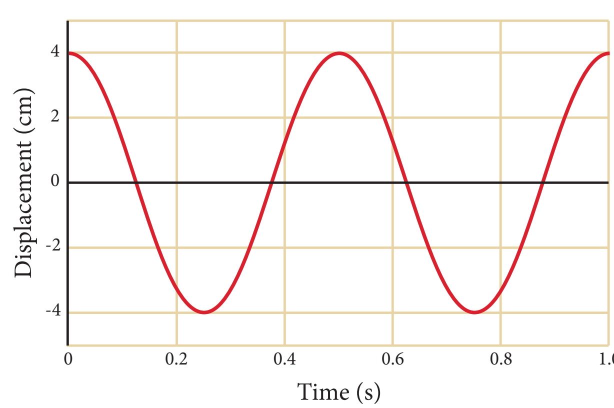


Figure 907

Using information from this graph calculate the

(a) value of the spring constant.
(b) maximum kinetic energy of the magnet.

7. A system is oscillating with SHM as described by the graph in the Figure 908 below.

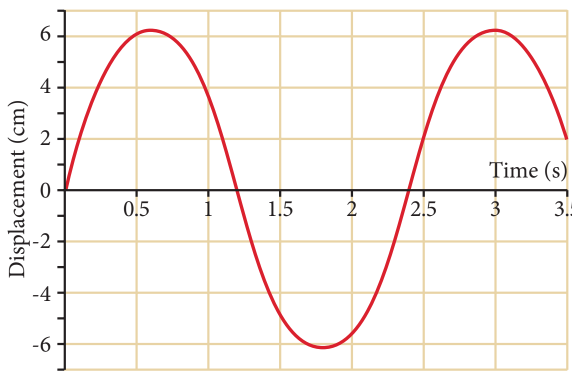


Figure 908

(a) Use the graph to determine the

(i) period of oscillation
(ii) amplitude of oscillation
(iii) maximum speed
(iv) speed at t = 1.3 s
(v) maximum acceleration

(b) State two values of t for when the magnitude of the velocity is a maximum and two values of t for when the magnitude of the acceleration is a maximum.

9. Figure 909 shows the relationship that exists between the acceleration and the displacement from the equilibrium position for a harmonic oscillator.

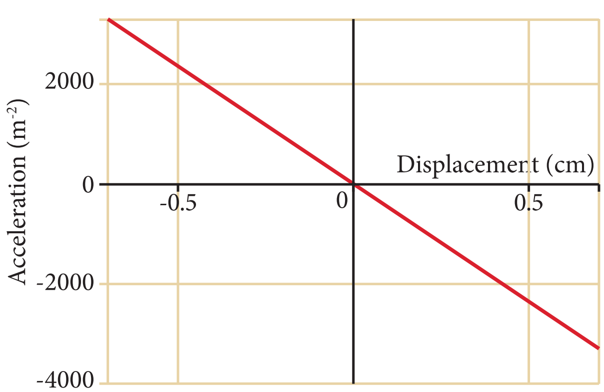


Figure 909

(a) State and explain two reasons why the graph opposite indicates that the object is executing simple harmonic motion.

(b) Determine the frequency of oscillation.

9.2 Single-slit diffraction

NATURE OF SCIENCE:

Development of theories: When light passes through an aperture the summation of all parts of the wave leads to an intensity pattern that is far removed from the geometrical shadow that simple theory predicts. (1.9)

Essential idea: Single-slit diffraction occurs when a wave is incident upon a slit of approximately the same size as the wavelength.

Understandings:
- The nature of single-slit diffraction

The nature of single-slit diffraction

Single slit diffraction intensity distribution

When plane wavefronts pass through a small aperture they spread out as discussed in chapter 4. This is an example of the phenomenon called diffraction. Light waves are no exception to this and ways for observing the diffraction of light have also been discussed previously.

However, when we look at the diffraction pattern produced by light we observe a fringe pattern, that is, on the screen there is a bright central maximum with "secondary" maxima either side of it. There are also regions where there is no illumination and these minima separate the maxima. If we were to actually plot how the intensity of illumination varies along the screen then we would obtain a graph similar to that as in Figure 910.

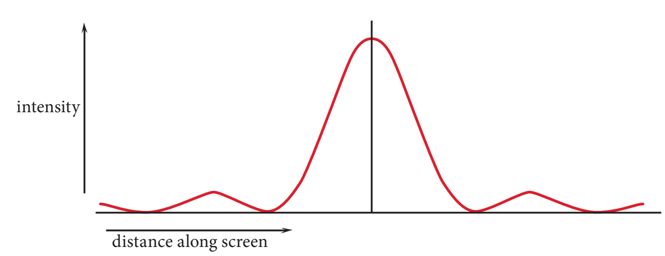


Figure 910 Intensity distribution for single-slit diffraction

We would get the same intensity distribution if we were to plot the intensity against the angle of diffraction θ. (See next section).

This intensity pattern arises from the fact that each point on the slit acts, in accordance with Huygen's principle, as a source of secondary wavefronts. It is the interference between these secondary wavefronts that produces the typical diffraction pattern.

Obtaining an expression for the intensity distribution is mathematically a little tricky and it is not something that we are going to attempt here. However, we can deduce a useful relationship from a simple argument. In this argument we deal with a phenomenon called **Fraunhofer diffraction**, that is the light source and the screen are an infinite distance away form the slit. This can be achieved with the set up shown in Figure 911.

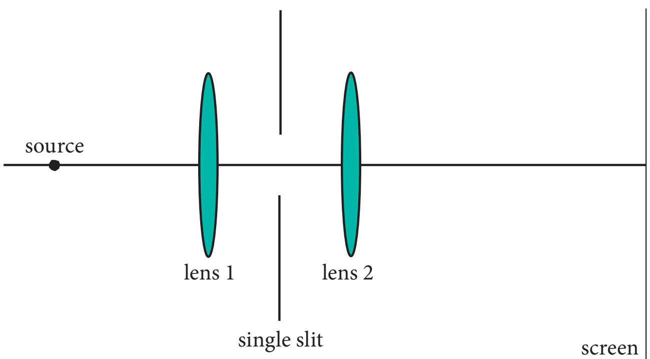


Figure 911 Apparatus for viewing Fraunhofer diffraction

The source is placed at the principal focus of lens 1 and the screen is placed at the principal focus of lens 2. Lens 1 ensures that parallel wavefronts fall on the single slit and lens 2 ensures that the parallel rays are brought to a focus on the screen. The same effect can be achieved using a laser and placing the screen some distance from the slit. If the light and screen are not an infinite distance from the slit then we are dealing with a phenomenon called **Fresnel diffraction** and such diffraction is very difficult to analyse mathematically. To obtain a good idea of how the single slit pattern comes about we consider the diagram Figure 912.

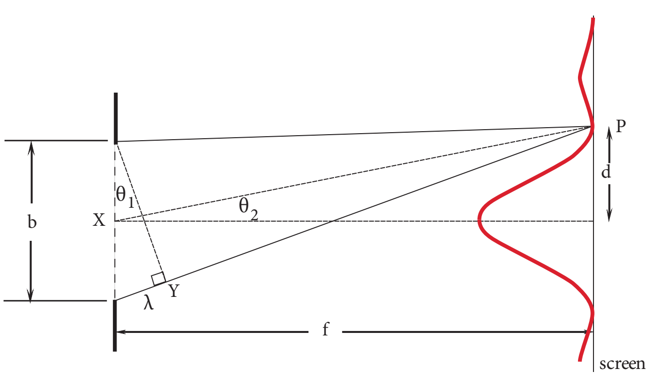


Figure 912 Single slit diffraction

In particular we consider the light from one edge of the slit to the point P where this point is just one wavelength further from the lower edge of the slit than it is from the upper edge. The secondary wavefront from the upper edge will travel a distance $\lambda/2$ further than a secondary wavefront from a point at the centre of the slit. Hence when these wavefronts arrive at P they will be out of phase and will interfere destructively. The wavefronts from the next point below the upper edge will similarly interfere destructively with the wavefront from the next point below the centre of the slit. In this way we can pair the sources across the whole width of the slit. If the screen is a long way from the slit then the angles θ_1 and θ_2 become nearly equal. (*If the screen is at infinity then they are equal and the two lines PX and XY are at right angles to each other*). From Figure 1116 we see therefore that there will be a minimum at P if

$\lambda = b\sin\theta_1$ where b is the width of the slit.

However, both angles are very small, equal to θ say, where θ is the angle of diffraction.

So it can be written $\theta = \frac{\lambda}{b}$

This actually gives us the half-angular width of the central maximum. We can calculate the actual width of the maximum along the screen if we know the focal length of the lens focussing the light onto the screen. If this is f then we have that

$\theta = \frac{d}{f}$ Such that $d = \frac{f\lambda}{b}$

To obtain the position of the next maximum in the pattern we note that the path difference is $\frac{3}{2}\lambda$. We therefore divide the slit into three equal parts, two of which will produce wavefronts that will cancel and the other producing wavefronts that reinforce. The intensity of the second maximum is therefore much less than the intensity of the central maximumof the order of 0.05 or 1/20 of the original intensity. (Much less than one third in fact since the wavefronts that reinforce will have differing phases).

We can also see now how diffraction effects become more and more noticeable the narrower the slit becomes. If light of wavelength 430 nm was to pass through a slit of width say 10 cm and fall on a screen 3.0 m away, then the half angular width of the central maximum would be 0.13 μm.

$$d = \frac{f\lambda}{b} = \frac{3 \times 430 \times 10^{-9}}{0.1} = 0.13\ \mu\text{m}$$

There will be lots of maxima of nearly the same intensity and the maxima will be packed very closely together. (The first minimum occurs at a distance of 0.12 μm from the centre of the central maximum and the next occurs effectively at a distance of 0.24 μm). We effectively observe the geometric pattern. *Refer to Example.*

We also see now how for diffraction effects to be noticeable the wavelength must be of the order of the slit width. The width of the pattern increases in proportion to the wavelength and decreases inversely with the width of the slit. If the slit width is much greater than the wavelength then the width of the central maxima is very small.

In summary,

for **destructive interference** and **minima**

$d \sin\theta = n\lambda$ where n = 1, 2, 3…

for **constructive interference** and **maxima**

$d \sin\theta = (n + ½)\lambda$ where n = 1, 2, 3…

Note that waves from point sources along the slit arrive in phase at the centre of the fringe pattern and constructively interfere to produce the central maximum, called the **zero order maximum** (n = 0).

Example

Light from a laser is used to form a single slit diffraction pattern. The width of the slit is 0.10 mm and the screen is placed 3.0 m from the slit. The width of the central maximum is measured as 2.6 cm. Calculate the wavelength of the laser light?

Solution

Since the screen is a long way from the slit we can use the small angle approximation such that the f is equal to 3.0 m.

The half-width of the central maximum is 1.3 cm so we have

$$\lambda = \frac{(1.3 \times 10^{-2}) \times (1.0 \times 10^{-4})}{3.0}$$

To give λ = 430 nm.

This example demonstrates why the image of a point source formed by a thin converging lens will always have a finite width.

Exercise 9.2

1. A parallel beam of light of wavelength 500 nm is incident on a slit of width 0.25 mm. The light is brought to focus on a screen placed 1.50 m from the slit. Calculate the angular width and the linear width of the central diffraction maximum.

2. Light from a laser is used to form a single slit diffraction pattern on a screen. The width of the slit is 0.10 mm and the screen is 3.0 m from the slit. The width of the central diffraction maximum is 2.6 cm. Calculate the wavelength of the laser light.

3. Determine the angle that you would expect to find constructive interference (for n = 0, n = 1 and n = 2) from a single slit of width 2.0 μm and light of wavelength 600 nm.

9.3 Interference

NATURE OF SCIENCE:

Curiosity: Observed patterns of iridescence in animals, such as the shimmer of peacock feathers, led scientists to develop the theory of thin film interference. (1.5)

Serendipity: The first laboratory production of thin films was accidental. (1.5)

Essential idea: Interference patterns from multiple slits and thin films produce accurately repeatable patterns.

Understandings:
- Young's double-slit experiment
- Modulation of two-slit interference pattern by one-slit diffraction effect
- Multiple slit and diffraction grating interference patterns
- Thin film interference

AHL

Young's double slit experiment

Please refer to interference in chapter 4. However, we will reiterate the condition for the interference of waves from two sources to be observed. The two sources must be **coherent**, that is they must have the same phase or the phase difference between them must remain constant.

Also, to reinforce topics concerning the principle of superposition, and path difference and phase difference, included in Figure 913 is another example of two source interference.

We can obtain evidence for the wave nature of sound by showing that sound produces an interference pattern. Figure 913 shows the set up for demonstrating this.

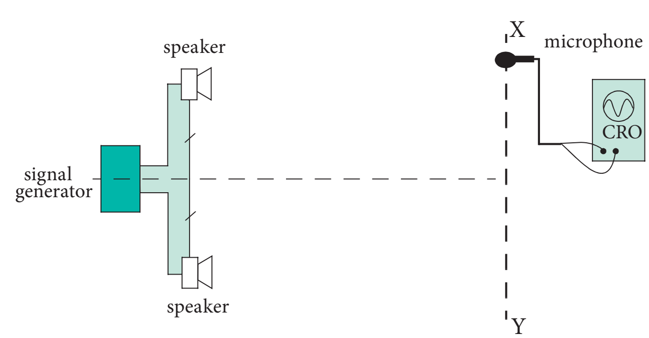


Figure 913 Interference using sound waves

The two speakers are connected to the same output of the signal generator and placed about 50 cm apart. The signal generator frequency is set to about 600 Hz and the microphone is moved along the line XY, that is about one metre from the speakers. As the microphone is moved along XY the trace on the cathode ray oscilloscope is seen to go through a series of maxima and minima corresponding to points of constructive and destructive interference of the sound waves.

An interesting investigation is to find how the separation of the points of maximum interference depends on the frequency of the source and also the distance apart of the speakers.

In the demonstration of the interference between two sound sources described above, if we were to move one of the speakers from side to side or backwards and forwards the sound emitted from the two speakers would no longer be in phase. No permanent points of constructive or destructive interference will now be located since the phase difference between the waves from the two sources is no longer constant, i.e. the sources are no longer coherent.

Light from an incandescent source is emitted with a completely random phase. Although the light from two separate sources will interfere, because of the randomly changing phase no permanent points of constructive or destructive interference will be observed. This is why a single slit is needed in the Young's double slit experiment. By acting as a point source, it essentially becomes a coherent light source. The light emitted from a laser is also very nearly coherent and this is why it is so easy to demonstrate optical interference and diffraction with a laser.

Young's double slit experiment is one of the great classic experiments of physics and did much to reinforce the wave theory of light. Thomas Young carried out the experiment in about 1830.

It is essentially the demonstration with the ripple tank and the sound experiment previously described, but using light. The essential features of the experiment are shown in Figure 914.

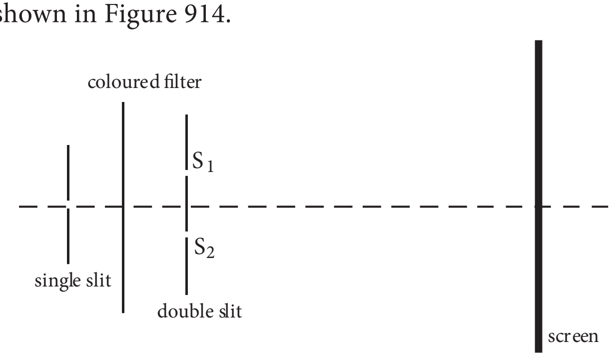


Figure 914 Young's double slit experiment

Young allowed sunlight to fall onto a narrow single slit. A few centimetres from the single slit he placed a double slit. The slits are very narrow and separated by a distance equal to about fourteen slit widths. A screen is placed about a metre from the double slits. Young observed a pattern of multicoloured "fringes" in the screen. When he placed a coloured filter between the single slit and double slit he obtained a pattern that consisted of bright coloured fringes separated by darkness.

The single slit essentially ensures that the light falling on the double slit is coherent. The two slits then act as the two speakers in the sound experiment or the two dippers in the ripple tank. The light waves from each slit interfere and produce the interference pattern on the screen. Without the filter a pattern is formed for each wavelength present in the sunlight. Hence the multicoloured fringe pattern.

You can demonstrate optical interference for yourself. Smoking a small piece of glass and then drawing two parallel lines on it can make a double slit. If you then look through the double slit at a single tungsten filament lamp you will see the fringe pattern. By placing filters between the lamps and the slits you will see the monochromatic fringe pattern.

You can also see the effects of optical interference by looking at net curtains. Each 'hole' in the net acts as a point source and the light from all these separate sources interferes and produces quite a complicated interference pattern.

A laser can also be used to demonstrate optical interference. Since the light from the laser is coherent it is very easy to demonstrate interference. Just point the laser at a screen and place a double slit in the path of the laser beam.

Let us now look at the Young's double slit experiment in more detail. The geometry of the situation is shown in Figure 915.

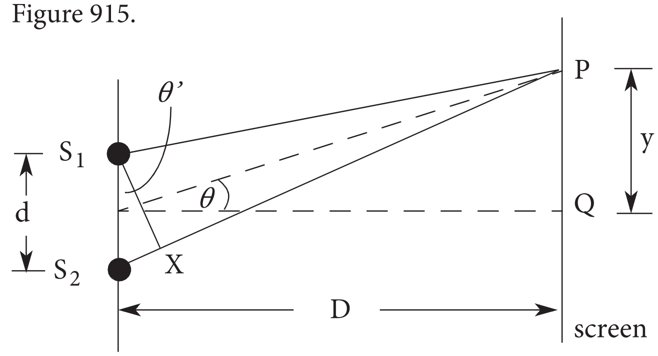


Figure 915 The geometry of Young's double slit experiment

S_1 and S_2 are the two narrow slits that we shall regard as two coherent, monochromatic point sources. The distance from the sources to the screen is D and the distance between the slits is d.

The waves from the two sources will be in phase at Q and there will be a bright fringe here. We wish to find the condition for there to be a bright fringe at P distance y from Q.

We note that that D (≈ 1 metre) is very much greater than either y or d. (≈ few millimetres) . This means that both θ and θ' are very small angles and for intents and purposes equal.

From the diagram we have that

$$\theta = \frac{y}{D}$$

And

$$\theta' = \frac{S_2X}{d}$$

(Remember, the angles are very small)

Since $\theta \approx \theta'$ then

$$\frac{y}{D} = \frac{S_2X}{d}$$

But S_2X is the path difference between the waves as they travel to P. We have therefore that

$$\text{path difference} = \frac{yd}{D}$$

There will therefore be a bright fringe at P if

$$\frac{yd}{D} = n\lambda$$

Suppose that there is a bright fringe at $y = y_1$ corresponding to $n = n_1$ then

$$\frac{y_1 d}{D} = n_1\lambda$$

If the next bright fringe occurs at $y = y_2$ this will correspond to $n = n_1 + 1$. Hence

$$\frac{y_2 d}{D} = (n_1 + 1)\lambda$$

This means that the spacing between the fringes $y_2 - y_1$ is given by

$$y_2 - y_1 = \frac{D}{d}\lambda$$

Young actually use this expression to measure the wavelength of the light he used and it is a method still used today.

We see for instance that if in a given set up we move the slits closer together then the spacing between the fringes will get greater. Effectively our interference pattern spreads out, that is there will be fewer fringes in a given distance. We can also increase the fringe spacing by increasing the distance between the slits and the screen. You will also note that for a given set up using light of different wavelengths, then "red" fringes will space further apart than "blue" fringes.

In this analysis we have assumed that the slits act as point sources and as such the fringes will be uniformly spaced and of equal intensity. A more thorough analysis should take into account the finite width of the slits.

Returning to Figure 915 we see that we can write the path difference S_2X as

$$S_2X = d\tan\theta'$$

But since θ' is a small angle the sine and tangent will be nearly equal so that

$$S_2X = d\sin\theta'$$

And since $\theta' \approx \theta$ then

$$S_2X = d\sin\theta$$

The condition therefore for a bright fringe to be found at a point of the screen can therefore be written as

$$d\sin\theta = n\lambda$$

In summary,

for **constructive interference** and **minima**

$d \sin\theta = n\lambda$ where n = 1, 2, 3…

for destructive interference and maxima

$d \sin\theta = (n + ½)\lambda$ where n = 1, 2, 3…

Figure 916 shows the intensity distribution of the fringes on the screen when the separation of the slits is large compared to their width. The fringes are of equal intensity and of equal separation.

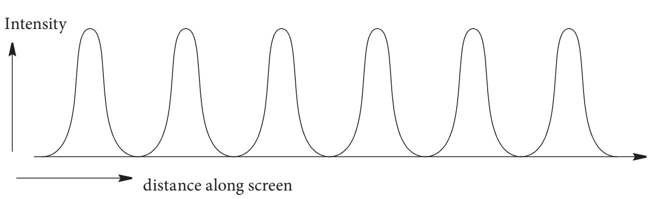


Figure 916 The intensity distribution of the fringes

It is worth noting that if the slits are close together, the intensity of the fringes is modulated by the intensity distribution of the diffraction pattern of one of the slits.

AHL

Example

Light of wavelength 500 nm is incident on two small parallel slits separated by 1.0 mm. Determine the angle where the first maximum is formed?

If after passing through the slits the light is brought to a focus on a screen 1.5 m from the slits calculate the observed fringe spacing on the screen.

Solution

Using the small angle approximation we have

$$\theta = \frac{\lambda}{d} = \frac{5\times10^{-7}}{10^{-3}}$$

$= 5 \times 10^{-4}$ *rad*

The fringe spacing is given by

$$y = \frac{D\lambda}{d} = \frac{1.5\times5\times10^{-7}}{10^{-3}}$$

$= 0.75$ *mm*

Modulation of two-slit interference pattern by one-slit diffraction effect

In practice the intensity pattern of the maxima as shown in Figure 917 is not constant, but fluctuates while decreasing symmetrically on either side of the central maximum as shown in Figure 917. The intensity pattern is a combination of both the single-slit **diffraction envelope** and the double slit pattern. That is the amplitude of the two-slit interference pattern is **modulated** (i.e. adjusted to) by a single slit diffraction envelope.

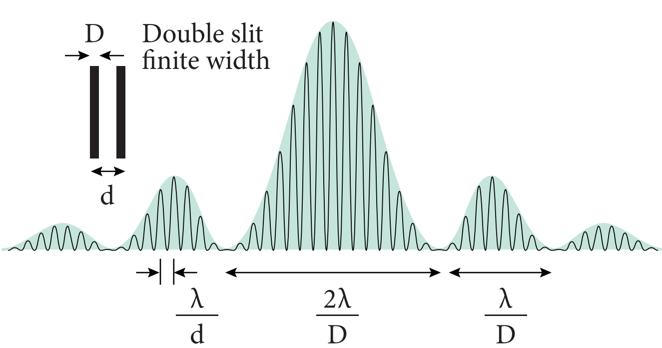


Figure 917 Double slit pattern

For a given slit separation d, wavelength of light, and fixed slit to screen distance, the variation in intensity depends on the width of the slit D. Although increasing the width of the slits increases the intensity of light in the fringes, it also makes them less sharp and they become blurred. As the slit width is widened, the fringes gradually disappear because numerous point sources along the widened slit give rise to their own dark and bright fringes that then overlap with each other.

Exercise 9.3

1. In Figure 913, the distance between the speakers is 0.50 m and the distance between the line of the speakers and the screen is 2.0 m. As the microphone is moved along the line XY, the distance between successive points of maximum sound intensity is 0.30 m. The frequency of the sound waves is 4.4 × 103 Hz. Calculate a value for the speed of sound.

2. Laser light of wavelength 610 nm falls on two slits 1.0 x 10-5 apart. Determine the separation of the first order maximum formed on a screen 2.0 m away.

3. Two parallel slits 0.12 mm apart are illuminated with red light with a wavelength of 600 nm. An interference pattern falls on a screen 1.5 m away.

   Calculate

   (a) the distance from the central maximum to the first bright fringe

   (b) the distance to the second dark line.

Multiple slit and diffraction grating interference patterns

If we examine the interference pattern produced when monochromatic light passes through a different number of slits we notice that as the number of slits increases the number of observed fringes decreases, the spacing between them increases and the individual fringes become much sharper. We can get some idea of how this comes about by looking at the way light behaves when a parallel beam passes through a large number of slits. The diagram for this is shown in Figure 918.

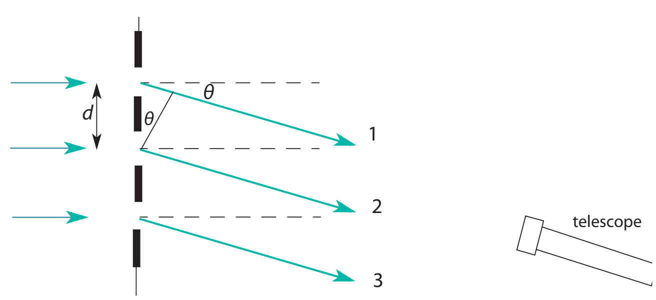


Figure 918 A parallel beam passing through several slits

The slits are very small so that they can be considered to act as point sources. They are also very close together such that d is small (10^{-6} m). Each slit becomes a source of circular wave fronts and the waves from each slit will interfere. Let us consider the light that leaves the slit at an angle θ as shown. The path difference between wave 1 and wave 2 is $d\sin\theta$ and if this is equal to an integral number of wavelengths then the two waves will interfere constructively in this direction. Similarly wave 2 will interfere constructively with wave 3 at this angle, and wave 3 with 4 etc., across the whole grating. Hence if we look at the light through a telescope, that is bring it to a focus, then when the telescope makes an angle θ to the grating a bright fringe will be observed. The condition for observing a bright fringe is therefore

$d\sin\theta = m\lambda$

Suppose we use light of wavelength 500 nm and suppose that $d = 1.6 \times 10^{-6}$ m.

Obviously we will see a bright fringe in the straight on position $\theta = 0$ (the *zero order*).

The next position will be when $m = 1$ (the *first order*) and substitution in the above equation gives $\theta = 18°$.

The next position will be when $m = 2$ (the *second order*) and this give $\theta = 38°$.

For $m = 3$, $\sin\theta$ is greater than 1 so with this set up we only obtain 5 fringes, one zero order and two either side of the zero order.

The calculation shows that the separation of the orders is relatively large. At any angles other than 18° or 38° the light leaving the slits interferes destructively. We can see that the fringes will be sharp since if we move just a small angle away from 18° the light from the slits will interfere destructively.

An array of narrow slits is usually made by cutting narrow transparent lines very close together into the emulsion on a photographic plate (typically 200 lines per millimetre). Such an arrangement is called a **diffraction grating**.

The diffraction grating is of great use in examining the spectral characteristics of light sources.

All elements have their own characteristic spectrum. An element can be made to emit light either by heating it until it is incandescent or by causing an electric discharge through it when it is in a gaseous state. Your school probably has some discharge tubes and diffraction gratings. If it has, then look at the glowing discharge tube through a diffraction grating. If for example, the element that you are looking has three distinct wavelengths then each wavelength will be diffracted by a different amount in accordance with the equation $d\sin\theta = n\lambda$.

Also if laser light is shone through a grating on to a screen, you will see just how sharp and spaced out are the maxima. By measuring the line spacing and the distance of the screen from the laser, the wavelength of the laser can be measured.

If your school has a set of multiple slits say from a single slit to eight slits, then it is also a worthwhile exercise to examine how the diffraction pattern changes when laser light is shone through increasing numbers of slits.

Consider a grating of N parallel equidistant slits of width a separated by an opaque region of width D. We know that light diffracted from one slit will be superimposed and interfere with light diffracted from all the other slits. The **intensity profile** of the resultant wave will have a shape determined by

- the intensity of light incident on the slits
- the wavelength of this light
- the slit width
- the slit separation

Typical intensity profiles formed when a plane wave monochromatic light at normal incidence on one, two and six slits are shown in Figure 919. In this example the slit width is one quarter of the grating spacing.

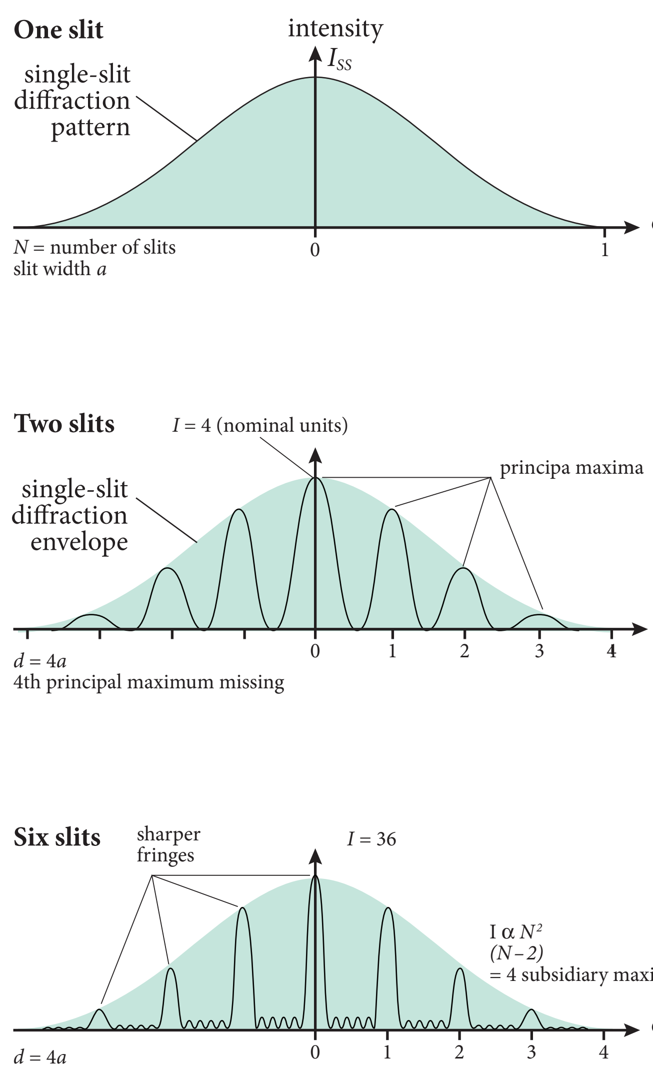


Figure 919 Intensity profiles

1. For $N \geq 2$, the fringe pattern contains principal maxima that are modulated by the diffraction envelope.

2. For $N > 2$, the fringe pattern contains $(N - 2)$ subsidiary maxima.

3. Where the grating spacing d is four time the slit width a (4 x $a = d$), then the fourth principal maximum is missing from the fringe pattern. Where the grating spacing d is three time the slit width a (3 x $a = d$), then the third principal maximum is missing from the fringe pattern, and so on.

4. Increasing N increases the absolute intensities of the diffraction pattern ($I = N^2$)

If white light is shone through a grating then the central image will be white but for the other orders each will be spread out into a continuous spectrum composed of an infinite number of adjacent images of the slit formed by the wavelength of the different wavelengths present in the white light. At any given point in the continuous spectrum the light will be very nearly monochromatic because of the narrowness of the images of the slit formed by the grating. This is in contrast to the double slit where if white light is used, the images are broad and the spectral colours are not separated.

In summary, as is the case of interference at a double slit, there is a path difference between the rays from adjacent slits of d sinθ. Where there is a whole number of wavelengths, constructive interference occurs

For **constructive interference** and **minima**

d sinθ = $\boldsymbol{m\lambda}$ where n = 1, 2, 3…

When there is a path difference between adjacent rays of half a wavelength, destructive interference occurs.

For and **maxima**

d sinθ = ($\boldsymbol{m + ½}$ **)λ** where n = 1, 2, 3…

Exercise 9.4

1. Light from a laser is shone through a diffraction grating on to a screen. The screen is a distance of 2.0 m from the laser. The distance between the central diffraction maximum and the first principal maximum formed on the screen is 0.94 m. The diffraction grating has 680 lines per mm. Estimate the wavelength of the light emitted by the laser.

2. Monochromatic light from a laser is normally incident on a six-slit grating. The slit spacing is three times the slit width. An interference pattern is formed on a screen 2 m on the other side of the grating.

   (a) Determine and explain the order of the principal maxima missing from the fringe pattern.

   (b) How many subsidiary maxima are there between the principal maxima?

   (c) How many principal maxima are there

      (i) Between the two first order minima in the diffraction envelope?
      (ii) Between the first and second order minima in the diffraction envelope?

   (d) Draw the pattern formed by the bright fringes in (c).

   (e) Sketch the intensity profile of the fringe pattern.

Thin film interference

You might well be familiar with the coloured pattern of fringes that can be seen when light is reflected off the surface of water upon which a thin oil film has been spilt or from light reflected from bubbles. We can see how these patterns arise by looking at Figure 920.

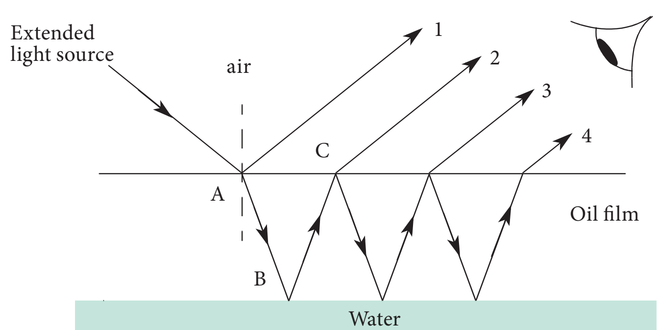


Figure 920 Reflection from an oil film

Consider light from an extended source incident on a thin film. We also consider a wave from one point of the source whose direction is represented by the ray shown. Some of this light will be reflected at A and some transmitted through the film where some will again be reflected at B (some will also be transmitted into the air). Some of the light reflected at B will then be transmitted at C and some reflected and so on. If we consider just rays 1 and 2 then these will be not be in phase when the are brought to a focus by the eye since they have travelled different distances. If the path difference between them is an integral number of half wavelengths then we might expect the two waves to be out of phase. However, ray 1 undergoes a phase change of π on reflection but ray 2 does not since it is reflected at a boundary between a more dense and less dense medium. (See chapter 4) Hence if the path difference is an integral number of half-wavelengths rays 1 and 2 will reinforce i.e. ray 1 and 2 are in phase. However, rays 3, 5, 7 etc. will be out of phase with rays 2, 4, 6 etc. but since ray 2 is more intense than ray 3 and ray 4 more intense than ray 5, these pairs will not cancel out so there will be a maximum of intensity.

If the path difference is such that wave 1 and 2 are out of phase, since wave 1 is more intense than wave 2, they will not completely annul. However, it can be shown that the intensities of waves 2, 3, 4, 5… add to equal the intensity of wave 1. Since waves 3, 4. 5… are in phase with wave 2 there will be complete cancellation.

The path difference will be determined by the angle at which ray 1 is incident and also on the thickness (and the actual refractive index as well) of the film. Since the source is an extended source, the light will reach the eye from many different angles and so a system of bright and dark fringes will be seen. You will only see fringes if the film is very thin (except if viewed at normal incidence) since increasing the thickness of the film will cause the reflected rays to get so far apart that they will not be collected by the pupil of the eye.

From the argument above, to find the conditions for constructive and destructive **interference** we need only find the path difference between ray 1 and ray 2. Figure 921 shows the geometry of the situation.

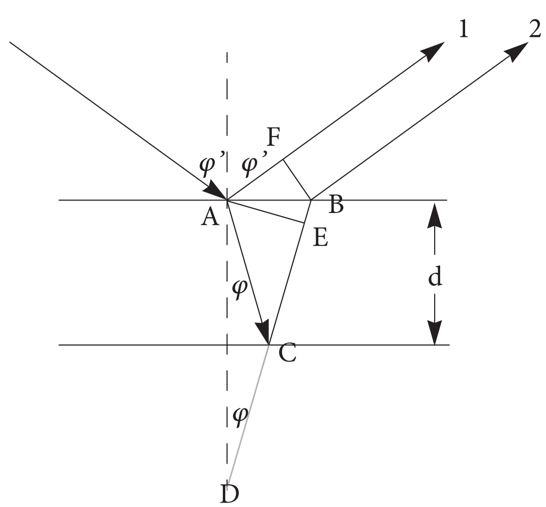


Figure 921 The geometry of interference

The film is of thickness d and refractive index n and the light has wavelength λ. If the line BF is perpendicular to ray 1 then the **optical path difference (opd)** between ray 1 and ray 2 when brought to a focus is

$$opd = n(\mathrm{AC} + \mathrm{CB}) - \mathrm{AF}$$

We have to multiple by the refractive index for the path travelled by the light in the film since the light travels more slowly in the film. If the light travels say a distance x in a material of refractive index n then in the time that it takes to travel this distance, the light would travel a distance nx in air.

If the line CE is at right angles to ray 2 then we see that

$$\mathrm{AF} = n\mathrm{BE}$$

From the diagram AC = CD so we can write

$$opd = n(\mathrm{CD} + \mathrm{CB}) - n\mathrm{BE} = n\mathrm{DE}$$

Also from the diagram we see that, where ϕ is the angle of refraction

$$\mathrm{DE} = 2\cos\phi$$

From which $opd = 2nd\cos\phi$

Bearing in mind the change in phase of ray 1 on reflection we have therefore that the condition for constructive interference is

$$2nd\cos\phi = \left(m + \frac{1}{2}\right)\lambda,\ m = 1, 2, \ldots$$

And for destructive interference $2nd\cos\phi = m\lambda$

Each fringe corresponds to a particular *opd* for a particular value of the integer m and for any fringe the value of the angle ϕ is fixed. This means that it will be in the form of an arc of a circle with the centre of the circle at the point where the perpendicular drawn from the eye meets the surface of the film. Such fringes are called fringes of equal inclination. Since the eye has a small aperture these fringes, unless viewed at near to normal incidence ($\phi = 0$), will only be observed if the film is very thin. This is because as the thickness of the film increases the reflected rays will get further and further apart and so very few will enter the eye.

If white light is shone onto the film then we can see why we get multi-coloured fringes since a series of maxima and minima will be formed for each wavelength present in the white light. However, when viewed at normal incidence, it is possible that only light of one colour will under go constructive interference and the film will take on this colour.

AHL

Thickness of oil films

The exercise to follow will help explain this use.

Non-reflecting films

A very important but simple application of thin film interference is in the production of non-reflecting surfaces.

A thin film of thickness d and refractive index n_1 is coated onto glass of refractive index n where $n_1 < n$. Light of wavelength λ that is reflected at normal incidence will undergo destructive interference if $2n_1 d = \frac{\lambda}{2}$, that is

$$d = \frac{\lambda}{4n_1}$$

(remember that there will now no phase change at the glass interface i.e. we have a rare to dense reflection)

The use of such films can greatly reduce the loss of light by reflection at the various surfaces of a system of lenses or prisms. Optical parts of high quality systems are usually all coated with non-reflecting films in order to reduce stray reflections. The films are usually made by evaporating calcium or magnesium fluoride onto the surfaces in vacuum, or by chemical treatment with acids that leave a thin layer of silica on the surface of the glass. The coated surfaces have a purplish hue by reflected light. This is because the condition for destructive interference from a particular film thickness can only be obtained for one wavelength. The wavelength chosen is one that has a value corresponding to light near the middle of the visible spectrum. This means that reflection of red and violet light is greater combining to give the purple colour. Because of the factor $\cos\phi$, at angles other than normal incidence, the path difference will change but not significantly until say about 30° (e.g. cos 25° = 0.90).

It should be borne in mind that no light is actually lost by a non-reflecting film; the decrease of reflected intensity is compensated by increase of transmitted intensity.

Non-reflecting films can be painted onto aircraft to suppress reflection of radar. The thickness of the film is determined by $nd = \frac{\lambda}{4}$ where λ is the wavelength of the radar waves and n the refractive index of the film at this wavelength.

Example

A plane-parallel glass plate of thickness 2.0 mm is illuminated with light from an extended source. The refractive index of the glass is 1.5 and the wavelength of the light is 600 nm. Calculate how many fringes are formed.

Solution

We assume that the fringes are formed by light incident at all angles from normal to grazing incidence.

At normal incidence we have $2nd = m\lambda$

From which,

$$m = \frac{2 \times 1.5 \times 2 \times 10^{-3}}{6 \times 10^{-7}} = 10,000$$

At grazing incidence the angle of refraction ϕ *is in fact the critical angle.*

Therefore, $\phi = \arcsin\left(\frac{1}{1.5}\right) = 42°$

i.e. $\cos\phi = 0.75$

At grazing incidence $2nd\cos\phi = m'\lambda$

From which (and using $\cos(\sin^{-1}(1/1.5)) = 0.75$*),*

$$m = \frac{2 \times 1.5 \times 2 \times 10^{-3} \times 0.75}{6 \times 10^{-7}} = 7500$$

The total number of fringes seen is equal to $m - m' = 2500$.

Exercise 9.5

When viewed from above, the colour of an oil film on water appears red in colour. Use the data below to estimate the minimum thickness of the oil film.

average wavelength of red light = 630 nm

refractive index of oil for red light = 1.5

refractive index of water for red light = 1.3

9.4 Resolution

NATURE OF SCIENCE:

Improved technology: The Rayleigh criterion is the limit of resolution. Continuing advancement in technology such as large diameter dishes or lenses or the use of smaller wavelength lasers pushes the limits of what we can resolve. (1.8)

Essential idea: Resolution places an absolute limit on the extent to which an optical or other system can separate images of objects.

Understandings:

- The size of a diffracting aperture
- The resolution of simple monochromatic two-source systems

The size of a diffracting aperture

Our discussion concerning diffraction so far has been for rectangular slits. When light from a point source enters a small circular aperture, it does not produce a bright dot as an image, but a circular disc known as Airy's disc surrounded by fainter concentric circular rings as shown in Figure 922.

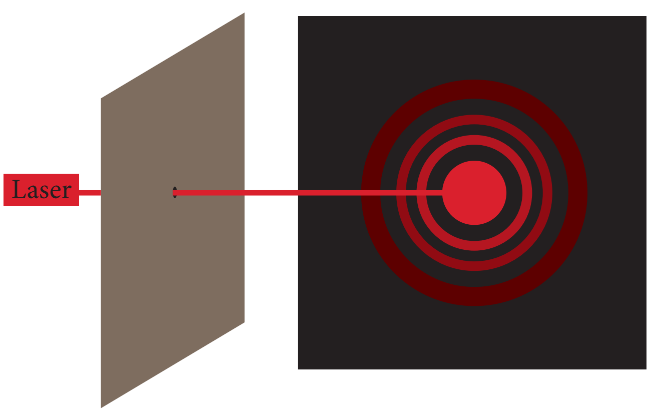


Figure 922 Diffraction at a circular aperture

Diffraction at an aperture is of great importance because the eye and many optical instruments have circular apertures.

What is the half-angular width of the central maximum of the diffraction formed by a circular aperture? This is not easy to calculate since it involves some advanced mathematics. The problem was first solved by the English Astronomer Royal, George Airy, in 1835 who showed that for circular apertures

$\theta = \frac{1.22\lambda}{b}$ where b is the diameter of the aperture.

Example 2

In the following diagram, parallel light from a distant point source (such as a star) is brought to focus on the screen S by a converging lens (*the lens is shown as a vertical arrow*).

The focal length (distance from lens to screen) is 25 cm and the diameter of the lens is 3.0 cm. The wavelength of the light from the star is 560 nm. Calculate the diameter of the diameter of the image on the screen.

Solution

The lens actually acts as a circular aperture of diameter 3.0 cm. The half angular width of central maximum of the diffraction pattern that it forms on the screen is

$$\theta = \frac{1.22\lambda}{b} = \frac{1.22 \times 5.6 \times 10^{-7}}{3.0 \times 10^{-2}} = 2.3 \times 10^{-5}\ rad$$

The diameter of the central maxima is therefore

$25 \times 10^{-2} \times 2.3 \times 10^{-5}$

$= 5.7 \times 10^{-6}$ *m.*

Although this is small, it is still finite and is effectively the image of the star as the intensity of the secondary maxima are small compared to that of the central maximum.

The resolution of simple monochromatic two-source systems

The astronomers tell us that many of the stars that we observe with the naked eye are in fact binary stars. That is, what we see as a single star actually consists of two stars in orbit about a common centre. Furthermore, the astronomers tell us that if we use a "good" telescope then we will actually see the two stars. That is, we will resolve the single point source into its two component parts. So what is it that determines whether or not we see the two stars as a single point source i.e. what determines whether or not two sources can be resolved? It can't just be that the telescope magnifies the stars since if they are acting as point sources magnifying them is not going to make a great deal of difference.

In each of our eyes there is an aperture, the pupil, through which the light enters. This light is then focused by the eye lens onto the retina. But we have seen that when light passes through an aperture it is diffracted and so when we look at a point source, a diffraction pattern will be formed on the retina. If we look at two point sources then two diffraction patterns will be formed on the retina and these patterns will overlap. The width of our pupil and the wavelength of the light emitted by the sources will determine the amount by which they overlap. But the degree of overlap will also depend on the angular separation of the two point sources. We can see this from Figures 923.

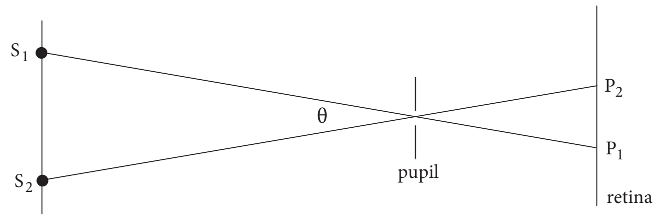
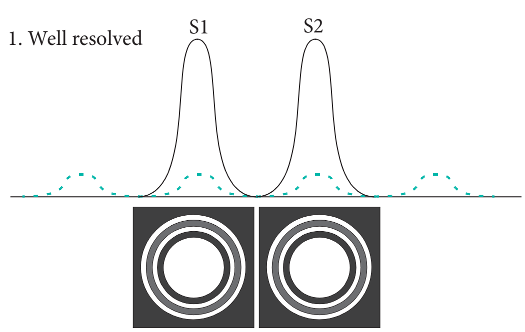


Figure 923

Light from the source S_1 enters the eye and is diffracted by the pupil such that the central maximum of the diffraction pattern is formed on the retina at P_1. Similarly, light from S_2 produces a maximum at P_2. If the two central maxima are well separated then there is a fair chance that we will see the two sources as separate sources. If they overlap then we will not be able to distinguish one source from another. From the diagram we see as the sources are moved closer to the eye, then the angle θ increases and so does the separation of the central maxima.

Figures 924, 925, 926 and 927 shows the different diffraction patterns and the intensity distribution, that might result on the retina as a result of light from two point sources



.Figure 924 Very well resolved

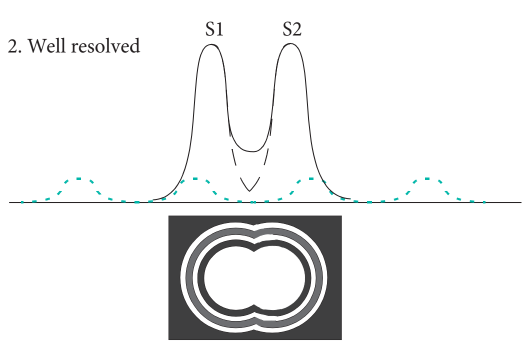


Figure 925 Well resolved

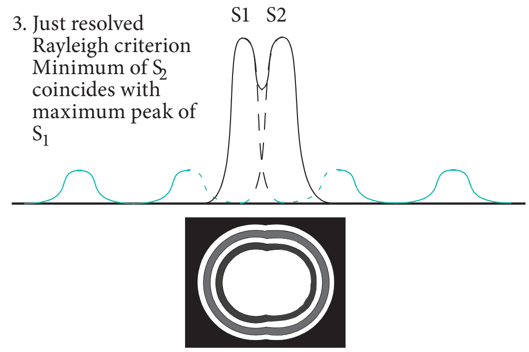


Figure 926 Just resolved

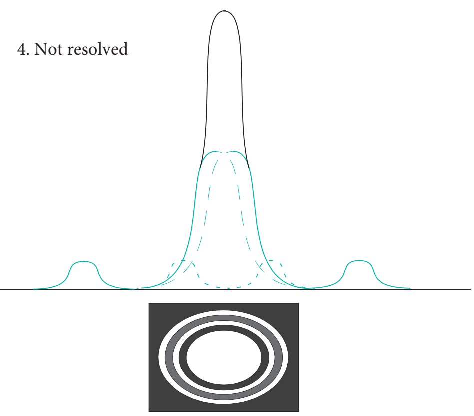


Figure 927 Not resolved

We have suggested that if the central maxima of the two diffraction patterns are reasonably separated then we should be able to resolve two point sources. In the late 19th century *Lord Rayleigh* suggested by how much they should be separated in order for the two sources to be just resolved. **If the central maximum of one diffraction pattern coincides with the first minima of the other diffraction pattern then the two sources will just be resolved.** This is known as the **Rayleigh Criterion.**

Figure 926 shows just this situation. The two sources are just resolved according to the Rayleigh criterion since the peak of the central maximum of one diffraction pattern coincides with the first minimum of the other diffraction pattern. This means that the angular separation of the peaks of the two central maxima formed by each source is just the half angular width of one central maximum i.e.

$$\theta = \frac{\lambda}{b}$$

where b is the width of the slit through which the light from the sources passes. However, we see from Figure 1117 that θ is the angle that the two sources subtend at the slit. Hence we conclude that two sources will be resolved by a slit if the angle that they subtend at the slit is greater than or equal to λ/b

So far we have been assuming that the eye is a rectangular slit whereas clearly it is a circular aperture and so we must use the formula

$$\theta = \frac{1.22\lambda}{b}$$

As mentioned above the angle θ is sometimes called the *resolving power* but should more accurately be called the **minimum angle of resolution** (θ_{min})

Clearly the smaller θ the greater will be the resolving power.

The diffraction grating is a useful for differentiating closely spaced lines in emission spectra. Like a prism spectrometer, it can disperse a spectrum into its components but it is better suited because it has higher resolution than the prism spectrometer.

If λ_1 and λ_2 are two nearly equal wavelengths that can barely be distinguished, the resolvance or resolving power of the grating is defined as:

$$R = \frac{\lambda}{(\lambda_2 - \lambda_1)} = \frac{\lambda}{\Delta\lambda}$$

Where $\lambda \approx \lambda_1 \approx \lambda_2$

Therefore, the diffraction grating with a high resolvance will be better suited in determining small differences in wavelength.

If the diffraction grating has N lines being illuminated, it can be shown that the resolving power of the mth order diffraction is given by

$$\boldsymbol{R = mN}$$

Obviously, the resolvance becomes greater with the order number m and with a greater number of illuminated slits.

Example

A benchmark for the resolving power of a grating is the sodium doublet in the sodium emission spectrum. Two yellow lines in its spectrum have wavelengths of 589.00 nm and 589.59 nm.

(a) Determine the resolvance of a grating if the given wavelengths are to be distinguished from each other.

(b) How many lines in the grating must be illuminated in order to resolve these lines in the second order spectrum?

Solution

(a) $$R = \frac{\lambda}{\Delta\lambda} = \frac{589.00 \text{ nm}}{(589.00 - 589.59) \text{ nm}}$$

$$= \frac{589 \text{ nm}}{0.59 \text{ nm}} = 1.0 \times 10^3$$

(b) $$N = \frac{R}{m} = \frac{1.0 \times 10^3}{2} = 5.0 \times 10^2 \text{ lines.}$$

It has been seen that diffraction effectively limits the resolving power of optical systems. This includes such systems as the eye, telescopes and microscopes. The resolving power of these systems is dealt with in the next section. This section looks at links between technology and resolving power when looking at the very distant and when looking at the very small.

Radio telescopes

The average diameter of the pupil of the human eye is about 2.5 mm. This means that the eye will just resolve two point sources emitting light of wavelength 500 nm if their angular separation at the eye is

$$\theta = 1.22 \times \frac{5.0 \times 10^{-7}}{2.5 \times 10^{-3}} = 2.4 \times 10^{-4} \text{ rad}$$

If the eye were to be able to detect radio waves of wavelength 0.15 m, then to have the same resolving power the pupil would have to have a diameter of about 600 m. Clearly this is nonsense, but it does illustrate a problem facing astronomers who wish to view very distant objects such as quasars and galaxies that emit radio waves. Conventional radio telescopes consist of a large dish, typically 25 m in diameter. Even with such a large diameter, the radio wavelength resolving power of the telescope is much less than the optical resolving power of the human eye. Let us look at an example.

AHL

Example

The Galaxy Cygnus A can be resolved optically as an elliptically shaped galaxy. However, it is also a strong emitter of radio waves of wavelength 0.15 m. The Galaxy is estimated to be 5.0×10^{24} m from Earth. Use of a radio telescope shows that the radio emission is from two sources separated by a distance of 3.0×10^{21} m. Estimate the diameter of the dish required to just resolve the sources.

Solution

The angle θ that the sources resolve at the telescope is given by

$$\theta = \frac{3.0 \times 10^{21}}{5.0 \times 10^{24}} = 6.0 \times 10^{-4} \textit{ rad}$$

$$\textit{and } d = \frac{1.22\lambda}{\theta} = \frac{1.22 \times 0.15}{6.0 \times 10^{-4}} = 3000 \textit{ m} = 3.0 \textit{ km.}$$

A radio telescope dish of this size would be impossible to make, let alone support. This shows that a single dish type radio telescope cannot be used to resolve the sources and yet they were resolved. To get round the problem, astronomers use two radio telescopes separated by a large distance. The telescopes view the same objects at the same time and the signals that each receive from the objects are simultaneously superimposed. The result of the superposition of the two signals is a two-slit interference pattern. The pattern has much narrower fringe spacing than that of the diffraction pattern produced by either telescope on its own, hence producing a much higher resolving power. When telescopes are used like this, they are called a **stellar interferometer.**

In Socorro in New Mexico there is a stellar interferometer that consists of 27 parabolic dishes each of diameter 25 m, arranged in a Y-shape that covers an area of 570 km^2. This is a so-called Very Large Array (VLA). Even higher resolution can be obtained by using an array of radio telescopes in observatories thousands of kilometres apart. A system that uses this so-called technique of 'very-long-baseline interferometry' (VLBI) is known as a 'very-long-baseline array' (VLBA). With VLBA, a radio wavelength resolving power can be achieved that is 100 times greater than the best optical telescopes. Even higher resolving power can be achieved by using a telescope that is in a satellite orbiting Earth. Japan's Institute of Space and Astronautical Science (ISAS) launched such a system in February 1997. The National Astronomical Observatory of Japan, the National Science Foundation's National Radio Astronomy Observatory (NRAO); the Canadian Space Agency; the Australia Telescope National Facility; the European VLBI Network and the Joint Institute back this project for Very Long Baseline Interferometry in Europe. This project is a very good example of how Internationalism can operate in Physics.

AHL

Electron microscope

Telescopes are used to look at very distant objects that are very large but, because of their distance from us, appear very small. Microscopes on the other hand, are used to look at objects that are close to us but are physically very small. As we have seen, just magnifying objects, that is making them appear larger, is not sufficient on its own to gain detail about the object; for detail, high resolution is needed.

Figure 928 is a schematic of how an **optical microscope** is used to view an object and Figure 929 is a schematic of a transmission **electron microscope** (TEM).

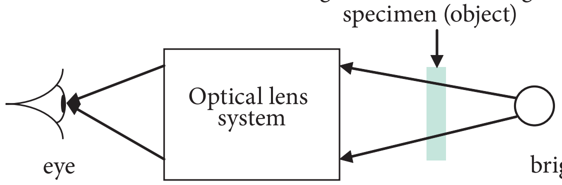


Figure 928 The principle of a light microscope

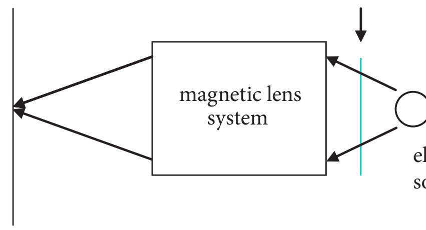


Figure 929 The principle of an electron microscope

In the optical microscope, the resolving power is determined by the particular lens system employed and the wavelength λ of the light used. For example, two points in the sample separated by a distance d will just be resolved if

$$d = \frac{\lambda}{2m}$$

where m is a property of the lens system know as the **numerical aperture**. In practice the largest value of m obtainable is about 1.6. Hence, if the microscope slide is illuminated with light of wavelength 480 nm, a good microscope will resolve two points separated by a distance $d \approx 1.5 \times 10^{-7}$ m ≈ 0.15 μm. Points closer together than this will not be resolved. However, this is good enough to distinguish some viruses such as the Ebola virus.

Clearly, the smaller λ the higher the resolving power and this is where the electron microscope comes to the fore. The electron microscope makes use of the wave nature of electrons (see 13.1.5). In the TEM, electrons pass through a wafer thin sample and are then focused by a magnetic field onto a fluorescent screen or CCD (charge coupled device see 14.2). Electrons used in an electron microscope have wavelengths typically of about 5×10^{-12} m. However, the numerical aperture of electron microscopes is considerably smaller than that of an optical microscope, typically about 0.02. Nonetheless, this means that a TEM can resolve two points that are about 0.25 nm apart. This resolving power is certainly high enough to make out the shape of large molecules.

Another type of electron microscope uses a technique by which electrons are scattered from the surface of the sample. The scattered electrons and then focused as in the TEM to form an image of the surface. These so-called scanning electron microscopes (SEM) have a lower resolving power than TEM's but give very good three dimensional images.

The eye

We saw in the last section that the resolving power of the human eye is about 2×10^{-4} rad. Suppose that you are looking at car headlights on a dark night and the car is a distance D away. If the separation of the headlight is say 1.5 m then the headlights will subtend an angle $\theta = \frac{1.5}{D}$ at your eye. Assuming an average wavelength of 500 nm, your eye will resolve the headlights into two separate sources if this angle equals 2×10^{-4} rad. This gives $D = 7.5$ km. In other words if the car is approaching you on a straight road then you will be able to distinguish the two headlights as separate sources when the car is 7.5 km away from you. Actually because of the structure of the retina and optical defects the resolving power of the average eye is about 6×10^{-4} rad. This means that the car is more likely to be 2.5 km away before you resolve the headlights.

Astronomical telescope

Let us return to the example of the binary stars discussed at the beginning of this section on resolution. The stars Kruger A and B form a binary system. The average separation of the stars is 1.4×10^{12} m and their average distance from Earth is 1.2×10^{17} m. When viewed through a telescope on Earth, the system will therefore subtend an angle.

$$\theta = \frac{1.4 \times 10^{12}}{1.2 \times 10^{17}}$$

$= 1.2 \times 10^{-5}$ rad at the objective lens of the telescope. Assuming that the average wavelength of the light emitted by the stars is 500 nm, then if the telescope is to resolve the system into two separate images it must have a minimum diameter D where $1.2 \times 10^{-5} = \frac{1.22 \times 5.00 \times 10^{-7}}{D}$.

This gives $D = 0.050$m, which is about 5 cm. So this particular system is easily resolved with a small astronomical telescope.

Exercise 9.6

1. It is suggested that using the ISAS, VLBA, it would be possible to "see" a grain of rice at a distance of 5000 km. Estimate the resolving power of the VLBA.

2. The distance from the eye lens to the retina is 20 mm. The light receptors in the central part of the retina are about 5×10^{-6} apart. Determine whether the spacing of the receptors will allow for the eye to resolve the headlights in the above discussion when they are 2.5 km from the eye.

3. The diameter of Pluto is 2.3×10^{6} m and its average distance from Earth is 6.0×10^{12} m. Estimate the minimum diameter of the objective of a telescope that will enable Pluto to be seen as a disc as opposed to a point source.

9.5 Doppler effect

NATURE OF SCIENCE:

Technology: Although originally based on physical observations of the pitch of fast moving sources of sound, the Doppler effect has an important role in many different areas such as evidence for the expansion of the universe and generating images used in weather reports and in medicine. (5.5)

Essential idea: The Doppler effect describes the phenomenon of wavelength/frequency shift when relative motion occurs.

Understandings:
- The Doppler effect for sound waves and light waves

The Doppler effect for sound waves and light waves

Consider two observers A and B at rest with respect to a sound source that emits a sound of constant frequency f. Clearly both observers will hear a sound of the same frequency. However, suppose that the source now moves at constant speed towards A. A will now hear a sound of frequency f_A that is greater than f and B will hear a sound of frequency f_B that is less than f. This phenomenon is known as the **Doppler Effect** or Doppler Principle after *C. J. Doppler* (1803-1853).

The same effect arises for an observer who is either moving towards or away from a stationary source.

Figure 930 shows the waves spreading out from a stationary source that emits a sound of constant frequency f. The observers A and B hear a sound of the same frequency.

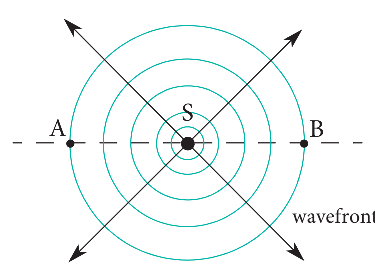


Figure 930 Sound waves from a stationary source

Suppose now that the source moves towards A with constant speed v. Figure 9310 (a) shows a snapshot of the new wave pattern.

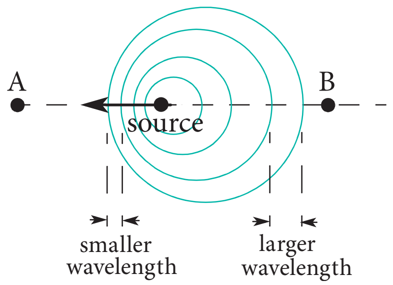


Figure 931 (a) *Sound waves from a moving source*

The wavefronts are now crowded together in the direction of travel of the source and stretched out in the opposite direction. This is why now the two observers will now hear notes of different frequencies. How much the waves bunch together and how much they stretch out will depend on the speed v. Essentially, $f_A = \frac{c}{\lambda_A}$ and $f_B = \frac{c}{\lambda_B}$ where $\lambda_A < \lambda_B$ and v is the speed of sound.

If the source is stationary and A is moving towards it, then the waves from the source incident on A will be bunched up. If A is moving away from the stationary source then the waves from the source incident on A will be stretched out.

Christian Doppler (1803–1853) actually applied the principle (incorrectly as it happens) to try and explain the colour of stars. However, the Doppler effect does apply to light as well as to sound. If a light source emits a light of frequency f then if it is moving away from an observer the observer will measure the light emitted as having a lower frequency than f. Since the sensation of colour vision is related to the frequency of light (blue light is of a higher frequency than red light), light emitted by objects moving way from an observer is often referred to as being red-shifted whereas if the object is moving toward the observer it is referred to as blue-shifted. This idea is used in Option E (Chapter 16).

We do not need to consider here the situations where either the source or the observer is accelerating. In a situation for example where an accelerating source is approaching a stationary observer, then the observer will hear a sound of ever increasing frequency. This sometimes leads to confusion in describing what is heard when a source approaches, passes and then recedes from a stationary observer. Suppose for example that you are standing on a station platform and a train sounding its whistle is approaching at constant speed. What will you hear as the train approaches and then passes through the station? As the train approaches you will hear a sound of *constant* pitch but increasing loudness. The pitch of the sound will be greater than if the train were stationary. As the train passes through the station you will hear the pitch change at the moment the train passes you, to a sound, again of constant pitch. The pitch of this sound will be lower than the sound of the approaching train and its intensity will decrease as the train recedes from you. What you do not hear is a sound of increasing pitch and then decreasing pitch.

The Doppler equations for sound

Although you will not be expected in an IB examination to derive the equations associated with aspects of the Doppler effect, you will be expected to apply them. For completeness therefore, the derivation of the equations associated with the Doppler effect as outlined above is given here.

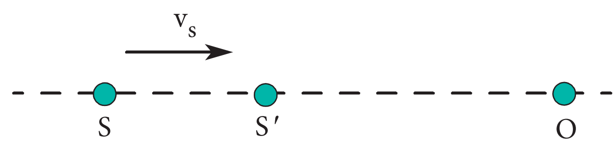


Figure 931 (b)

In Figure 931 (b) the observer O is at rest with respect to a source of sound S is moving with constan speed v_s directly towards O. The source is emitting a note of constant frequency f and the speed of the emitted sound is v.

S' shows the position of the source Δt later. When the source is at rest, then in a time Δt the observer will receive $f\Delta t$ waves and these waves will occupy a distance $v\Delta t$. i.e

$$\lambda = \frac{v\Delta t}{f\Delta t} = \frac{v}{f}$$

(*Because of the motion of the source this number of waves will now occupy a distance* ($v\Delta t - v_s\Delta t$). The 'new' wavelength is therefore

$$\lambda' = \frac{v\Delta t - v_s\Delta t}{f\Delta t} = \frac{v - v_s}{f}$$

If f' is the frequency heard by O then

$$f' = \frac{v}{\lambda'} \text{ or } \lambda' = \frac{v}{f'} = \frac{v - v_s}{f}$$

From which

$$f' = \frac{v}{v - v_s} f$$

Dividing through by v gives

$$f' = f\left(\frac{1}{1 - \frac{v_s}{v}}\right)$$ **Equation 11.1**

If the source is moving away from the observer then we have

$$f' = \frac{v}{v - v_s} f = f\left(\frac{1}{1 + \frac{v_s}{v}}\right)$$ **Equation 11.2**

We now consider the case where the source is stationary and the **observer is moving towards the source** with speed v_0. In this situation the speed of the sound waves as measured by the observer will be $v_0 + v$. We therefore have that

$$v_0 + v = \frac{f}{\lambda} = f^{/} \times \frac{v}{\lambda}$$

From which

$$f^{/} = \left(1 + \frac{v_o}{v}\right) f \qquad \textbf{Equation 11.3}$$

If the **observer is moving away from the source** then

$$f^{/} = \left(1 - \frac{v_o}{v}\right) f$$

From equation (11.3), we have that

$$\Delta f = f^{/} - f = \left(1 + \frac{v_o}{v}\right) f - f = \frac{v_o}{v} f \qquad \textbf{Equation 11.4}$$

The velocities that we refer to in the above equations are the velocities with respect to the medium in which the waves from the source travel. However, when we are dealing with a light source it is the relative velocity between the source and the observer that we must consider. The reason for this is that light is unique in the respect that the speed of the light waves does not depend on the speed of the source. All observers irrespective of their speed or the speed of the source will measure the same velocity for the speed of light. This is one of the cornerstones of the Special Theory of Relativity which is discussed in more detail in Option H (Chapter18). When applying the Doppler effect to light we are mainly concerned with the motion of the source. We look here only at the situation where the speed of the source v is much smaller than the speed of light c in free space. ($v << c$). Under these circumstances, when the source is moving towards the observer, equation 11.1 becomes

$$f^{/} - f = \Delta f = \frac{v}{c} f \qquad \textbf{Equation 11.5}$$

and when the source is moving away from the observer, equation 11.2 becomes $f^{/} - f = \Delta f = -\frac{v}{c} f$

Provided that $v << c$, these same equations apply for a stationary source and moving observer

We look at the following example and exercise.

Example

A source emits a sound of frequency 440 Hz. It moves in a straight line towards a stationary observer with a speed of 30 m s-1. The observer hears a sound of frequency 484 Hz. Calculate the speed of sound in air.

Solution

We use equation 11.1 and substitute f = 484 Hz, f = 440 Hz and v_s = 30 m s^{-1}.

therefore $484 = 440\left(\frac{1}{1 - \frac{30}{v}}\right)$ *such that* $1 - \frac{30}{v} = \frac{440}{484}$ *to give v = 330 m s^{-1}.*

Example

A particular radio signal from a galaxy is measured as having a frequency of 1.39×10^9 Hz. The same signal from a source in a laboratory has a frequency of 1.42×10^9 Hz.

Suggest why the galaxy is moving away from Earth and calculate its recession speed (i.e. the speed with which it is moving away from Earth).

Solution

The fact that the frequency from the moving source is less than that when it is stationary indicates that it is moving away from the stationary observer i.e. Earth.

Using $\Delta f = \frac{v}{c} f$ *we have*

$$v = \frac{c\Delta f}{f} = \frac{3\times10^8 \times (1.42 - 1.39) \times 10^9}{1.42 \times 10^9} = 6.34 \times 10^6 \text{ m s}^{-1}$$

It is usual when dealing with the Doppler effect of light to express speeds as a fraction of c. So in this instance we have v = 0.021 c

Using the Doppler effect

We have seen in the above example and exercise how the Doppler effect may be used to measure the recession speed of distant galaxies. The effect is also used to measure speed in other situations. Here we will look at the general principle involved in using the Doppler effect to measure speed. Figure 932 shows a source (the transmitter) that emits either sound or em waves of constant frequency *f*. The waves

AHL

from the source are incident on a reflector that is moving towards the transmitter with speed v. The reflected waves are detected by the receiver placed alongside the transmitter.

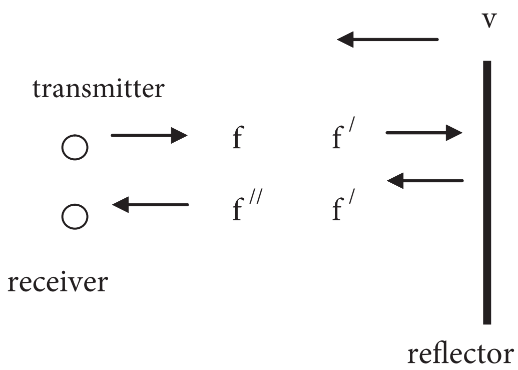


Figure 932 Using the Doppler effect to measure speed

We shall consider the situation where $v << c$ where c is the speed of the waves from the transmitter.

For the reflector receiving waves from the transmitter, it is effectively an observer moving towards a stationary source. From equation (11.4), it therefore receives waves that have been Doppler shifted by an amount

$$f' - f = \frac{v}{c} f \qquad \textbf{Equation 11.6}$$

For the receiver receiving waves from the reflector, it is effectively a stationary observer receiving waves from a moving source. From equation (11.5), it therefore receives waves that have been Doppler shifted by an amount

$$f'' - f' = \frac{v}{c} f' \qquad \textbf{Equation 11.7}$$

If we add equations (11.6) and (11.7) we get that the total Doppler shift at the receiver Δf is

$$f'' - f = \Delta f = f' \frac{v}{c} + f \frac{v}{c}$$

But $f' = \left(1 + \frac{v}{c}\right) f$ hence

$$\Delta f = f\left(1 + \frac{v}{c}\right)\frac{v}{c} + \frac{v}{c} f$$

But since $v << c$, we can ignore the term $\frac{v^2}{c^2}$ when we expand the bracket in the above equation.

Therefore we have

$$\Delta f = \frac{2v}{c} f \qquad \textbf{Equation 11.8}$$

If $v \approx c$ then we must use the full Doppler equations. However, for em radiation we will always only consider situations in which $v << c$.

Example

The speed of sound in blood is 1.500×10^3 m s^{-1}. Ultrasound of frequency 1.00 MHz is reflected from blood flowing in an artery. The frequency of the reflected waves received back at the transmitter is 1.05 MHz. Estimate the speed of the blood flow in the artery.

Solution

Using equation (11.8) we have

$$0.05 \times 10^6 = \frac{2v}{1.5 \times 10^3} \times 10^6$$

to give $v \approx 36$ m s^{-1}. (We have assumed that the ultrasound is incident at right angles to the blood flow.)

Exercise 9.7

1. Judy is standing on the platform of a station. A high speed train is approaching the station in a straight line at constant speed and is sounding its whistle. As the train passes by Judy, the frequency of the sound emitted by the whistle as heard by Judy, changes from 640 Hz to 430 Hz. Determine

   (a) the speed of the train

   (b) the frequency of the sound emitted by the whistle as heard by a person on the train. (Speed of sound = 330 m s^{-1})

2. A galaxy is moving away from Earth with a speed of 0.0500c. The wavelength of a particular spectral line in light emitted by atomic hydrogen in a laboratory is 6.56×10^{-7} m. Calculate the value of the wavelength of this line, measured in the laboratory, in light emitted from a source of atomic hydrogen in the galaxy.

AHL

10. Fields

Contents

Essential Ideas

Electric charges and masses each influence the space around them and that influence can be represented through the concept of fields.

Similar approaches can be taken in analysing electrical and gravitational potential problems.

10.1 Describing fields

NATURE OF SCIENCE:

Paradigm shift: The move from direct, observable actions being responsible for influence on an object to acceptance of a field's "action at a distance" required a paradigm shift in the world of science. (2.3)

Essential idea: Electric charges and masses each influence the space around them and that influence can be represented through the concept of fields.

Understandings:

- Gravitational fields
- Electrostatic fields
- Electric potential and gravitational potential
- Field lines
- Equipotential surfaces

Gravitational potential

We have seen that if we lift an object of mass *m* to a height *h* above the surface of the Earth then its gain in gravitational potential energy is *mgh*. However, this is by no means the full story. For a start we now know that *g* varies with *h* and also the expression really gives a difference in potential energy between the value that the object has at the Earth's surface and the value that it has at height *h*. So what we really need is a zero point. Can we find a point where the potential energy is zero and use this point from which to measure changes in potential energy?

The point that is chosen is in fact infinity. At infinity the gravitational field strength of any object will be zero. So let us see if we can deduce an expression for the gain in potential energy of an object when it is "lifted" from the surface of the Earth to infinity. This in effect means finding the work necessary to perform this task.

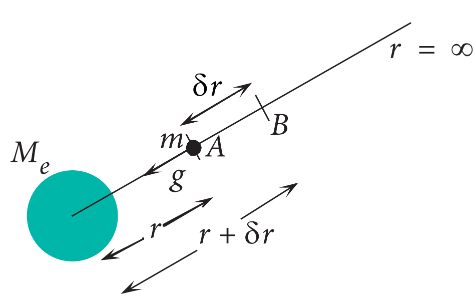


Figure 1001 Gravitational forces

In the diagram we consider the work necessary to move the particle of mass *m* a distance δr in the gravitational field of the Earth.

The force on the particle at *A* is $F = \frac{GM_e m}{r^2}$

If the particle is moved to *B*, then since δr is very small, we can assume that the field remains constant over the distance *AB*. The work δW done *against* the gravitational field of the Earth in moving the distance AB is

$$\delta W = -\frac{GM_e m}{r^2}\delta r$$

(remember that work done against a force is negative)

To find the total work done, *W*, in going from the surface of the Earth to infinity we have to add all these little bits of work. This is done mathematically by using integral calculus.

$$W = \int_R^{\infty}\left(-\frac{GM_e m}{r^2}\right)dr = -GM_e m\int_R^{\infty}\frac{1}{r^2}dr = -GM_e m\left[-\frac{1}{r}\right]_R^{\infty}$$

$$= -GM_e m\left[0 - \left(-\frac{1}{R}\right)\right]$$

$$= -\frac{GM_e m}{R}$$

Hence we have, where *R* is the radius of the Earth, that the work done by the gravitational field in moving an object of mass *m* from *R* (surface of the Earth) to infinity, is given by

$$W = -\frac{GM_e m}{R}$$

We can generalise the result by calculating the work necessary per unit mass to take a small mass from the surface of the Earth to infinity. This we call the **gravitational potential,** *V*, i.e.,

$$V = \frac{W}{m}$$

We would get exactly the same result if we calculated the work done by the field to bring the point mass from infinity to the surface of Earth. In this respect the formal definition of gravitational potential at a point in a gravitational field is therefore defined as **the work done per unit mass in bringing a point mass from infinity to that point.**

Clearly then, the gravitational potential at any point in the Earth's field distance *r* from the centre of the Earth (providing $r > R$) is

$$V = -\frac{GM_e}{r}$$

The potential is therefore a measure of the amount of work that has to be done to move particles between points in a gravitational field and its unit is the J kg^{-1}. We also note that the potential is negative so that the potential energy as we move away from the Earth's surface increases until it reaches the value of zero at infinity.

If the gravitational field is due to a point mass of mass m, then we have the same expression as above except that M_e is replaced by m and must also exclude the value of the potential at the point mass itself i.e. at $r = 0$.

We can express the gravitational potential due to the Earth (or due to any spherical mass) in terms of the gravitational field strength at its surface.

At the surface of the Earth we have

$$-g_0 R_e = -\frac{GM_e}{R_e}$$

So that,

$$g_0 R_e^2 = GM_e$$

Hence at a distance r from the centre of the Earth the gravitational potential V can be written as

$$V = -\frac{GM_e}{r} = -\frac{g_0 R_e^2}{r}$$

The potential at the surface of the Earth

($r = R_e$) is therefore $-g_0 R_e$

It is interesting to see how the expression for the gravitational potential ties in with the expression mgh. The potential at the surface of the Earth is $-g_0 R_e$ (see the example above) and at a height h will be $-g_0(R_e + h)$ if we assume that g_0 does not change over the distance h. The difference in potential between the surface and the height h is therefore $g_0 h$. So the work needed to raise an object of mass m to a height h is mgh , i.e., $m \times$ difference in gravitational potential

This we have referred to as the gain in gravitational potential energy (see 2.3.5).

However, this expression can be extended to any two points in any gravitational field such that if an object of mass m moves between two points whose potentials are V_1 and V_2 respectively, then the change in gravitational potential energy of the object is $m(V_1 - V_2)$.

Gravitational potential gradient

Let us consider now a region in space where the gravitational field is constant. In Figure 912 the two points A and B are separated by the distance Δx.

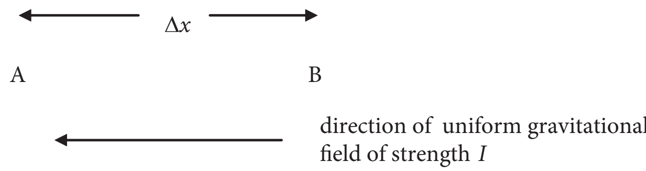


Figure 1002 The gravitational potential gradient

The gravitational field is of strength I and is in the direction shown. The gravitational potential at A is V and at B is $V + \Delta V$.

The work done is taking a point mass m from A to B is $F\Delta x = mI\Delta x$.

However, by definition this work is also equal to $-m\Delta V$.

Therefore $mI\Delta x = -m\Delta V$

or $I = -\dfrac{\Delta V}{\Delta x}$

Effectively this says that the magnitude of the gravitational field strength is equal to the negative gradient of the potential. If I is constant then V is a linear function of x and I is equal to the negative gradient of the straight line graph formed by plotting V against x. If I is not constant (as usually the case), then the magnitude of I at any point in the field can be found by find the gradient of the V-x graph at that point. An example of such a calculation can be found in Section 9.2.9.

For those of you who do HL maths the relationship between field and potential is seen to follow from the expression for the potential of a point mass viz:

$$V = -G\frac{m}{r}$$

$$-\frac{dV}{dr} = +G\frac{m}{r^2} = I$$

Potential due to one or more point masses

Gravitational potential is a scalar quantity so calculating the potential due to more than one point mass is a matter of simple addition. So for example, the potential V due to the Moon and Earth and a distance x from the centre of Earth, on a straight line between them, is given by the expression

$$V = -G\left(\frac{M_E}{x} + \frac{M_M}{r - x}\right)$$

where M_E = mass of Earth, M_M= mass of Moon and r = distance between centre of Earth and Moon.

AHL

Equipotentials and field lines

If the gravitational potential has the same value at all points on a surface, the surface is said to be an **equipotential surface.** So for example, if we imagine a spherical shell about Earth whose centre coincides with the centre of Earth, this shell will be an equipotential surface. Clearly, if we represent the gravitational field strength by field lines, since the lines "radiate" out from the centre of Earth, then these lines will be at right angles to the surface If the field lines were not normal to the equipotential surface then there would be a component of the field parallel to the surface. This would mean that points on the surface would be at different potentials and so it would no longer be an equipotential surface. This of course holds true for any equipotential surface.

Figure 1003 shows the field lines and equipotentials for two point masses *m*.

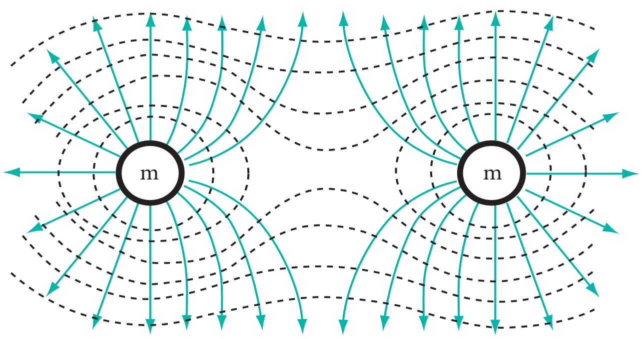


Figure 1003 Equipotentials for two point masses

It is worth noting that we would get exactly the same pattern if we were to replace the point masses with two equal point charges. (See 9.3.5)

Escape speed

The potential at the surface of Earth is $-G\frac{M}{R}$ which means that the energy required to take a particle of mass *m* from the surface to infinity is equal to $-G\frac{Mm}{R}$

But what does it actually mean to take something to infinity? When the particle is on the surface of the Earth we can think of it as sitting at the bottom of a "potential well" as in figure 1004.

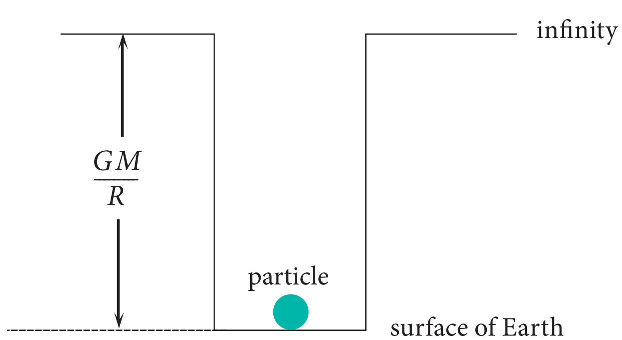


Figure 1004 A potential well

The "depth" of the well is $\frac{GM}{R}$ and if the particle gains an amount of kinetic energy equal to $\frac{GMm}{R}$ where *m* is its mass then it will have just enough energy to "lift" it out of the well.

In reality it doesn't actually go to infinity it just means that the particle is effectively free of the gravitational attraction of the Earth. We say that it has "escaped" the Earth's gravitational pull. We meet this idea in connection with molecular forces. Two molecules in a solid will sit at their equilibrium position, the separation where the repulsive force is equal to the attractive force. If we supply just enough energy to increase the separation of the molecules such that they are an infinite distance apart then intermolecular forces no longer affect the molecules and the solid will have become a liquid. There is no increase in the kinetic energy of the molecules and so the solid melts at constant temperature.

We can calculate the escape speed of an object very easily by equating the kinetic energy to the potential energy such that

$$\frac{1}{2}mv_{escape}^2 = \frac{GM_em}{R_e}$$

$$\Rightarrow v_{escape} = \sqrt{\frac{2GM_e}{R_e}} = \sqrt{2g_0R_e}$$

Substituting for g_0 and R_e gives a value for v_{escape} of about 11 km s^{-1} from the surface of the Earth.

You will note that the escape speed does not depend on the mass of the object since both kinetic energy and potential energy are proportional to the mass.

In theory, if you want to get a rocket to the moon it can be done without reaching the escape speed. However, this would necessitate an enormous amount of fuel and it is likely that the rocket plus fuel would be so heavy that it would never get off the ground. It is much more practical to accelerate the rocket to the escape speed from Earth orbit and then, in theory, just launch it to the Moon.

Example

Use the following data to determine the potential at the surface of Mars and the magnitude of the acceleration of free fall

mass of Mars = 6.4×10^{23} kg

radius of Mars = 3.4×10^{6} m

Determine also the gravitational field strength at a distance of 6.8×10^{6} m above the surface of Mars.

Solution

$$V = -G\frac{M}{R} = -6.7\times10^{-11}\times\frac{6.4\times10^{23}}{3.4\times10^{6}} = -1.3\times10^{7}\ \text{N kg}^{-1}$$

But $V = -g_0 R$

Therefore $g_0 = -\frac{V}{R} = \frac{1.3\times10^{7}}{3.4\times10^{6}} = 3.8\ m\ s^{-2}$

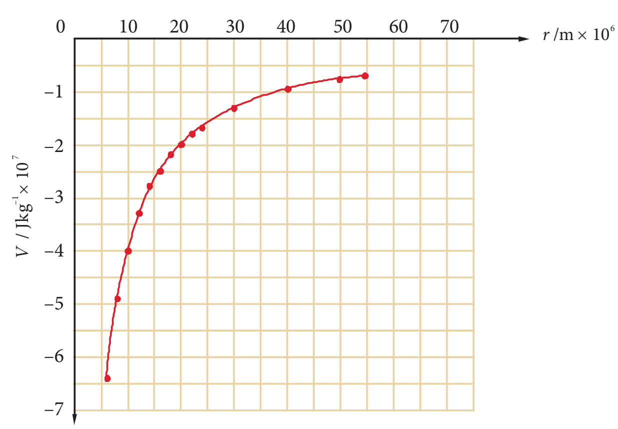
To determine the field strength g_h *at* 6.8×10^{6} *m above the surface, we use the fact that* $g_0 = G\frac{M}{R^2}$ *such that* $GM = g_0 R^2$

Therefore $g_h = \frac{GM}{R_h^{\ 2}} = \frac{g_0 R^2}{R_h^{\ 2}} = \frac{3.8\times(3.4)^2}{(10.2)^2} = 0.42\ \text{m s}^{-2}$

(the distance from the centre is $3.4\times10^{6} + 6.8\times10^{6} = 10.2\times10^{6}$ *m)*

Exercise 10.1

1. The graph below shows how the gravitational potential outside of the Earth varies with distance from the centre.



(a) Use the graph to determine the gain in gravitational potential energy of a satellite of mass 200 kg as it moves from the surface of the Earth to a height of 3.0×10^{7} m above the Earth's surface.

(b) Calculate the energy required to take it to infinity?

(c) Determine the slope of the graph at the surface of the Earth, m? Comment on your answer.

Electric potential energy

The concept of electric potential energy was developed with gravitational potential energy in mind. Just as an object near the surface of the Earth has potential energy because of its gravitational interaction with the Earth, so too there is electrical potential energy associated with interacting charges.

Let us first look at a case of two positive point charges each of 1μC that are initially bound together by a thread in a vacuum in space with a distance between them of 10 cm as shown in Figure 1005. When the thread is cut, the point charges, initially at rest would move in opposite directions, moving with velocities $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ along the direction of the electrostatic force of repulsion.

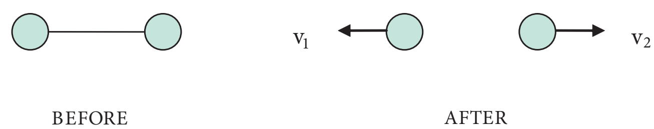


Figure 1005 Interaction of two positive particles

The **electric potential energy** between two point charges can be found by simply adding up the energy associated with each pair of point charges. For a pair of interacting charges, the electric potential energy is given by:

$$\Delta U = \Delta E_p + \Delta E_k = \Delta W = \Delta F_r = \frac{kqQ}{r^2}\times r = \frac{kqQ}{r}$$

Because no external force is acting on the system, the energy and momentum must be conserved. Initially, $E_k = 0$ and $E_p = k\,qQ / r = 9\times10^{9}\times1\times10^{-12} / 0.1\ \text{m} = 0.09$ J. When they are a great distance from each other, E_p will be negligible. The final energy will be equal to $\frac{1}{2}m\boldsymbol{v}_1^{\ 2} + \frac{1}{2}m\boldsymbol{v}_2^{\ 2} = 0.09$ J. Momentum is also conserved and the velocities would be the same magnitude but in opposite directions.

Electric potential energy is more often defined in terms of a point charge moving in an electric field as:

'the electric potential energy between any two points in an electric field is defined as negative of the work done by an electric field in moving a point electric charge between two locations in the electric field.'

$$\Delta U = \Delta Ep = -\Delta W = -Fd = qEx$$

where x is the distance moved along (or opposite to) the direction of the electric field.

Electric potential energy is measured in joule (J). Just as work is a scalar quantity, so too electrical potential energy is a **scalar quantity**. **The negative of the work done by an electric field in moving a unit electric charge between two points is independent of the path taken.** In physics, we say the electric field is a 'conservative' field.

AHL

Suppose an external force such as your hand moves a small positive point test charge in the direction of a uniform electric field. As it is moving it must be gaining kinetic energy. If this occurs, then the electric potential energy of the unit charge is changing.

In Figure 1006 a point charge $+q$ is moved between points A and B through a distance x in a uniform electric field.

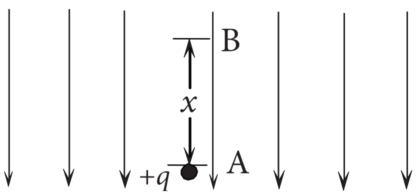


Figure 1006 Movement of a positive point charge in a uniform field

In order to move a positive point charge from point A to point B, an external force must be applied to the charge equal to qE ($\boldsymbol{F} = q\boldsymbol{E}$).

Since the force is applied through a distance x, then negative work has to be done to move the charge because energy is gained, meaning there is an increase **electric potential energy** between the two points. Remember that the work done is equivalent to the energy gained or lost in moving the charge through the electric field. The concept of electric potential energy is only meaningful as the electric field which generates the force in question is conservative.

$$W = F \times x = Eq \times x$$

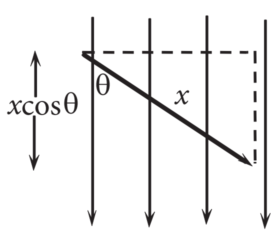


Figure 1007 Charge moved at an angle to the field

If a charge moves at an angle θ to an electric field, the component of the displacement parallel to the electric field is used as shown in Figure 918

$$W = Fx = Eq \times x \cos\theta$$

The electric potential energy is stored in the electric field, and the electric field will return the energy to the point charge when required so as not to violate the Law of conservation of energy.

Electric potential

The electric potential at a point in an electric field is defined as being the work done per unit charge in bringing a small positive point charge from infinity to that point.

$$\Delta V = V_{\infty} - V_f = -\frac{W}{q}$$

If we designate the potential energy to be zero at infinity then it follows that electric potential must also be zero at infinity and the electric potential at any point in an electric field will be:

$$\Delta V = -\frac{W}{q}$$

Now suppose we apply an external force to a small positive test charge as it is moved towards an isolated positive charge. The external force must do work on the positive test charge to move it towards the isolated positive charge and the work must be positive while the work done by the electric field must therefore be negative. So the electric potential at that point must be positive according to the above equation. If a negative isolated charge is used, the electric potential at a point on the positive test charge would be negative. Positive point charges of their own accord, move from a place of high electric potential to a place of low electric potential. Negative point charges move the other way, from low potential to high potential. In moving from point A to point B in the diagram, the positive charge $+q$ is moving from a low electric potential to a high electric potential.

In the definition given, the term "work per unit charge" has significance. If the test charge is $+1.6 \times 10^{-19}$C where the charge has a potential energy of 3.2×10^{-17} J, then the potential would be 3.2×10^{-17}J / $+1.6 \times 10^{-19}$ C = 200 JC^{-1}. Now if the charge was doubled, the potential would become 6.4×10^{-17} J. However, the potential per unit charge would be the same.

Electric potential is a scalar quantity and it has units JC^{-1} or volts where 1 volt equals one joule per coloumb. The volt allows us to adopt a unit for the electric field in terms of the volt.

Previously, the unit for the electric field was NC^{-1}.

$W = qV$ and $F = qE$, so $\frac{W}{V} = \frac{F}{E}$

$E = \frac{FV}{W} = \frac{FV}{Fm}$ $V\,m^{-1}$.

That is, the units of the electric field, E, can also be expressed as $V\,m^{-1}$.

The work done per unit charge in moving a point charge between two points in an electric field is again independent of the path taken.

Electric potential due to a point charge

Let us take a point r metres from a charged object. The potential at this point can be calculated using the following:

$W = -Fr = -qV$ and $F = -\dfrac{q_1 q_2}{4\pi\varepsilon_0 r^2}$

Therefore,

$$W = -\frac{q_1 q_2}{4\pi\varepsilon_0 r^2} \times r = -\frac{q_1 q_2}{4\pi\varepsilon_0 r} = -q_1 \times \frac{q_2}{4\pi\varepsilon_0 r} = -q_1 V$$

That is

$$V = \frac{q}{4\pi\varepsilon_0 r}$$

Or, simply

$$V = \frac{kq}{r}$$

Example

Determine how much work is done by the electric field of point charge 15.0 μC when a charge of 2.00 μC is moved from infinity to a point 0.400 m from the point charge. (Assume no acceleration of the charges).

Solution

The work done by the electric field is $W = -qV$
$= -1/4\pi\varepsilon_0 \times q \times (Q/r_\infty - Q/r_{0.400})$

$W = (-2.00 \times 10^{-6}\ C \times 9.00 \times 10^{9}\ NmC^{-2} \times 15.0 \times 10^{-6}\ C) \div 0.400\ m = -0.675\ J$

An external force would have to do +0.675 J of work.

Electric field strength and electric potential gradient

Let us look back at Figure 1006. Suppose again that the charge +q is moved a small distance by a force F from A to B so that the force can be considered constant. The work done is given by:

$$\Delta W = F \times \Delta x$$

The force F and the electric field E are oppositely directed, and we know that:

$\mathbf{F} = -q\mathbf{E}$ *and* $\Delta W = q\,\Delta V$

Therefore, the work done can be given as:

$q\,\Delta V = -q\,\mathbf{E}\,\Delta x$

Therefore $$E = -\frac{\Delta V}{\Delta x}$$

The rate of change of potential ΔV at a point with respect to distance Δx in the direction in which the change is maximum is called the ***potential gradient****. We say that the electric field = - the potential gradient and the units are* Vm^{-1}. *From the equation we can see that in a graph of electric potential versus distance, the gradient of the straight line equals the electric field strength.*

In reality, if a charged particle enters a uniform electric field, it will be accelerated uniformly by the field and its kinetic energy will increase. This is why we had to assume no acceleration in the last worked example.

$$\Delta E_k = \frac{1}{2}mv^2 = q \cdot E \cdot x = q \cdot \frac{V}{x} \cdot x = q \cdot V$$

Example

Determine how far apart two parallel plates must be situated so that a potential difference of 1.50×10^2 V produces an electric field strength of 1.00×10^3 NC^{-1}.

Solution

Using $E = -\dfrac{\Delta V}{x} \Leftrightarrow x = \dfrac{\Delta V}{E} = \dfrac{1.5 \times 10^2\ V}{1.00 \times 10^3\ N\,C^{-1}}$

$= 1.50 \times 10^{-1}$

The plates are $\mathbf{1.50 \times 10^{-1}\ m}$ *apart.*

The electric field and the electric potential at a point due to an evenly distributed charge +q on a sphere can be represented graphically as in Figure 1008.

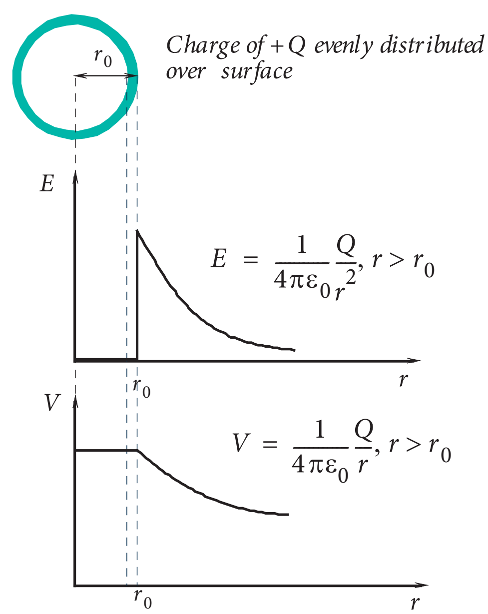


Figure 1008 Electric field and potential due to a charged sphere

When the sphere becomes charged, we know that the charge distributes itself evenly over the surface. Therefore every part of the material of the conductor is at the same potential. As the electric potential at a point is defined as being numerically equal to the work done in bringing a unit positive charge from infinity to that point, it has a constant value in every part of the material of the conductor.

Since the potential is the same at all points on the conducting surface, then $\Delta V / \Delta x$ is zero. But $E = -\Delta V / \Delta x$. Therefore, the electric field inside the conductor is zero. There is no electric field inside the conductor.

Some further observations of the graphs in Figure 1009 are:

- Outside the sphere, the graphs obey the relationships given as $\boldsymbol{E} \propto 1 / r^2$ and $V \propto 1 / r$
- At the surface, $r = r_0$. Therefore, the electric field and potential have the minimum value for r at this point and this infers a maximum field and potential.
- Inside the sphere, the electric field is zero.
- Inside the sphere, no work is done to move a charge from a point inside to the surface. Therefore, there is no potential **difference** and the potential is the same as it is when $r = r_0$.

Similar graphs can be drawn for the electric field intensity and the electric potential as a function of distance from conducting parallel plates and surfaces, and these are given in Figure 1009.

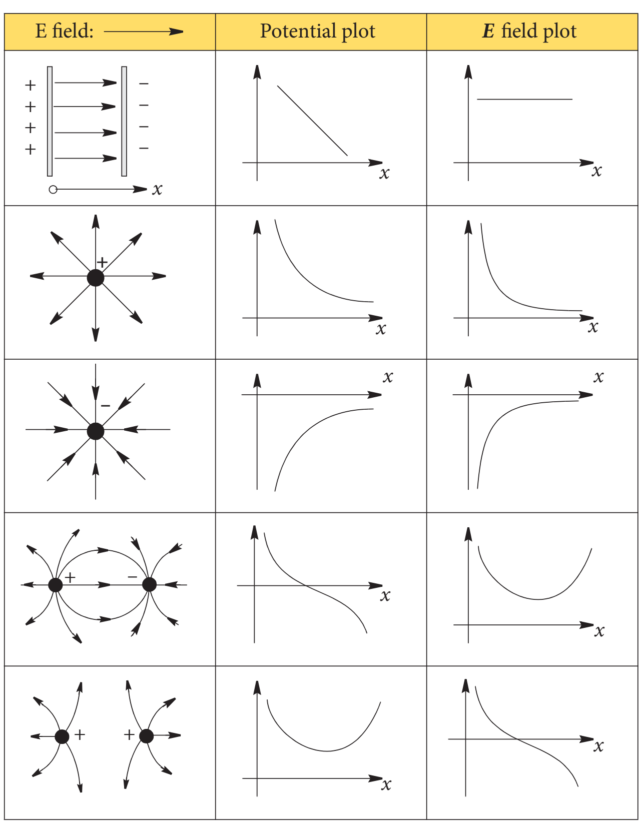


Figure 1009 Electric field and electric potential at a distance from a charged surface

Potential due to one or more point charges

The potential due to one point charge can be determined by using the equation formula

$V = kq / r.$

Example 1

Determine the electric potential at a point 2.0×10^{-1} m from the centre of an isolated conducting sphere with a point charge of 4.0 pC in air.

Solution

Using the formula

$V = kq / r$ *, we have*

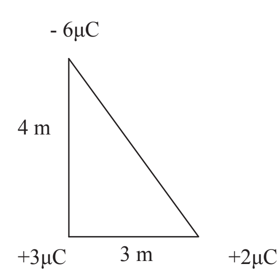
$$V = \frac{(9.0 \times 10^{9}) \times (4.0 \times 10^{-12})}{(2.0 \times 10^{-1})} = 0.18 \text{ V}$$

the potential at the point is $\mathbf{1.80 \times 10^{-1}}$ **V**.

The potential due to a number of point charges can be determined by adding up the potentials due to individual point charges because the electric potential at any point outside a conducting sphere will be the same as if all the charge was concentrated at its centre.

Example 2

Three point charges of are placed at the vertices of a right-angled triangle as shown in the diagram below. Determine the absolute potential at the + 2.0 μC charge, due to the two other charges.



Solution

The electric potential of the +2 μC charge due to the – 6 μC charge is:

$V = (9 \times 10^9 \ Nm^2C^{-2} \times -6 \times 10^{-6} \ C) \div (\sqrt{3^2 + 4^2}) \ m =$
$- 1.08 \times 10^4 \ V$

The electric potential of the +2 μC charge due to the +3 μC charge is:

$V = (9 \times 10^9\ Nm^2C^{-2} \times 3 \times 10^{-6}\ C) \div 3m = 9 \times 10^3\ V$

The net absolute potential is the sum of the 2 potentials
$-\ 1.08 \times 10^4\ V + 9 \times 10^3\ V =$
$-\ 1.8 \times 10^3\ V$

The absolute potential at the point is - ***1.8 × 10³ V****.*

Equipotential surfaces

Regions in space where the electric potential of a charge distribution has a constant value are called equipotentials**.** The places where the potential is constant in three dimensions are called **equipotential surfaces,** and where they are constant in two dimensions they are called **equipotential lines**.

They are in some ways analogous to the contour lines on topographic maps. In this case, the gravitational potential energy is constant as a mass moves around the contour lines because the mass remains at the same elevation above the Earth's surface. The gravitational field strength acts in a direction perpendicular to a contour line.

Similarly, because the electric potential on an equipotential line has the same value, an electric force can do no work when a test charge moves on an equipotential. Therefore, the electric field cannot have a component along an equipotential, and thus it must be everywhere perpendicular to the equipotential surface or equipotential line. This fact makes it easy to plot equipotentials if the lines of force or lines of electric flux of an electric field are known.

For example, there are a series of equipotential lines between two parallel plate conductors that are perpendicular to the electric field. There will be a series of concentric circles (each circle further apart than the previous one) that map out the equipotentials around an isolated positive sphere. The lines of force and some equipotential lines for an isolated positive sphere are shown in Figure 1010.

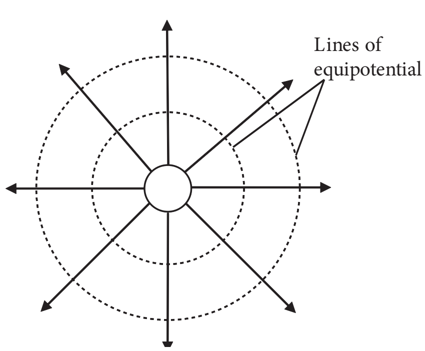


Figure 1010 Equipotentials around an isolated positive sphere

In summary, we can conclude that

- **No work is done to move a charge along an equipotential.**
- **Equipotentials are always perpendicular to the electric lines of force.**

Figure 1011 and 1012 show some equipotential lines for two oppositely charged and identically positive spheres separated by a distance.

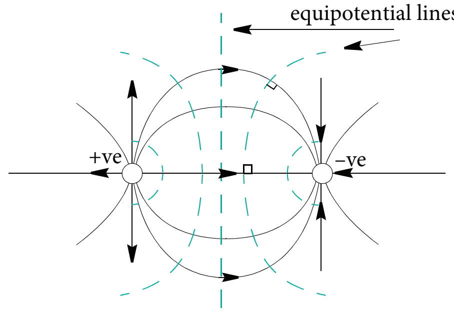


Figure 1011 Equipotential lines between two opposite charges

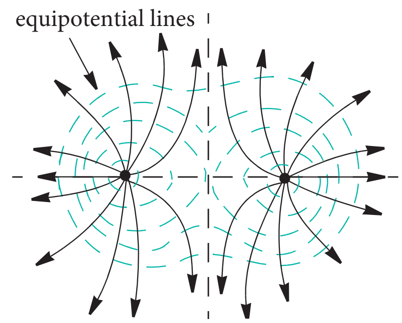


Figure 1012 Equipotential lines between two charges which are the same

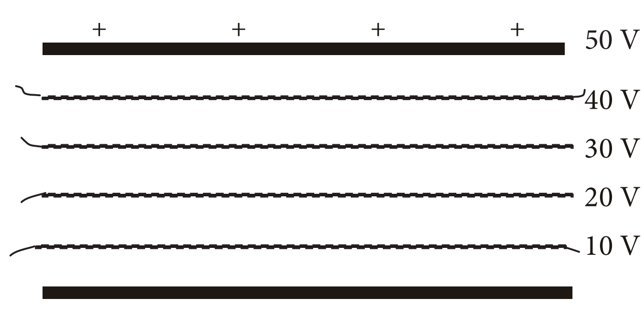


Figure 1013 Equipotential lines between charged parallel plates

Figure 1013 shows the equipotential lines between charged parallel plates. Throughout this chapter the similarities and differences between gravitational fields and electric fields have been discussed. The relationships that exists between gravitational and electric quantities and the effects of point masses and charges is summarised in Table 1014

	Gravitational quantity	Electrical quantity
Quantities	$V = \frac{W}{m}$	$V = \frac{W}{q}$
	$g = \frac{F}{m}$	$E = \frac{F}{q}$
	$g = -\frac{\Delta V}{\Delta x}$	$E = -\frac{\Delta V}{\Delta x}$
Point masses and charges	$V = -G\frac{m}{r}$	$V = \frac{1}{4\pi\varepsilon_0}\frac{q}{r}$
	$g = -G\frac{m}{r^2}$	$E = \frac{1}{4\pi\varepsilon_0}\frac{q}{r^2}$
	$F = G\frac{m_1 m_2}{r^2}$	$F = \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{r^2}$

Figure 1014 Formulas

Example

Deduce the electric potential on the surface of a gold nucleus that has a radius of 6.2 fm.

Solution

Using the formula

$V = kq / r$, and knowing the atomic number of gold is 79. We will assume the nucleus is spherical and it behaves as if it were a point charge at its centre (relative to outside points).

$V = 9.0 \times 10^9\ Nm^2C^{-2} \times 79 \times 1.6 \times 10^{-19}\ C \div 6.2 \times 10^{-15}\ m = 1.8 \times 10^7\ V$

*The potential at the point is **18 MV**.*

Example

Deduce the ionisation energy in electron-volts of the electron in the hydrogen atom if the electron is in its ground state and it is in a circular orbit at a distance of 5.3×10^{-11} m from the proton.

Solution

This problem is an energy, coulombic, circular motion question based on Bohr's model of the atom (not the accepted quantum mechanics model). The ionisation energy is the energy required to remove the electron from the ground state to infinity. The electron travels in a circular orbit and therefore has a centripetal acceleration. The ionisation energy will counteract the coulombic force and the movement of the electron will be in the opposite direction to the centripetal force.

Total energy = E_k electron + E_p due to the proton-electron interaction

$\Sigma \mathbf{F} = kqQ / r^2 = m\mathbf{v}^2 / r$ and as such $m\mathbf{v}^2 = = kqQ / r$.

Therefore, E_k electron = ½ kqQ / r.

E_p due to the proton-electron interaction = - kqQ / r.

Total energy = ½ kqQ / r + - kqQ / r = -½ kqQ / r

$= -\ 9.0 \times 10^9\ Nm^2C^{-2} \times (1.6 \times 10^{-19}\ C)^2 \div 5.3 \times 10^{-11}\ m = -2.17 \times 10^{-18}\ J$

$= -2.17 \times 10^{-18}\ J \div 1.6 \times 10^{-19} = -13.6\ eV$.

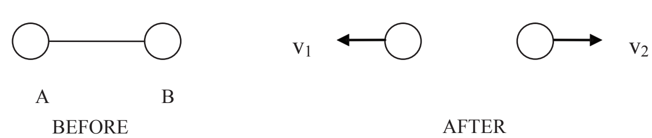
*The ionisation energy is **13.6 eV**.*

Exercise 10.2

1. A point charge P is placed midway between two identical negative charges. Which one of the following is correct with regards to electric field and electric potential at point P?

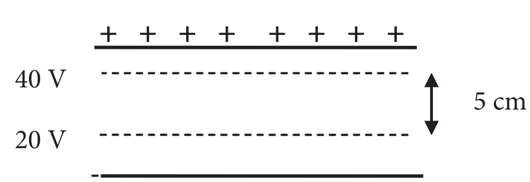
	Electric field	Electric potential
A	non-zero	zero
B	zero	non-zero
C	non-zero	non-zero
D	zero	zero

2. Two positive charged spheres are tied together in a vacuum somewhere in space where there are no external forces. A has a mass of 25 g and a charge of 2.0 μC and B has a mass of 15 g and a charge of 3.0 μC. The distance between them is 4.0 cm. They are then released as shown in the diagram.



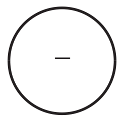
(a) Determine their initial electric potential energy in the before situation.
(b) Determine the speed of sphere B after release.

3. The diagram below represents two equipotential lines in separated by a distance of 5 cm in a uniform electric field.



Determine the strength of the electric field.

4. This question is about the electric field due to a charged sphere and the motion of electrons in that field. The diagram below shows an isolated, metal sphere in a vacuum that carries a negative electric charge of 6.0 μC.



(a) On the diagram draw the conventional way to represent the electric field pattern due to the charged sphere and lines to represent three equipotential surfaces in the region outside the sphere.

(b) Explain how the lines representing the equipotential surfaces that you have sketched indicate that the strength of the electric field is decreasing with distance from the centre of the sphere.

(c) The electric field strength at the surface of the sphere and at points outside the sphere can be determined by assuming that the sphere acts as a point charge of magnitude 6.0 μC at its centre. The radius of the sphere is 2.5×10^{-2} m. Deduce that the magnitude of the field strength at the surface of the sphere is 8.6×10^{7} Vm^{-1}.

An electron is initially at rest on the surface of the sphere.

(d) (i) Describe the path followed by the electron as it leaves the surface of the sphere.

(ii) Calculate the initial acceleration of the electron.

5. Determine the amount of work that is done in moving a charge of 10.0 nC through a potential difference of 1.50×10^{2} V.

6. Three identical 2.0 μC conducting spheres are placed at the corners of an equilateral triangle of sides 25 cm. The triangle has one apex C pointing up the page and 2 base angles A and B. Determine the absolute potential at B .

7. Determine how far apart two parallel plates must be situated so that a potential difference of 2.50×10^{2} V produces an electric field strength of 2.00×10^{3} NC^{-1}.

8. The gap between two parallel plates is 1.0×10^{-3} m, and there is a potential difference of 1.0×10^{4} V between the plates. Calculate

i. the work done by an electron in moving from one plate to the other
ii. the speed with which the electron reaches the second plate if released from rest.
iii. the electric field intensity between the plates.

9. An electron gun in a picture tube is accelerated by a potential 2.5×10^{3} V. Determine the kinetic energy gained by the electron in electron-volts.

10. Determine the electric potential 2.0×10^{-2} m from a charge of -1.0×10^{-5} C.

11. Determine the electric potential at a point mid-way between a charge of –20 pC and another of + 5 pC on the line joining their centres if the charges are 10 cm apart.

12. During a thunderstorm the electric potential difference between a cloud and the ground is 1.0×10^{9} V. Determine the magnitude of the change in electric potential energy of an electron that moves between these points in electron-volts.

13. A charge of 1.5 μC is placed in a uniform electric field of two oppositely charged parallel plates with a magnitude of 1.4×10^{3} NC^{-1}.

(a) Determine the work that must be done against the field to move the point charge a distance of 5.5 cm.

(b) Calculate the potential difference between the final and initial positions of the charge.

(c) Determine the potential difference between the plates if their separation distance is 15 cm.

14. During a flash of lightning, the potential difference between a cloud and the ground was 1.2×10^{9} V and the amount of transferred charge was 32 C.

(a) Determine the change in energy of the transferred charge.

(b) If the energy released was all used to accelerate a 1 tonne car, deduce its final speed.

(c) If the energy released could be used to melt ice at 0 °C, deduce the amount of ice that could be melted.

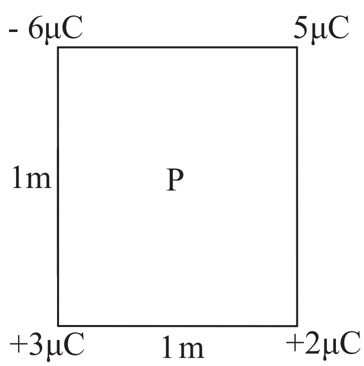
AHL

15. Suppose that when an electron moved from A to B in the diagram along an electric field line that the electric field does 3.6×10^{-19} J of work on it.

Determine the differences in electric potential:

(a) $V_B - V_A$
(b) $V_C - V_A$
(c) $V_C - V_B$

16. Determine the potential at point P that is located at the centre of the square as shown in the diagram below.



10.2 Fields at work

NATURE OF SCIENCE:

Communication of scientific explanations: The ability to apply field theory to the unobservable (charges) and the massively scaled (motion of satellites) required scientists to develop new ways to investigate, analyse and report findings to a general public used to scientific discoveries based on tangible and discernible evidence. (5.1)

Essential idea: Similar approaches can be taken in analysing electrical and gravitational potential problems.

Understandings:
- Potential and potential energy
- Potential gradient
- Potential difference
- Escape speed
- Orbital motion, orbital speed and orbital energy
- Forces and inverse-square law behaviour

Orbital motion, orbital speed and orbital energy

Although orbital motion may be circular, elliptical or parabolic, this sub-topic only deals with circular orbits. This sub-topic is not fundamentally new physics, but an application that synthesizes ideas from gravitation, circular motion, dynamics and energy.

The Moon orbits the Earth and in this sense it is often referred to as a satellite of the Earth. Before 1957 it was the only Earth satellite. However, in 1957 the Russians launched the first man made satellite, Sputnik 1. Since this date many more satellites have been launched and there are now literally thousands of them orbiting the Earth. Some are used to monitor the weather, some used to enable people to find accurately their position on the surface of the Earth, many are used in communications, and no doubt some are used to spy on other countries. Figure 1015 shows how, in principle, a satellite can be put into orbit.

The person (whose size is greatly exaggerated with respect to Earth) standing on the surface on the Earth throws some stones. The greater the speed with which a stone is thrown the further it will land from her. The paths followed by the thrown stones are parabolas. By a stretch of the imagination we can visualise a situation in which a stone is thrown with such a speed that, because of the curvature of the Earth, it will not land on the surface of the Earth but go into "orbit". (Path 4 on Figure 1015).

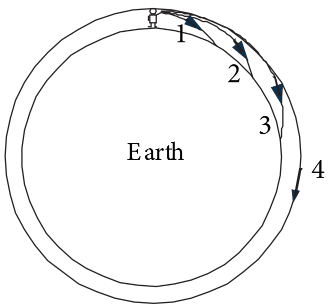


Figure 1015 Throwing a stone into orbit

The force that causes the stones to follow a parabolic path and to fall to Earth is gravity and similarly the force that keeps the stone in orbit is gravity. For circular motion to occur we have seen that a force must act at right angles to the velocity of an object, that is there must be a centripetal force. Hence in the situation we describe here the centripetal force for circular orbital motion about the Earth is provided by gravitational attraction of the Earth.

We can calculate the speed with which a stone must be thrown in order to put it into orbit just above the surface of the Earth.

If the stone has mass m and speed v then we have from Newton's 2nd law

$$\frac{mv^2}{R_E} = G\frac{M_E m}{R_E^2}$$

where R_E is the radius of the Earth and M_E is the mass of the Earth.

Bearing in mind that $g_0 = G\frac{M_E}{R_E^2}$, then

$$v = \sqrt{gR_E} = \sqrt{10 \times 6.4 \times 10^6} = 8 \times 10^3$$

That is, the stone must be thrown at 8×10^3m s^{-1}.

Clearly we are not going to get a satellite into orbit so close to the surface of the Earth. Moving at this speed the friction due to air resistance would melt the satellite before it had travelled a couple of kilometres. In reality therefore a satellite is put into orbit about the Earth by sending it, attached to a rocket, beyond the Earth's atmosphere and then giving it a component of velocity perpendicular to a radial vector from the Earth. See Figure 1016.

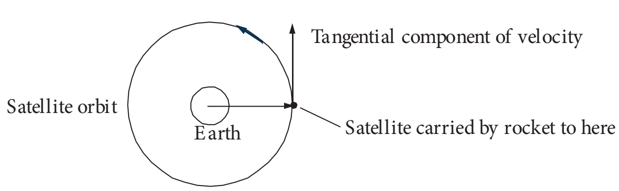


Figure 1016 Getting a satellite into orbit

Kepler's third law

(*This work of Kepler and Newton's synthesis of the work is an excellent example of the scientific method and makes for a good TOK discussion)*

In 1627 *Johannes Kepler* (1571-1630) published his laws of planetary motion. The laws are empirical in nature and were deduced from the observations of the Danish astronomer *Tycho de Brahe* (1546-1601). The third law gives a relationship between the radius of orbit R of a planet and its period T of revolution about the Sun. The law is expressed mathematically as

$$\frac{T^2}{R^3} = \text{constant}$$

We shall now use Newton's Law of Gravitation to show how it is that the planets move in accordance with Kepler's third law.

In essence Newton was able to use his law of gravity to predict the motion of the planets since all he had to do was factor the F given by this law into his second law, $\boldsymbol{F} = m\boldsymbol{a}$, to find their accelerations and hence their future positions.

In Figure 1017 the Earth is shown orbiting the Sun and the distance between their centres is R.

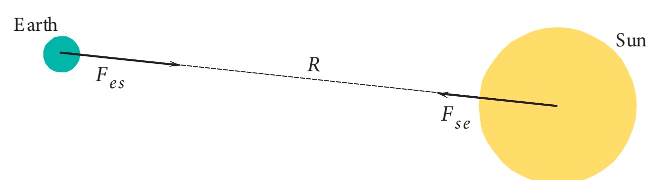


Figure 1017 Planets move according to Kepler's third law

F_{es} is the force that the Earth exerts on the Sun and F_{se} is the force that the Sun exerts on the Earth. The forces are equal and opposite and the Sun and the Earth will actually orbit about a common centre. However since the Sun is so very much more massive than the Earth this common centre will be close to the centre of the Sun and so we can regard the Earth as orbiting about the centre of the Sun. The other thing that we shall assume is that we can ignore the forces that the other planets exert on the Earth. (This would not be a wise thing to do if you were planning to send a space ship to the Moon for example.). We shall also assume that we have followed Newton's example and indeed proved that a sphere will act as a point mass situated at the centre of the sphere.

Kepler had postulated that the orbits of the planets are elliptical but since the eccentricity of the Earth's orbit is small we shall assume a circular orbit.

AHL

The acceleration of the Earth towards the Sun is $a = R\omega^2$

where $\omega = \dfrac{2\pi}{T}$

Hence, $$a = R \times \left(\frac{2\pi}{T}\right)^2 = \frac{4\pi^2 R}{T^2}$$

But the acceleration is given by Newton's Second Law, $F = ma$, where F is now given by the Law of Gravitation. So in this situation

$$F = ma = \frac{GM_s M_e}{R^2}$$, but, we also have that

$$a = \frac{4\pi^2 R}{T^2}$$ and $m = M_e$ so that

$$\frac{GM_s M_e}{R^2} = M_e \times \frac{4\pi^2 R}{T^2} \Rightarrow \frac{GM_s}{4\pi^2} = \frac{R^3}{T^2}$$

But the quantity $\dfrac{GM_s}{4\pi^2}$ is a constant that has the same value for each of the planets so we have for all the planets, not just Earth, that

$$\frac{R^3}{T^2} = k$$

where k is a constant. Which is of course Kepler's third law.

This is indeed an amazing breakthrough. It is difficult to refute the idea that all particles attract each other in accordance with the Law of Gravitation when the law is able to account for the observed motion of the planets about the Sun.

The gravitational effects of the planets upon each other should produce perturbations in their orbits. Such is the predictive power of the Universal Gravitational Law that it enabled physicists to compute these perturbations. The telescope had been invented in 1608 and by the middle of the 18th Century had reached a degree of perfection in design that enabled astronomers to actually measure the orbital perturbations of the planets. Their measurements were always in agreement with the predictions made by Newton's law. However, in 1781 a new planet, Uranus was discovered and the orbit of this planet did not fit with the orbit predicted by Universal Gravitation. Such was the physicist's faith in the Newtonian method that they suspected that the discrepancy was due to the presence of a yet undetected planet. Using the Law of Gravitation the French astronomer *J.Leverrier* and the English astronomer. *J. C. Adams* were able to calculate just how massive this new planet must be and also where it should be. In 1846 the planet Neptune was discovered just where they had predicted. In a similar way, discrepancies in the orbit of Neptune led to the prediction and subsequent discovery in 1930 of the planet Pluto. Newton's Law of Gravitation had passed the ultimate test of any theory; it is not only able to explain existing data but also to make predictions.

Satellite energy

When a satellite is in orbit about a planet it will have both kinetic energy and gravitational potential energy. Suppose we consider a satellite of mass m that is in an orbit of radius r about a planet of mass M.

The gravitational potential due to the planet at distance r from its centre is

$$-\frac{GM_e}{r}.$$

The gravitational potential energy of the satellite V_{sat} is therefore $-\dfrac{GM_e m}{r}$.

That is, $$V_{sat} = -\frac{GM_e m}{r}.$$

The gravitational field strength at the surface of the planet is given by

$$g_0 = \frac{GM_e}{R_e^2}$$

Hence we can write

$$V_{sat} = -\frac{g_0 R_e^2 m}{r}$$

The kinetic energy of the satellite K_{sat} is equal to $\frac{1}{2}mv^2$, where v is its orbital speed.

By equating the gravitational force acting on the satellite to its centripetal acceleration we have

$$\frac{GM_e m}{r^2} = \frac{mv^2}{r} \Leftrightarrow mv^2 = \frac{GM_e m}{r}.$$

$$\frac{1}{2}mv^2 = \frac{1}{2} \times \frac{GM_e m}{r}$$

$$= \frac{g_0 R_e^2 m}{2r}$$

Which is actually quite interesting since it shows that, irrespective of the orbital radius the KE is numerically equal to half the PE, Also the **total** energy E_{tot} of the satellite is always negative since

$$E_{tot} = K_{sat} + V_{sat} = \frac{1}{2} \times \frac{GM_e m}{r} + \left(-\frac{GM_e m}{r}\right) = -\frac{1}{2}\frac{GM_e m}{r}$$

AHL

The energies of an orbiting satellite as a function of radial distance from the centre of a planet are shown plotted in Figure 1018.

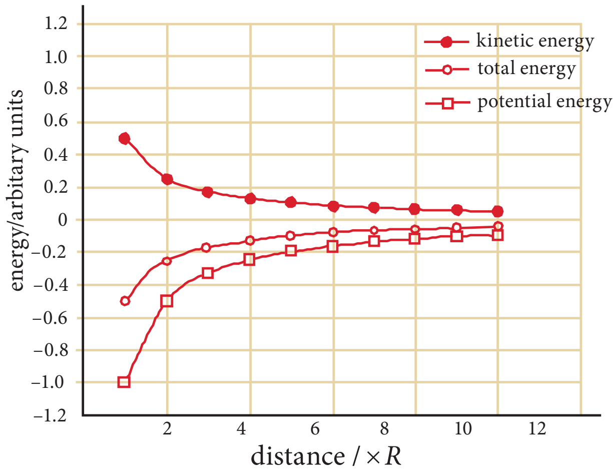


Figure 1018 Energy of an orbiting satellite as a function of distance from the centre of a planet

Weightlessness

Suppose that you are in an elevator (lift) that is descending at constant speed and you let go of a book that you are holding in your hand. The book will fall to the floor with acceleration equal to the acceleration due to gravity. If the cable that supports the elevator were to snap (a situation that I trust will never happen to any of you) and you now let go the book that you are holding in your other hand, this book will not fall to the floor - it will stay exactly in line with your hand. This is because the book is now falling with the same acceleration as the elevator and as such the book cannot "catch" up with the floor of the elevator. Furthermore, if you happened to be standing on a set of bathroom scales, the scales would now read zero - you would be apparently weightless. It is this idea of free fall that explains the apparent weightlessness of astronauts in an orbiting satellite. These astronauts are in free fall in the sense that they are accelerating towards the centre of the Earth.

It is actually possible to define the weight of a body in several different ways. We can define it for example as the gravitational force exerted on the body by a specified object such as the Earth. This we have seen that we do in lots of situations where we define the weight as being equal to mg. If we use this definition, then an object in free fall cannot by definition be weightless since it is still in a gravitational field. However, if we define the weight of an object in terms of a "weighing" process such as the reading on a set of bathroom scales, which in effect measures the contact force between the object and the scales, then clearly objects in free fall are weightless. One now has to ask the question whether or not it is possible. For example, to measure the gravitational force acting on an astronaut in orbit about the Earth.

We can also define weight in terms of the net gravitational force acting on a body due to several different objects. For example for an object out in space, its weight could be defined in terms of the resultant of the forces exerted on it by the Sun, the Moon, the Earth and all the other planets in the Solar System. If this resultant is zero at a particular point then the body is weightless at this point.

In view of the various definitions of weight that are available to us it is important that when we use the word "weight" we are aware of the context in which it is being used.

Example

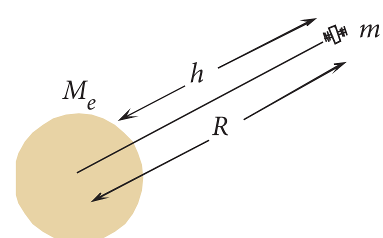
Calculate the height above the surface of the Earth at which a geo-stationary satellite orbits.

Solution

A geo-stationary satellite is one that orbits the Earth in such a way that it is stationary with respect to a point on the surface of the Earth. This means that its orbital period must be the same as the time for the Earth to spin once on its axis i.e. 24 hours.

From Kepler's third law we have $\frac{GM_s}{4\pi^2} = \frac{R^3}{T^2}$.

That is,



using the fact that the force of attraction between the satellite and the Earth is given by

$$F = \frac{GM_e m}{R^2}$$

and that $F = ma$

where m is the mass of the satellite and $a = \frac{4\pi^2 R}{T^2}$

we have,

$$\frac{GM_e m}{R^2} = m \times \frac{4\pi^2 R}{T^2} \Rightarrow \frac{GM_e}{4\pi^2} = \frac{R^3}{T^2}$$

Now, the mass of the Earth is 6.0×10^{24} kg and the period, T, measured in seconds is given by T = 86,400 s.

So substitution gives $R = 42 \times 10^6$ m

The radius of the Earth is 6.4×10^6 m so that the orbital height, h, is about 3.6×10^7 m.

Example

Calculate the minimum energy required to put a satellite of mass 500 kg into an orbit that is as a height equal to the Earth's radius above the surface of the Earth.

Solution

We have seen that when dealing with gravitational fields and potential it is useful to remember that

$$g_0 = \frac{GM}{R_e^2} \text{ or, } g_0 R_e^2 = GM$$

The gravitational potential at the surface of the Earth is

$$-g_0 R_e = -\frac{GM}{R_e}.$$

The gravitational potential at a distance R

from the centre of the Earth is $-\dfrac{GM}{R}$

$$= -\frac{g_0 R_e^2}{R}$$

The difference in potential between the surface of the Earth and a point distance R from the centre is therefore

$$\Delta V = g_0 R_e\left(1 - \frac{R_e}{R}\right)$$

If $R = 2R_e$ *then* $\Delta V = \dfrac{g_0 R_e}{2}$

This means that the work required to "lift" the satellite into orbit is $g_0 Rm$ where m is the mass of the satellite. This is equal to

$10 \times 3.2 \times 10^6 \times 500 = 16000$ *MJ.*

However, the satellite must also have kinetic energy in order to orbit Earth. This will be equal to

$$\frac{g_0 m R_e^2}{2R} = \frac{g_0 m R_e^2}{4} = 8000\ MJ$$

The minimum energy required is therefore

24000 MJ.

Exercise 10.3

1. The speed needed to put a satellite in orbit does not depend on

   A. the radius of the orbit.
   B. the shape of the orbit.
   C. the value of g at the orbit.
   D. the mass of the satellite.

2. Estimate the speed of an Earth satellite whose orbit is 400 km above the Earth's surface. Also determine the period of the orbit.

3. Calculate the speed of a 200 kg satellite, orbiting the Earth at a height of 7.0×10^6 m.

   Assume that g = 8.2 m s^{-2} for this orbit.

4. The radii of two satellites, X and Y, orbiting the Earth are 2*r* and 8*r* where *r* is the radius of the Earth. Calculate the ratio of the periods of revolution of X to Y.

5. A satellite of mass *m* kg is sent from Earth's surface into an orbit of radius 5*R*, where *R* is the radius of the Earth. Write down an expression for

   (a) the potential energy of the satellite in orbit.

   (b) the kinetic energy of the satellite in orbit.

   (c) the minimum work required to send the satellite from rest at the Earth's surface into its orbit.

6. A satellite in an orbit of 10*r*, falls back to Earth (radius *r*) after a malfunction. Determine the speed with which it will hit the Earth's surface?

7. The radius of the moon is ¼ that of the Earth Assuming Earth and the Moon to have the same density, compare the accelerations of free fall at the surface of Earth to that at the surface of the Moon.

8. Use the following data to determine the gravitational field strength at the surface of the Moon and hence determine the escape speed from the surface of the Moon.

   Mass of the Moon = 7.3×10^{22} kg,

   Radius of the Moon = 1.7×10^6 m