FHSST Authors

The Free High School Science Texts: Textbooks for High School Students Studying the Sciences
Physics
Grades 10 - 12

Version 0
November 9, 2008

STOP!!!!

Did you notice the **FREEDOMS** we've granted you?

Our copyright license is **different!** It grants freedoms rather than just imposing restrictions like all those other textbooks you probably own or use.

- We know people copy textbooks illegally but we would LOVE it if you copied our's - go ahead copy to your hearts content, **legally!**
- Publishers' revenue is generated by controlling the market, we don't want any money, go ahead, distribute our books far and wide - we DARE you!
- Ever wanted to change your textbook? Of course you have! Go ahead, change ours, make your own version, get your friends together, rip it apart and put it back together the way you like it. That's what we really want!
- Copy, modify, adapt, enhance, share, critique, adore, and contextualise. Do it all, do it with your colleagues, your friends, or alone but get involved! Together we can overcome the challenges our complex and diverse country presents.
- So what is the catch? The only thing you can't do is take this book, make a few changes and then tell others that they can't do the same with your changes. It's share and share-alike and we know you'll agree that is only fair.
- These books were written by volunteers who want to help support education, who want the facts to be freely available for teachers to copy, adapt and re-use. Thousands of hours went into making them and they are a gift to everyone in the education community.

FHSST Core Team

Mark Horner ; Samuel Halliday ; Sarah Blyth ; Rory Adams ; Spencer Wheaton

FHSST Editors

Jaynie Padayachee ; Joanne Boulle ; Diana Mulcahy ; Annette Nell ; René Toerien ; Donovan Whitfield

FHSST Contributors

Rory Adams ; Prashant Arora ; Richard Baxter ; Dr. Sarah Blyth ; Sebastian Bodenstein ; Graeme Broster ; Richard Case ; Brett Cocks ; Tim Crombie ; Dr. Anne Dabrowski ; Laura Daniels ; Sean Dobbs ; Fernando Durrell ; Dr. Dan Dwyer ; Frans van Eeden ; Giovanni Franzoni ; Ingrid von Glehn ; Tamara von Glehn ; Lindsay Glesener ; Dr. Vanessa Godfrey ; Dr. Johan Gonzalez ; Hemant Gopal ; Umeshree Govender ; Heather Gray ; Lynn Greeff ; Dr. Tom Gutierrez ; Brooke Haag ; Kate Hadley ; Dr. Sam Halliday ; Asheena Hanuman ; Neil Hart ; Nicholas Hatcher ; Dr. Mark Horner ; Robert Hovden ; Mfandaidza Hove ; Jennifer Hsieh ; Clare Johnson ; Luke Jordan ; Tana Joseph ; Dr. Jennifer Klay ; Lara Kruger ; Sihle Kubheka ; Andrew Kubik ; Dr. Marco van Leeuwen ; Dr. Anton Machacek ; Dr. Komal Maheshwari ; Kosma von Maltitz ; Nicole Masureik ; John Mathew ; JoEllen McBride ; Nikolai Meures ; Riana Meyer ; Jenny Miller ; Abdul Mirza ; Asogan Moodaly ; Jothi Moodley ; Nolene Naidu ; Tyrone Negus ; Thomas O'Donnell ; Dr. Markus Oldenburg ; Dr. Jaynie Padayachee ; Nicolette Pekeur ; Sirika Pillay ; Jacques Plaut ; Andrea Prinsloo ; Joseph Raimondo ; Sanya Rajani ; Prof. Sergey Rakityansky ; Alastair Ramlakan ; Razvan Remsing ; Max Richter ; Sean Riddle ; Evan Robinson ; Dr. Andrew Rose ; Bianca Ruddy ; Katie Russell ; Duncan Scott ; Helen Seals ; Ian Sherratt ; Roger Sieloff ; Bradley Smith ; Greg Solomon ; Mike Stringer ; Shen Tian ; Robert Torregrosa ; Jimmy Tseng ; Helen Waugh ; Dr. Dawn Webber ; Michelle Wen ; Dr. Alexander Wetzler ; Dr. Spencer Wheaton ; Vivian White ; Dr. Gerald Wigger ; Harry Wiggins ; Wendy Williams ; Julie Wilson ; Andrew Wood ; Emma Wormauld ; Sahal Yacoob ; Jean Youssef

Contributors and editors have made a sincere effort to produce an accurate and useful resource. Should you have suggestions, find mistakes or be prepared to donate material for inclusion, please don't hesitate to contact us. We intend to work with all who are willing to help make this a continuously evolving resource!

www.fhsst.org

Contents

Part IV

Grade 12 - Physics

Chapter 21

Motion in Two Dimensions - Grade 12

21.1 Introduction

In Chapter 3, we studied motion in one dimension and briefly looked at vertical motion. In this chapter we will discuss vertical motion and also look at motion in two dimensions. In Chapter 12, we studied the conservation of momentum and looked at applications in one dimension. In this chapter we will look at momentum in two dimensions.

21.2 Vertical Projectile Motion

In Chapter 4, we studied the motion of objects in free fall and we saw that an object in free fall falls with gravitational acceleration g. Now we can consider the motion of objects that are thrown upwards and then fall back to the Earth. We call this *projectile motion* and we will only consider the situation where the object is thrown straight upwards and then falls straight downwards - this means that there is no horizontal displacement of the object, only a vertical displacement.

21.2.1 Motion in a Gravitational Field

When an object is in a gravitational field, it always accelerates downwards with a constant acceleration g whether the object is moving upward or downward. This is shown in Figure 21.1.

Important: Projectiles moving upwards or downwards always accelerate downwards with a constant acceleration g.

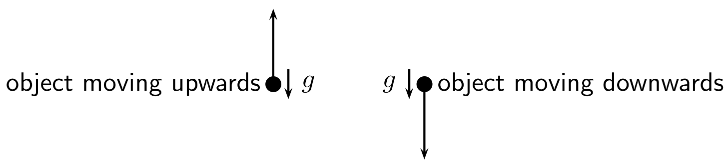


Figure 21.1: Objects moving upwards or downwards, always accelerate downwards.

This means that if an object is moving upwards, it decreases until it stops ($v_f = 0$ m·s^{-1}). This is the maximum height that the object reaches, because after this, the object starts to fall.

Important: Projectiles have zero velocity at their greatest height.

Consider an object thrown upwards from a vertical height h_o. We have seen that the object will travel upwards with decreasing velocity until it stops, at which point it starts falling. The time that it takes for the object to fall down to height h_o is the same as the time taken for the object to reach its maximum height from height h_o.

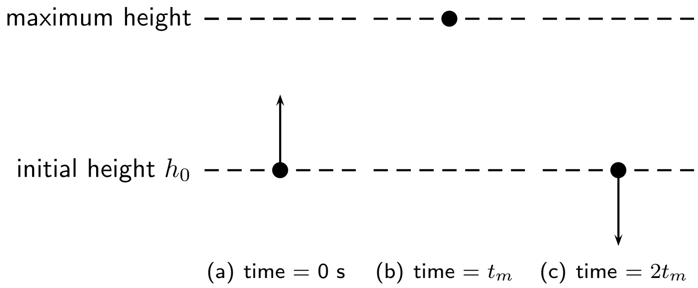


Figure 21.2: (a) An object is thrown upwards from height h_0. (b) After time t_m, the object reaches its maximum height, and starts to fall. (c) After a time $2t_m$ the object returns to height h_0.

Important: Projectiles take the same the time to reach their greatest height from the point of upward launch as the time they take to fall back to the point of launch.

21.2.2 Equations of Motion

The equations of motion that were used in Chapter 4 to describe free fall can be used for projectile motion. These equations are the same as those equations that were derived in Chapter 3, but with $a = g$. We use $g = 9{,}8\,\mathrm{m \cdot s^{-2}}$ for our calculations.

$$\begin{aligned}
v_i &= \text{initial velocity (m·s}^{-1}\text{) at } t = 0 \text{ s} \\
v_f &= \text{final velocity (m·s}^{-1}\text{) at time } t \\
\Delta x &= \text{height above ground (m)} \\
t &= \text{time (s)} \\
\Delta t &= \text{time interval (s)} \\
g &= \text{acceleration due to gravity (m·s}^{-2}\text{)}
\end{aligned}$$

$$v_f = v_i + gt \tag{21.1}$$

$$\Delta x = \frac{(v_i + v_f)}{2}t \tag{21.2}$$

$$\Delta x = v_i t + \frac{1}{2}gt^2 \tag{21.3}$$

$$v_f^2 = v_i^2 + 2g\Delta x \tag{21.4}$$

Worked Example 132: Projectile motion

Question: A ball is thrown upwards with an initial velocity of 10 m·s^{-1}.

1. Determine the maximum height reached above the thrower's hand.
2. Determine the time it takes the ball to reach its maximum height.

Answer

Step 1 : Identify what is required and what is given
We are required to determine the maximum height reached by the ball and how long it takes to reach this height. We are given the initial velocity $v_i = 10$ m·s^{-1}and the acceleration due to gravity g = 9,8 m·s^{-2}.
Step 2 : Determine how to approach the problem
Choose down as positive. We know that at the maximum height the velocity of the ball is 0 m·s^{-1}. We therefore have the following:

- $v_i = -10\,\text{m}\cdot\text{s}^{-1}$ (it is negative because we chose upwards as positive)
- $v_f = 0\,\text{m}\cdot\text{s}^{-1}$
- $g = +9{,}8\,\text{m}\cdot\text{s}^{-2}$

Step 3 : Identify the appropriate equation to determine the height.
We can use:

$$v_f^2 = v_i^2 + 2g\Delta x$$

to solve for the height.
Step 4 : Substitute the values in and find the height.

$$\begin{aligned} v_f^2 &= v_i^2 + 2g\Delta x \\ (0)^2 &= (-10)^2 + (2)(9{,}8)(\Delta x) \\ -100 &= 19{,}6\Delta x \\ \Delta x &= 5{,}102...m \end{aligned}$$

The value for the displacement will be negative because the displacement is upwards and we have chosen downward as positive (and upward as negative). The height will be a positive number, $h = 5.10m$.
Step 5 : Identify the appropriate equation to determine the time.
We can use:

$$v_f = v_i + gt$$

to solve for the time.
Step 6 : Substitute the values in and find the time.

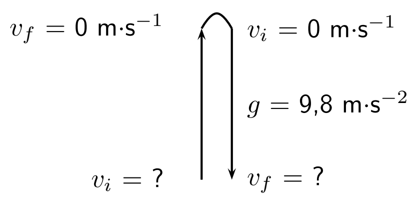
$$\begin{aligned} v_f &= v_i + gt \\ 0 &= -10 + 9{,}8t \\ 10 &= 9{,}8t \\ t &= 1{,}02...s \end{aligned}$$

Step 7 : Write the final answer.
The ball reaches a maximum height of 5,10 m.
The ball takes 1,02 s to reach the top.

Worked Example 133: Height of a projectile

Question: A cricketer hits a cricket ball from the ground so that it goes directly upwards. If the ball takes, 10 s to return to the ground, determine its maximum height.
Answer
Step 1 : Identify what is required and what is given
We need to find how high the ball goes. We know that it takes 10 seconds to go up and down. We do not know what the initial velocity of the ball (v_i) is.
Step 2 : Determine how to approach the problem

A problem like this one can be looked at as if there are two motions. The first is the ball going up with an initial velocity and stopping at the top (final velocity is zero). The second motion is the ball falling, its initial velocity is zero and its final velocity is unknown.



Choose down as positive. We know that at the maximum height, the velocity of the ball is 0 m·s^{-1}. We also know that the ball takes the same time to reach its maximum height as it takes to travel from its maximum height to the ground. This time is half the total time. We therefore have the following for the motion of the ball going down:

- $t = 5\,\mathrm{s}$, half of the total time
- $v_{top} = v_i = 0\,\mathrm{m\cdot s^{-1}}$
- $g = 9{,}8\,\mathrm{m\cdot s^{-2}}$
- $\Delta x = ?$

Step 3 : Find an appropriate equation to use
We are not given the initial velocity of the ball going up and therefore we do not have the final velocity of the ball coming down. We need to choose an equation that does not have v_f in it. We can use the following equation to solve for Δx:

$$\Delta x = v_i t + \frac{1}{2} g t^2$$

Step 4 : Substitute values and find the height.

$$\begin{aligned} \Delta x &= (0)(5) + \frac{1}{2}(9{,}8)(5)^2 \\ \Delta x &= 0 + 122{,}5\mathrm{m} \\ \text{height} &= 122{,}5\mathrm{m} \end{aligned}$$

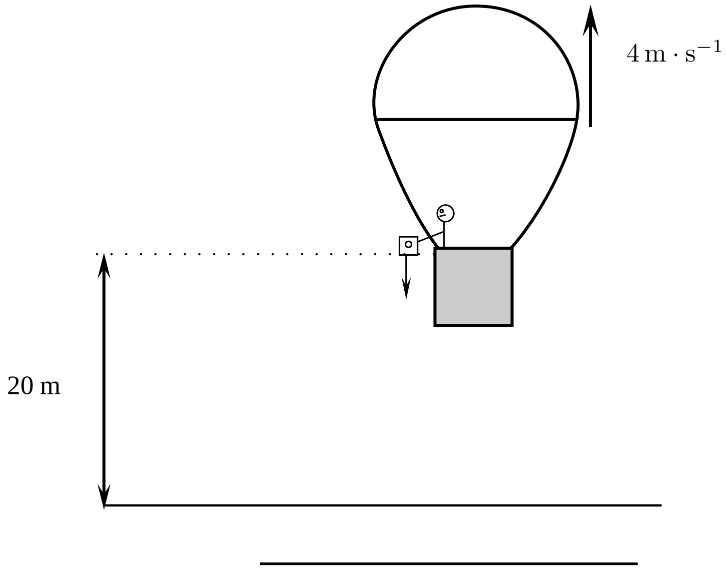
Step 5 : Write the final answer
The ball reaches a maximum height of 122,5 m.

?

Exercise: Equations of Motion

1. A cricketer hits a cricket ball straight up into the air. The cricket ball has an initial velocity of 20 m·s^{-1}.
   A What height does the ball reach before it stops to fall back to the ground.
   B How long has the ball been in the air for?
2. Zingi throws a tennis ball up into the air. It reaches a height of 80 cm.
   A Determine the initial velocity of the tennis ball.
   B How long does the ball take to reach its maximum height?

3. A tourist takes a trip in a hot air balloon. The hot air balloon is ascending (moving up) at a velocity of 4 m·s^{-1}. He accidentally drops his camera over the side of the balloon's basket, at a height of 20 m. Calculate the velocity with which the camera hits the ground.



21.2.3 Graphs of Vertical Projectile Motion

Vertical projectile motion is similar to motion at constant acceleration. In Chapter 3 you learned about the graphs for motion at constant acceleration. The graphs for vertical projectile motion are therefore identical to the graphs for motion under constant acceleration.
When we draw the graphs for vertical projectile motion, we consider two main situations: an object moving upwards and an object moving downwards.
If we take the upwards direction as positive then for an object moving upwards we get the graphs shown in Figure 21.9.

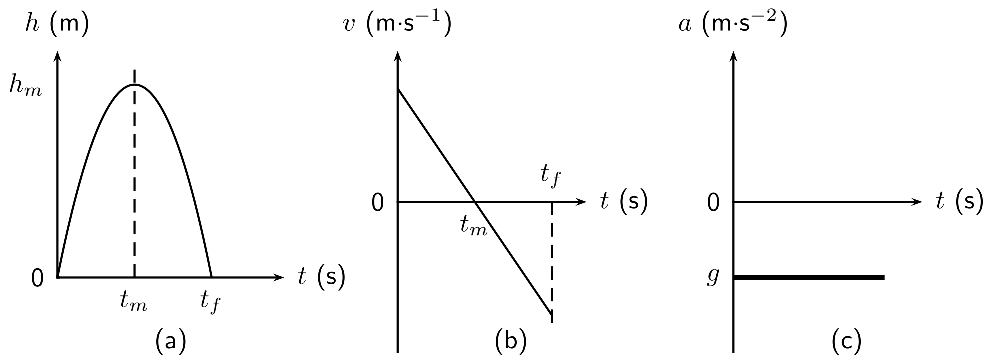


Figure 21.3: Graphs for an object thrown upwards with an initial velocity v_i. The object takes t_m s to reach its maximum height of h_m m after which it falls back to the ground. (a) position vs. time graph (b) velocity vs. time graph (c) acceleration vs. time graph.

Worked Example 134: Drawing Graphs of Projectile Motion

Question: Stanley is standing on the a balcony 20 m above the ground. Stanley tosses up a rubber ball with an initial velocity of 4,9 m·s^{-1}. The ball travels upwards and then falls to the ground. Draw graphs of position vs. time, velocity vs. time and acceleration vs. time. Choose upwards as the positive direction.

Answer

Step 1 : Determine what is required

We are required to draw graphs of

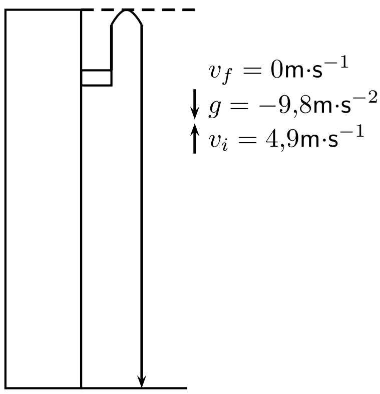
1. Δx vs. t
2. v vs. t
3. a vs. t

Step 2 : Analysis of problem

There are two parts to the motion of the ball:

1. ball travelling upwards from the building
2. ball falling to the ground

We examine each of these parts separately. To be able to draw the graphs, we need to determine the time taken and displacement for each of the motions.



Step 3 : Find the height and the time taken for the first motion.

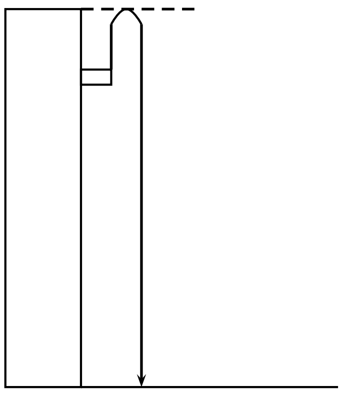
For the first part of the motion we have:

- $v_i = +4{,}9\,\text{m}\cdot\text{s}^{-1}$
- $v_f = 0\,\text{m}\cdot\text{s}^{-1}$
- $g = -9{,}8\,\text{m}\cdot\text{s}^{-2}$

Therefore we can use $v_f^2 = v_i^2 + 2g\Delta x$ to solve for the height and $v_f = v_i + gt$ to solve for the time.

$$\begin{aligned} v_f^2 &= v_i^2 + 2g\Delta x \\ (0)^2 &= (4{,}9)^2 + 2 \times (-9{,}8) \times \Delta x \\ 19{,}6\Delta x &= (4{,}9)^2 \\ \Delta x &= 1{,}225\ m \end{aligned}$$

$$\begin{aligned} v_f &= v_i + gt \\ 0 &= 4{,}9 + (-9{,}8) \times t \\ 9{,}8t &= 4{,}9 \\ t &= 0{,}5\ s \end{aligned}$$



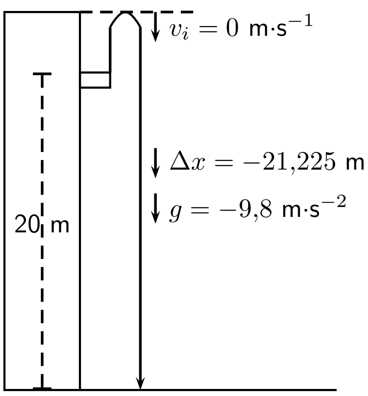
Step 4 : Find the height and the time taken for the second motion.

For the second part of the motion we have:

- $v_i = 0\,\mathrm{m \cdot s^{-1}}$
- $\Delta x = -(20 + 1{,}225)$ m
- $g = -9{,}8\,\mathrm{m \cdot s^{-2}}$

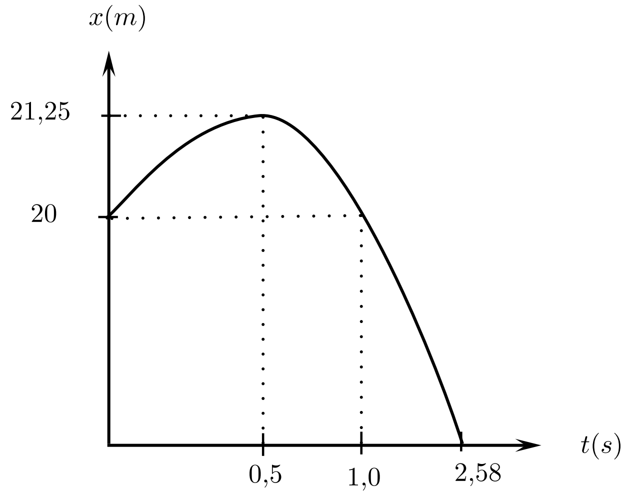
Therefore we can use $\Delta x = v_i t + \frac{1}{2}gt^2$ to solve for the time.

$$\begin{aligned} \Delta x &= v_i t + \frac{1}{2}gt^2 \\ -(20 + 1{,}225) &= (0) \times t + \frac{1}{2} \times (-9{,}8) \times t^2 \\ -21{,}225 &= 0 - 4{,}9t^2 \\ t^2 &= 4{,}33163... \\ t &= 2{,}08125...\ s \end{aligned}$$



Step 5 : Graph of position vs. time

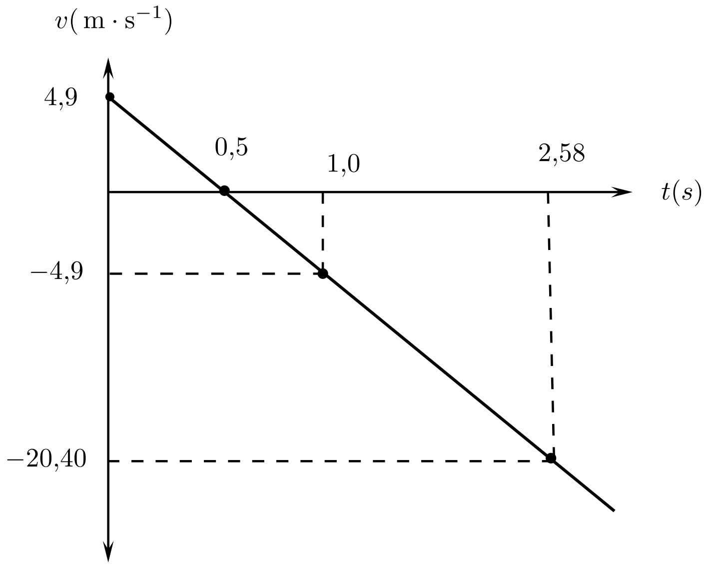
The ball starts from a position of 20 m (at t = 0 s) from the ground and moves upwards until it reaches (20 + 1,225) m (at t = 0,5 s). It then falls back to 20 m (at t = 0,5 + 0,5 = 1,0 s) and then falls to the ground, Δ x = 0 m at (t = 0,5 + 2,08 = 2,58 s).



Step 6 : Graph of velocity vs. time

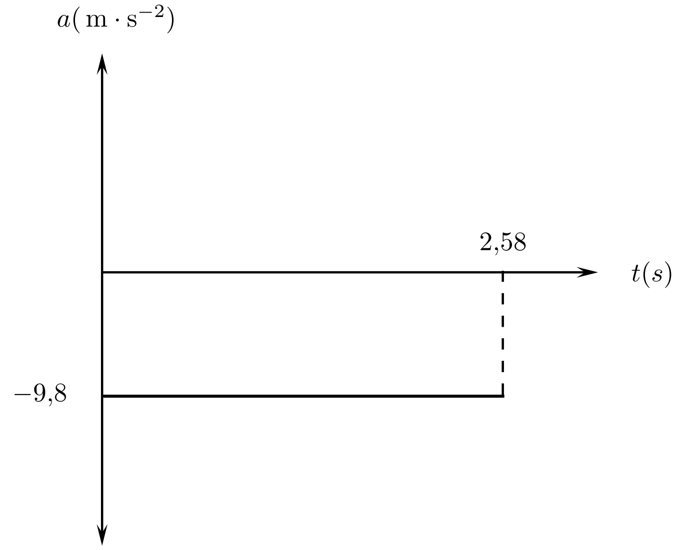
The ball starts off with a velocity of +4,9 m·s^{-1}at t = 0 s, it then reaches a velocity of 0 m·s^{-1}at t = 0,5 s. It stops and falls back to the Earth. At t = 1,0 it has a velocity of -4,9 m·s^{-1}. This is the same as the initial upwards velocity but it is downwards. It carries on at constant acceleration until t = 2,58 s. In other words, the velocity graph will be a straight line. The final velocity of the ball can be calculated as follows:

$$\begin{aligned} v_f &= v_i + gt \\ &= 0 + (-9{,}8)(2{,}08...) \\ &= -20{,}396...\,\mathrm{m \cdot s^{-1}} \end{aligned}$$



Step 7 : Graph of a vs t

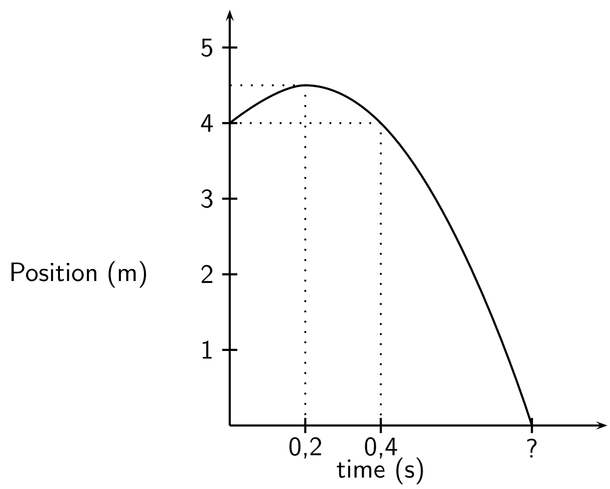
We chose upwards to be positive. The acceleration of the ball is downward. $g = -9.8\,\mathrm{m \cdot s^{-2}}$. Because the acceleration is constant throughout the motion, the graph looks like this:



Worked Example 135: Analysing Graphs of Projectile Motion

Question: The graph below (not drawn to scale) shows the motion of tennis ball that was thrown vertically upwards from an open window some distance from the ground. It takes the ball 0,2 s to reach its highest point before falling back to the ground. Study the graph given and calculate

1. how high the window is above the ground.
2. the time it takes the ball to reach the maximum height.
3. the initial velocity of the ball.
4. the maximum height that the ball reaches.
5. the final velocity of the ball when it reaches the ground.



Answer

Step 1 : Find the height of the window.

The initial position of the ball will tell us how high the window is. From the y-axis on the graph we can see that the ball is 4 m from the ground.

The window is therefore 4 m above the ground.

Step 2 : Find the time taken to reach the maximum height.

The maximum height is where the position-time graph show the maximum position - the top of the curve. This is when t = 0,2 s.

It takes the ball 0,2 seconds to reach the maximum height.

Step 3 : Find the initial velocity (v_i) of the ball.

To find the initial velocity we only look at the first part of the motion of the ball. That is from when the ball is released until it reaches its maximum height. We have the following for this: Choose upwards as positive.

$$\begin{aligned} t &= 0{,}2 \text{ s} \\ g &= -9{,}8 \ \text{m}\cdot\text{s}^{-2} \\ v_f &= 0 \ \text{m}\cdot\text{s}^{-1} \text{(because the ball stops)} \end{aligned}$$

To calculate the initial velocity of the ball (v_i), we use:

$$\begin{aligned} v_f &= v_i + gt \\ 0 &= v_i + (-9{,}8)(0{,}2) \\ v_i &= 1{,}96\,\text{m}\cdot\text{s}^{-1} \end{aligned}$$

The initial velocity of the ball is 1,96 m·s^{-1}upwards.

Step 4 : Find the maximum height (Δx) of the ball.

To find the maximum height we look at the initial motion of the ball. We have the following:

$$\begin{aligned} t &= 0{,}2 \text{ s} \\ g &= -9{,}8 \ \text{m}\cdot\text{s}^{-2} \\ v_f &= 0 \ \text{m}\cdot\text{s}^{-1} \text{(because the ball stops)} \\ v_i &= +1{,}96\,\text{m}\cdot\text{s}^{-1} \text{(calculated above)} \end{aligned}$$

To calculate the maximum height (Δx) we use:

$$\begin{aligned} \Delta x &= v_i t + \frac{1}{2} g t^2 \\ \Delta x &= (1{,}96)(0{,}2) + \frac{1}{2}(-9{,}8)(0{,}2)^2 \\ \Delta x &= 0{,}196\text{m} \end{aligned}$$

The maximum height of the ball is (4 + 0,196) = 4,196 m above the ground.

Step 5 : Find the final velocity (v_f) of the ball.

To find the final velocity of the ball we look at the second part of the motion. For this we have:

$$\begin{aligned} \Delta x &= -4{,}196 \text{ m (because upwards is positive)} \\ g &= -9{,}8 \ \text{m}\cdot\text{s}^{-2} \\ v_i &= 0 \ \text{m}\cdot\text{s}^{-1} \end{aligned}$$

We can use $(v_f)^2 = (v_i)^2 + 2g\Delta x$ to calculate the final velocity of the ball.

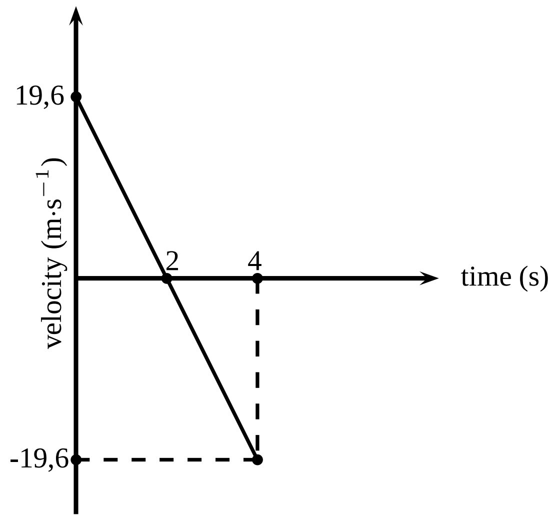
$$\begin{aligned} (v_f)^2 &= (v_i)^2 + 2g\Delta x \\ (v_f)^2 &= (0)^2 + 2(-9{,}8)(-4{,}196) \\ (v_f)^2 &= 82{,}2416 \\ v_f &= 9{,}0687...\,\text{m}\cdot\text{s}^{-1} \end{aligned}$$

The final velocity of the ball is 9,07 m·s^{-1}downwards.

Worked Example 136: Describing Projectile Motion

Question: A cricketer hits a cricket ball from the ground and the following graph of velocity vs. time was drawn. Upwards was taken as positive. Study the graph and answer the following questions:

1. Describe the motion of the ball according to the graph.
2. Draw a sketch graph of the corresponding displacement-time graph. Label the axes.
3. Draw a sketch graph of the corresponding acceleration-time graph. Label the axes.



Answer

Step 1 : Describe the motion of the ball.

We need to study the velocity-time graph to answer this question. We will break the motion of the ball up into two time zones: t = 0 s to t = 2 s and t = 2 s to t = 4 s.

From t = 0 s to t = 2 s the following happens:

The ball starts to move at an initial velocity of 19,6 m·s^{-1}and decreases its velocity to 0 m·s^{-1}at t = 2 s. At t = 2 s the velocity of the ball is 0 m·s^{-1}and therefore it stops.

From t = 2 s to t = 4 s the following happens:

The ball moves from a velocity of 0 m·s^{-1}to 19,6 m·s^{-1}in the opposite direction to the original motion.

If we assume that the ball is hit straight up in the air (and we take upwards as positive), it reaches its maximum height at t = 2 s, stops, turns around and falls back to the Earth to reach the ground at t = 4 s.

Step 2 : Draw the displacement-time graph.

To draw this graph, we need to determine the displacements at t = 2 s and t = 4 s.

At t = 2 s:

The displacement is equal to the area under the graph:

Area under graph = Area of triangle

Area = $\frac{1}{2}$bh

Area = $\frac{1}{2} \times 2 \times 19{,}6$

Displacement = 19,6 m

At t = 4 s:

The displacement is equal to the area under the whole graph (top and bottom). Remember that an area under the time line must be substracted:

Area under graph = Area of triangle 1 + Area of triangle 2

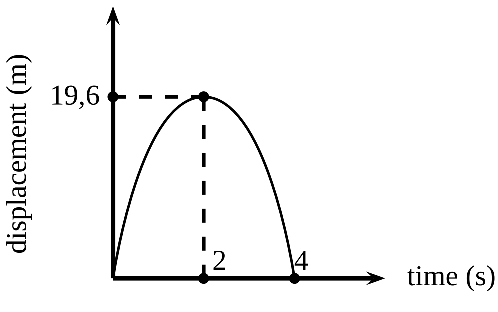
Area = $\frac{1}{2}$bh + $\frac{1}{2}$bh

Area = $(\frac{1}{2} \times 2 \times 19{,}6) + (\frac{1}{2} \times 2 \times (-19{,}6))$

Area = 19,6 - 19,6

Displacement = 0 m

The displacement-time graph for motion at constant acceleration is a curve. The graph will look like this:

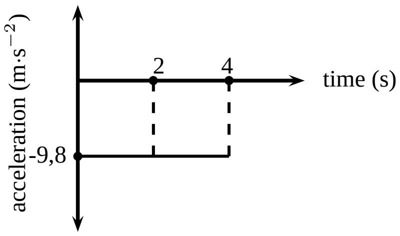


Step 3 : Draw the acceleration-time graph.

To draw the acceleration vs. time graph, we need to know what the acceleration is. The velocity-time graph is a straight line which means that the acceleration is constant. The gradient of the line will give the acceleration.

The line has a negative slope (goes down towards the left) which means that the acceleration has a negative value.

Calculate the gradient of the line:
gradient = $\frac{\Delta v}{t}$
gradient = $\frac{0-19{,}6}{2-0}$
gradient = $\frac{-19{,}6}{2}$
gradient = -9,8
acceleration = 9,8 m·s^{-2}downwards



Exercise: Graphs of Vertical Projectile Motion

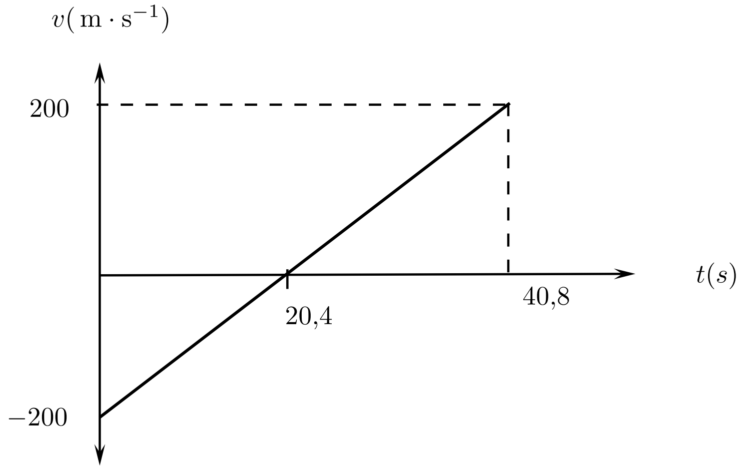
1. Amanda throws a tennisball from a height of $1{,}5m$ straight up into the air and then lets it fall to the ground. Draw graphs of Δx vs t; v vs t and a vs t for the motion of the ball. The initial velocity of the tennisball is $2\,\mathrm{m}\cdot\mathrm{s}^{-1}$. Choose upwards as positive.

2. A bullet is shot from a gun. The following graph is drawn. Downwards was chosen as positive

   a Describe the motion of the bullet

   b Draw a displacement - time graph

   c Draw a acceleration - time graph



21.3 Conservation of Momentum in Two Dimensions

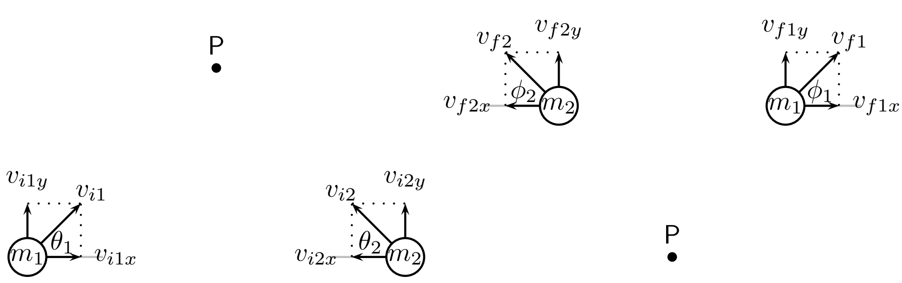
We have seen in Chapter **??** that the momentum of a system is conserved when there are no external forces acting on the system. Conversely, an external force causes a change in momentum Δp, with the impulse delivered by the force, F acting for a time Δt given by:

$$\Delta p = F \cdot \Delta t$$

The same principles that were studied in applying the conservation of momentum to problems in one dimension, can be applied to solving problems in two dimensions.

The calculation of momentum is the same in two dimensions as in one dimension. The calculation of momentum in two dimensions is broken down into determining the x and y components of momentum and applying the conservation of momentum to each set of components.

Consider two objects moving towards each other as shown in Figure 21.4. We analyse this situation by calculating the x and y components of the momentum of each object.



(a) Before the collision (b) After the collision

Figure 21.4: Two balls collide at point P.

Before the collision

Total momentum:

$$\begin{aligned} p_{i1} &= m_1 v_{i1} \\ p_{i2} &= m_2 v_{i2} \end{aligned}$$

x-component of momentum:

$$\begin{aligned} p_{i1x} &= m_1 v_{i1x} = m_1 v_{i1} \cos\theta_1 \\ p_{i2x} &= m_2 u_{i2x} = m_2 v_{i2} \sin\theta_2 \end{aligned}$$

y-component of momentum:

$$\begin{aligned} p_{i1y} &= m_1 v_{i1y} = m_1 v_{i1} \cos\theta_1 \\ p_{i2y} &= m_2 v_{i2y} = m_2 v_{i2} \sin\theta_2 \end{aligned}$$

After the collision

Total momentum:

$$\begin{aligned} p_{f1} &= m_1 v_{f1} \\ p_{f2} &= m_2 v_{f2} \end{aligned}$$

x-component of momentum:

$$\begin{aligned} p_{f1x} &= m_1 v_{f1x} = m_1 v_{f1} \cos\phi_1 \\ p_{f2x} &= m_2 v_{f2x} = m_2 v_{f2} \sin\phi_2 \end{aligned}$$

y-component of momentum:

$$\begin{aligned} p_{f1y} &= m_1 v_{f1y} = m_1 v_{f1} \cos\phi_1 \\ p_{f2y} &= m_2 v_{f2y} = m_2 v_{f2} \sin\phi_2 \end{aligned}$$

Conservation of momentum

The initial momentum is equal to the final momentum:

$$p_i = p_f$$

$$\begin{aligned} p_i &= p_{i1} + p_{i2} \\ p_f &= p_{f1} + p_{f2} \end{aligned}$$

This forms the basis of analysing momentum conservation problems in two dimensions.

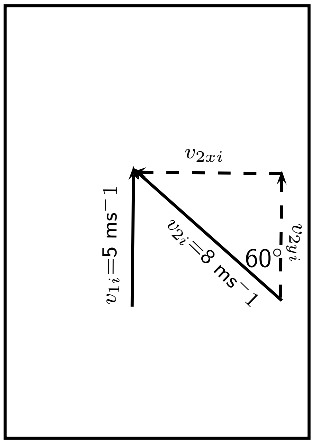
Worked Example 137: 2D Conservation of Momentum

Question: In a rugby game, Player 1 is running with the ball at 5 m·s^{-1} straight down the field parallel to the edge of the field. Player 2 runs at 6 m·s^{-1} an angle of $60°$ to the edge of the field and tackles Player 1. In the tackle, Player 2 stops completely while Player 1 bounces off Player 2. Calculate the velocity (magnitude and direction) at which Player 1 bounces off Player 2. Both the players have a mass of 90 kg.

Answer

Step 1 : Understand what is given and what is being asked

The first step is to draw the picture to work out what the situation is. Mark the initial velocities of both players in the picture.



We also know that $m_1 = m_2 = 90$ kg and $v_{f2} = 0$ ms^{-1}.
We need to find the final velocity and angle at which Player 1 bounces off Player 2.

Step 2 : Use conservation of momentum to solve the problem. First find the initial total momentum:

Total initial momentum = Total final momentum. But we have a two dimensional problem, and we need to break up the initial momentum into x and y components.

$$\begin{aligned} p_{ix} &= p_{fx} \\ p_{iy} &= p_{fy} \end{aligned}$$

For Player 1:

$$\begin{aligned} p_{ix1} &= m_1 v_{i1x} = 90 \times 0 = 0 \\ p_{iy1} &= m_1 v_{i1y} = 90 \times 5 \end{aligned}$$

For Player 2:

$$\begin{aligned} p_{ix2} &= m_2 v_{i2x} = 90 \times 8 \times \sin 60^\circ \\ p_{iy2} &= m_2 v_{i2y} = 90 \times 8 \times \cos 60^\circ \end{aligned}$$

Step 3 : Now write down what we know about the final momentum:

For Player 1:

$$\begin{aligned} p_{fx1} &= m_1 v_{fx1} = 90 \times v_{fx1} \\ p_{fy1} &= m_1 v_{fy1} = 90 \times v_{fy1} \end{aligned}$$

For Player 2:

$$\begin{aligned} p_{fx2} &= m_2 v_{fx2} = 90 \times 0 = 0 \\ p_{fy2} &= m_2 v_{fy2} = 90 \times 0 = 0 \end{aligned}$$

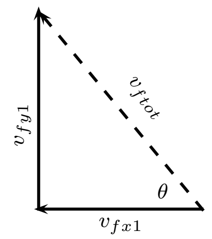
Step 4 : Use conservation of momentum:

The initial total momentum in the x direction is equal to the final total momentum in the x direction.
The initial total momentum in the y direction is equal to the final total momentum in the y direction.
If we find the final x and y components, then we can find the final $total$ momentum.

$$\begin{aligned} p_{ix1} + p_{ix2} &= p_{fx1} + p_{fx2} \\ 0 + 90 \times 8 \times \sin 60^\circ &= 90 \times v_{fx1} + 0 \\ v_{fx1} &= \frac{90 \times 8 \times \sin 60^\circ}{90} \\ v_{fx1} &= 6.928 \text{ms}^{-1} \end{aligned}$$

$$
\begin{aligned}
p_{iy1} + p_{iy2} &= p_{fy1} + p_{fy2} \\
90 \times 5 + 90 \times 8 \times \cos 60^\circ &= 90 \times v_{fy1} + 0 \\
v_{fy1} &= \frac{90 \times 5 + 90 \times 8 \times \cos 60^\circ}{90} \\
v_{fy1} &= 9.0\text{ms}^{-1}
\end{aligned}
$$

Step 5 : Using the x and y components, calculate the final total v
Use Pythagoras's theorem to find the total final velocity:



$$
\begin{aligned}
v_{ftot} &= \sqrt{v_{fx1}^2 + v_{fx2}^2} \\
&= \sqrt{6.928^2 + 9^2} \\
&= 11.36
\end{aligned}
$$

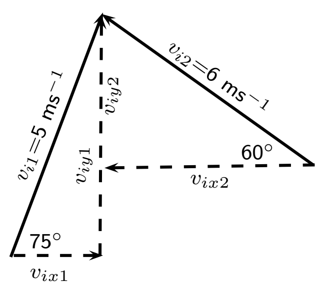
Calculate the angle θ to find the direction of Player 1's final velocity:

$$
\begin{aligned}
\sin\theta &= \frac{v_{fxy1}}{v_{ftot}} \\
\theta &= 52.4^\circ
\end{aligned}
$$

Therefore Player 1 bounces off Player 2 with a final velocity of 11.36 m·s^{-1} at an angle of 52.4° from the horizontal.

Worked Example 138: 2D Conservation of Momentum: II

Question: In a soccer game, Player 1 is running with the ball at 5 m·s^{-1} across the pitch at an angle of 75° from the horizontal. Player 2 runs towards Player 1 at 6 m·s^{-1} an angle of 60° to the horizontal and tackles Player 1. In the tackle, the two players bounce off each other. Player 2 moves off with a velocity in the opposite x-direction of 0.3 m·s^{-1} and a velocity in the y-direction of 6 m·s^{-1}. Both the players have a mass of 80 kg. What is the final total velocity of Player 1?
Answer
Step 1 : Understand what is given and what is being asked
The first step is to draw the picture to work out what the situation is. Mark the initial velocities of both players in the picture.



We also know that $m_1 = m_2 = 80$ kg. And v_{fx2}=-0.3 ms^{-1} and v_{fy2}=6 ms^{-1}.
We need to find the final velocity and angle at which Player 1 bounces off Player 2.

Step 2 : Use conservation of momentum to solve the problem. First find the initial total momentum:

Total initial momentum = Total final momentum. But we have a two dimensional problem, and we need to break up the initial momentum into x and y components.

$$\begin{aligned} p_{ix} &= p_{fx} \\ p_{iy} &= p_{fy} \end{aligned}$$

For Player 1:

$$\begin{aligned} p_{ix1} &= m_1 v_{i1x} = 80 \times 5 \times \cos 75° \\ p_{iy1} &= m_1 v_{i1y} = 80 \times 5 \times \sin 75° \end{aligned}$$

For Player 2:

$$\begin{aligned} p_{ix2} &= m_2 v_{i2x} = 80 \times 6 \times \cos 60° \\ p_{iy2} &= m_2 v_{i2y} = 80 \times 6 \times \sin 60° \end{aligned}$$

Step 3 : Now write down what we know about the final momentum:

For Player 1:

$$\begin{aligned} p_{fx1} &= m_1 v_{fx1} = 80 \times v_{fx1} \\ p_{fy1} &= m_1 v_{fy1} = 80 \times v_{fy1} \end{aligned}$$

For Player 2:

$$\begin{aligned} p_{fx2} &= m_2 v_{fx2} = 80 \times (-0.3) \times \cos 60° \\ p_{fy2} &= m_2 v_{fy2} = 80 \times 6 \times \sin 60° \end{aligned}$$

Step 4 : Use conservation of momentum:

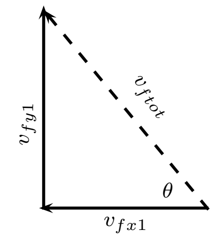
The initial total momentum in the x direction is equal to the final total momentum in the x direction.
The initial total momentum in the y direction is equal to the final total momentum in the y direction.
If we find the final x and y components, then we can find the final $total$ momentum.

$$\begin{aligned} p_{ix1} + p_{ix2} &= p_{fx1} + p_{fx2} \\ 80 \times 5 \cos 75° + 80 \times \cos 60° &= 80 \times v_{fx1} + 80 \times (-0.3) \\ v_{fx1} &= \frac{80 \times 5 \cos 75° + 80 \times \cos 60° + 80 \times (-0.3)}{80} \\ v_{fx1} &= 2.0 \text{ms}^{-1} \end{aligned}$$

$$\begin{aligned} p_{iy1} + p_{iy2} &= p_{fy1} + p_{fy2} \\ 80 \times 5 \sin 75° + 80 \times \sin 60° &= 80 \times v_{fy1} + 80 \times 6 \\ v_{fy1} &= \frac{80 \times 5 \sin 75° + 80 \times \sin 60° - 80 \times 6}{80} \\ v_{fy1} &= 4.0 \text{ms}^{-1} \end{aligned}$$

Step 5 : Using the x and y components, calculate the final total v

Use Pythagoras's theorem to find the total final velocity:



$$\begin{aligned} v_{ftot} &= \sqrt{v_{fx1}^2 + v_{fx2}^2} \\ &= \sqrt{2^2 + 4^2} \\ &= 4.5 \end{aligned}$$

Calculate the angle θ to find the direction of Player 1's final velocity:

$$\begin{aligned} \tan\theta &= \frac{v_{fy1}}{v_{fx1}} \\ \theta &= 26.6^\circ \end{aligned}$$

Therefore Player 1 bounces off Player 2 with a final velocity of 4.5 m·s^{-1} at an angle of 26.6° from the horizontal.

21.4 Types of Collisions

Two types of collisions are of interest:

- elastic collisions
- inelastic collisions

In both types of collision, total momentum is *always* conserved. Kinetic energy is conserved for elastic collisions, but not for inelastic collisions.

21.4.1 Elastic Collisions

Definition: Elastic Collisions
An elastic collision is a collision where total momentum and total kinetic energy are both conserved.

This means that in an elastic collision the total momentum *and* the total kinetic energy before the collision is the same as after the collision. For these kinds of collisions, the kinetic energy is not changed into another type of energy.

Before the Collision

Figure 21.5 shows two balls rolling toward each other, about to collide:

Before the balls collide, the total momentum of the system is equal to all the individual momenta added together. Ball 1 has a momentum which we call p_{i1} and ball 2 has a momentum which we call p_{i2}, it means the total momentum before the collision is:

$$p_i = p_{i1} + p_{i2}$$

Figure 21.5: Two balls before they collide.

We calculate the total kinetic energy of the system in the same way. Ball 1 has a kinetic energy which we call KE_{i1} and the ball 2 has a kinetic energy which we call KE_{i2}, it means that the total kinetic energy before the collision is:

$$KE_i = KE_{i1} + KE_{i2}$$

After the Collision

Figure 21.6 shows two balls after they have collided:

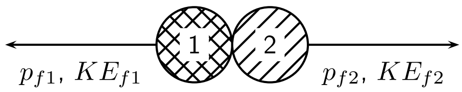


Figure 21.6: Two balls after they collide.

After the balls collide and bounce off each other, they have new momenta and new kinetic energies. Like before, the total momentum of the system is equal to all the individual momenta added together. Ball 1 now has a momentum which we call p_{f1} and ball 2 now has a momentum which we call p_{f2}, it means the total momentum after the collision is

$$p_f = p_{f1} + p_{f2}$$

Ball 1 now has a kinetic energy which we call KE_{f1} and ball 2 now has a kinetic energy which we call KE_{f2}, it means that the total kinetic energy after the collision is:

$$KE_f = KE_{f1} + KE_{f2}$$

Since this is an *elastic* collision, the total momentum before the collision equals the total momentum after the collision **and** the total kinetic energy before the collision equals the total kinetic energy after the collision. Therefore:

$$\begin{array}{rcl}
\text{Initial} & & \text{Final} \\
p_i & = & p_f \qquad (21.5) \\
p_{i1} + p_{i2} & = & p_{f1} + p_{f2} \\
 & \textbf{and} & \\
KE_i & = & KE_f \qquad (21.6) \\
KE_{i1} + KE_{i2} & = & KE_{f1} + KE_{f2}
\end{array}$$

Worked Example 139: An Elastic Collision

Question: Consider a collision between two pool balls. Ball 1 is at rest and ball 2 is moving towards it with a speed of 2 m·s^{-1}. The mass of each ball is 0.3 kg. After the balls collide *elastically*, ball 2 comes to an immediate stop and ball 1 moves off. What is the final velocity of ball 1?

Answer

Step 1 : Determine how to approach the problem

We are given:

- mass of ball 1, $m_1 = 0.3$ kg
- mass of ball 2, $m_2 = 0.3$ kg
- initial velocity of ball 1, $v_{i1} = 0$ m·s^{-1}
- initial velocity of ball 2, $v_{i2} = 2$ m·s^{-1}
- final velocity of ball 2, $v_{f2} = 0$ m·s^{-1}
- the collision is elastic

All quantities are in SI units. We are required to determine the final velocity of ball 1, v_{f1}. Since the collision is elastic, we know that

- momentum is conserved, $m_1 v_{i1} + m_2 v_{i2} = m_1 v_{f1} + m_2 v_{f2}$
- energy is conserved, $\frac{1}{2}(m_1 v_{i1}^2 + m_2 v_{i2}^2 = m_1 v_{f1}^2 + m_2 v_{f2}^2)$

Step 2 : Choose a frame of reference

Choose to the right as positive.

Step 3 : Draw a rough sketch of the situation

Before collision After collision

Step 4 : Solve the problem

Momentum is conserved. Therefore:

$$\begin{aligned} p_i &= p_f \\ m_1 v_{i1} + m_2 v_{i2} &= m_1 v_{f1} + m_2 v_{f2} \\ (0{,}3)(0) + (0{,}3)(2) &= (0{,}3) v_{f1} + 0 \\ v_{f1} &= 2\,\mathrm{m \cdot s^{-1}} \end{aligned}$$

Step 5 : Quote the final answer

The final velocity of ball 1 is 2 m·s^{-1}in the same direction as ball 2.

Worked Example 140: Another Elastic Collision

Question: Consider two 2 marbles. Marble 1 has mass 50 g and marble 2 has mass 100 g. Edward rolls marble 2 along the ground towards marble 1 in the positive x-direction. Marble 1 is initially at rest and marble 2 has a velocity of 3 m·s^{-1} in the positive x-direction. After they collide *elastically*, both marbles are moving. What is the final velocity of each marble?

Answer

Step 1 : Decide how to approach the problem

We are given:

- mass of marble 1, m_1=50 g
- mass of marble 2, m_2=100 g
- initial velocity of marble 1, v_{i1}=0 m·s^{-1}
- initial velocity of marble 2, v_{i2}=3 m·s^{-1}
- the collision is elastic

The masses need to be converted to SI units.

$$\begin{aligned} m_1 &= 0{,}05 \text{ kg} \\ m_2 &= 0{,}1 \text{ kg} \end{aligned}$$

We are required to determine the final velocities:

- final velocity of marble 1, v_{f1}
- final velocity of marble 2, v_{f2}

Since the collision is elastic, we know that

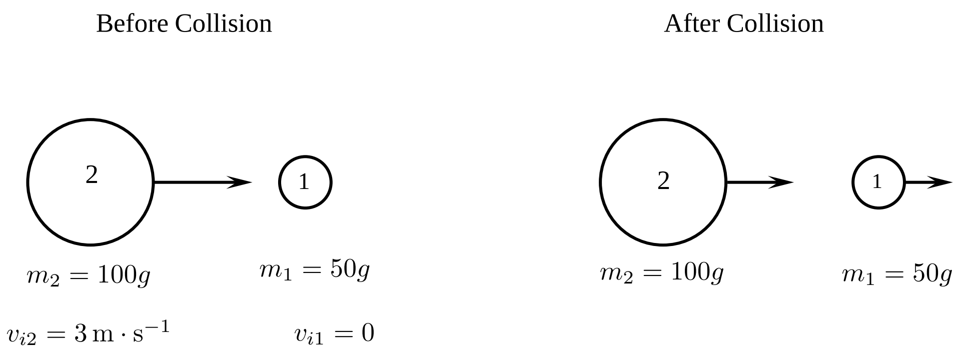
- momentum is conserved, $p_i = p_f$.
- energy is conserved, KE_i=KE_f.

We have two equations and two unknowns (v_1, v_2) so it is a simple case of solving a set of simultaneous equations.

Step 2 : Choose a frame of reference

Choose to the right as positive.

Step 3 : Draw a rough sketch of the situation



Step 4 : Solve problem

Momentum is conserved. Therefore:

$$\begin{aligned} p_i &= p_f \\ p_{i1} + p_{i2} &= p_{f1} + p_{f2} \\ m_1 v_{i1} + m_2 v_{i2} &= m_1 v_{f1} + m_2 v_{f2} \\ (0{,}05)(0) + (0{,}1)(3) &= (0{,}05) v_{f1} + (0{,}1) v_{f2} \\ 0{,}3 &= 0{,}05 v_{f1} + 0{,}1 v_{f2} \end{aligned} \tag{21.7}$$

Energy is also conserved. Therefore:

$$\begin{aligned} KE_i &= KE_f \\ KE_{i1} + KE_{i2} &= KE_{f1} + KE_{f2} \\ \frac{1}{2} m_1 v_{i1}^2 + \frac{1}{2} m_2 v_{i2}^2 &= \frac{1}{2} m_1 v_{f1}^2 + \frac{1}{2} m_2 v_{f2}^2 \\ (\frac{1}{2})(0{,}05)(0)^2 + (\frac{1}{2})(0{,}1)(3)^2 &= \frac{1}{2}(0{,}05)(v_{f1})^2 + (\frac{1}{2})(0{,}1)(v_{f2})^2 \\ 0{,}45 &= 0{,}025 v_{f1}^2 + 0{,}05 v_{f2}^2 \end{aligned} \tag{21.8}$$

Substitute Equation 21.7 into Equation 21.8 and solve for v_{f2}.

$$
\begin{aligned}
m_2 v_{i2}^2 &= m_1 v_{f1}^2 + m_2 v_{f2}^2 \\
&= m_1 \left(\frac{m_2}{m_1}(v_{i2} - v_{f2}) \right)^2 + m_2 v_{f2}^2 \\
&= m_1 \frac{m_2^2}{m_1^2} (v_{i2} - v_{f2})^2 + m_2 v_{f2}^2 \\
&= \frac{m_2^2}{m_1} (v_{i2} - v_{f2})^2 + m_2 v_{f2}^2 \\
v_{i2}^2 &= \frac{m_2}{m_1} (v_{i2} - v_{f2})^2 + v_{f2}^2 \\
&= \frac{m_2}{m_1} \left(v_{i2}^2 - 2 \cdot v_{i2} \cdot v_{f2} + v_{f2}^2 \right) + v_{f2}^2 \\
0 &= \left(\frac{m_2}{m_1} - 1 \right) v_{i2}^2 - 2\frac{m_2}{m_1} v_{i2} \cdot v_{f2} + \left(\frac{m_2}{m_1} + 1 \right) v_{f2}^2 \\
&= \left(\frac{0.1}{0.05} - 1 \right) (3)^2 - 2\frac{0.1}{0.05}(3) \cdot v_{f2} + \left(\frac{0.1}{0.05} + 1 \right) v_{f2}^2 \\
&= (2-1)(3)^2 - 2 \cdot 2(3) \cdot v_{f2} + (2+1) v_{f2}^2 \\
&= 9 - 12 v_{f2} + 3 v_{f2}^2 \\
&= 3 - 4 v_{f2} + v_{f2}^2 \\
&= (v_{f2} - 3)(v_{f2} - 1)
\end{aligned}
$$

Substituting back into Equation 21.7, we get:

$$
\begin{aligned}
v_{f1} &= \frac{m_2}{m_1}(v_{i2} - v_{f2}) \\
&= \frac{0.1}{0.05}(3-3) \\
&= 0 \ \mathrm{m \cdot s^{-1}} \\
&\text{or} \\
v_{f1} &= \frac{m_2}{m_1}(v_{i2} - v_{f2}) \\
&= \frac{0.1}{0.05}(3-1) \\
&= 4 \ \mathrm{m \cdot s^{-1}}
\end{aligned}
$$

But according to the question, ball 1 is moving after the collision, therefore ball 1 moves to the right at 4 m·s^{-1}and ball 2 moves to the left with a velocity of 1 m·s^{-1}.

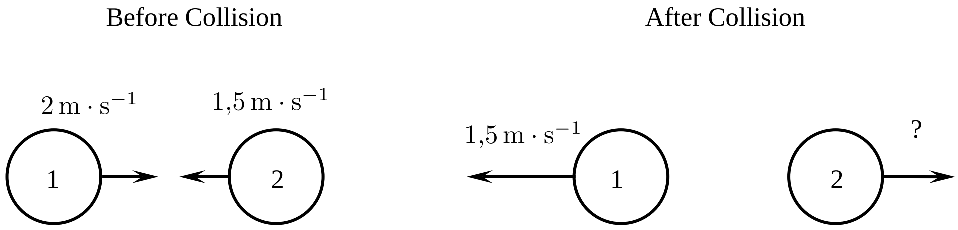
Worked Example 141: Colliding Billiard Balls

Question: Two billiard balls each with a mass of $150g$ collide head-on in an elastic collision. Ball 1 was travelling at a speed of $2\,\mathrm{m \cdot s^{-1}}$ and ball 2 at a speed of $1{,}5\,\mathrm{m \cdot s^{-1}}$. After the collision, ball 1 travels away from ball 2 at a velocity of $1{,}5\,\mathrm{m \cdot s^{-1}}$.

1. Calculate the velocity of ball 2 after the collision.
2. Prove that the collision was elastic. Show calculations.

Answer

1. **Step 1 : Draw a rough sketch of the situation**



Step 2 : Decide how to approach the problem
Since momentum is conserved in all kinds of collisions, we can use conservation of momentum to solve for the velocity of ball 2 after the collision.
Step 3 : Solve problem

$$\begin{aligned}
p_{before} &= p_{after} \\
m_1 v_{i1} + m_2 v_{i2} &= m_1 v_{f1} + m_2 v_{f2} \\
\left(\frac{150}{1000}\right)(2) + \left(\frac{150}{1000}\right)(-1{,}5) &= \left(\frac{150}{1000}\right)(-1{,}5) + \left(\frac{150}{1000}\right)(v_{f2}) \\
0{,}3 - 0{,}225 &= -0{,}225 + 0{,}15 v_{f2} \\
v_{f2} &= 3\,\mathrm{m}\cdot\mathrm{s}^{-1}
\end{aligned}$$

So after the collision, ball 2 moves with a velocity of $3\,\mathrm{m}\cdot\mathrm{s}^{-1}$.

2. The fact that characterises an elastic collision is that the total kinetic energy of the particles before the collision is the same as the total kinetic energy of the particles after the collision. This means that if we can show that the initial kinetic energy is equal to the final kinetic energy, we have shown that the collision is elastic.
Calculating the initial total kinetic energy:

$$\begin{aligned}
EK_{before} &= \frac{1}{2} m_1 v_{i1}^2 + \frac{1}{2} m_2 v_{i2}^2 \\
&= \left(\frac{1}{2}\right)(0{,}15)(2)^2 + \left(\frac{1}{2}\right)(0{,}15)(-1{,}5)^2 \\
&= 0.469....J
\end{aligned}$$

Calculating the final total kinetic energy:

$$\begin{aligned}
EK_{after} &= \frac{1}{2} m_1 v_{f1}^2 + \frac{1}{2} m_2 v_{f2}^2 \\
&= \left(\frac{1}{2}\right)(0{,}15)(-1{,}5)^2 + \left(\frac{1}{2}\right)(0{,}15)(2)^2 \\
&= 0.469....J
\end{aligned}$$

So $EK_{before} = EK_{after}$ and hence the collision is elastic.

21.4.2 Inelastic Collisions

Definition: Inelastic Collisions
An inelastic collision is a collision in which **total momentum** is conserved but **total kinetic energy** is *not* conserved. The kinetic energy is *transformed* into other kinds of energy.

So the total momentum before an inelastic collisions is the same as after the collision. But the total *kinetic* energy before and after the inelastic collision is *different*. Of course this does not mean that total energy has not been conserved, rather the energy has been *transformed* into another type of energy.

As a rule of thumb, inelastic collisions happen when the colliding objects are distorted in some way. Usually they change their shape. The modification of the shape of an object requires energy and this is where the "missing" kinetic energy goes. A classic example of an inelastic collision is a motor car accident. The cars change shape and there is a noticeable change in the kinetic energy of the cars before and after the collision. This energy was used to bend the metal and deform the cars. Another example of an inelastic collision is shown in Figure 21.7.

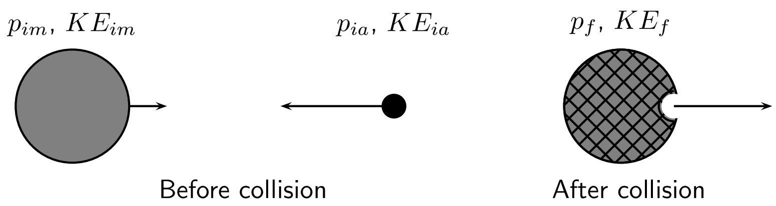


Figure 21.7: Asteroid moving towards the Moon.

An asteroid is moving through space towards the Moon. Before the asteroid crashes into the Moon, the total momentum of the system is:

$$p_i = p_{im} + p_{ia}$$

The total kinetic energy of the system is:

$$KE_i = KE_{im} + KE_{ia}$$

When the asteroid collides **inelastically** with the Moon, its kinetic energy is transformed mostly into heat energy. If this heat energy is large enough, it can cause the asteroid and the area of the Moon's surface that it hits, to melt into liquid rock! From the force of impact of the asteroid, the molten rock flows outwards to form a crater on the Moon.

After the collision, the total momentum of the system will be the same as before. But since this collision is *inelastic*, (and you can see that a change in the shape of objects has taken place!), total kinetic energy is **not** the same as before the collision.
Momentum is conserved:

$$p_i = p_f$$

But the total kinetic energy of the system is not conserved:

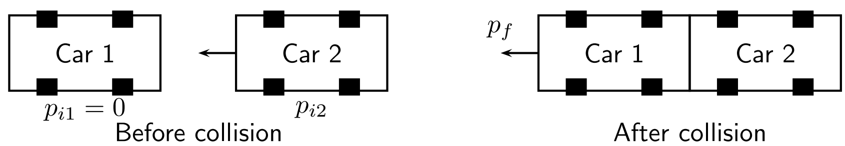
$$KE_i \neq KE_f$$

Worked Example 142: An Inelastic Collision

Question: Consider the collision of two cars. Car 1 is at rest and Car 2 is moving at a speed of 2 m·s^{-1} in the negative x-direction. Both cars each have a mass of 500 kg. The cars collide *inelastically* and stick together. What is the resulting velocity of the resulting mass of metal?

Answer

Step 1 : Draw a rough sketch of the situation



Step 2 : Determine how to approach the problem

We are given:

- mass of car 1, $m_1 = 500$ kg
- mass of car 2, $m_2 = 500$ kg
- initial velocity of car 1, $v_{i1} = 0$ m·s^{-1}
- initial velocity of car 2, $v_{i2} = 2$ m·s^{-1}to the left
- the collision is inelastic

All quantities are in SI units. We are required to determine the final velocity of the resulting mass, v_f.
Since the collision is inelastic, we know that

- momentum is conserved, $m_1 v_{i1} + m_2 v_{i2} = m_1 v_{f1} + m_2 v_{f2}$
- kinetic energy is not conserved

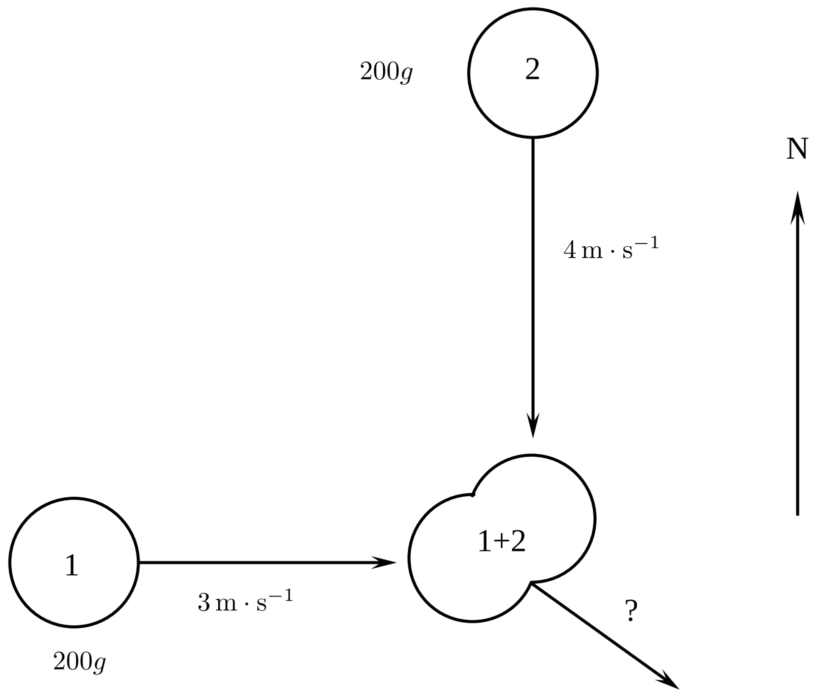
Step 3 : Choose a frame of reference
Choose to the left as positive.
Step 4 : Solve problem
So we must use conservation of momentum to solve this problem.

$$\begin{aligned} p_i &= p_f \\ p_{i1} + p_{i2} &= p_f \\ m_1 v_{i1} + m_2 v_{i2} &= (m_1 + m_2) v_f \\ (500)(0) + (500)(2) &= (500 + 500) v_f \\ 1000 &= 1000 v_f \\ v_f &= 1 \ \mathrm{m \cdot s^{-1}} \end{aligned}$$

Therefore, the final velocity of the resulting mass of cars is 1 m·s^{-1}to the left.

Worked Example 143: Colliding balls of clay

Question: Two balls of clay, $200g$ each, are thrown towards each other according to the following diagram. When they collide, they stick together and move off together. All motion is taking place in the horizontal plane. Determine the velocity of the clay after the collision.



Answer

Step 1 : Analyse the problem

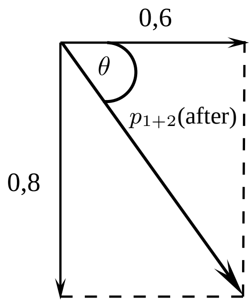
This is an inelastic collision where momentum is conserved.
The momentum before = the momentum after.
The momentum after can be calculated by drawing a vector diagram.

Step 2 : Calculate the momentum before the collision

$$p_1(\text{before}) = \mathrm{m_1 v_{i1}} = (0{,}2)(3) = 0{,}6\ \text{kg} \cdot \mathrm{m{\cdot}s^{-1}}\text{east}$$
$$p_2(\text{before}) = \mathrm{m_2 v_{i2}} = (0{,}2)(4) = 0{,}8\ \text{kg} \cdot \mathrm{m{\cdot}s^{-1}}\text{south}$$

Step 3 : Calculate the momentum after the collision.

Here we need to draw a diagram:



$$\begin{aligned} p_{1+2}(\text{after}) &= \sqrt{(0{,}8)^2 + (0{,}6)^2} \\ &= 1 \end{aligned}$$

Step 4 : Calculate the final velocity

First we have to find the direction of the final momentum:

$$\begin{aligned} \tan\theta &= \frac{0{,}8}{0{,}6} \\ \theta &= 53{,}13° \end{aligned}$$

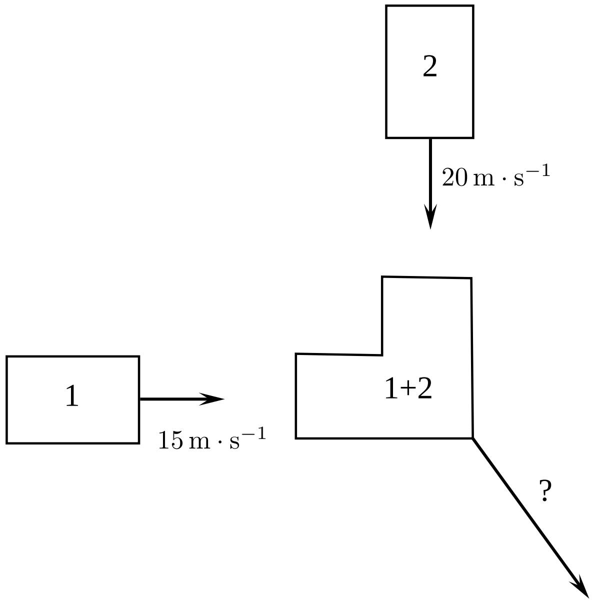
Now we have to find the magnitude of the final velocity:

$$\begin{aligned} p_{1+2} &= m_{1+2} v_f \\ 1 &= (0{,}2 + 0{,}2) v_f \\ v_f &= 2{,}5\,\mathrm{m \cdot s^{-1}}\mathrm{E}53{,}13°\mathrm{S} \end{aligned}$$

?

Exercise: Collisions

1. A truck of mass 4500 kg travelling at 20 m·s^{-1}hits a car from behind. The car (mass 1000 kg) was travelling at 15 m·s^{-1}. The two vehicles, now connected carry on moving in the same direction.
   a Calculate the final velocity of the truck-car combination after the collision.
   b Determine the kinetic energy of the system before and after the collision.
   c Explain the difference in your answers for b).
   d Was this an example of an elastic or inelastic collision? Give reasons for your answer.
2. Two cars of mass 900 kg each collide and stick together at an angle of 90°. Determine the final velocity of the cars if
   car 1 was travelling at 15m·s^{-1}and
   car 2 was travelling at 20m·s^{-1}.



Extension: Tiny, Violent Collisions

Author: Thomas D. Gutierrez

Tom Gutierrez received his Bachelor of Science and Master degrees in Physics from San Jose State University in his home town of San Jose, California. As a Master's student he helped work on a laser spectrometer at NASA Ames Research Centre. The instrument measured the ratio of different isotopes of carbon in CO_2 gas and could be used for such diverse applications as medical diagnostics and space exploration. Later, he received his PhD in physics from the University of California, Davis where he performed calculations for various reactions in high energy physics collisions. He currently lives in Berkeley, California where he studies proton-proton collisions seen at the STAR experiment at Brookhaven National Laboratory on Long Island, New York.

High Energy Collisions

Take an orange and expand it to the size of the earth. The atoms of the earth-sized orange would themselves be about the size of regular oranges and would fill the entire "earth-orange". Now, take an atom and expand it to the size of a football field. The nucleus of that atom would be about the size of a tiny seed in the middle of the field. From this analogy, you can see that atomic nuclei are very small objects by human standards. They are roughly 10^{-15} meters in diameter – one-hundred thousand times smaller than a typical atom. These nuclei cannot be seen or studied via any conventional means such as the naked eye or microscopes. So how do scientists study the structure of very small objects like atomic nuclei?

The simplest nucleus, that of hydrogen, is called the proton. Faced with the inability to isolate a single proton, open it up, and directly examine what is inside, scientists must resort to a brute-force and somewhat indirect means of exploration: high energy collisions. By colliding protons with other particles (such as other protons or electrons) at very high energies, one hopes to learn about what they are made of and how they work. The American physicist Richard Feynman once compared this process to slamming delicate watches together and figuring out how they work by only examining the broken debris. While this analogy may seem pessimistic, with sufficient mathematical models and experimental precision, considerable information can be extracted from the debris of such high energy subatomic collisions. One can learn about both the nature of the forces at work and also about the sub-structure of such systems.

The experiments are in the category of “high energy physics” (also known as “subatomic” physics). The primary tool of scientific exploration in these experiments is an extremely violent collision between two very, very small subatomic objects such as nuclei. As a general rule, the higher the energy of the collisions, the more detail of the original system you are able to resolve. These experiments are operated at laboratories such as CERN, SLAC, BNL, and Fermilab, just to name a few. The giant machines that perform the collisions are roughly the size of towns. For example, the RHIC collider at BNL is a ring about 1 km in diameter and can be seen from space. The newest machine currently being built, the LHC at CERN, is a ring 9 km in diameter!

Activity :: Casestudy : Atoms and its Constituents
Questions:

1. What are isotopes? (2)
2. What are atoms made up of? (3)
3. Why do you think protons are used in the experiments and not atoms like carbon? (2)
4. Why do you think it is necessary to find out what atoms are made up of and how they behave during collisions? (2)
5. Two protons (mass $1{,}67 \times 10^{-27}$ kg) collide and somehow stick together after the collision. If each proton travelled with an initial velocity of $5{,}00 \times 10^{7}\,\mathrm{m \cdot s^{-1}}$ and they collided at an angle of 90°, what is the velocity of the combination after the collision. (9)

21.5 Frames of Reference

21.5.1 Introduction

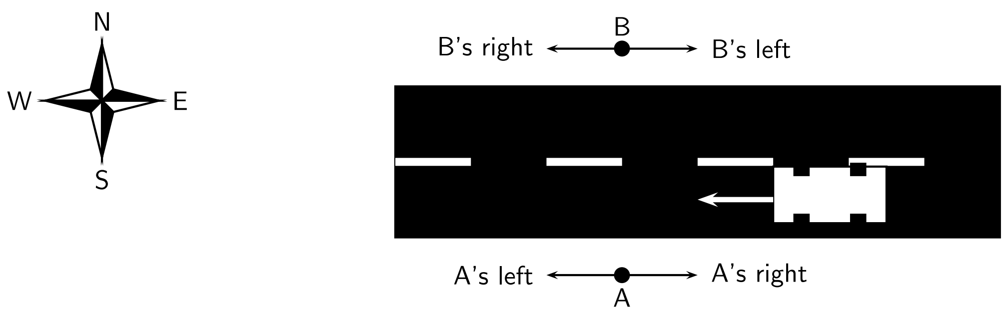


Figure 21.8: Top view of a road with two people standing on opposite sides. A car drives past.

Consider two people standing, facing each other on either side of a road. A car drives past them, heading West. For the person facing South, the car was moving toward the right. However, for the person facing North, the car was moving toward the left. This discrepancy is due to the fact that the two people used two different *frames of reference* from which to investigate this system. If each person were asked in what direction the car were moving, they would give a different answer. The answer would be relative to their frame of reference.

21.5.2 What is a frame of reference?

Definition: Frame of Reference
A *frame of reference* is the point of view from which a system is observed.

In practical terms, a frame of reference is a set of axes (specifying directions) with an origin. An observer can then measure the position and motion of all points in a system, as well as the orientation of objects in the system relative to the frame of reference.

There are two types of reference frames: inertial and non-inertial. An inertial frame of reference travels at a constant velocity, which means that Newton's first law (inertia) holds true. A non-inertial frame of reference, such as a moving car or a rotating carousel, accelerates. Therefore, Newton's first law does not hold true in a non-inertial reference frame, as objects appear to accelerate without the appropriate forces.

21.5.3 Why are frames of reference important?

Frames of reference are important because (as we have seen in the introductory example) the velocity of a car can differ depending on which frame of reference is used.

Extension: Frames of Reference and Special Relativity
Frames of reference are especially important in special relativity, because when a frame of reference is moving at some significant fraction of the speed of light, then the flow of time in that frame does not necessarily apply in another reference frame. The speed of light is considered to be the only true constant between moving frames of reference.

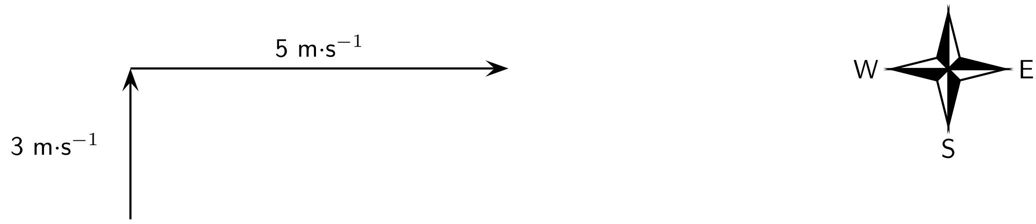
The next worked example will explain this.

21.5.4 Relative Velocity

The velocity of an object is frame dependent. More specifically, the perceived velocity of an object depends on the velocity of the observer. For example, a person standing on shore would observe the velocity of a boat to be different than a passenger on the boat.

Worked Example 144: Relative Velocity

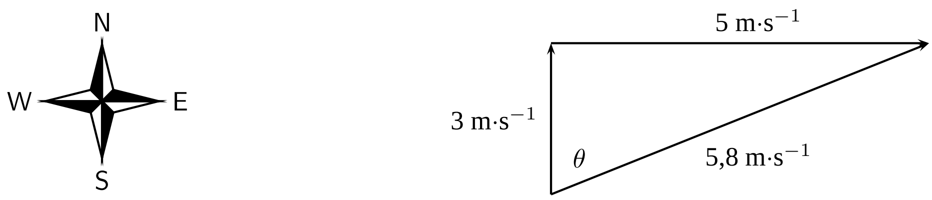
Question: The speedometer of a motor boat reads 5 m·s^{-1}. The boat is moving East across a river which has a current traveling 3 m·s^{-1} North. What would the velocity of the motor boat be according to an observer on the shore?
Answer
Step 1 : First, draw a diagram showing the velocities involved.



Step 2 : Use the Theorem of Pythagoras to solve for the resultant of the two velocities.

$$
\begin{aligned}
R &= \sqrt{(3)^2 + (5)^2} \\
&= \sqrt{34} \\
&= 5{,}8 \ \mathrm{m \cdot s^{-1}}
\end{aligned}
$$

$$
\begin{aligned}
\tan\theta &= \frac{5}{3} \\
\theta &= 59{,}04°
\end{aligned}
$$



The observer on the shore sees the boat moving with a velocity of 5,8 m·s^{-1} at N59,04°E due to the current pushing the boat perpendicular to its velocity. This is contrary to the perspective of a passenger on the boat who perceives the velocity of the boat to be 5 m·s^{-1} due East. Both perspectives are correct as long as the frame of the observer is considered.

Extension:

Worked Example 145: Relative Velocity 2

Question: It takes a man 10 seconds to ride down an escalator. It takes the same man 15 s to walk back up the escalator against its motion. How long will it take the man to walk down the escalator at the same rate he was walking before?

Answer

Step 1 : Determine what is required and what is given

We are required to determine the time taken for a man to walk down an escalator, with its motion.

We are given the time taken for the man to ride down the escalator and the time taken for the man to walk up the escalator, against it motion.

Step 2 : Determine how to approach the problem

Select down as positive and assume that the escalator moves at a velocity v_e. If the distance of the escalator is x_e then:

$$v_e = \frac{x_e}{10\,\mathrm{s}} \tag{21.9}$$

Now, assume that the man walks at a velocity v_m. Then we have that:

$$v_e - v_m = \frac{x_e}{15\,\mathrm{s}} \tag{21.10}$$

We are required to find t in:

$$v_e + v_m = \frac{x_e}{t} \tag{21.11}$$

Step 3 : Solve the problem

We find that we have three equations and three unknowns (v_e, v_m and t).

Add (21.10) to (21.11) to get:

$$2v_e = \frac{x_e}{15\,\mathrm{s}} + \frac{x_e}{t}$$

Substitute from (21.9) to get:

$$2\frac{x_e}{10\,\mathrm{s}} = \frac{x_e}{15\,\mathrm{s}} + \frac{x_e}{t}$$

Since x_e is not equal to zero we can divide throughout by x_e.

$$\frac{2}{10\,\mathrm{s}} = \frac{1}{15\,\mathrm{s}} + \frac{1}{t}$$

Re-write:

$$\frac{2}{10\,\mathrm{s}} - \frac{1}{15\,\mathrm{s}} = \frac{1}{t}$$

Multiply by t:

$$t(\frac{2}{10\,\mathrm{s}} - \frac{1}{15\,\mathrm{s}}) = 1$$

Solve for t:

$$t = \frac{1}{\frac{2}{10\,\mathrm{s}} - \frac{1}{15\,\mathrm{s}}}$$

to get:

$$t = \frac{2}{15}\,\mathrm{s}$$

Step 4 : Write the final answer

The man will take $\frac{1}{15}\,\mathrm{s} + \frac{2}{15}\,\mathrm{s} = \frac{1}{5}\,\mathrm{s}$.

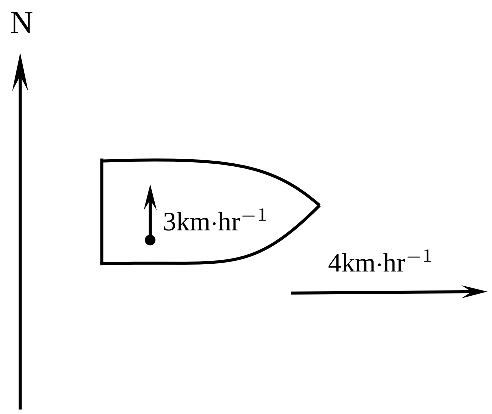
?

Exercise: Frames of Reference

1. A woman walks north at 3 km·hr^{-1} on a boat that is moving east at 4 km·hr^{-1}. This situation is illustrated in the diagram below.

   A How fast is the woman moving according to her friend who is also on the boat?

   B What is the woman's velocity according to an observer watching from the river bank?



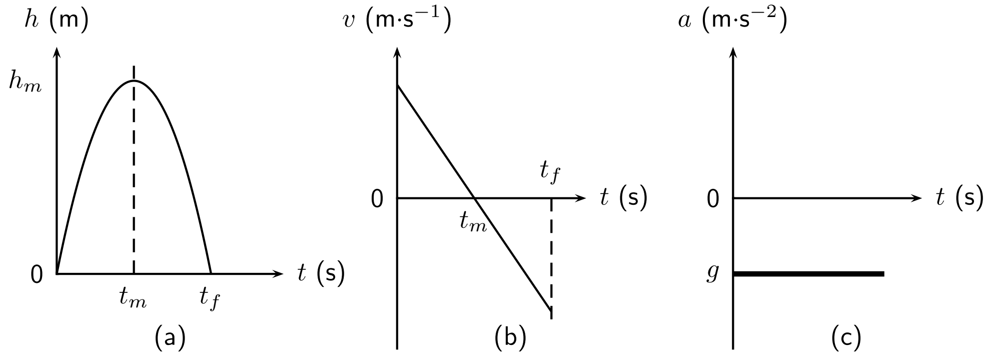
2. A boy is standing inside a train that is moving at 10 m·s^{-1}to the left. The boy throws a ball in the air with a velocity of 4 m·s^{-1}. What is the resultant velocity of the ball
   A according to the boy?
   B according to someone outside the train?

21.6 Summary

1. Projectiles are objects that move through the air.
2. Objects that move up and down (vertical projectiles) accelerate with a constant acceleration g which is more or less equal to 9,8 m·s^{-2}.
3. The equations of motion can be used to solve vertical projectile problems.

$$\begin{aligned} v_f &= v_i + gt \\ \Delta x &= \frac{(v_i + v_f)}{2}t \\ \Delta x &= v_i t + \frac{1}{2}gt^2 \\ v_f^2 &= v_i^2 + 2g\Delta x \end{aligned}$$

4. Graphs can be drawn for vertical projectile motion and are similar to the graphs for motion at constant acceleration. If upwards is taken as positive the Δx vs t, v vs t ans a vs t graphs for an object being thrown upwards look like this:



5. Momentum is conserved in one and two dimensions

$$\begin{aligned} p &= mv \\ \Delta p &= m\Delta v \\ \Delta p &= F\Delta t \end{aligned}$$

6. An elastic collision is a collision where both momentum and kinetic energy is conserved.

$$\begin{aligned} p_{\text{before}} &= p_{\text{after}} \\ KE_{\text{before}} &= KE_{\text{after}} \end{aligned}$$

7. An inelastic collision is where momentum is conserved but kinetic energy is not conserved.

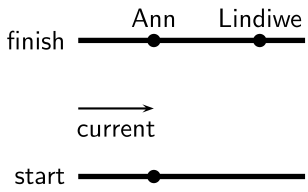
$$p_{\text{before}} = p_{\text{after}}$$

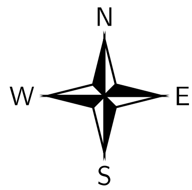
KE$_{\text{before}} \neq KE_{\text{after}}$

8. The frame of reference is the point of view from which a system is observed.

21.7 End of chapter exercises

1. [IEB 2005/11 HG] Two friends, Ann and Lindiwe decide to race each other by swimming across a river to the other side. They swim at identical speeds relative to the water. The river has a current flowing to the east.

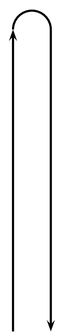




Ann heads a little west of north so that she reaches the other side directly across from the starting point. Lindiwe heads north but is carried downstream, reaching the other side downstream of Ann. Who wins the race?

   A Ann
   B Lindiwe
   C It is a dead heat
   D One cannot decide without knowing the velocity of the current.

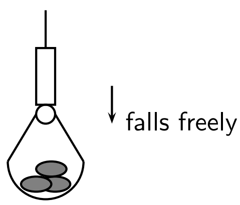
2. [SC 2001/11 HG1] A bullet fired vertically upwards reaches a maximum height and falls back to the ground.



Which **one** of the following statements is **true** with reference to the acceleration of the bullet during its motion, if air resistance is ignored?

   A is always downwards
   B is first upwards and then downwards
   C is first downwards and then upwards
   D decreases first and then increases

3. [SC 2002/03 HG1] Thabo suspends a bag of tomatoes from a spring balance held vertically. The balance itself weighs 10 N and he notes that the balance reads 50 N. He then lets go of the balance and the balance and tomatoes fall freely. What would the reading be on the balance while falling?



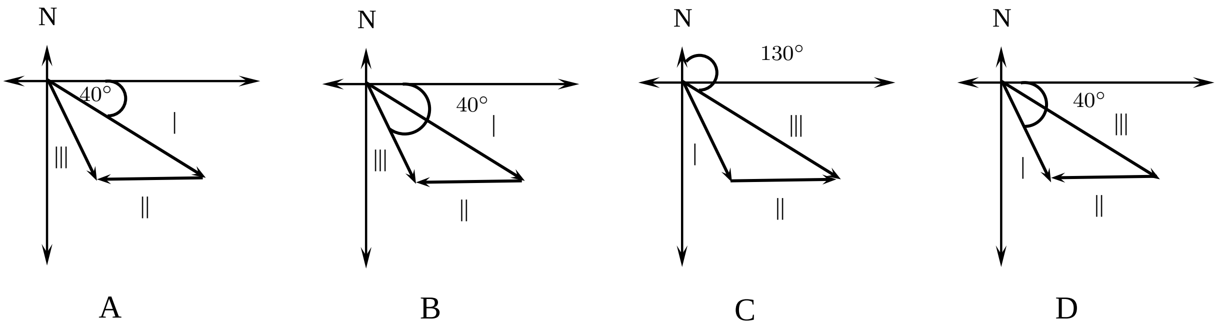
A 50 N
B 40 N
C 10 N
D 0 N

4. [IEB 2002/11 HG1] Two balls, P and Q, are simultaneously thrown into the air from the same height above the ground. P is thrown vertically upwards and Q vertically downwards with the same initial speed. Which of the following is true of both balls just before they hit the ground? (Ignore any air resistance. Take downwards as the positive direction.)

	Velocity	**Acceleration**
A	The same	The same
B	P has a greater velocity than Q	P has a negative acceleration; Q has a positive acceleration
C	P has a greater velocity than Q	The same
D	The same	P has a negative acceleration; Q has a positive acceleration

5. [IEB 2002/11 HG1] An observer on the ground looks up to see a bird flying overhead along a straight line on bearing 130° (40° S of E). There is a steady wind blowing from east to west. In the vector diagrams below, I, II and III represent the following:

I the velocity of the bird relative to the air
II the velocity of the air relative to the ground
III the resultant velocity of the bird relative to the ground
Which diagram correctly shows these three velocities?

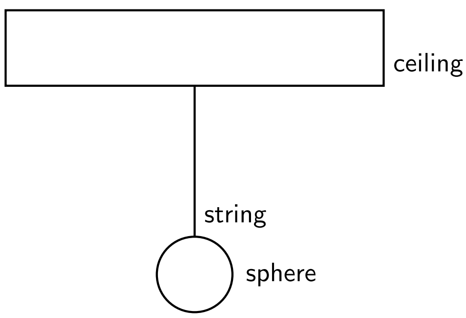


6. [SC 2003/11] A ball X of mass m is projected vertically upwards at a speed u_x from a bridge 20 m high. A ball Y of mass $2m$ is projected vertically downwards from the same bridge at a speed of u_y. The two balls reach the water at the same speed. Air friction can be ignored.

Which of the following is true with reference to the speeds with which the balls are projected?

A $u_x = \frac{1}{2}u_y$
B $u_x = u_y$
C $u_x = 2u_y$
D $u_x = 4u_y$

7. [SC 2001/11 HG1] A sphere is attached to a string, which is suspended from a horizontal ceiling.



The reaction force to the gravitational force exerted by the earth on the sphere is ...

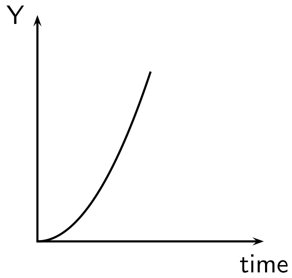
A the force of the sphere on the earth.

B the force of the ceiling on the string.

C the force of the string on the sphere.

D the force of the ceiling on the sphere.

8. [SC 2002/03 HG1] A stone falls freely from rest from a certain height. Which on eof the following quantities could be represented on the y-axis of the graph below?



A velocity

B acceleration

C momentum

D displacement

9. A man walks towards the back of a train at 2 m·s^{-1}while the train moves forward at 10 m·s^{-1}. The magnitude of the man's velocity with respect to the ground is

A 2 m·s^{-1}

B 8 m·s^{-1}

C 10 m·s^{-1}

D 12 m·s^{-1}

10. A stone is thrown vertically upwards and it returns to the ground. If friction is ignored, its acceleration as it reaches the highest point of its motion is

A greater than just after it left the throwers hand.

B less than just before it hits the ground.

C the same as when it left the throwers hand.

D less than it will be when it strikes the ground.

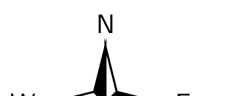
11. An exploding device is thrown vertically upwards. As it reaches its highest point, it explodes and breaks up into three pieces of **equal mass**. Which one of the following combinations is possible for the motion of the three pieces if they all move in a vertical line?

	Mass 1	Mass 2	Mass 3
A	v downwards	v downwards	v upwards
B	v upwards	2v downwards	v upwards
C	2v upwards	v downwards	v upwards
D	v upwards	2v downwards	v downwards

12. [IEB 2004/11 HG1] A stone is thrown vertically up into the air. Which of the following graphs best shows the resultant force exerted on the stone against time while it is in the air? (Air resistance is negligible.)

13. What is the velocity of a ball just as it hits the ground if it is thrown upward at 10 m·s^{-1}from a height 5 meters above the ground?

14. [IEB 2005/11 HG1] A breeze of 50 km·hr^{-1} blows towards the west as a pilot flies his light plane from town A to village B. The trip from A to B takes 1 h. He then turns west, flying for $\frac{1}{2}$ h until he reaches a dam at point C. He turns over the dam and returns to town A. The diagram shows his flight plan. It is not to scale.



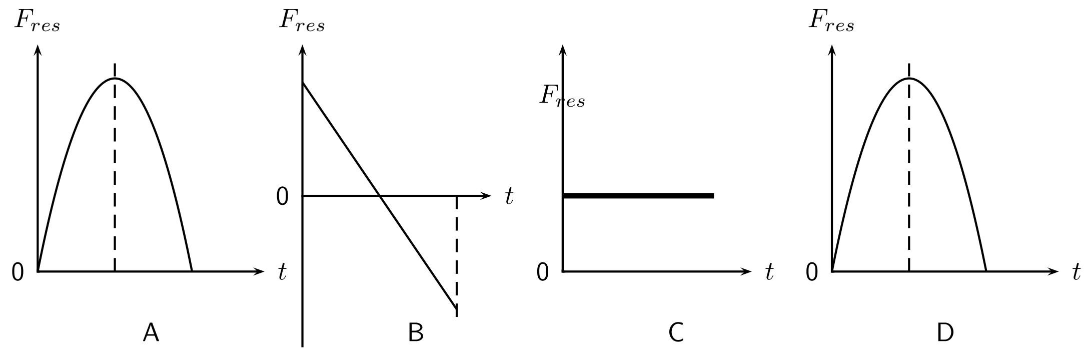
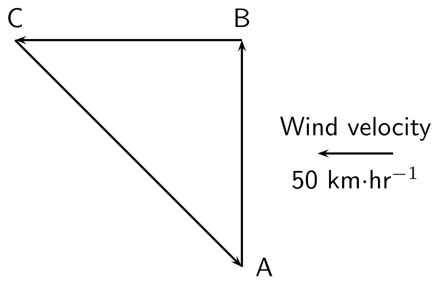


Figure 21.9: Graphs for an object thrown upwards with an initial velocity v_i. The object takes t_m s to reach its maximum height of h_m m after which it falls back to the ground. (a) position vs. time graph (b) velocity vs. time graph (c) acceleration vs. time graph.



The pilot flies at the same altitude at a constant speed of 130 km.h^{-1} relative to the air throughout this flight.

a Determine the magnitude of the pilot's resultant velocity from the town A to the village B.

b How far is village B from town A?

c What is the plane's speed relative to the ground as it travels from village B to the dam at C?

d Determine the following, by calculation or by scale drawing:

i. The distance from the village B to the dam C.

ii. The displacement from the dam C back home to town A.

15. A cannon (assumed to be at ground level) is fired off a flat surface at an angle, θ above the horizontal with an initial speed of v_0.

a What is the initial horizontal component of the velocity?

b What is the initial vertical component of the velocity?

c What is the horizontal component of the velocity at the highest point of the trajectory?

d What is the vertical component of the velocity at that point?

e What is the horizontal component of the velocity when the projectile lands?

f What is the vertical component of the velocity when it lands?

16. [IEB 2004/11 HG1] Hailstones fall vertically on the hood of a car parked on a horizontal stretch of road. The average terminal velocity of the hailstones as they descend is 8,0 m.s^{-1} and each has a mass of 1,2 g.

a Explain why a hailstone falls with a **terminal** velocity.

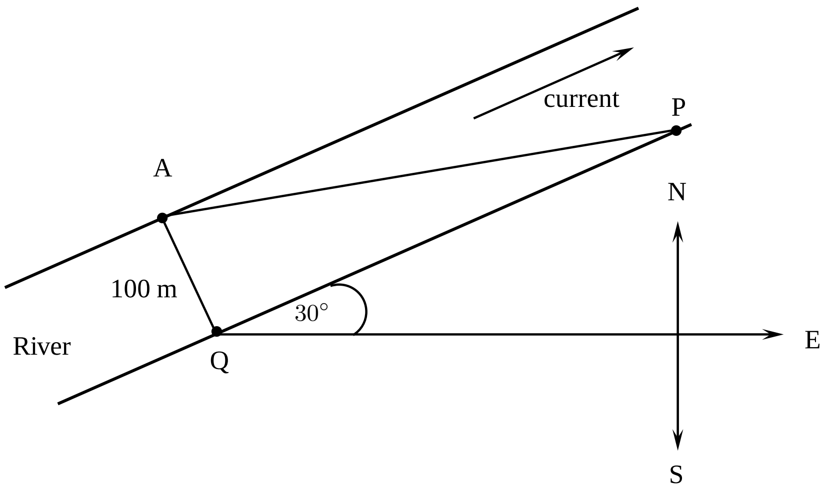
b Calculate the magnitude of the momentum of a hailstone just before it strikes the hood of the car.

c If a hailstone rebounds at 6,0 m.s^{-1} after hitting the car's hood, what is the magnitude of its change in momentum?

d The hailstone is in contact with the car's hood for 0,002 s during its collision with the hood of the car. What is the magnitude of the resultant force exerted on the hood if the hailstone rebounds at 6,0 m.s^{-1}?

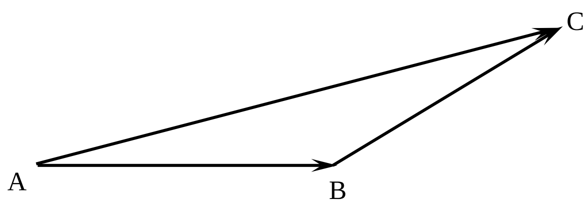
e A car's hood can withstand a maximum impulse of 0,48 N·s without leaving a permanent dent. Calculate the minimum mass of a hailstone that will leave a dent in the hood of the car, if it falls at 8,0 m.s^{-1} and rebounds at 6,0 m.s^{-1} after a collision lasting 0,002 s.

17. [IEB 2003/11 HG1 - Biathlon] Andrew takes part in a biathlon race in which he first swims across a river and then cycles. The diagram below shows his points of entry and exit from the river, A and P, respectively.



During the swim, Andrew maintains a constant velocity of 1,5 m.s^{-1} East relative to the water. The water in the river flows at a constant velocity of 2,5 m.s^{-1} in a direction 30° North of East. The width of the river is 100 m.

The diagram below is a velocity-vector diagram used to determine the resultant velocity of Andrew relative to the river bed.



a Which of the vectors (AB, BC and AC) refer to each of the following?
  i. The velocity of Andrew relative to the water.
  ii. The velocity of the water relative to the water bed.
  iii. The resultant velocity of Andrew relative to the river bed.

b Determine the magnitude of Andrew's velocity relative to the river bed either by calculations or by scale drawing, showing your method clearly.

c How long (in seconds) did it take Andrew to cross the river?

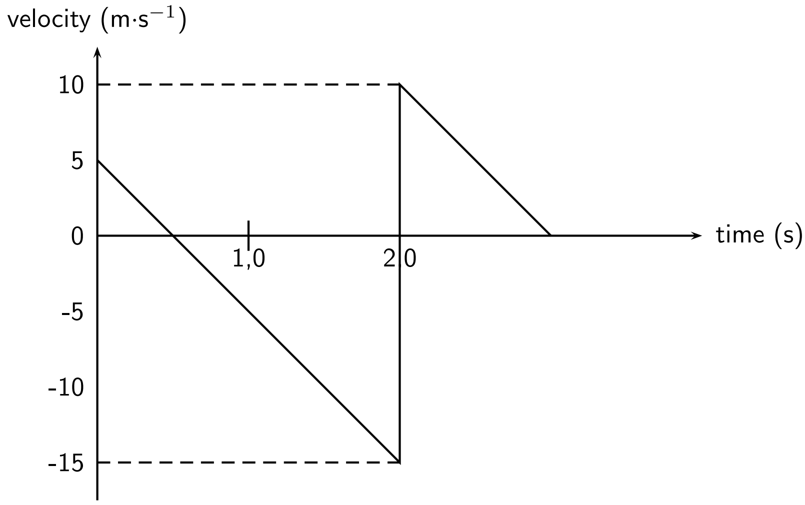
d At what distance along the river bank (QP) should Peter wait with Andrew's bicycle ready for the next stage of the race?

18. [IEB 2002/11 HG1 - Bouncing Ball]

A ball bounces vertically on a hard surface after being thrown vertically up into the air by a boy standing on the ledge of a building.

Just before the ball hits the ground for the first time, it has a velocity of magnitude 15 m.s^{-1}. Immediately, after bouncing, it has a velocity of magnitude 10 m.s^{-1}.

The graph below shows the velocity of the ball as a function of time from the moment it is thrown upwards into the air until it reaches its maximum height after bouncing once.



a At what velocity does the boy throw the ball into the air?

b What can be determined by calculating the gradient of the graph during the first two seconds?

c Determine the gradient of the graph over the first two seconds. State its units.

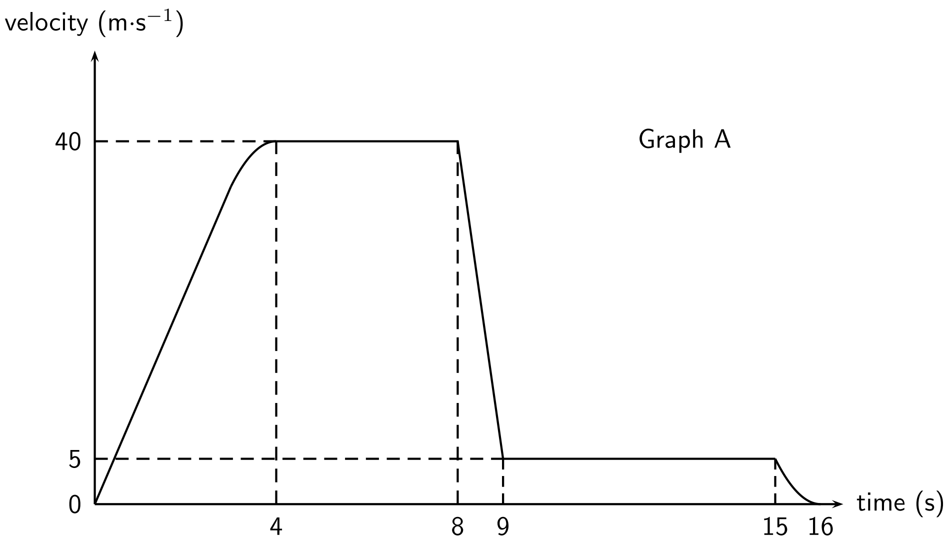
d How far below the boy's hand does the ball hit the ground?

e Use an equation of motion to calculate how long it takes, from the time the ball was thrown, for the ball to reach its maximum height after bouncing.

f What is the position of the ball, measured from the boy's hand, when it reaches its maximum height after bouncing?

19. [IEB 2001/11 HG1] - **Free Falling?**

A parachutist steps out of an aircraft, flying high above the ground. She falls for the first few seconds before opening her parachute. A graph of her velocity is shown in Graph A below.

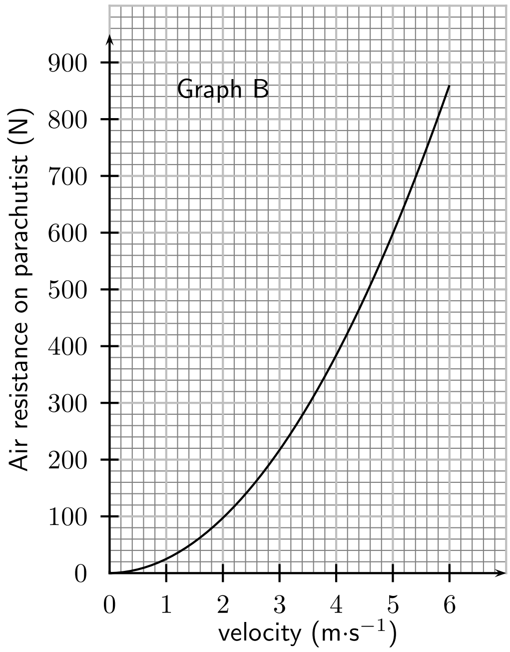


a Describe her motion between A and B.

b Use the information from the graph to calculate an approximate height of the aircraft when she stepped out of it (to the nearest 10 m).

c What is the magnitude of her velocity during her descent with the parachute fully open?

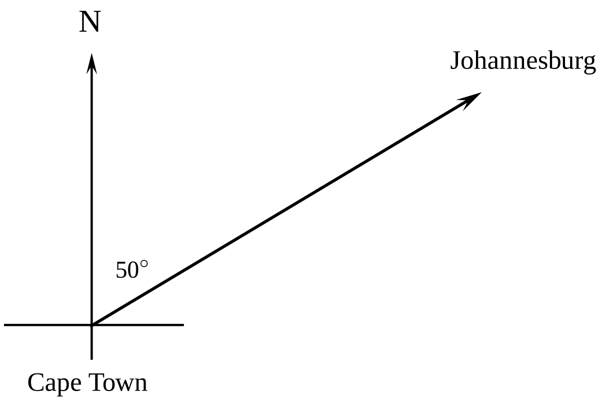
The air resistance acting on the parachute is related to the speed at which the parachutist descends. Graph B shows the relationship between air resistance and velocity of the parachutist descending with the parachute open.



d Use Graph B to find the magnitude of the air resistance on her parachute when she was descending with the parachute open.

e Assume that the mass of the parachute is negligible. Calculate the mass of the parachutist showing your reasoning clearly.

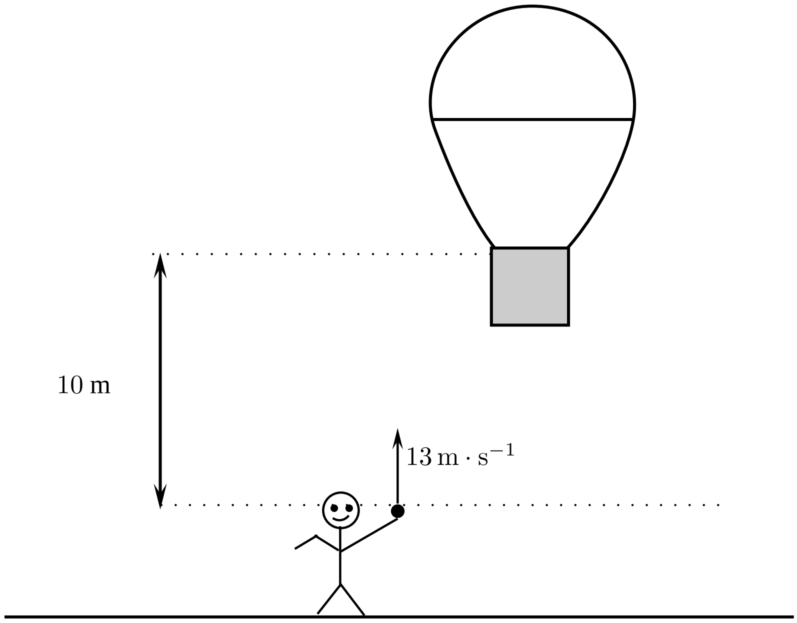
20. An aeroplane travels from Cape Town and the pilot must reach Johannesburg, which is situated 1300 km from Cape Town on a bearing of 50° in 5 hours. At the height at which the plane flies, a wind is blowing at 100 km·hr^{-1}on a bearing of 130 ° for the whole trip.



a Calculate the magnitude of the average resultant velocity of the aeroplane, in km·hr^{-1}, if it is to reach its destination on time.

b Calculate ther average velocity, in km·hr^{-1}, in which the aeroplane should be travelling in order to reach Johannesburg in the prescribed 5 hours. Include a labelled, rough vector diagram in your answer.
(If an accurate scale drawing is used, a scale of 25 km·hr^{-1}= 1 cm must be used.)

21. Niko, in the basket of a hot-air balloon, is stationary at a height of 10 m above the level from where his friend, Bongi, will throw a ball. Bongi intends throwing the ball upwards and Niko, in the basket, needs to **descend** (move downwards) to catch the ball at its maximum height.



Bongi throws the ball upwards with a velocity of 13 m·s^{-1}. Niko starts his descent at the same instant the ball is thrown upwards, by letting air escape from the balloon, causing it to accelerate downwards. Ignore the effect of air friction on the ball.

a Calculate the maximum height reached by the ball.

b Calculate the magnitude of the minimum average acceleration the balloon must have in order for Niko to catch the ball, if it takes 1,3 s for the ball to rach its maximum height.

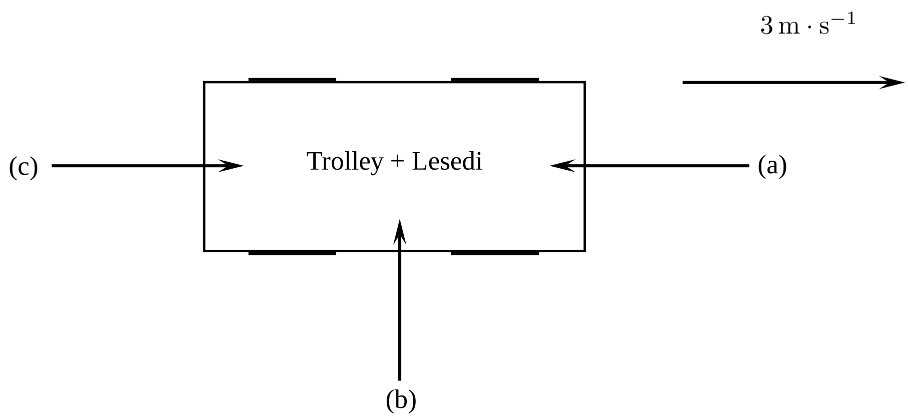
22. Lesedi (mass 50 kg) sits on a massless trolley. The trolley is travelling at a constant speed of 3 m·s^{-1}. His friend Zola (mass 60 kg) jumps on the trolley with a velocity of 2 m·s^{-1}. What is the final velocity of the combination (lesedi, Zola and trolley) if Zola jumps on the trolley from

a the front

b behind

c the side

(Ignore all kinds of friction)



Chapter 22

Mechanical Properties of Matter - Grade 12

22.1 Introduction

In this chapter we will look at some mechanical (physical) properties of various materials that we use. The mechanical properties of a material are those properties that are affected by forces being applied to the material. These properties are important to consider when we are constructing buildings, structures or modes of transport like an aeroplane.

22.2 Deformation of materials

22.2.1 Hooke's Law

Deformation (change of shape) of a solid is caused by a force that can either be compressive or tensile when applied in one direction (plane). Compressive forces try to compress the object (make it smaller or more compact) while tensile forces try to tear it apart. We can study these effects by looking at what happens when you compress or expand a spring.

Hooke's Law describes the relationship between the force applied to a spring and its extension.

Historical Note: Hooke's Law
Hooke's law is named after the seventeenth century physicist Robert Hooke who discovered it in 1660 (18 July 1635 - 3 March 1703).

Definition: Hooke's Law
In an elastic spring, the extension varies linearly with the force applied.
$F = -kx$ where F is the force in newtons (N), k is the spring constant in $N \cdot m^{-1}$ and x is the extension in metres (m).

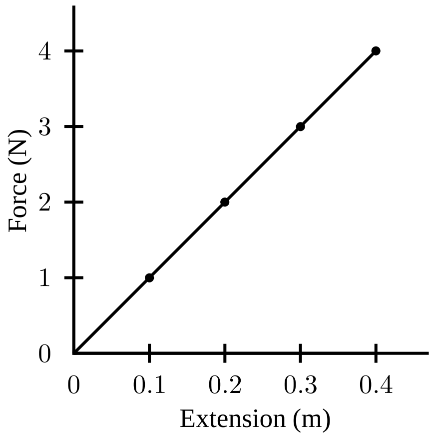


Figure 22.1: Hooke's Law - the relationship between extension of a spring and the force applied to it.

Activity :: Experiment : Hooke's Law

Aim:
Verify Hooke's Law.

Apparatus:

- weights
- spring
- ruler

Method:

1. Set up a spring vertically in such a way that you are able to hang weights from it.
2. Measure the extension of the spring for a range of different weights.
3. Draw a table of force (weight) in newtons and corresponding extension.
4. Draw a graph of force versus extension for your experiment.

Conclusions:

1. What do you observe about the relationship between the applied force and the extension?
2. Determine the gradient of the graph.
3. Hence, calculate the spring constant for your spring.

Worked Example 146: Hooke's Law I

Question: A spring is extended by 7 cm by a force of 56 N. Calculate the spring constant for this spring.

Answer

$$\begin{aligned} \mathrm{F} &= -\mathrm{kx} \\ 56 &= -\mathrm{k}.0{,}07 \end{aligned}$$

$$
\begin{aligned}
k &= \frac{-56}{0{,}07} \\
&= -800 \text{ N.m}^{-1}
\end{aligned}
$$

Worked Example 147: Hooke's Law II

Question: A spring of length 20cm stretches to 24cm when a load of 0,6N is applied to it.

1. Calculate the spring constant for the spring.
2. Determine the extension of the spring if a load of 0,5N is applied to it.

Answer

1.

$$
\begin{aligned}
x &= 24 \text{ cm} - 20 \text{ cm} \\
&= 4 \text{ cm} \\
&= 0{,}04 \text{ m}
\end{aligned}
$$

$$
\begin{aligned}
F &= -kx \\
0{,}6 &= -k.\, 0{,}04
\end{aligned}
$$

$$k = -15 \text{ N.m}^{-1}$$

2.

$$F = -kx$$

$$
\begin{aligned}
x &= \frac{F}{-k} \\
&= \frac{0{,}5}{15} \\
&= 0{,}033 \text{ m} \\
&= 3{,}3 \text{ cm}
\end{aligned}
$$

Worked Example 148: Hooke's Law III

Question: A spring has a spring constant of -400 N.m^{-1}. By how much will it stretch if a load of 50 N is applied to it?

Answer

$$
\begin{aligned}
F &= -kx \\
50 &= -(-400)x
\end{aligned}
$$

$$\begin{aligned} x &= \frac{50}{400} \\ &= 0{,}125 \text{ m} \\ &= 12{,}5 \text{ cm} \end{aligned}$$

22.2.2 Deviation from Hooke's Law

We know that if you have a small spring and you pull it apart too much it stops 'working'. It bends out of shape and loses its springiness. When this happens Hooke's Law no longer applies, the spring's behaviour deviates from Hooke's Law.

Depending on what type of material we are dealing the manner in which it deviates from Hooke's Law is different. We give classify materials by this deviation. The following graphs show the relationship between force and extension for different materials and they all deviate from Hooke's Law. Remember that a straight line show proportionality so as soon as the graph is no longer a straight line, Hooke's Law no longer applies.

Brittle material

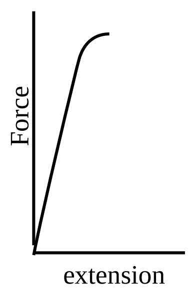


Figure 22.2: A hard, brittle substance

This graph shows the relationship between force and extension for a brittle, but strong material. Note that there is very little extension for a large force but then the material suddenly fractures. Brittleness is the property of a material that makes it break easily without bending.

Have you ever dropped something made of glass and seen it shatter? Glass does this because of its brittleness.

Plastic material

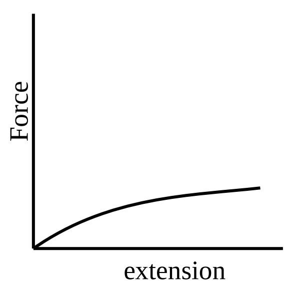


Figure 22.3: A plastic material's response to an applied force.

Here the graph shows the relationship between force and extension for a plastic material. The material extends under a small force but it does not fracture.

Ductile material

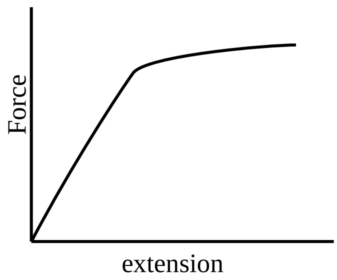


Figure 22.4: A ductile substance.

In this graph the relationship between force and extension is for a material that is ductile. The material shows plastic behaviour over a range of forces before the material finally fractures. Ductility is the ability of a material to be stretched into a new shape without breaking. Ductility is one of the characteristic properties of metals.

A good example of this is aluminium, many things are made of aluminium. Aluminium is used for making everything from cooldrink cans to aeroplane parts and even engine blocks for cars. Think about squashing and bending a cooldrink can.

Brittleness is the opposite of ductility.

When a material reaches a point where Hooke's Law is no longer valid, we say it has reached its *limit of proportionality*. After this point, the material will not return to its original shape after the force has been removed. We say it has reached its *elastic limit*.

Definition: Elastic limit
The elastic limit is the point beyond which permanent deformation takes place.

Definition: Limit of proportionality
The limit of proportionality is the point beyond which Hooke's Law is no longer obeyed.

Exercise: Hooke's Law and deformation of materials

1. What causes deformation?
2. Describe Hooke's Law in words and mathematically.
3. List similarities and differences between ductile, brittle and polymeric materials, with specific reference to their force-extension graphs.
4. Describe what is meant by the *elastic limit*.
5. Describe what is meant by the *limit of proportionality*.
6. A spring of length 15 cm stretches to 27 cm when a load of 0,4 N is applied to it.
   A Calculate the spring constant for the spring.
   B Determine the extension of the spring if a load of 0,35 N is applied to it.
7. A spring has a spring constant of -200 N.m^{-1}. By how much will it stretch if a load of 25 N is applied to it?
8. A spring of length 20 cm stretches to 24 cm when a load of 0,6 N is applied to it.

A Calculate the spring constant for the spring.
B Determine the extension of the spring if a load of 0,8 N is applied to it.

22.3 Elasticity, plasticity, fracture, creep

22.3.1 Elasticity and plasticity

Materials are classified as plastic or elastic depending on how they respond to an applied force. It is important to note that plastic substances are not necessarily a type of plastic (polymer) they only behave like plastic. Think of them as being like plastic which you will be familiar with.

A rubber band is a material that has elasticity. It returns to its original shape after an applied force is removed, providing that the material is not stretched beyond its elastic limit.

Plasticine is an example of a material that is plastic. If you flatten a ball of plasticine, it will stay flat. A plastic material does not return to its original shape after an applied force is removed.

- Elastic materials return to their original shape.
- Plastic materials deform easily and do not return to their original shape.

22.3.2 Fracture, creep and fatigue

Some materials are neither plastic nor elastic. These substances will break or fracture when a large enough force is applied to them. The brittle glass we mentioned earlier is an example.

Creep occurs when a material deforms over a long period of time because of an applied force. An example of creep is the bending of a shelf over time when a heavy object is put on it. Creep may eventually lead to the material fracturing. The application of heat may lead to an increase in creep in a material.

Fatigue is similar to creep. The difference between the two is that fatigue results from the force being applied and then removed repeatedly over a period of time. With metals this results in failure because of metal fatigue.

- Fracture is an abrupt breaking of the material.
- Creep is a slow deformation process due to a continuous force over a long time.
- Fatigue is weakening of the material due to short forces acting many many times.

?

Exercise: Elasticity, plasticity, fracture and creep

1. List the similarities and differences between elastic and plastic deformation.
2. List the similarities and differences between creep and fracture as modes of failure in material.

22.4 Failure and strength of materials

22.4.1 The properties of matter

The strength of a material is defined as the stress (the force per unit cross-sectional area) that it can withstand. Strength is measured in newtons per square metre ($N \cdot m^{-2}$).

Stiffness is a measure of how flexible a material is. In Science we measure the stiffness of a material by calculating its Young's Modulus. The Young's modulus is a ratio of how much it bends to the load applied to it. Stiffness is measure in newtons per metre ($N \cdot m^{-1}$).

Hardness of a material can be measured by determining what force will cause a permanent deformation in the material. Hardness can also be measured using a scale like Mohs hardness scale. On this scale, diamond is the hardest at 10 and talc is the softest at 1.

Interesting Fact

Remembering that the Mohs scale is the hardness scale and that the softest substance is talc will often come in handy for general knowledge quizes.

The *toughness* of a material is a measure of how it can resist breaking when it is stressed. It is scientifically defined as the amount of energy that a material can absorb before breaking.

A ductile material is a substance that can undergo large plastic deformation without fracturing. Many metals are very ductile and they can be drawn into wires, e.g. copper, silver, aluminium and gold.

A *malleable* material is a substance that can easily undergo plastic deformation by hammering or rolling. Again, metals are malleable substances, e.g. copper can be hammered into sheets and aluminium can be rolled into aluminium foil.

A brittle material fractures with very little or no plastic deformation. Glassware and ceramics are brittle.

22.4.2 Structure and failure of materials

Many substances fail because they have a weakness in their atomic structure. There are a number of problems that can cause these weaknesses in structure. These are vacancies, dislocations, grain boundaries and impurities.

Vacancies occur when there are spaces in the structure of a crystalline solid. These vacancies cause weakness and the substance often fail at these places. Think about bricks in a wall, if you started removing bricks the wall would get weaker.

Dislocations occur when there are no strong bonds between two rows in a crystal lattice. The crystal will fail along this boundary when sufficient force is applied. The two pieces of the crystal keep their shape and structure but move along the boundary.

Impurities in a crystal structure can cause a weak spot in the crystal lattice around the impurity. Like vacancies, the substance often fail from these places in the lattice. This you can think of as bricks in a wall which don't fit properly, they are the wrong kind of bricks (atoms) to make the structure strong.

A difference in *grain size* in a crystal lattice will result in rusting or oxidation at the boundary which again will result in failure when sufficient force is applied.

22.4.3 Controlling the properties of materials

There are a number of processes that can be used to ensure that materials are less likely to fail. We shall look at a few methods in this section.

Cold working

Cold working is a process in which a metal is *strengthened* by repeatedly being reshaped. This is carried out at a temperature below the melting point of the metal. The repeated shaping of the metal result in dislocations which then prevent further dislocations in the metal. Cold working increases the strength of the metal but in so doing, the metal loses its ductility. We say the metal is *work-hardened*.

Annealing

Annealing is a process in which a metal is heated strongly to a temperature that is about half of its melting point. When the metal cools, it recrystallises which removes vacancies and dislocations in the metal. Annealing is often used before cold working. In annealing the metal cools very very slowly.

Alloying

An alloy is a mixture of a metal with other substances. The other substances can be metal or non-metal. An alloy often has properties that are very different to the properties of the substances from which it is made. The added substances strengthen the metal by preventing dislocations from spreading. Ordinary steel is an alloy of iron and carbon. There are many types of steel that also include other metals with iron and carbon. Brass is an alloy of copper and Zinc. Bronze is an alloy of copper and tin. Gold and silver that is used in coins or jewellery are also alloyed.

Tempering

Tempering is a process in which a metal is melted then quickly cooled. The rapid cooling is called quenching. Usually tempering is done a number of times before a metal has the correct properties that are needed for a particular application.

Sintering

Sintering is used for making ceramic objects among other things. In this process the substance is heated so that its particles stick together. It is used with substances that have a very high melting point. The resulting product is often very pure and it is formed in the process into the shape that is wanted. Unfortunately, sintered products are brittle.

22.4.4 Steps of Roman Swordsmithing

- Purifying the iron ore.
- Heating the iron blocks in a furnace with charcoal.
- Hammering and getting into the needed shape. The smith used a hammer to pound the metal into blade shape. He usually used tongs to hold the iron block in place.
- Reheating. When the blade cooled, the smith reheated it to keep it workable. While reheated and hammered repeatedly.
- *Quenching* which involved the process of white heating and cooling in water. Quenching made the blade harder and stronger. At the same time it made the blade quite brittle, which was a considerable problem for the sword smiths.
- *Tempering* was then done to avoid brittleness the blade was tempered. In another words it was reheated a final time to a very specific temperature. How the Romans do balanced the temperature? The smith was guided only by the blade's color and his own experience.

?

Exercise: Failure and strength of materials

1. List the similarities and differences between the brittle and ductile modes of failure.
2. What is meant by the following terms:
   A vacancies
   B dislocations
   C impurities
   D grain boundaries
3. What four terms can be used to describe a material's mechanical properties?
4. What is meant by the following:
   A cold working
   B annealing
   C tempering
   D introduction of impurities
   E alloying
   F sintering

22.5 Summary

1. Hooke's Law gives the relationship between the extension of a spring and the force applied to it. The law says they are proportional.
2. Materials can be classified as plastic or elastic depending on how they respond to an applied force.
3. Materials can fracture or undergo creep or fatigue when forces are applied to them.
4. Materials have the following mechanical properties to a greater or lesser degree: strength, hardness, ductility, malleability, brittleness, stiffness.
5. Materials can be weakened by have the following problems in their crystal lattice: vacancies, dislocations, impurities, difference in grain size.
6. Materials can have their mechanical properties improved by one or more of the following processes: cold working, annealing, adding impurities, tempering, sintering.

22.6 End of chapter exercise

1. State Hooke's Law in words.
2. What do we mean by the following terms with respect to Hooke's Law?
   A elastic limit
   B limit of proportionality
3. A spring is extended by 18 cm by a force of 90 N. Calculate the spring constant for this spring.
4. A spring of length 8 cm stretches to 14 cm when a load of 0,8 N is applied to it.
   A Calculate the spring constant for the spring.
   B Determine the extension of the spring if a load of 0,7 N is applied to it.

5. A spring has a spring constant of -150 N.m^{-1}. By how much will it stretch if a load of 80 N is applied to it?

6. What do we mean by the following terms when speaking about properties of materials?

   A hardness

   B toughness

   C ductility

   D malleability

   E stiffness

   F strength

7. What is Young's modulus?

8. In what different ways can we improve the material properties of substances?

9. What is a metal alloy?

10. What do we call an alloy of:

    A iron and carbon

    B copper and zinc

    C copper and tin

11. Do some research on what added substances can do to the properties of steel. Present you findings in a suitable table.

Chapter 23

Work, Energy and Power - Grade 12

(NOTE TO SELF: Status: Content is complete. More exercises, worked examples and activities are needed.)

23.1 Introduction

Imagine a vendor carrying a basket of vegetables on her head. Is she doing any work? One would definitely say yes! However, in Physics she is not doing any work! Again, imagine a boy pushing against a wall? Is he doing any work? We can see that his muscles are contracting and expanding. He may even be sweating. But in Physics, he is not doing any work!

If the vendor is carrying a very heavy load for a long distance, we would say she has lot of energy. By this, we mean that she has a lot of stamina. If a car can travel very fast, we describe the car as powerful. So, there is a link between power and speed. However, power means something different in Physics. This chapter describes the links between work, energy and power and what these mean in Physics.

You will learn that work and energy are closely related. You shall see that the energy of an object is its capacity to do work and doing work is the process of transferring energy from one object or form to another. In other words,

- an object with lots of energy can do lots of work.
- when work is done, energy is lost by the object doing work and gained by the object on which the work is done.

Lifting objects or throwing them requires that you do work on them. Even making electricity flow requires that something do work. Something must have energy and transfer it through doing work to make things happen.

23.2 Work

Definition: Work
When a force exerted on an object causes it to move, work is done on the object (except if the force and displacement are at right angles to each other).

This means that in order for work to be done, an object must be moved a distance d by a force F, such that there is some non-zero component of the force in the direction of the displacement. Work is calculated as:

$$W = F \cdot \Delta x \cos\theta. \tag{23.1}$$

where F is the applied force, Δx is the displacement of the object and θ is the angle between the applied force and the direction of motion.

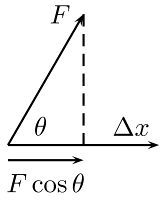


Figure 23.1: The force F causes the object to be displaced by Δx at angle θ.

It is very important to note that for work to be done there must be a component of the applied force in the direction of motion. Forces perpendicular to the direction of motion do no work.

For example work is done on the object in Figure 23.2,

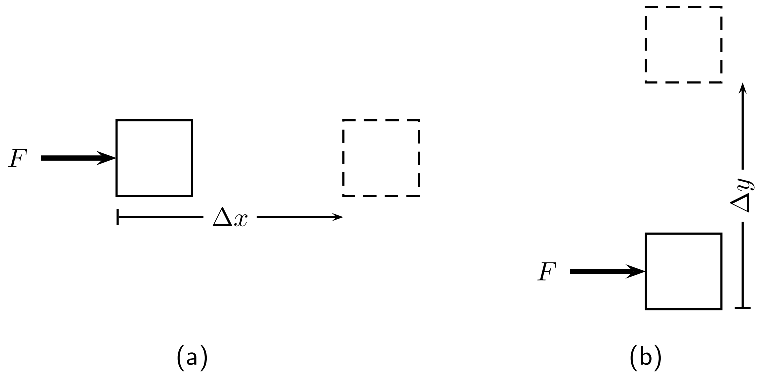


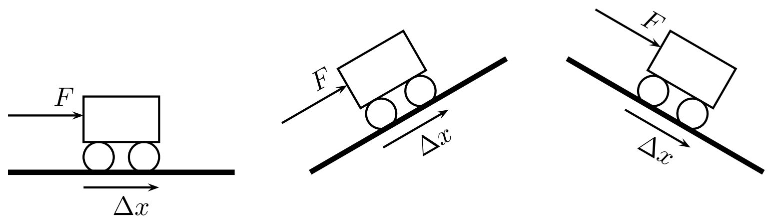
Figure 23.2: (a) The force F causes the object to be displaced by Δx in the same direction as the force. $\theta = 180°$ and $\cos\theta = 1$. Work is done in this situation. (b) A force F is applied to the object. The object is displaced by Δy at right angles to the force. $\theta = 90°$ and $\cos\theta = 0$. Work is not done in this situation.

Activity :: Investigation : Is work done?

Decide whether on not work is done in the following situations. Remember that for work to be done a force must be applied in the direction of motion and there must be a displacement. Give reasons for your answer.

1. Max applies a force to a wall and becomes tired.
2. A book falls off a table and free falls to the ground.
3. A rocket accelerates through space.
4. A waiter carries a tray full of meals above his head by one arm straight across the room at constant speed. (Careful! This is a very difficult question.)

Important: The Meaning of θ The angle θ is the angle between the force vector and the displacement vector. In the following situations, $\theta = 0°$.



As with all physical quantities, work must have units. Following from the definition, work is measured in N·m. The name given to this combination of S.I. units is the joule (symbol J).

Definition: Joule
1 joule is the work done when an object is moved 1 m under the application of a force of 1 N in the direction of motion.

The work done by an object can be positive or negative. Since force ($F_{||}$) and displacement (s) are both vectors, the result of the above equation depends on their directions:

- If $F_{||}$ acts in the same direction as the motion then positive work is being done. In this case the object on which the force is applied gains energy.
- If the direction of motion and $F_{||}$ are opposite, then negative work is being done. This means that energy is transferred in the opposite direction. For example, if you try to push a car uphill by applying a force up the slope and instead the car rolls down the hill you are doing negative work on the car. Alternatively, the car is doing positive work on you!

Important: The everyday use of the word "work" differs from the physics use. In physics, only the component of the applied force that is parallel to the motion does work on an object. So, for example, a person holding up a heavy book does no work on the book.

Worked Example 149: Calculating Work Done I

Question: If you push a box 20 m forward by applying a force of 15 N in the forward direction, what is the work you have done on the box?
Answer
Step 1 : Analyse the question to determine what information is provided

- The force applied is F=15 N.
- The distance moved is s=20 m.
- The applied force and distance moved are in the same direction. Therefore, $F_{||}$=15 N.

These quantities are all in the correct units, so no unit conversions are required.
Step 2 : Analyse the question to determine what is being asked

- We are asked to find the work done on the box. We know from the definition that work done is $W = F_{||}s$

Step 3 : Next we substitute the values and calculate the work done

$$\begin{aligned} W &= F_{\|}s \\ &= (15\ \text{N})(20\ \text{m}) \\ &= 300\ \text{J} \end{aligned}$$

Remember that the answer must be *positive* as the applied force and the motion are in the same direction (forwards). In this case, you (the pusher) lose energy, while the box gains energy.

Worked Example 150: Calculating Work Done II

Question: What is the work done by you on a car, if you try to push the car up a hill by applying a force of 40 N directed up the slope, but it slides downhill 30 cm?

Answer

Step 1 : Analyse the question to determine what information is provided

- The force applied is F=40 N
- The distance moved is s=30 cm. This is expressed in the wrong units so we must convert to the proper S.I. units (meters):

$$\begin{aligned} 100\,\text{cm} &= 1\,\text{m} \\ 1\,\text{cm} &= \frac{1}{100}\,\text{m} \\ \therefore 30 \times 1\,\text{cm} &= 30 \times \frac{1}{100}\,\text{m} \\ &= \frac{30}{100}\,\text{m} \\ &= 0{,}3\,\text{m} \end{aligned}$$

- The applied force and distance moved are in opposite directions. Therefore, if we take s=0.3 m, then $F_{\|}$=-40 N.

Step 2 : Analyse the question to determine what is being asked

- We are asked to find the work done on the car by you. We know that work done is $W = F_{\|}s$

Step 3 : Substitute the values and calculate the work done

Again we have the applied force and the distance moved so we can proceed with calculating the work done:

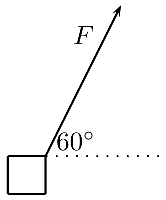
$$\begin{aligned} W &= F_{\|}s \\ &= (-40\,\text{N})(0.3\,\text{m}) \\ &= -12\text{J} \end{aligned}$$

Note that the answer must be *negative* as the applied force and the motion are in opposite directions. In this case the car does work on the person trying to push.

What happens when the applied force and the motion are not parallel? If there is an angle between the direction of motion and the applied force then to determine the work done we have to calculate the *component* of the applied force *parallel* to the direction of motion. Note that this means a force perpendicular to the direction of motion can do no work.

Worked Example 151: Calculating Work Done III

Question: Calculate the work done on a box, if it is pulled 5 m along the ground by applying a force of F=10 N at an angle of $60°$ to the horizontal.



Answer

Step 1 : Analyse the question to determine what information is provided

- The force applied is F=10 N
- The distance moved is s=5 m along the ground
- The angle between the applied force and the motion is $60°$

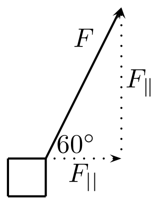
These quantities are in the correct units so we do not need to perform any unit conversions.

Step 2 : Analyse the question to determine what is being asked

- We are asked to find the work done on the box.

Step 3 : Calculate the component of the applied force in the direction of motion

Since the force and the motion are not in the same direction, we must first calculate the component of the force in the direction of the motion.



From the force diagram we see that the component of the applied force parallel to the ground is

$$\begin{aligned} F_{||} &= F \cdot \cos(60°) \\ &= 10\,\text{N} \cdot \cos(60°) \\ &= 5\,\text{N} \end{aligned}$$

Step 4 : Substitute and calculate the work done

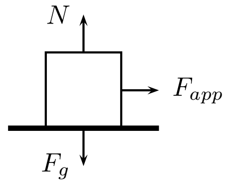
Now we can calculate the work done on the box:

$$\begin{aligned} W &= F_{||}s \\ &= (5\,\text{N})(5\,\text{m}) \\ &= 25\,\text{J} \end{aligned}$$

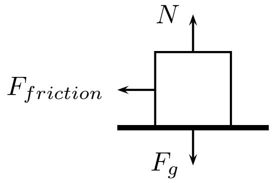
Note that the answer is positive as the component of the force $F_{||}$ is in the same direction as the motion.

Exercise: Work

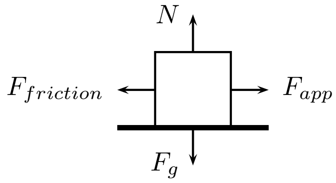
1. A 10 N force is applied to push a block across a friction free surface for a displacement of 5.0 m to the right. The block has a weight of 20 N. Determine the work done by the following forces: normal force, weight, applied force.



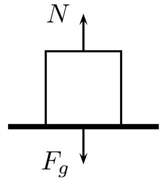
2. A 10 N frictional force slows a moving block to a stop after a displacement of 5.0 m to the right. The block has a weight of 20 N. Determine the work done by the following forces: normal force, weight, frictional force.



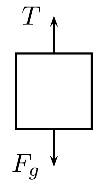
3. A 10 N force is applied to push a block across a frictional surface at constant speed for a displacement of 5.0 m to the right. The block has a weight of 20 N and the frictional force is 10 N. Determine the work done by the following forces: normal force, weight, frictional force.



4. A 20 N object is sliding at constant speed across a friction free surface for a displacement of 5 m to the right. Determine if there is any work done.



5. A 20 N object is pulled upward at constant speed by a 20 N force for a vertical displacement of 5 m. Determine if there is any work done.

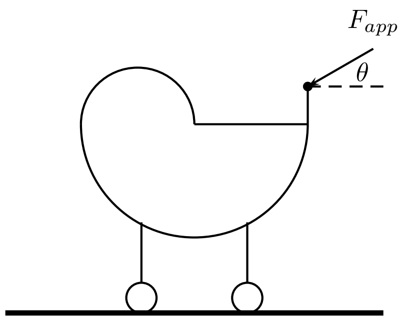


6. Before beginning its descent, a roller coaster is always pulled up the first hill to a high initial height. Work is done on the roller coaster to achieve this initial height. A coaster designer is considering three different incline angles **of the hill** at which to drag the 2 000 kg car train to the top of the 60 m high hill. In each case, the force applied to the car will be applied parallel to the hill. Her critical question is: which angle would require the least work? Analyze the data, determine the work done in each case, and answer this critical question.

Angle of Incline	**Applied Force**	**Distance**	**Work**
35°	$1.1 \times 10^4\,\mathrm{N}$	100 m	
45°	$1.3 \times 10^4\,\mathrm{N}$	90 m	
55°	$1.5 \times 10^4\,\mathrm{N}$	80 m	

7. Big Bertha carries a 150 N suitcase up four flights of stairs (a total height of 12 m) and then pushes it with a horizontal force of 60 N at a constant speed of 0.25 m·s^{-1} for a horizontal distance of 50 m on a frictionless surface. How much work does Big Bertha do on the suitcase during this entire trip?

8. A mother pushes down on a pram with a force of 50 N at an angle of 30°. The pram is moving on a frictionless surface. If the mother pushes the pram for a horizontal distance of 30 m, how much does she do on the pram?



9. How much work is done by an applied force to raise a 2 000 N lift 5 floors vertically at a constant speed? Each floor is 5 m high.
10. A student with a mass of 60 kg runs up three flights of stairs in 15 s, covering a vertical distance of 10 m. Determine the amount of work done by the student to elevate her body to this height. Assume that her speed is constant.
11. (NOTE TO SELF: exercises are needed.)

23.3 Energy

23.3.1 External and Internal Forces

In Grade 10, you saw that mechanical energy was conserved in the absence of external forces. It is important to know whether a force is an internal force or an external force, because this is related to whether the force can change an object's total mechanical energy when it does work upon an object.

Activity :: Investigations : External Forces
(NOTE TO SELF: need an activity that helps the learner investigate how energy is lost when external forces do work on an object.)

When an external force (for example friction, air resistance, applied force) does work on an object, the total mechanical energy (KE + PE) of that object changes. If positive work is done, then the object will gain energy. If negative work is done, then the object will lose energy. The gain or loss in energy can be in the form of potential energy, kinetic energy, or both. However, the work which is done is equal to the change in mechanical energy of the object.

Activity :: Investigation : Internal Forces and Energy Conservation
(NOTE TO SELF: need an activity that helps the learner investigate how energy changes form when an internal force does work on an object.)

When an internal force does work on an object by an (for example, gravitational and spring forces), the total mechanical energy (KE + PE) of that object remains constant but the object's energy can change form. For example, as an object falls in a gravitational field from a high elevation to a lower elevation, some of the object's potential energy is changed into kinetic energy. However, the sum of the kinetic and potential energies remain constant. When the only forces doing work are internal forces, energy changes forms - from kinetic to potential (or vice versa); yet the total amount of mechanical is conserved.

23.3.2 Capacity to do Work

Energy is the capacity to do work. When positive work is done on an object, the system doing the work loses energy. In fact, **the energy lost by a system is exactly equal to the work done by the system.** An object with larger potential energy has a greater capacity to do work.

Worked Example 152: Work Done on a System

Question: Show that a hammer of mass 2 kg does more work when dropped from a height of 10 m than when dropped from a height of 5 m. Confirm that the hammer has a greater potential energy at 10 m than at 5 m.
Answer
Step 5 : Determine what is given and what is required
We are given:

- the mass of the hammer, m =2 kg
- height 1, h_1=10 m
- height 2, h_2=5 m

We are required to show that the hammer does more work when dropped from h_1 than from h_2. We are also required to confirm that the hammer has a greater potential energy at 10 m than at 5 m.
Step 6 : Determine how to approach the problem

1. Calculate the work done by the hammer, W_1, when dropped from h_1 using:
$$W_1 = F_g \cdot h_1.$$
2. Calculate the work done by the hammer, W_2, when dropped from h_2 using:
$$W_2 = F_g \cdot h_2.$$
3. Compare W_1 and W_2
4. Calculate potential energy at h_1 and h_2 and compare using:
$$PE = m \cdot g \cdot h. \tag{23.2}$$

Step 7 : Calculate W_1

$$\begin{aligned} W_1 &= F_g \cdot h_1 \\ &= m \cdot g \cdot h_1 \\ &= (2\,\mathrm{kg})(9.8\,\mathrm{m \cdot s^{-2}})(10\,\mathrm{m}) \\ &= 196\,\mathrm{J} \end{aligned}$$

Step 8 : Calculate W_2

$$\begin{aligned} W_2 &= F_g \cdot h_2 \\ &= m \cdot g \cdot h_2 \\ &= (2\,\mathrm{kg})(9.8\,\mathrm{m \cdot s^{-2}})(5\,\mathrm{m}) \\ &= 98\,\mathrm{J} \end{aligned}$$

Step 9 : Compare W_1 and W_2
We have W_1=196 J and W_2=98 J. $W_1 > W_2$ as required.
Step 10 : Calculate potential energy
From 23.2, we see that:

$$\begin{aligned} PE &= m \cdot g \cdot h \\ &= F_g \cdot h \\ &= W \end{aligned}$$

This means that the potential energy is equal to the work done. Therefore, $PE_1 > PE_2$, because $W_1 > W_2$.

This leads us to the work-energy theorem.

Definition: Work-Energy Theorem
The work-energy theorem states that the work done on an object is equal to the change in its kinetic energy:

$$W = \Delta KE = KE_f - KE_i$$

The work-energy theorem is another example of the conservation of energy which you saw in Grade 10.

Worked Example 153: Work-Energy Theorem

Question: A ball of mass 1 kg is dropped from a height of 10 m. Calculate the work done on the ball at the point it hits the ground assuming that there is no air resistance?
Answer
Step 1 : Determine what is given and what is required
We are given:

- mass of the ball: m=1 kg
- initial height of the ball: h_i=10 m
- final height of the ball: h_f=0 m

We are required to determine the work done on the ball as it hits the ground.
Step 2 : Determine how to approach the problem
The ball is falling freely, so energy is conserved. We know that the work done is equal to the difference in kinetic energy. The ball has no kinetic energy at the moment it is dropped, because it is stationary. When the ball hits the ground, all the ball's potential energy is converted to kinetic energy.
Step 3 : Determine the ball's potential energy at h_i

$$\begin{aligned} PE &= m \cdot g \cdot h \\ &= (1\,\mathrm{kg})(9{,}8\,\mathrm{m \cdot s^{-2}})(10\,\mathrm{m}) \\ &= 98\,\mathrm{J} \end{aligned}$$

Step 4 : Determine the work done on the ball
The ball had 98 J of potential energy when it was released and 0 J of kinetic energy. When the ball hit the ground, it had 0 J of potential energy and 98 J of kinetic energy. Therefore KE_i=0 J and KE_f=98 J.

From the work-energy theorem:

$$\begin{aligned} W &= \Delta KE \\ &= KE_f - KE_i \\ &= 98\,\mathrm{J} - 0\,\mathrm{J} \\ &= 98\,\mathrm{J} \end{aligned}$$

Step 5 : Write the final answer
98 J of work was done on the ball.

Worked Example 154: Work-Energy Theorem 2

Question: The driver of a 1 000 kg car traveling at a speed of 16,7 m·s^{-1} applies the car's brakes when he sees a red robot. The car's brakes provide a frictional force of 8000 N. Determine the stopping distance of the car.
Answer
Step 1 : Determine what is given and what is required
We are given:

- mass of the car: m=1 000 kg
- speed of the car: v=16,7 m·s^{-1}
- frictional force of brakes: F=8 000 N

We are required to determine the stopping distance of the car.
Step 2 : Determine how to approach the problem
We apply the work-energy theorem. We know that all the car's kinetic energy is lost to friction. Therefore, the change in the car's kinetic energy is equal to the work done by the frictional force of the car's brakes.
Therefore, we first need to determine the car's kinetic energy at the moment of braking using:

$$KE = \frac{1}{2}mv^2$$

This energy is equal to the work done by the brakes. We have the force applied by the brakes, and we can use:

$$W = F \cdot d$$

to determine the stopping distance.
Step 3 : Determine the kinetic energy of the car

$$\begin{aligned} KE &= \frac{1}{2}mv^2 \\ &= \frac{1}{2}(1\;000\,\mathrm{kg})(16{,}7\,\mathrm{m\cdot s^{-1}})^2 \\ &= 139\;445\,\mathrm{J} \end{aligned}$$

Step 4 : Determine the work done
Assume the stopping distance is d_0. Then the work done is:

$$W = F \cdot d$$

$$= (-8\;000\,\mathrm{N})(\mathrm{d_0})$$

The force has a negative sign because it acts in a direction opposite to the direction of motion.
Step 5 : Apply the work-enemy theorem
The change in kinetic energy is equal to the work done.

$$\begin{aligned}
\Delta KE &= W \\
KE_f - KE_i &= (-8\,000\,\mathrm{N})(\mathrm{d_0}) \\
0\,\mathrm{J} - 139\,445\,\mathrm{J} &= (-8\,000\,\mathrm{N})(\mathrm{d_0}) \\
\therefore d_0 &= \frac{139\,445\,\mathrm{J}}{8\,000\,\mathrm{N}} \\
&= 17{,}4\,\mathrm{m}
\end{aligned}$$

Step 6 : Write the final answer

The car stops in 17,4 m.

Important: A force only does work on an object for the time that it is in contact with the object. For example, a person pushing a trolley does work on the trolley, but the road does no work on the tyres of a car if they turn without slipping (the force is not applied over any distance because a different piece of tyre touches the road every instant.

Energy is conserved!

Important: Energy Conservation

In the absence of friction, the work done on an object by a system is equal to the energy gained by the object.

Work Done = Energy Transferred

In the presence of friction, only some of the energy lost by the system is transferred to useful energy. The rest is lost to friction.

Total Work Done = Useful Work Done + Work Done Against Friction

In the example of a falling mass the potential energy is known as *gravitational potential energy* as it is the gravitational force exerted by the earth which causes the mass to accelerate towards the ground. The gravitational field of the earth is what does the work in this case.

Another example is a rubber-band. In order to stretch a rubber-band we have to do work on it. This means we transfer energy to the rubber-band and it gains potential energy. This potential energy is called *elastic potential energy*. Once released, the rubber-band begins to move and elastic potential energy is transferred into kinetic energy.

Extension: Other forms of Potential Energy

1. elastic potential energy - potential energy is stored in a compressed or extended spring or rubber band. This potential energy is calculated by:

   $$\frac{1}{2}kx^2$$

   where k is a constant that is a measure of the stiffness of the spring or rubber band and x is the extension of the spring or rubber band.
2. Chemical potential energy is related to the making and breaking of chemical bonds. For example, a battery converts chemical energy into electrical energy.
3. The electrical potential energy of an electrically charged object is defined as the work that must be done to move it from an infinite distance away to its present location, in the absence of any non-electrical forces on the object.

This energy is non-zero if there is another electrically charged object nearby otherwise it is given by:

$$k\frac{q_1 q_2}{d}$$

where k is Coulomb's constant. For example, an electric motor lifting an elevator converts electrical energy into gravitational potential energy.

4. Nuclear energy is the energy released when the nucleus of an atom is split or fused. A nuclear reactor converts nuclear energy into heat.

Some of these forms of energy will be studied in later chapters.

Activity :: Investigation : Energy Resources

Energy can be taken from almost anywhere. Power plants use many different types of energy sources, including oil, coal, nuclear, biomass (organic gases), wind, solar, geothermal (the heat from the earth's rocks is very hot underground and is used to turn water to steam), tidal and hydroelectric (waterfalls). Most power stations work by using steam to turn turbines which then drive generators and create an electric current.

Most of these sources are dependant upon the sun's energy, because without it we would not have weather for wind and tides. The sun is also responsible for growing plants which decompose into fossil fuels like oil and coal. All these sources can be put under 2 headings, renewable and non-renewable. Renewable sources are sources which will not run out, like solar energy and wind power. Non-renewable sources are ones which will run out eventually, like oil and coal.

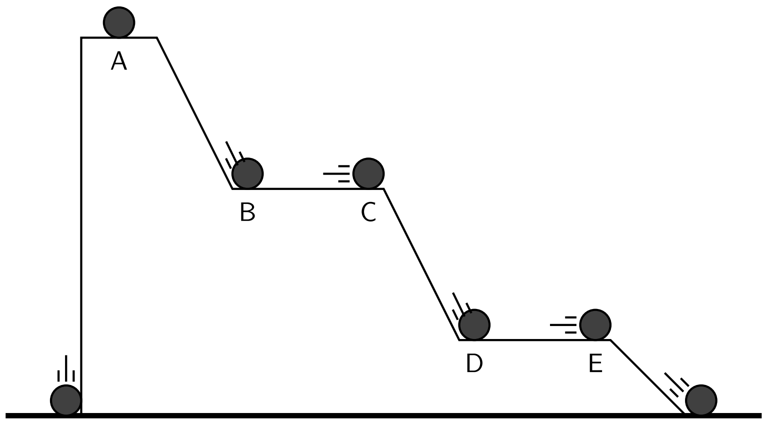
It is important that we learn to appreciate conservation in situations like this. The planet has a number of linked systems and if we don't appreciate the long-term consequences of our actions we run the risk of doing damage now that we will only suffer from in many years time.

Investigate two types of renewable and two types of non-renewable energy resources, listing advantages and disadvantages of each type. Write up the results as a short report.

?

Exercise: Energy

1. Fill in the table with the missing information using the positions of the ball in the diagram below combined with the work-energy theorem.



position	KE	PE	v
A		50 J	
B		30 J	
C			
D		10 J	
E			
F			
G			

2. A falling ball hits the ground at 10 m·s^{-1} in a vacuum. Would the speed of the ball be increased or decreased if air resistance were taken into account. Discuss using the work-energy theorem.
3. (NOTE TO SELF: Exercises are needed.)

23.4 Power

Now that we understand the relationship between work and energy, we are ready to look at a quantity that defines how long it takes for a certain amount of work to be done. For example, a mother pushing a trolley full of groceries can take 30 s or 60 s to push the trolley down an aisle. She does the same amount of work, but takes a different length of time. We use the idea of *power* to describe the rate at which work is done.

Definition: Power
Power is defined as the rate at which work is done or the rate at which energy is expended. The mathematical definition for power is:

$$P = F \cdot v \tag{23.3}$$

(23.3) is easily derived from the definition of work. We know that:

$$W = F \cdot d.$$

However, power is defined as the rate at which work is done. Therefore,

$$P = \frac{\Delta W}{\Delta t}.$$

This can be written as:

$$\begin{aligned} P &= \frac{\Delta W}{\Delta t} \\ &= \frac{\Delta(F \cdot d)}{\Delta t} \\ &= F\frac{\Delta d}{\Delta t} \\ &= F \cdot v \end{aligned}$$

The unit of power is watt (symbol W).

Activity :: Investigation : Watt
Show that the W is equivalent to $\mathrm{J} \cdot \mathrm{s}^{-1}$.

Interesting Fact

The unit watt is named after Scottish inventor and engineer James Watt (19 January 1736 - 19 August 1819) whose improvements to the steam engine were fundamental to the Industrial Revolution. A key feature of it was that it brought the engine out of the remote coal fields into factories.

Activity :: Research Project : James Watt

Write a short report 5 pages on the life of James Watt describing his many other inventions.

Interesting Fact

Historically, the *horsepower* (symbol hp) was the unit used to describe the power delivered by a machine. One horsepower is equivalent to approximately 750 W. The horsepower is sometimes used in the motor industry to describe the power output of an engine. Incidentally, the horsepower was derived by James Watt to give an indication of the power of his steam engine in terms of the power of a horse, which was what most people used to for example, turn a mill wheel.

Worked Example 155: Power Calculation 1

Question: Calculate the power required for a force of 10 N applied to move a 10 kg box at a speed of 1 ms over a frictionless surface.

Answer

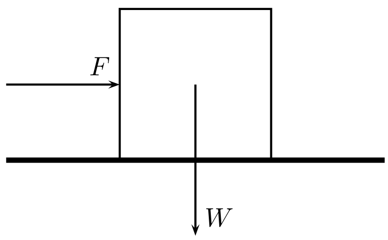
Step 1 : Determine what is given and what is required.

We are given:

- we are given the force, F=10 N
- we are given the speed, v=1 m·s^{-1}

We are required to calculate the power required.

Step 2 : Draw a force diagram



Step 3 : Determine how to approach the problem

From the force diagram, we see that the weight of the box is acting at right angles to the direction of motion. The weight does not contribute to the work done and does not contribute to the power calculation.

We can therefore calculate power from:

$$P = F \cdot v$$

Step 4 : Calculate the power required

$$\begin{aligned} P &= F \cdot v \\ &= (10\,\mathrm{N})(1\,\mathrm{m} \cdot \mathrm{s}^{-1}) \\ &= 10\,\mathrm{W} \end{aligned}$$

Step 5 : Write the final answer

10 W of power are required for a force of 10 N to move a 10 kg box at a speed of 1 ms over a frictionless surface.

Machines are designed and built to do work on objects. All machines usually have a power rating. The power rating indicates the rate at which that machine can do work upon other objects.

A car engine is an example of a machine which is given a power rating. The power rating relates to how rapidly the car can accelerate. Suppose that a 50 kW engine could accelerate the car from 0 $\mathrm{km} \cdot \mathrm{hr}^{-1}$ to 60$\mathrm{km} \cdot \mathrm{hr}^{-1}$ in 16 s. Then a car with four times the power rating (i.e. 200 kW) could do the same amount of work in a quarter of the time. That is, a 200 kW engine could accelerate the same car from 0 $\mathrm{km} \cdot \mathrm{hr}^{-1}$ to 60$\mathrm{km} \cdot \mathrm{hr}^{-1}$ in 4 s.

Worked Example 156: Power Calculation 2

Question: A forklift lifts a crate of mass 100 kg at a constant velocity to a height of 8 m over a time of 4 s. The forklift then holds the crate in place for 20 s. Calculate how much power the forklift exerts in lifting the crate? How much power does the forklift exert in holding the crate in place?

Answer

Step 1 : Determine what is given and what is required

We are given:

- mass of crate: m=100 kg
- height that crate is raised: h=8 m
- time taken to raise crate: t_r=4 s
- time that crate is held in place: t_s=20 s

We are required to calculate the power exerted.

Step 2 : Determine how to approach the problem

We can use:

$$P = F\frac{\Delta x}{\Delta t}$$

to calculate power. The force required to raise the crate is equal to the weight of the crate.

Step 3 : Calculate the power required to raise the crate

$$\begin{aligned} P &= F\frac{\Delta x}{\Delta t} \\ &= m \cdot g\frac{\Delta x}{\Delta t} \\ &= (100\,\mathrm{kg})(9{,}8\,\mathrm{m}\cdot\mathrm{s}^{-2})\frac{8\,\mathrm{m}}{4\,\mathrm{s}} \\ &= 1\ 960\,\mathrm{W} \end{aligned}$$

Step 4 : Calculate the power required to hold the crate in place
While the crate is being held in place, there is no displacement. This means there is no work done on the crate and therefore there is no power exerted.
Step 5 : Write the final answer
1 960 W of power is exerted to raise the crate and no power is exerted to hold the crate in place.

Activity :: Experiment : Simple measurements of human power
You can perform various physical activities, for example lifting measured weights or climbing a flight of stairs to estimate your output power, using a stop watch. Note: the human body is not very efficient in these activities, so your actual power will be much greater than estimated here.

?

Exercise: Power

1. [IEB 2005/11 HG] Which of the following is equivalent to the SI unit of power:
   A V·A
   B V·A^{-1}
   C $\mathrm{kg}\cdot$m·s^{-1}
   D $\mathrm{kg}\cdot\mathrm{m}\cdot\mathrm{s}^{-2}$
2. Two students, Bill and Bob, are in the weight lifting room of their local gum. Bill lifts the 50 kg barbell over his head 10 times in one minute while Bob lifts the 50 kg barbell over his head 10 times in 10 seconds. Who does the most work? Who delivers the most power? Explain your answers.
3. Jack and Jill ran up the hill. Jack is twice as massive as Jill; yet Jill ascended the same distance in half the time. Who did the most work? Who delivered the most power? Explain your answers.
4. Alex (mass 60 kg) is training for the Comrades Marathon. Part of Alex's training schedule involves push-ups. Alex does his push-ups by applying a force to elevate his center-of-mass by 20 cm. Determine the number of push-ups that Alex must do in order to do 10 J of work. If Alex does all this work in 60 s, then determine Alex's power.
5. When doing a chin-up, a physics student lifts her 40 kg body a distance of 0.25 m in 2 s. What is the power delivered by the student's biceps?
6. The unit of power that is used on a monthly electricity account is *kilowatt-hours* (symbol kWh). This is a unit of energy delivered by the flow of l kW of electricity for 1 hour. Show how many joules of energy you get when you buy 1 kWh of electricity.

7. An escalator is used to move 20 passengers every minute from the first floor of a shopping mall to the second. The second floor is located 5-meters above the first floor. The average passenger's mass is 70 kg. Determine the power requirement of the escalator in order to move this number of passengers in this amount of time.
8. (NOTE TO SELF: need a worked example - for example the minimum power required of an electric motor to pump water from a borehole of a particular depth at a particular rate)
9. (NOTE TO SELF: need a worked example -for example the power of different kinds of cars operating under different conditions.)
10. (NOTE TO SELF: Some exercises are needed.)

23.5 Important Equations and Quantities

Units				
Quantity	Symbol	Unit	S.I. Units	Direction
velocity	$\vec{v}$	—	$\frac{m}{s}$ **or** $m.s^{-1}$	✓
momentum	$\vec{p}$	—	$\frac{kg.m}{s}$ **or** $kg.m.s^{-1}$	✓
energy	E	J	$\frac{kg.m^2}{s^2}$ **or** $\text{kg}.m^2s^{-2}$	—
Work	W	J	$N.m$ **or** $kg.m^2.s^{-2}$	—
Kinetic Energy	E_K	J	$N.m$ **or** $kg.m^2.s^{-2}$	—
Potential Energy	E_P	J	$N.m$ **or** $kg.m^2.s^{-2}$	—
Mechanical Energy	U	J	$N.m$ **or** $kg.m^2.s^{-2}$	—

Table 23.1: Units commonly used in **Collisions and Explosions**

Momentum:

$$\vec{p} = m\vec{v} \tag{23.4}$$

Kinetic energy:

$$E_k = \frac{1}{2}m\vec{v}^2 \tag{23.5}$$

Principle of Conservation of Energy: Energy is never created nor destroyed, but is merely transformed from one form to another.

Conservation of Mechanical Energy: In the absence of friction, the total mechanical energy of an object is conserved.

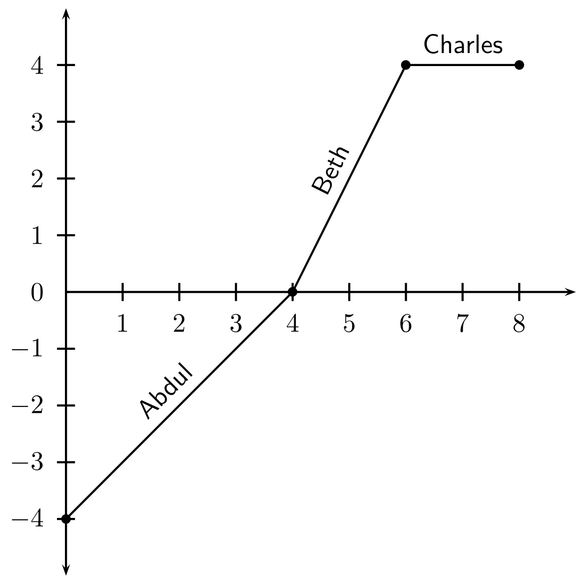
When a force moves in the direction along which it acts, work is done.

Work is the process of converting energy.

Energy is the ability to do work.

23.6 End of Chapter Exercises

1. The force vs. displacement graph shows the amount of force applied to an object by three different people. Abdul applies force to the object for the first 4 m of its displacement, Beth applies force from the 4 m point to the 6 m point, and Charles applies force from the 6 m point to the 8 m point. Calculate the work done by each person on the object? Which of the three does the most work on the object?



2. How much work does a person do in pushing a shopping trolley with a force of 200 N over a distance of 80 m in the direction of the force?

3. How much work does the force of gravity do in pulling a 20 kg box down a 45° frictionless inclined plane of length 18 m?

4. [IEB 2001/11 HG1] Of which one of the following quantities is kg.m^2.s^{-3} the base S.I. unit?

   A Energy

   B Force

   C Power

   D Momentum

5. [IEB 2003/11 HG1] A motor is used to raise a mass m through a vertical height h in time t.
   What is the power of the motor while doing this?

   A $mght$

   B $\frac{mgh}{t}$

   C $\frac{mgt}{h}$

   D $\frac{ht}{mg}$

6. [IEB 2002/11 HG1] An electric motor lifts a load of mass M vertically through a height h at a constant speed v. Which of the following expressions can be used to correctly calculate the power transferred by the motor to the load while it is lifted at a constant speed?

   A Mgh

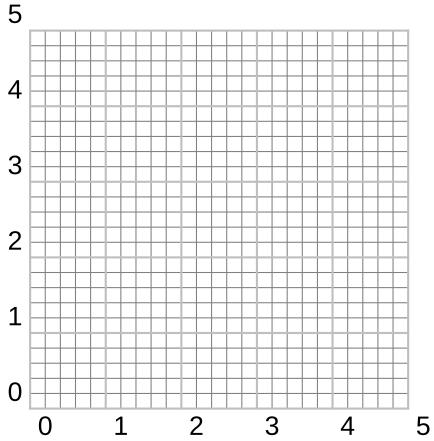
   B $Mgh + \frac{1}{2}\text{Mv}^2$

   C Mgv

   D $Mgv + \frac{1}{2}\,\frac{\text{Mv}^3}{\text{h}}$

7. [IEB 2001/11 HG1] An escalator is a moving staircase that is powered by an electric motor. People are lifted up the escalator at a constant speed of v through a vertical height h.

   What is the energy gained by a person of mass m standing on the escalator when he is lifted from the bottom to the top?



A mgh

B mgh $\sin\theta$

C $\frac{\text{mgh}}{\sin\theta}$

D $\frac{1}{2}\text{mv}^2$

8. [IEB 2003/11 HG1] In which of the following situations is there no work done on the object?

   A An apple falls to the ground.

   B A brick is lifted from the ground to the top of a building.

   C A car slows down to a stop.

   D A box moves at constant velocity across a frictionless horizontal surface.

9. (NOTE TO SELF: exercises are needed.)

Chapter 24

Doppler Effect - Grade 12

24.1 Introduction

Have you noticed how the pitch of a car hooter changes as the car passes by or how the pitch of a radio box on the pavement changes as you drive by? This effect is known as the **Doppler Effect** and will be studied in this chapter.

The Doppler Effect is named after Johann Christian Andreas Doppler (29 November 1803 - 17 March 1853), an Austrian mathematician and physicist who first explained the phenomenon in 1842.

24.2 The Doppler Effect with Sound and Ultrasound

As seen in the introduction, there are two situations which lead to the Doppler Effect:

1. When the source moves relative to the observer, for example the pitch of a car hooter as it passes by.
2. When the observer moves relative to the source, for example the pitch of a radio on the pavement as you drive by.

Definition: Doppler Effect
The Doppler effect is the apparent change in frequency and wavelength of a wave when the observer and the source of the wave move relative to each other.

We experience the Doppler effect quite often in our lives, without realising that it is science taking place. The changing sound of a taxi hooter or ambulance as it drives past are examples of this as you have seen in the introduction.
The question is how does the Doppler effect take place. Let us consider a source of sound waves with a constant frequency and amplitude. The sound waves can be drawn as concentric circles where each circle represents another wavefront, like in figure 24.1 below.

The sound source is the dot in the middle and is stationary. For the Doppler effect to take place, the source must be moving. Let's consider the following situation: The source (dot) emits one peak (represented by a circle) that moves away from the source at the same rate in all directions.

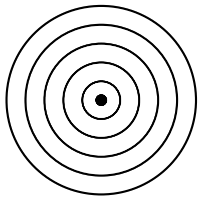
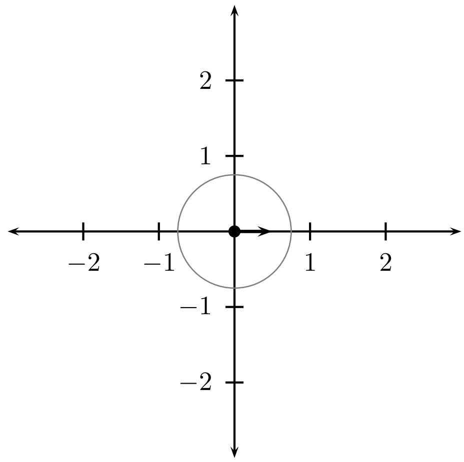
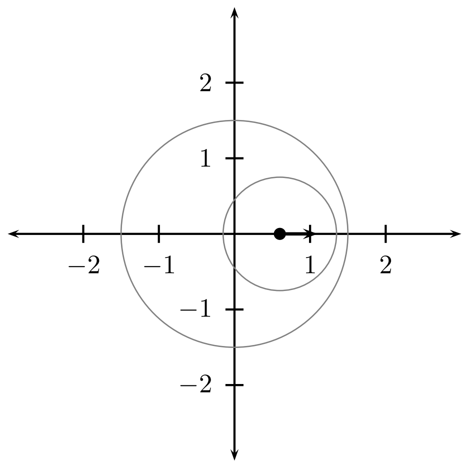


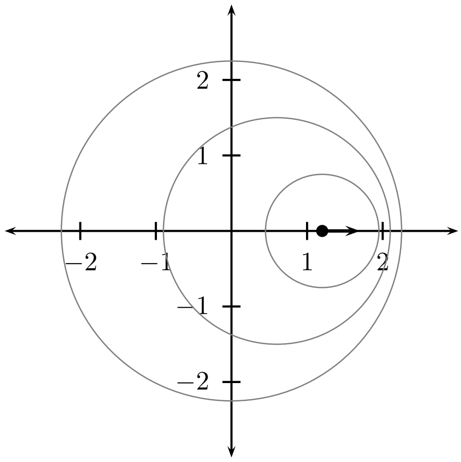
Figure 24.1: Stationary sound source



As this peak moves away, the source also moves and then emits the second peak. Now the two circles are not concentric any more, but on the one side they are closer together and on the other side they are further apart. This is shown in the next diagram.



If the source continues moving at the same speed in the same direction (i.e. with the same velocity which you will learn more about later). then the distance between peaks on the right of the source is the constant. The distance between peaks on the left is also constant but they are different on the left and right.



This means that the time between peaks on the right is less so the frequency is higher. It is higher than on the left and higher than if the source were not moving at all.

On the left hand side the peaks are further apart than on the right and further apart than if the source were at rest - this means the frequency is lower.
When a car appoaches you, the sound waves that reach you have a shorter wavelength and a higher frequency. You hear a higher sound. When the car moves away from you, the sound waves that reach you have a longer wavelength and lower frequency. You hear a lower sound. This change in frequency can be calculated by using:

$$f_L = \frac{v \pm v_L}{v \mp v_S} f_S \tag{24.1}$$

where f_L is the frequency perceived by the listener,
f_S is the frequency of the source,
v is the speed of the waves,
v_L the speed of the listener and
v_S the speed of the source.

Worked Example 157: The Doppler Effect for Sound

Question: The siren of an ambulance has a frequency of 700 Hz. You are standing on the pavement. If the ambulance drives past you at a speed of 20 m·s^{-1}, what frequency will you hear, when

a) the ambulance is approaching you

b) the ambulance is driving away from you

Take the speed of sound to be 340 m·s^{-1}.
Answer
Step 1 : Determine how to appoach the problem based on what is given

$$f_L = \frac{v \pm v_L}{v \mp v_S} f_S$$

$$\begin{aligned} f_s &= 700\mathrm{Hz} \\ v &= 340\,\mathrm{m\cdot s^{-1}} \\ v_L &= 0 \\ v_S &= -20\,\mathrm{m\cdot s^{-1}}\,\mathrm{for\,(a)\,and} \\ v_S &= 20\,\mathrm{m\cdot s^{-1}}\,\mathrm{for\,(b)} \end{aligned}$$

Step 2 : Determine f_L when ambulance is appoaching

$$\begin{aligned} f_L &= \frac{340+0}{340-20}(700) \\ &= 743{,}75\text{Hz} \end{aligned}$$

Step 3 : Determine f_L when ambulance has passed

$$\begin{aligned} f_L &= \frac{340+0}{340+20}(700) \\ &= 661{,}11\text{Hz} \end{aligned}$$

Worked Example 158: The Doppler Effect for Sound 2

Question: What is the frequency heard by a person driving at 15 m·s^{-1} toward a factory whistle that is blowing at a frequency of 800 Hz. Assume that the speed of sound is 340 m·s^{-1}.

Answer

Step 1 : Determine how to approach the problem based on what is given

We can use

$$f_L = \frac{v \pm v_L}{v \mp v_S} f_S$$

with:

$$\begin{aligned} v &= 340{,}6\,\mathrm{m \cdot s^{-1}} \\ v_L &= +15\,\mathrm{m \cdot s^{-1}} \\ v_S &= 0\,\mathrm{m \cdot s^{-1}} \\ f_S &= 800\ \mathrm{Hz} \\ f_L &= ? \end{aligned}$$

The listener is moving towards the source, so v_L is positive.

Step 2 : Calculate the frequency

$$\begin{aligned} f_L &= \frac{v \pm v_L}{v \mp v_S} f_S \\ &= \frac{340{,}6\,\mathrm{m \cdot s^{-1}} + 15\,\mathrm{m \cdot s^{-1}}}{340{,}6\,\mathrm{m \cdot s^{-1}} + 0\,\mathrm{m \cdot s^{-1}}}(800\ \mathrm{Hz}) \\ &= 835\ \mathrm{Hz} \end{aligned}$$

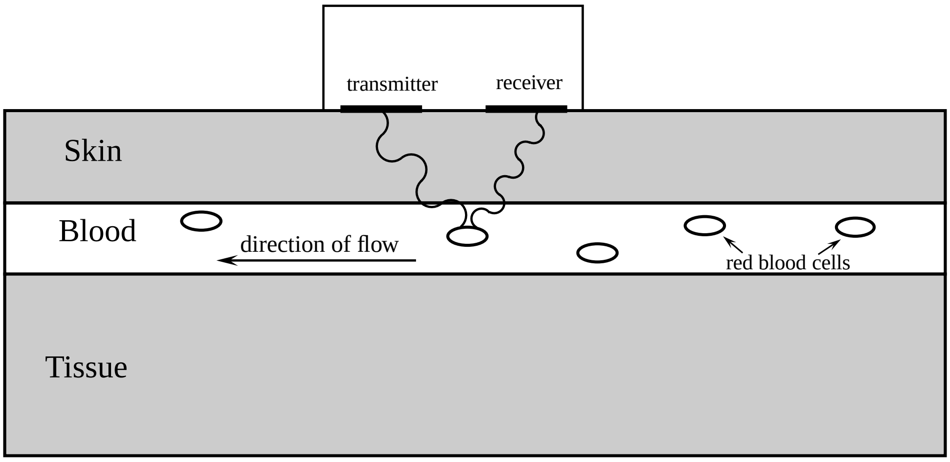
Step 3 : Write the final answer

The driver hears a frequency of 835 Hz.

Radar-based speed-traps use the Doppler Effect. The radar gun emits radio waves of a specific frequency. When the car is standing still, the waves reflected waves are the same frequency as the waves emitted by the radar gun. When the car is moving the Doppler frequency shift can be used to determine the speed of the car.

24.2.1 Ultrasound and the Doppler Effect

Ultrasonic waves (ultrasound) are sound waves with a frequency greater than 20 000 Hz (the upper limit of hearing). These waves can be used in medicine to determine the direction of blood flow. The device, called a Doppler flow meter, sends out sound waves. The sound waves can travle through skin and tissue and will be reflected by moving objects in the body (like blood). The reflected waves return to the flow meter where its frequency (received frequency) is compared to the transmitted frequency. Because of the Doppler effect, blood that is moving towards the flow meter will change the sound to a higher frequency (blue shift) and blood that is moving away from the flow meter will cause a lower frequency (red shift).



Ultrasound can be used to determine whether blood is flowing in the right direction in the circulation system of unborn babies, or identify areas in the body where blood flow is restricted due to narrow veins. The use of ultrasound equipment in medicine is called sonography or ultrasonography.

?

Exercise: The Doppler Effect with Sound

1. Suppose a train is approaching you as you stand on the platform at the station. As the train approaches the station, it slows down. All the while, the engineer is sounding the hooter at a constant frequency of 400 Hz. Describe the pitch and the changes in pitch that you hear.
2. Passengers on a train hear its whistle at a frequency of 740 Hz. Anja is standing next to the train tracks. What frequency does Anja hear as the train moves directly toward her at a speed of 25 m·s^{-1}?
3. A small plane is taxiing directly away from you down a runway. The noise of the engine, as the pilot hears it, has a frequency 1,15 times the frequency that you hear. What is the speed of the plane?
4. A Doppler flow meter detected a blue shift in frequency while determining the direction of blood flow. What does a "blue shift" mean and how does it take place?

24.3 The Doppler Effect with Light

Light is a wave and earlier you learnt how you can study the properties of one wave and apply the same ideas to another wave. The same applies to sound and light. We know the Doppler effect affects sound waves when the source is moving. Therefore, if we apply the Doppler effect to light, the frequency of the emitted light should change when the source of the light is moving relative to the observer.

When the frequency of a sound wave changes, the sound you hear changes. When the frequency of light changes, the colour you would see changes.
This means that the Doppler effect can be observed by a change in sound (for sound waves) and a change in colour (for light waves). Keep in mind that there are sounds that we cannot hear (for example ultrasound) and light that we cannot see (for example ultraviolet light).

We can apply all the ideas that we learnt about the Doppler effect to light. When talking about light we use slightly different names to describe what happens. If you look at the colour spectrum (more details Chapter 30) then you will see that blue light has shorter wavelengths than red light. If you are in the middle of the visible colours then longer wavelengths are more red and shorter wavelengths are more blue. So we call shifts towards longer wavelengths "red-shifts" and shifts towards shorter wavelengths "blue-shifts".

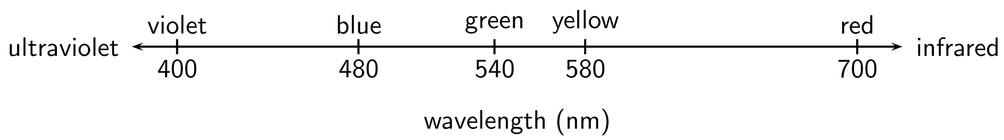


Figure 24.2: Blue light has shorter wavelengths than red light.

A shift in wavelength is the same as a shift in frequency. Longer wavelengths of light have lower frequencies and shorter wavelengths have higher frequencies. From the Doppler effect we know that when things move towards you any waves they emit that you measure are shifted to shorter wavelengths (blueshifted). If things move away from you, the shift is to longer wavelengths (redshifted).

24.3.1 The Expanding Universe

Stars emit light, which is why we can see them at night. Galaxies are huge collections of stars. An example is our own Galaxy, the Milky Way, of which our sun is only one of the millions of stars! Using large telescopes like the Southern African Large Telescope (SALT) in the Karoo, astronomers can measure the light from distant galaxies. The spectrum of light (see Chapter **??**) can tell us what elements are in the stars in the galaxies because each element emits/absorbs light at particular wavelengths (called spectral lines). If these lines are observed to be shifted from their usual wavelengths to shorter wavelengths, then the light from the galaxy is said to be *blueshifted*. If the spectral lines are shifted to longer wavelengths, then the light from the galaxy is said to be *redshifted*. If we think of the blueshift and redshift in Doppler effect terms, then a blueshifted galaxy would appear to be moving *towards* us (the observers) and a redshifted galaxy would appear to be moving *away* from us.

Important:

- If the light source is moving away from the observer (positive velocity) then the observed frequency is lower and the observed wavelength is greater (redshifted).
- If the source is moving towards (negative velocity) the observer, the observed frequency is higher and the wavelength is shorter (blueshifted).

Edwin Hubble (20 November 1889 - 28 September 1953) measured the Doppler shift of a large sample of galaxies. He found that the light from distant galaxies is *redshifted* and he discovered that there is a proportionality relationship between the *redshift* and the *distance* to the galaxy. Galaxies that are further away always appear more redshifted than nearby galaxies. Remember that a redshift in Doppler terms means a velocity of the light source *away* from the observer. So why do all distant galaxies appear to be moving away from our Galaxy?

The reason is that the universe is expanding! The galaxies are not actually moving themselves, rather the *space* between them is expanding!

24.4 Summary

1. The Doppler Effect is the apparent change in frequency and wavelength of a wave when the observer and source of the wave move relative to each other.
2. The following equation can be used to calculate the frequency of the wave according to the observer or listener:
$$f_L = \frac{v \pm v_L}{v \mp v_S} f_S$$
3. If the direction of the wave from the listener to the source is chosen as positive, the velocities have the following signs.

Source moves towards listener	v_S : negative
Source moves away from listener	v_S : positive
Listener moves towards source	v_L : positive
Listener moves away from source	v_L : negative

4. The Doppler Effect can be observed in all types of waves, including ultrasound, light and radiowaves.
5. Sonography makes use of ultrasound and the Doppler Effect to determine the direction of blood flow.
6. Light is emitted by stars. Due to the Doppler Effect, the frequency of this light decreases and the starts appear red. This is called a red shift and means that the stars are moving away from the Earth. This means that the Universe is expanding.

24.5 End of Chapter Exercises

1. Write a definition for each of the following terms.
   A Doppler Effect
   B Red-shift
   C Ultrasound
2. Explain how the Doppler Effect is used to determine the direction of blood flow in veins.
3. The hooter of an appoaching taxi has a frequency of 500 Hz. If the taxi is travelling at 30 m·s^{-1}and the speed of sound is 300 m·s^{-1}, calculate the frequency of sound that you hear when
   A the taxi is approaching you.
   B the taxi passed you and is driving away.
4. A truck approaches you at an unknown speed. The sound of the trucks engine has a frequency of 210 Hz, however you hear a frequency of 220 Hz. The speed of sound is 340 m·s^{-1}.
   A Calculate the speed of the truck.
   B How will the sound change as the truck passes you? Explain this phenomenon in terms of the wavelength and frequency of the sound.
5. A police car is driving towards a fleeing suspect. The frequency of the police car's siren is 400 Hz at $\frac{v}{35}$, where v is the speed of sound. The suspect is running away at $\frac{v}{68}$. What frequency does the suspect hear?
6. A Why are ultrasound waves used in sonography and not sound waves?
   B Explain how the Doppler effect is used to determine the direction of flow of blood in veins.

Chapter 25

Colour - Grade 12

25.1 Introduction

We call the light that we humans can see 'visible light'. Visible light is actually just a small part of the large spectrum of electromagnetic radiation which you will learn more about in Chapter 30. We can think of electromagnetic radiation and visible light as transverse waves. We know that transverse waves can be described by their amplitude, frequency (or wavelength) and velocity. The velocity of a wave is given by the product of its frequency and wavelength:

$$v = f \times \lambda \tag{25.1}$$

However, electromagnetic radiation, including visible light, is special because, no matter what the frequency, it all moves at a **constant velocity** (in vacuum) which is known as the speed of light. The speed of light has the symbol c and is:

$$c = 3 \times 10^8 \text{ m.s}^{-1}$$

Since the *speed of light* is c, we can then say:

$$c = f \times \lambda \tag{25.2}$$

25.2 Colour and Light

Our eyes are sensitive to visible light over a range of wavelengths from 390 nm to 780 nm (1 nm $= 1 \times 10^{-9}$ m). The different **colours** of light we see are related to specific *frequencies* (and *wavelengths*) of visible light. The wavelengths and frequencies are listed in table 25.1.

Colour	Wavelength range (nm)	Frequency range (Hz)
violet	390 - 455	769 - 659 $\times 10^{12}$
blue	455 - 492	659 - 610 $\times 10^{12}$
green	492 - 577	610 - 520 $\times 10^{12}$
yellow	577 - 597	520 - 503 $\times 10^{12}$
orange	597 - 622	503 - 482 $\times 10^{12}$
red	622 - 780	482 - 385 $\times 10^{12}$

Table 25.1: Colours, wavelengths and frequencies of light in the visible spectrum.

You can see from table 25.1 that **violet** light has the *shortest wavelengths* and *highest frequencies* while **red** light has the *longest wavelengths* and *lowest frequencies*.

Worked Example 159: Calculating the frequency of light given the wavelength

Question: A streetlight emits light with a wavelength of 520 nm.

1. What colour is the light? (Use table 25.1 to determine the colour)
2. What is the frequency of the light?

Answer

Step 1 : What is being asked and what information are we given?

We need to determine the colour and frequency of light with a wavelength of $\lambda = 520$ nm $= 520 \times 10^{-9}$ m.

Step 2 : Compare the wavelength of the light to those given in table 25.1

We see from table 25.1 that light with wavelengths between 492 - 577 nm is green. 520 nm falls into this range, therefore the colour of the light is green.

Step 3 : Next we need to calculate the frequency of the light

We know that

$$c = f \times \lambda$$

We know c and we are given that $\lambda = 520 \times 10^{-9}$ m. So we can substitute in these values and solve for the frequency f. (**NOTE:** Don't forget to always change units into S.I. units! 1 nm $= 1 \times 10^{-9}$ m.)

$$\begin{aligned} f &= \frac{c}{\lambda} \\ &= \frac{3 \times 10^8}{520 \times 10^{-9}} \\ &= 577 \times 10^{12} \text{ Hz} \end{aligned}$$

The frequency of the green light is 577×10^{12} Hz

Worked Example 160: Calculating the wavelength of light given the frequency

Question: A streetlight also emits light with a frequency of 490$\times 10^{12}$ Hz.

1. What colour is the light? (Use table 25.1 to determine the colour)
2. What is the wavelength of the light?

Answer

Step 1 : What is being asked and what information are we given?

We need to find the colour and wavelength of light which has a frequency of 490$\times 10^{12}$ Hz and which is emitted by the streetlight.

Step 2 : Compare the wavelength of the light to those given in table 25.1

We can see from table 25.1 that orange light has frequencies between 503 - 482$\times 10^{12}$ Hz. The light from the streetlight has $f = 490 \times 10^{12}$ Hz which fits into this range. Therefore the light must be orange in colour.

Step 3 : Next we need to calculate the wavelength of the light

We know that

$$c = f \times \lambda$$

We know $c = 3 \times 10^8 \text{ m.s}^{-1}$ and we are given that $f = 490 \times 10^{12}$ Hz. So we can substitute in these values and solve for the wavelength λ.

$$\begin{aligned}
\lambda &= \frac{c}{f} \\
&= \frac{3 \times 10^{8}}{490 \times 10^{12}} \\
&= 6.122 \times 10-7 \ \ \mathrm{m} \\
&= 612 \times 10^{-9} \ \ \mathrm{m} \\
&= 612 \ \ \mathrm{nm}
\end{aligned}$$

Therefore the orange light has a wavelength of 612 nm.

Worked Example 161: Frequency of Green

Question: The wavelength of green light ranges between 500 nm an d 565 nm. Calculate the range of frequencies that correspond to this range of wavelengths.
Answer
Step 1 : Determine how to approach the problem
Use

$$c = f \times \lambda$$

to determine f.
Step 2 : Calculate frequency corresponding to upper limit of wavelength range

$$\begin{aligned}
c &= f \times \lambda \\
f &= \frac{c}{\lambda} \\
&= \frac{3 \times 10^{8}\,\mathrm{m \cdot s^{-1}}}{565 \times 10^{-9}\,\mathrm{m}} \\
&= 5{,}31 \times 10^{14}\ \mathrm{Hz}
\end{aligned}$$

Step 3 : Calculate frequency corresponding to lower limit of wavelength range

$$\begin{aligned}
c &= f \times \lambda \\
f &= \frac{c}{\lambda} \\
&= \frac{3 \times 10^{8}\,\mathrm{m \cdot s^{-1}}}{500 \times 10^{-9}\,\mathrm{m}} \\
&= 6{,}00 \times 10^{14}\ \mathrm{Hz}
\end{aligned}$$

Step 4 : Write final answer
The range of frequencies of green light is $5{,}31 \times 10^{14}$ Hz to $6{,}00 \times 10^{14}$ Hz.

Exercise: Calculating wavelengths and frequencies of light

1. Calculate the frequency of light which has a wavelength of 400 nm. (Remember to use S.I. units)

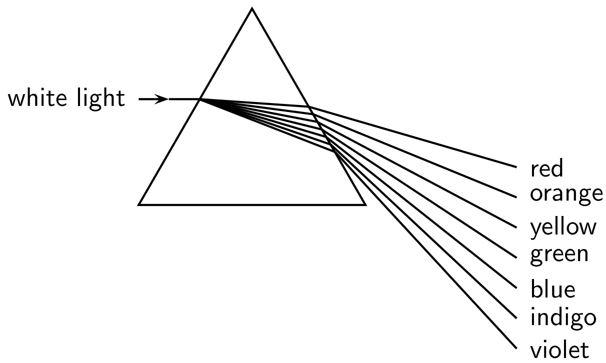
2. Calculate the wavelength of light which has a frequency of 550×10^{12} Hz.
3. What colour is light which has a wavelength of 470×10^{9} m and what is its frequency?
4. What is the wavelength of light with a frequency of 510×10^{12} Hz and what is its color?

25.2.1 Dispersion of white light

White light, like the light which comes from the sun, is made up of *all* the visible wavelengths of light. In other words, white light is a *combination* of all the colours of visible light.

In Chapter 7, you learnt that the speed of light is different in different substances. The speed of light in different substances depends on the frequency of the light. For example, when white light travels through glass, light of the different frequencies is slowed down by different amounts. The lower the frequency, the less the speed is reduced which means that red light (lowest frequency) is slowed down *less* than violet light (highest frequency). We can see this when white light is incident on a glass prism.

Have a look at the picture below. When the white light hits the edge of the prism, the light which travels through the glass is refracted as it moves from the less dense medium (air) to the more dense medium (glass).



- The **red** light which is slowed down the *least*, is refracted the *least*.
- The **violet** light which is slowed down the *most*, is refracted the *most*.

When the light hits the other side of the prism it is again refracted but the angle of the prism edge allows the light to remain separated into its different colours. White light is therefore separated into its different colours by the prism and we say that the white light has been **dispersed** by the prism.

The dispersion effect is also responsible for why we see rainbows. When sunlight hits drops of water in the atmosphere, the white light is dispersed into its different colours by the water.

25.3 Addition and Subtraction of Light

25.3.1 Additive Primary Colours

The primary colours of light are **red**, **green** and **blue**. When all the primary colours are superposed (added together), white light is produced. Red, green and blue are therefore called the *additive primary colours*. All the other colours can be produced by different combinations of red, green and blue.

25.3.2 Subtractive Primary Colours

The subtractive primary colours are obtained by subtracting one of the three additive primary colours from white light. The subtractive primary colours are **yellow**, **magenta** and **cyan**. Magenta appears as a pinkish-purplish colour and cyan looks greenish-blue. You can see how the primary colours of light add up to the different subtractive colours in the illustration below.

red + green + blue = white

PRIMARY COLOURS	SUBTRACTIVE PRIMARY COLOURS
red + green	= yellow
red + blue	= magenta
green + blue	= cyan

Activity :: Experiment : Colours of light

Aim:
To investigate the additive properties of colours and determine the complementary colours of light.

Apparatus:
You will need two battery operated torches with flat bulb fronts, a large piece of white paper, and some pieces of cellophane paper of the following colours: red, blue, green, yellow, cyan, magenta. (You should easily be able to get these from a newsagents.)

Make a table in your workbook like the one below:

Colour 1	Colour 2	Final colour prediction	Final colour measured
red	blue		
red	green		
green	blue		
magenta	green		
yellow	blue		
cyan	red		

Before you begin your experiment, use what you know about colours of light to write down in the third column "Final colour prediction", what you think the result of adding the two colours of light will be. You will then be able to test your predictions by making the following measurements:

Method:
Proceed according to the table above. Put the correct colour of cellophane paper over each torch bulb. e.g. the first test will be to put red cellophane on one torch and blue cellophane on the other. Switch on the torch with the red cellophane over it and shine it onto the piece of white paper.
What colour is the light?
Turn off that torch and turn on the one with blue cellophane and shine it onto the white paper.
What colour is the light?
Now shine both torches with their cellophane coverings onto the same spot on the white paper. What is the colour of the light produced? Write this down in the fourth column of your table.
Repeat the experiment for the other colours of cellophane so that you can complete your table.

Questions:

1. How did your predictions match up to your measurements?
2. Complementary colours of light are defined as the colours of light which, when added to one of the primary colours, produce white light. From your completed table, write down the complementary colours for red, blue and green.

25.3.3 Complementary Colours

Complementary colours are two colours of light which add together to give white.

Activity :: Investigation : Complementary colours for red, green and blue
Complementary colours are two colours which add together to give white. Place a tick in the box where the colours in the first column added to the colours in the top row give white.

	magenta (=red+blue)	**yellow** (=red+green)	**cyan** (=blue+green)
red			
green			
blue			

You should have found that the complementary colours for red, green and blue are:

- Red and Cyan
- Green and Magenta
- Blue and Yellow

25.3.4 Perception of Colour

The light-sensitive lining on the back inside half of the human eye is called the retina. The retina contains two kinds of light sensitive cells or *photoreceptors*: the rod cells (sensitive to low light) and the cone cells (sensitive to normal daylight) which enable us to see. The rods are not very sensitive to colour but work well in dimly lit conditions. This is why it is possible to see in a dark room, but it is hard to see any colours. Only your rods are sensitive to the low light levels and so you can only see in black, white and grey. The cones enable us to see colours. Normally, there are three kinds of cones, each containing a different pigment. The cones are activated when the pigments absorb light. The three types of cones are sensitive to (i.e. absorb) red, blue and green light respectively. Therefore we can perceive *all* the different colours in the visible spectrum when the different types of cones are stimulated by different amounts since they are just combinations of the three primary colours of light.

The rods and cones have different response times to light. The cones react quickly when bright light falls on them. The rods take a longer time to react. This is why it takes a while (about 10 minutes) for your eyes to adjust when you enter a dark room after being outside on a sunny day.

Color blindness in humans is the inability to perceive differences between some or all colors that other people can see. Most often it is a genetic problem, but may also occur because of eye, nerve, or brain damage, or due to exposure to certain chemicals. The most common forms of human color blindness result from problems with either the middle or long wavelength sensitive cone systems, and involve difficulties in discriminating reds, yellows, and greens from one another. This is called "red-green color blindness". Other forms of color blindness are much rarer. They include problems in discriminating blues from yellows, and the rarest forms of all, complete color blindness or monochromasy, where one cannot distinguish any color from grey, as in a black-and-white movie or photograph.

Worked Example 162: Seeing Colours

Question: When blue and green light fall on an eye, is cyan light being created? Discuss.

Answer

Cyan light is not created when blue and green light fall on the eye. The blue and green receptors are stimulated to make the brain believe that cyan light is being created.

25.3.5 Colours on a Television Screen

If you look very closely at a colour cathode-ray television screen or computer screen, you will see that there are very many small red, green and blue dots called *phosphors* on it. These dots are caused to fluoresce (glow brightly) when a beam of electrons from the cathode-ray tube behind the screen hits them. Since different combinations of the three primary colours of light can produce any other colour, only red, green and blue dots are needed to make pictures containing all the colours of the visible spectrum.

Exercise: Colours of light

1. List the three primary colours of light.
2. What is the term for the phenomenon whereby white light is split up into its different colours by a prism?
3. What is meant by the term "complementary colour" of light?
4. When white light strikes a prism which colour of light is refracted the most and which is refracted the least? Explain your answer in terms of the speed of light in a medium.

25.4 Pigments and Paints

We have learnt that white light is a combination of all the colours of the visible spectrum and that each colour of light is related to a different frequency. But what gives everyday objects around us their different colours?

Pigments are substances which give an object its colour by absorbing certain frequencies of light and reflecting other frequencies. For example, a red pigment absorbs all colours of light except red which it reflects. Paints and inks contain pigments which gives the paints and inks different colours.

25.4.1 Colour of opaque objects

Objects which you *cannot* see through (i.e. they are not transparent) are called **opaque**. Examples of some opaque objects are metals, wood and bricks. The colour of an opaque object is determined by the colours (therefore *frequencies*) of light which it *reflects*. For example, when white light strikes a blue opaque object such as a ruler, the ruler will absorb all frequencies of light *except* blue, which will be reflected. The reflected blue light is the light which makes it into our eyes and therefore the object will appear blue.

Opaque objects which appear white do not absorb any light. They reflect all the frequencies. Black opaque objects absorb all frequencies of light. They do not reflect at all and therefore appear to have no colour.

Worked Example 163: Colour of Opaque Objects

Question: If we shine white light on a sheet of paper that can only reflect green light, what is the colour of the paper?

Answer

Since the colour of an object is determined by that frequency of light that is *reflected*, the sheet of paper will appear green, as this is the only frequency that is reflected. All the other frequencies are absorbed by the paper.

Worked Example 164: Colour of an opaque object II

Question: The cover of a book appears to have a magenta colour. What colours of light does it reflect and what colours does it absorb?

Answer

We know that magenta is a combination of red and blue primary colours of light. Therefore the object must be reflecting blue and red light and absorb green.

25.4.2 Colour of transparent objects

If an object is **transparent** it means that you can see through it. For example, glass, clean water and some clear plastics are transparent. The colour of a transparent object is determined by the colours (frequencies) of light which it *transmits* (allows to pass through it). For example, a cup made of green glass will appear green because it absorbs all the other frequencies of light *except* green, which it transmits. This is the light which we receive in our eyes and the object appears green.

Worked Example 165: Colour of Transparent Objects

Question: If white light is shone through a glass plate that absorbs light of all frequencies except red, what is the colour of the glass plate?
Answer
Since the colour of an object is determined by that frequency of light that is *transmitted*, the glass plate will appear red, as this is the only frequency that is not absorbed.

25.4.3 Pigment primary colours

The primary pigments and paints are **cyan**, **magenta** and **yellow**. When pigments or paints of these three colours are mixed together in equal amounts they produce **black**. Any other colour of paint can be made by mixing the primary pigments together in different quantities. The primary pigments are related to the primary colours of light in the following way:

PRIMARY PIGMENTS

$$\boxed{\text{cyan} + \text{magenta} + \text{yellow} = \text{black}}$$

PRIMARY PIGMENTS		PRIMARY COLOURS OF LIGHT
cyan + magenta	=	blue
cyan + yellow	=	green
magenta + yellow	=	red

Colour printers only use 4 colours of ink: cyan, magenta, yellow and black. All the other colours can be mixed from these!

Worked Example 166: Pigments

Question: What colours of light are absorbed by a green pigment?
Answer
If the pigment is green, then green light must be *reflected*. Therefore, red and blue light are absorbed.

Worked Example 167: Primary pigments

Question: I have a ruler which reflects red light and absorbs all other colours of light. What colour does the ruler appear in white light? What primary pigments must have been mixed to make the pigment which gives the ruler its colour?
Answer
Step 1 : What is being asked and what are we given?
We need to determine the colour of the ruler and the pigments which were mixed to make the colour.
Step 2 : An opaque object appears the colour of the light it reflects
The ruler reflects red light and absorbs all other colours. Therefore the ruler appears to be red.
Step 3 : What pigments need to be mixed to get red?
Red pigment is produced when magenta and yellow pigments are mixed. Therefore magenta and yellow pigments were mixed to make the red pigment which gives the ruler its colour.

Worked Example 168: Paint Colours

Question: If cyan light shines on a dress that contains a pigment that is capable of absorbing blue, what colour does the dress appear?
Answer
Step 1 : Determine the component colours of cyan light
Cyan light is made up of blue and green light.
Step 2 : Determine solution
If the dress absorbs the blue light then the green light must be reflected, so the dress will appear green!

25.5 End of Chapter Exercises

1. Calculate the wavelength of light which has a frequency of 570×10^{12} Hz.
2. Calculate the frequency of light which has a wavelength of 580 nm.
3. Complete the following sentence: When white light is dispersed by a prism, light of the colour ? is refracted the most and light of colour ? is refracted the least.
4. What are the two types of photoreceptor found in the retina of the human eye called and which type is sensitive to colours?
5. What color do the following shirts appear to the human eye when the lights in a room are turned off and the room is completely dark?

   A red shirt

   B blue shirt

   C green shirt
6. Two light bulbs, each of a different colour, shine on a sheet of white paper. Each light bulb can be a primary colour of light - red, green, and blue. Depending on which primary colour of light is used, the paper will appear a different color. What colour will the paper appear if the lights are:

   A red and blue?

   B red and green?

C green and blue?

7. Match the primary colour of light on the left to its complementary colour on the right:

Column A	Column B
red	yellow
green	cyan
blue	magenta

8. Which combination of colours of light gives magenta?

    A red and yellow

    B green and red

    C blue and cyan

    D blue and red

9. Which combination of colours of light gives cyan?

    A yellow and red

    B green and blue

    C blue and magenta

    D blue and red

10. If yellow light falls on an object whose pigment absorbs green light, what colour will the object appear?

11. If yellow light falls on a blue pigment, what colour will it appear?

Chapter 26

2D and 3D Wavefronts - Grade 12

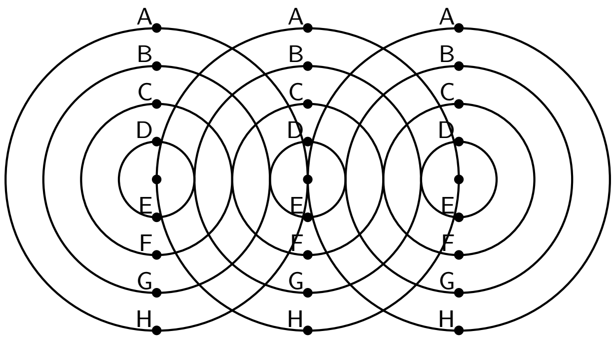
26.1 Introduction

You have learnt about the basic principles of reflection and refraction. In this chapter, you will learn about phenomena that arise with waves in two and three dimensions: interference and diffraction.

26.2 Wavefronts

Activity :: Investigation : Wavefronts

The diagram shows three identical waves being emitted by three point sources. All points marked with the same letter are in phase. Join all points with the same letter.



What type of lines (straight, curved, etc) do you get? How does this compare to the line that joins the sources?

Consider three point sources of waves. If each source emits waves isotropically (i.e. the same in all directions) we will get the situation shown in as shown in Figure 26.1.

We define a **wavefront** as the imaginary line that joins waves that are in phase. These are indicated by the grey, vertical lines in Figure 26.1. The points that are in phase can be peaks, troughs or anything in between, it doesn't matter which points you choose as long as they are in phase.

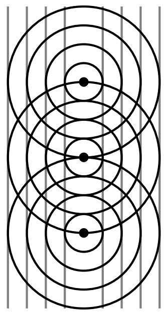


Figure 26.1: Wavefronts are imaginary lines joining waves that are in phase. In the example, the wavefronts (shown by the grey, vertical lines) join all waves at the crest of their cycle.

26.3 The Huygens Principle

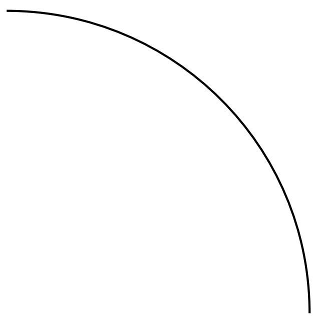
Christiaan Huygens described how to determine the path of waves through a medium.

Definition: The Huygens Principle
Each point on a wavefront acts like a point source of circular waves. The waves emitted from these point sources interfere to form another wavefront.

A simple example of the Huygens Principle is to consider the single wavefront in Figure 26.2.

Worked Example 169: Application of the Huygens Principle

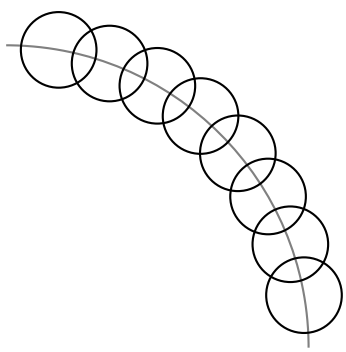
Question: Given the wavefront,



use the Huygens Principle to determine the wavefront at a later time.

Answer

Step 1 : Draw circles at various points along the given wavefront



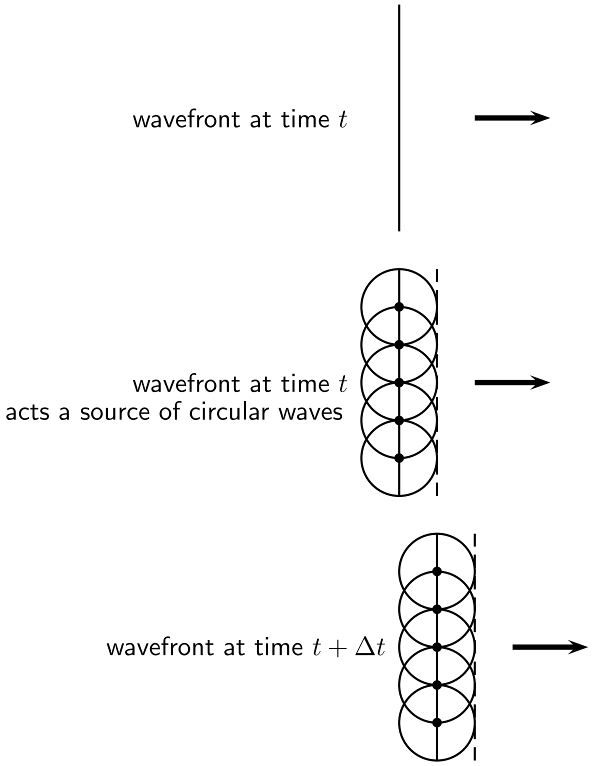
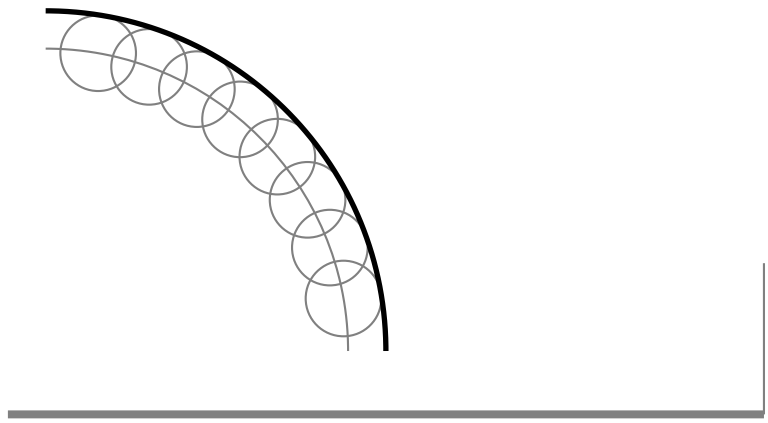


Figure 26.2: A single wavefront at time t acts as a series of point sources of circular waves that interfere to give a new wavefront at a time $t + \Delta t$. The process continues and applies to any shape of waveform.

Step 2 : Join the crests of each circle to get the wavefront at a later time



Interesting Fact

Christiaan Huygens (14 April 1629 - 8 July 1695), was a Dutch mathematician, astronomer and physicist; born in The Hague as the son of Constantijn Huygens. He studied law at the University of Leiden and the College of Orange in Breda before turning to science. Historians commonly associate Huygens with the scientific revolution.

Huygens generally receives minor credit for his role in the development of modern calculus. He also achieved note for his arguments that light consisted of waves; see: wave-particle duality. In 1655, he discovered Saturn's moon Titan. He also examined Saturn's planetary rings, and in 1656 he discovered that those rings consisted of rocks. In the same year he observed and sketched the Orion Nebula. He also discovered several interstellar nebulae and some double stars.

26.4 Interference

Interference occurs when two identical waves pass through the same region of space at the same time resulting in a superposition of waves. There are two types of interference which is of interest: **constructive** interference and **destructive** interference.

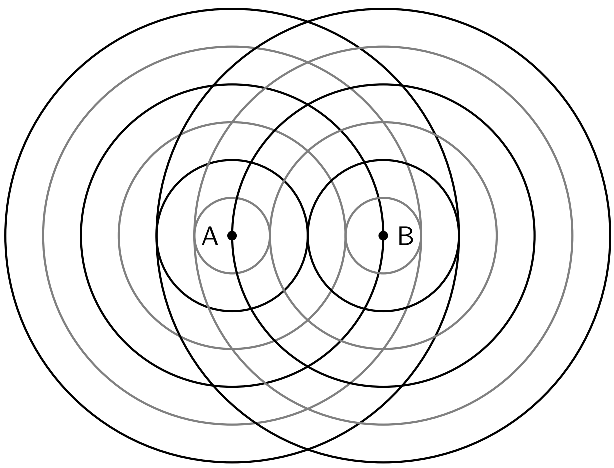
Constructive interference occurs when both waves have a displacement in the same direction, while destructive interference occurs when one wave has a displacement in the opposite direction to the other, thereby resulting in a cancellation. There is no displacement of the medium in destructive interference while for constructive interference the displacement of the medium is greater than the individual displacements.

Constructive interference occurs when both waves have a displacement in the same direction, this means they both have a peak or they both have a trough at the same place at the same time. If they both have a peak then the peaks add together to form a bigger peak. If they both have a trough then the trough gets deeper.

Destructive interference occurs when one wave has a displacement in the opposite direction to the other, this means that the one wave has a peak and the other wave has a trough. If the waves have identical magnitudes then the peak "fills" up the trough and the medium will look like there are no waves at that point. There will be no displacement of the medium. A place where destructive interference takes places is called a node.

Waves can interfere at places where there is never a trough and trough or peak and peak or trough and peak at the same time. At these places the waves will add together and the resultant displacement will be the sum of the two waves but they won't be points of maximum interference.

Consider the two identical waves shown in the picture below. The wavefronts of the peaks are shown as black lines while the wavefronts of the troughs are shown as grey lines. You can see that the black lines cross other black lines in many places. This means two peaks are in the same place at the same time so we will have constructive interference where the two peaks add together to form a bigger peak.



Two points sources (A and B) radiate identical waves. The wavefronts of the peaks (black lines) and troughs (grey lines) are shown. Constructive interference occurs where two black lines intersect or where two gray lines intersect. Destructive interference occurs where a black line intersects with a grey line.

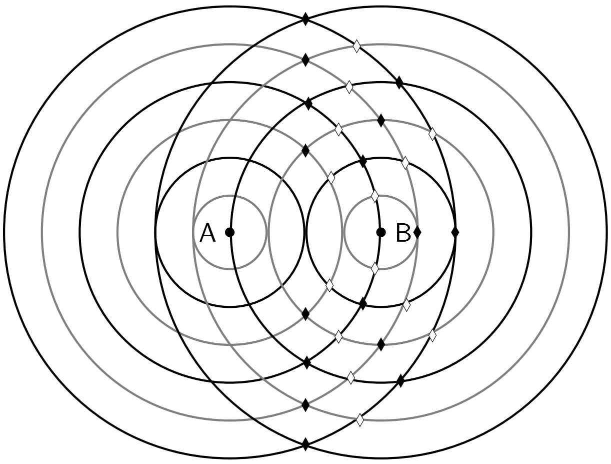
You can see that the black lines cross other black lines in many places. This means two peaks are in the same place at the same time so we will have constructive interference where the two peaks add together to form a bigger peak.

When the grey lines cross other grey lines there are two troughs are in the same place at the same time so we will have constructive interference where the two troughs add together to form a bigger trough.

In the case where a grey line crosses a black line we are seeing a trough and peak in the same place. These will cancel each other out and the medium will have no displacement at that point.

- black line + black line = peak + peak = constructive interference
- grey line + grey line = trough + trough = constructive interference
- black line + grey line = grey line + black line = peak + trough = trough + peak = destructive interference

On half the picture below, we have marked the constructive interference with a solid black diamond and the destructive interference with a hollow diamond.



To see if you understand it, cover up the half we have marked with diamonds and try to work out which points are constructive and destructive on the other half of the picture. The two halves are mirror images of each other so you can check yourself.

26.5 Diffraction

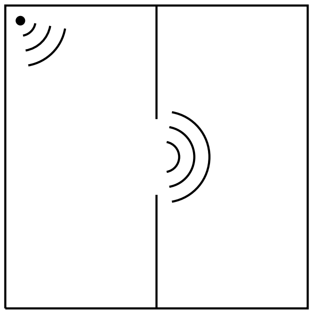
One of the most interesting, and also very useful, properties of waves is **diffraction**.

Definition: Diffraction
Diffraction is the ability of a wave to spread out in wavefronts as the wave passes through a small aperture or around a sharp edge.

Extension: Diffraction

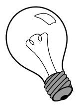
Diffraction refers to various phenomena associated with wave propagation, such as the bending, spreading and interference of waves emerging from an aperture. It occurs with any type of wave, including sound waves, water waves, electromagnetic waves such as light and radio waves. While diffraction always occurs, its effects are generally only noticeable for waves where the wavelength is on the order of the feature size of the diffracting objects or apertures.

For example, if two rooms are connected by an open doorway and a sound is produced in a remote corner of one of them, a person in the other room will hear the sound as if it originated at the doorway.



As far as the second room is concerned, the vibrating air in the doorway is the source of the sound. The same is true of light passing the edge of an obstacle, but this is not as easily observed because of the short wavelength of visible light.

This means that when waves move through small holes they appear to bend around the sides because there are not enough points on the wavefront to form another straight wavefront. This is bending round the sides we call *diffraction*.



Extension: Diffraction

Diffraction effects are more clear for water waves with longer wavelengths. Diffraction can be demonstrated by placing small barriers and obstacles in a ripple tank and observing the path of the water waves as they encounter the obstacles. The waves are seen to pass around the barrier into the regions behind it; subsequently the water behind the barrier is disturbed. The amount of diffraction (the sharpness of the bending) increases with increasing wavelength and decreases with decreasing wavelength. In fact, when the wavelength of the waves are smaller than the obstacle, no noticeable diffraction occurs.

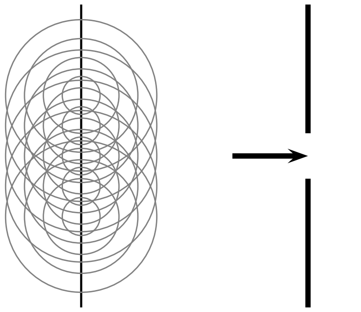
Activity :: Experiment : Diffraction

Water waves in a ripple tank can be used to demonstrate diffraction and interference.

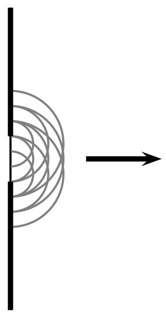
26.5.1 Diffraction through a Slit

When a wave strikes a barrier with a hole only part of the wave can move through the hole. If the hole is similar in size to the wavelength of the wave diffractions occurs. The waves that comes through the hole no longer looks like a straight wave front. It bends around the edges of the hole. If the hole is small enough it acts like a point source of circular waves.

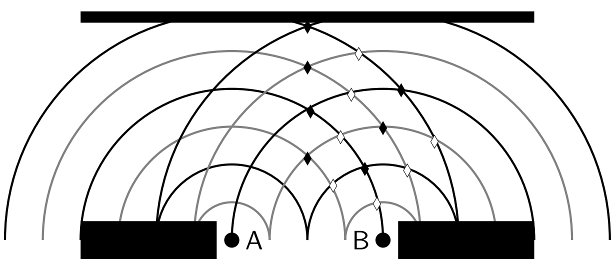
Now if allow the wavefront to impinge on a barrier with a hole in it, then only the points on the wavefront that move into the hole can continue emitting forward moving waves - but because a lot of the wavefront have been removed the points on the edges of the hole emit waves that bend round the edges.



If you employ Huygens' principle you can see the effect is that the wavefronts are no longer straight lines.

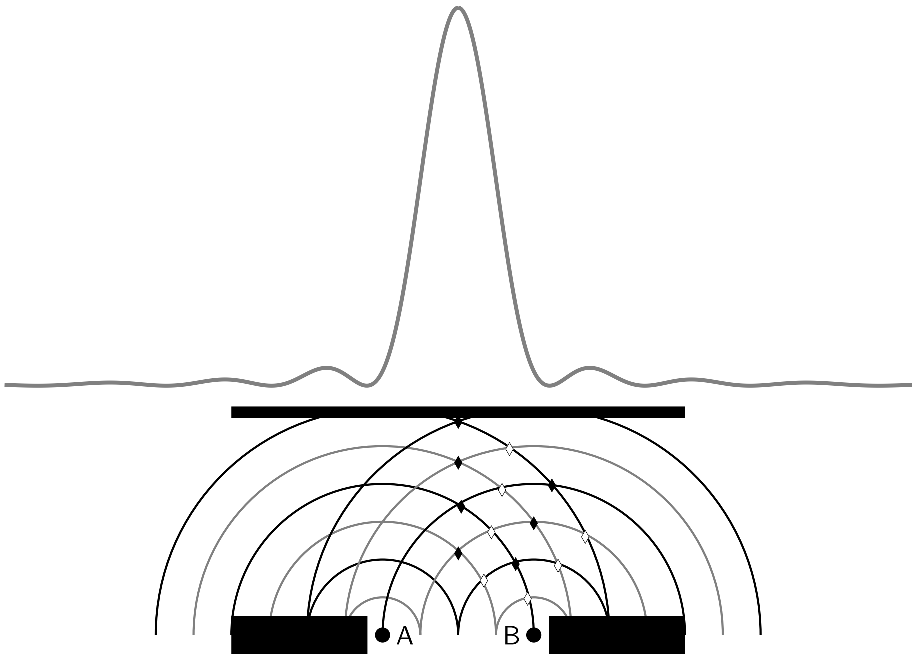


Each point of the slit acts like a point source. If we think about the two point sources on the edges of the slit and call them A and B then we can go back to the diagram we had earlier but with some parts block by the wall.



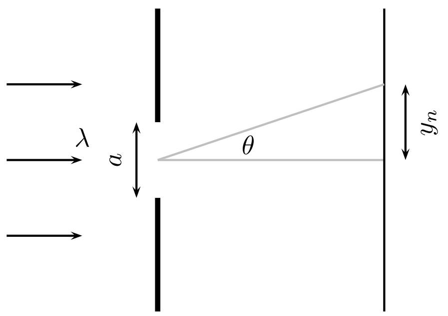
If this diagram were showing sound waves then the sound would be louder (constructive interference) in some places and quieter (destructive interference) in others. You can start to see that there will be a pattern (interference pattern) to the louder and quieter places. If we were studying light waves then the light would be brighter in some places than others depending on the interferences.

The intensity (how bright or loud) of the interference pattern for a single narrow slit looks like this:



The picture above shows how the waves add together to form the interference pattern. The peaks correspond to places where the waves are adding most intensely and the zeroes are places where destructive interference is taking place. When looking at interference patterns from light the spectrum looks like:

There is a formula we can use to determine where the peaks and minimums are in the interference spectrum. There will be more than one minimum. There are the same number of minima on either side of the central peak and the distances from the first one on each side are the same to the peak. The distances to the peak from the second minimum on each side is also the same, in fact the two sides are mirror images of each other. We label the first minimum that corresponds to a positive angle from the centre as $m = 1$ and the first on the other side (a negative angle from the centre as $m = -1$, the second set of minima are labelled $m = 2$ and $m = -2$ etc.



The equation for the angle at which the minima occur is

Definition: Interference Minima

The angle at which the minima in the interference spectrum occur is:

$$\sin\theta = \frac{m\lambda}{a}$$

where

θ is the angle to the minimum

λ is the wavelength of the impinging wavefronts

m is the order of the mimimum, $m = \pm1, \pm2, \pm3, ...$

Worked Example 170: Diffraction Minimum I

Question: A slit has a width of 2511 nm has red light of wavelength 650 nm impinge on it. The diffracted light interferers on a surface, at what angle will the first minimum be?

Answer

Step 1 : Check what you are given

We know that we are dealing with interference patterns from the diffraction of light passing through a slit. The slit has a width of 2511 nm which is 2511×10^{-9} m and we know that the wavelength of the light is 650 nm which is 650×10^{-9} m. We are looking to determine the angle to first minimum so we know that $m = 1$.

Step 2 : Applicable principles

We know that there is a relationship between the slit width, wavelength and interference minimum angles:

$$\sin\theta = \frac{m\lambda}{a}$$

We can use this relationship to find the angle to the minimum by substituting what we know and solving for the angle.

Step 3 : Substitution

$$\begin{aligned}
\sin\theta &= \frac{650 \times 10^{-9}}{2511 \times 10^{-9}} \\
\sin\theta &= \frac{650}{2511} \\
\sin\theta &= 0.258861012 \\
\theta &= \sin^{-1} 0.258861012 \\
\theta &= 15^{o}
\end{aligned}$$

The first minimum is at 15 degrees from the centre peak.

Worked Example 171: Diffraction Minimum II

Question: A slit has a width of 2511 nm has green light of wavelength 532 nm impinge on it. The diffracted light interferers on a surface, at what angle will the first minimum be?

Answer

Step 1 : Check what you are given

We know that we are dealing with interference patterns from the diffraction of light passing through a slit. The slit has a width of 2511 nm which is 2511×10^{-9} m and we know that the wavelength of the light is 532 nm which is 532×10^{-9} m. We are looking to determine the angle to first minimum so we know that $m = 1$.

Step 2 : Applicable principles

We know that there is a relationship between the slit width, wavelength and interference minimum angles:

$$\sin\theta = \frac{m\lambda}{a}$$

We can use this relationship to find the angle to the minimum by substituting what we know and solving for the angle.

Step 3 : Substitution

$$\begin{aligned}\sin\theta &= \frac{532\times10^{-9}}{2511\times10^{-9}}\\ \sin\theta &= \frac{532}{2511}\\ \sin\theta &= 0.211867782\\ \theta &= \sin^{-1}0.211867782\\ \theta &= 12.2^{o}\end{aligned}$$

The first minimum is at 12.2 degrees from the centre peak.

From the formula you can see that a smaller wavelength for the same slit results in a smaller angle to the interference minimum. This is something you just saw in the two worked examples. Do a sanity check, go back and see if the answer makes sense. Ask yourself which light had the longer wavelength, which light had the larger angle and what do you expect for longer wavelengths from the formula.

Worked Example 172: Diffraction Minimum III

Question: A slit has a width which is unknown and has green light of wavelength 532 nm impinge on it. The diffracted light interferers on a surface, and the first minimum is measure at an angle of 20.77 degrees?

Answer

Step 1 : Check what you are given

We know that we are dealing with interference patterns from the diffraction of light passing through a slit. We know that the wavelength of the light is 532 nm which is 532×10^{-9} m. We know the angle to first minimum so we know that $m = 1$ and $\theta = 20.77^{o}$.

Step 2 : Applicable principles

We know that there is a relationship between the slit width, wavelength and interference minimum angles:

$$\sin\theta = \frac{m\lambda}{a}$$

We can use this relationship to find the width by substituting what we know and solving for the width.

Step 3 : Substitution

$$\begin{aligned}\sin\theta &= \frac{532\times10^{-9}}{a}\\ \sin 20.77 &= \frac{532\times10^{-9}}{a}\\ a &= \frac{532\times10^{-9}}{0.354666667}\\ a &= 1500\times10^{-9}\\ a &= 1500\ \mathrm{nm}\end{aligned}$$

The slit width is 1500 nm.

26.6 Shock Waves and Sonic Booms

Now we know that the waves move away from the source at the speed of sound. What happens if the source moves at the same time as emitting sounds? Once a sound wave has been emitted it is no longer connected to the source so if the source moves it doesn't change the way the sound wave is propagating through the medium. This means a source can actually catch up with a sound waves it has emitted.

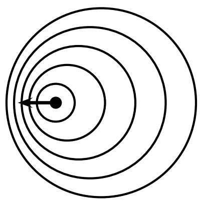
The speed of sound is very fast in air, about $340\ \mathrm{m \cdot s^{-1}}$, so if we want to talk about a source catching up to sound waves then the source has to be able to move very fast. A good source of sound waves to discuss is a jet aircraft. Fighter jets can move very fast and they are very noisy so they are a good source of sound for our discussion. Here are the speeds for a selection of aircraft that can fly faster than the speed of sound.

Aircraft	speed at altitude $(\mathrm{km \cdot h^{-1}})$	speed at altitude $(\mathrm{m \cdot s^{-1}})$
Concorde	2 330	647
Gripen	2 410	669
Mirage F1	2 573	990
Mig 27	1 885	524
F 15	2 660	739
F 16	2 414	671

26.6.1 Subsonic Flight

Definition: Subsonic
Subsonic refers to speeds slower than the speed of sound.

When a source emits sound waves and is moving but slower than the speed of sound you get the situation in this picture. Notice that the source moving means that the wavefronts and therefore peaks in the wave are actually closer together in the one direction and further apart in the other.



subsonic flight

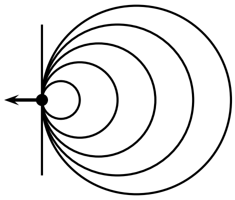
If you measure the waves on the side where the peaks are closer together you'll measure a different wavelength than on the other side of the source. This means that the noise from the source will sound different on the different sides. This is called the *Doppler Effect*.

Definition: Doppler Effect
when the wavelength and frequency measured by an observer are different to those emitted by the source due to movement of the source or observer.

26.6.2 Supersonic Flight

Definition: Supersonic
Supersonic refers to speeds faster than the speed of sound.

If a plane flies at exactly the speed of sound then the waves that it emits in the direction it is flying won't be able to get away from the plane. It also means that the next sound wave emitted will be exactly on top of the previous one, look at this picture to see what the wavefronts would look like:



shock wave at Mach 1

Sometimes we use the speed of sound as a reference to describe the speed of the object (aircraft in our discussion).

Definition: Mach Number
The Mach Number is the ratio of the speed of an object to the speed of sound in the surrounding medium.

Mach number is tells you how many times faster than sound the aircraft is moving.

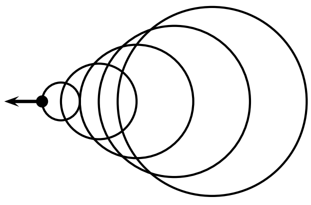
- Mach Number < 1 : aircraft moving slower than the speed of sound
- Mach Number $= 1$: aircraft moving at the speed of sound
- Mach Number > 1 : aircraft moving faster than the speed of sound

To work out the Mach Number divide the speed of the aircraft by the speed of sound.

$$\text{Mach Number} = \frac{\mathrm{v_{aircraft}}}{\mathrm{v_{sound}}}$$

Remember: the units must be the same before you divide.

If the aircraft is moving faster than the speed of sound then the wavefronts look like this:



supersonic shock wave

If the source moves faster than the speed of sound a cone of wave fronts is created. This is called a Mach cone. From constructive interference we know that two peaks that add together form a larger peak. In a Mach cone many, many peaks add together to form a very large peak, this is a sound wave so the large peak is a very very loud sound wave. This sounds like a huge "boom" and we call the noise a *sonic boom*.

Worked Example 173: Mach Speed I

Question: An aircraft flies at 1300 $\mathrm{km \cdot h^{-1}}$ and the speed of sound in air is $340\ \mathrm{m \cdot s^{-1}}$. What is the Mach Number of the aircraft?

Answer

Step 1 : Check what you are given

We know we are dealing with Mach Number. We are given the speed of sound in air, $340\ \mathrm{m \cdot s^{-1}}$, and the speed of the aircraft, 1300 $\mathrm{km \cdot h^{-1}}$. The speed of the aircraft is in different units to the speed of sound so we need to convert the units:

$$
\begin{aligned}
1300\mathrm{km \cdot h^{-1}} &= 1300\mathrm{km \cdot h^{-1}} \\
1300\mathrm{km \cdot h^{-1}} &= 1300 \times \frac{1000\mathrm{m}}{3600\mathrm{s}} \\
1300\mathrm{km \cdot h^{-1}} &= 361.1\ \mathrm{m \cdot s^{-1}}
\end{aligned}
$$

Step 2 : Applicable principles

We know that there is a relationship between the Mach Number, the speed of sound and the speed of the aircraft:

$$\text{Mach Number} = \frac{\mathrm{v_{aircraft}}}{\mathrm{v_{sound}}}$$

We can use this relationship to find the Mach Number.

Step 3 : Substitution

$$
\begin{aligned}
\text{Mach Number} &= \frac{v_{aircraft}}{v_{sound}} \\
\text{Mach Number} &= \frac{361.1}{340} \\
\text{Mach Number} &= 1.06
\end{aligned}
$$

The Mach Number is 1.06.

Definition: Sonic Boom

A sonic boom is the sound heard by an observer as a shockwave passes.

Exercise: Mach Number

In this exercise we will determine the Mach Number for the different aircraft in the table mentioned above. To help you get started we have calculated the Mach Number for the Concord with a speed of sound $v_{sound} = 340\ \mathrm{ms^{-1}}$.

For the Condorde we know the speed and we know that:

$$\text{Mach Number} = \frac{\mathrm{v_{aircraft}}}{\mathrm{v_{sound}}}$$

For the Concorde this means that

$$
\begin{aligned}
\text{Mach Number} &= \frac{647}{340} \\
&= 1.9
\end{aligned}
$$

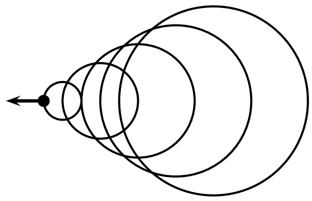
Aircraft	speed at altitude ($\mathrm{km \cdot h^{-1}}$)	speed at altitude ($\mathrm{m \cdot s^{-1}}$)	Mach Number
Concorde	2 330	647	1.9
Gripen	2 410	669	
Mirage F1	2 573	990	
Mig 27	1 885	524	
F 15	2 660	739	
F 16	2 414	671	

Now calculate the Mach Numbers for the other aircraft in the table.

26.6.3 Mach Cone

You can see that the shape of the Mach Cone depends on the speed of the aircraft. When the Mach Number is 1 there is no cone but as the aircraft goes faster and faster the angle of the cone gets smaller and smaller.

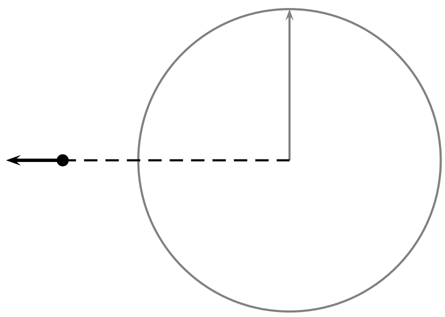
If we go back to the supersonic picture we can work out what the angle of the cone must be.



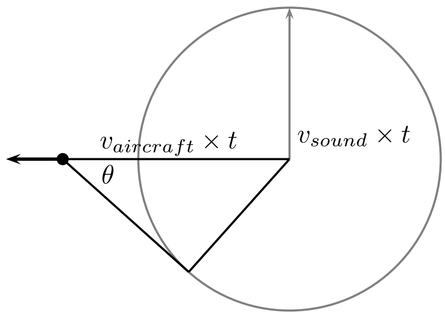
supersonic shock wave

We build a triangle between how far the plane has moved and how far a wavefront at right angles to the direction the plane is flying has moved:

An aircraft emits a sound wavefront. The wavefront moves at the speed of sound 340 $\mathrm{m \cdot s^{-1}}$ and the aircraft moves at Mach 1.5, which is $1.5 \times 340 = 510\ \mathrm{m \cdot s^{-1}}$. The aircraft travels faster than the wavefront. If we let the wavefront travel for a time t then the following diagram will apply:



We know how fast the wavefront and the aircraft are moving so we know the distances that they have traveled:



The angle between the cone that forms at the direction of the plane can be found from the right-angle triangle we have drawn into the figure. We know that $\sin\theta = \frac{opposite}{hypotenuse}$ which in this figure means:

$$\begin{aligned}\sin\theta &= \frac{opposite}{hypotenuse}\\ \sin\theta &= \frac{v_{sound}\times t}{v_{aircraft}\times t}\\ \sin\theta &= \frac{v_{sound}}{v_{aircraft}}\end{aligned}$$

In this case we have used sound and aircraft but a more general way of saying this is:

- aircraft = source
- sound = wavefront

We often just write the equation as:

$$\begin{aligned}\sin\theta &= \frac{v_{sound}}{v_{aircraft}}\\ v_{aircraft}\sin\theta &= v_{sound}\\ v_{source}\sin\theta &= v_{wavefront}\\ v_s\sin\theta &= v_w\end{aligned}$$

?

Exercise: Mach Cone

In this exercise we will determine the Mach Cone Angle for the different aircraft in the table mentioned above. To help you get started we have calculated the Mach Cone Angle for the Concorde with a speed of sound $v_{sound} = 340\ \mathrm{m\cdot s^{-1}}$.

For the Condorde we know the speed and we know that:

$$\sin\theta = \frac{v_{sound}}{v_{aircraft}}$$

For the Concorde this means that

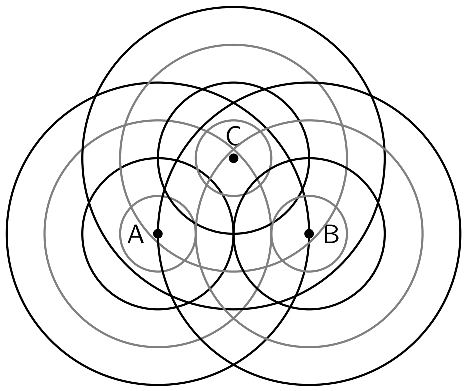
$$\begin{aligned}\sin\theta &= \frac{340}{647}\\ \theta &= \sin^{-1}\frac{340}{647}\\ \theta &= 31.7^{\circ}\end{aligned}$$

Aircraft	speed at altitude ($\mathrm{km\cdot h^{-1}}$)	speed at altitude ($\mathrm{m\cdot s^{-1}}$)	Mach Cone Angle (degrees)
Concorde	2 330	647	31.7
Gripen	2 410	669	
Mirage F1	2 573	990	
Mig 27	1 885	524	
F 15	2 660	739	
F 16	2 414	671	

Now calculate the Mach Cone Angles for the other aircraft in the table.

26.7 End of Chapter Exercises

1. In the diagram below the peaks of wavefronts are shown by black lines and the troughs by grey lines. Mark all the points where constructive interference between two waves is taking place and where destructive interference is taking place. Also note whether the interference results in a peak or a trough.



2. For an slit of width 1300 nm, calculate the first 3 minima for light of the following wavelengths:

   A blue at 475 nm

   B green at 510 nm

   C yellow at 570 nm

   D red at 650 nm

3. For light of wavelength 540 nm, determine what the width of the slit needs to be to have the first minimum at:

   A 7.76 degrees

   B 12.47 degrees

   C 21.1 degrees

4. For light of wavelength 635 nm, determine what the width of the slit needs to be to have the second minimum at:

   A 12.22 degrees

   B 18.51 degrees

   C 30.53 degrees

5. If the first minimum is at 8.21 degrees and the second minimum is at 16.6 degrees, what is the wavelength of light and the width of the slit? (**Hint:** solve simultaneously.)

6. Determine the Mach Number, with a speed of sound of 340 $\mathrm{m \cdot s^{-1}}$, for the following aircraft speeds:

   A 640 $\mathrm{m \cdot s^{-1}}$

   B 980 $\mathrm{m \cdot s^{-1}}$

   C 500 $\mathrm{m \cdot s^{-1}}$

   D 450 $\mathrm{m \cdot s^{-1}}$

   E 1300 $\mathrm{km \cdot h^{-1}}$

F 1450 $\text{km} \cdot \text{h}^{-1}$

G 1760 $\text{km} \cdot \text{h}^{-1}$

7. If an aircraft has a Mach Number of 3.3 and the speed of sound is 340 $\text{m} \cdot \text{s}^{-1}$, what is its speed?

8. Determine the Mach Cone angle, with a speed of sound of 340 $\text{m} \cdot \text{s}^{-1}$, for the following aircraft speeds:

   A 640 $\text{m} \cdot \text{s}^{-1}$

   B 980 $\text{m} \cdot \text{s}^{-1}$

   C 500 $\text{m} \cdot \text{s}^{-1}$

   D 450 $\text{m} \cdot \text{s}^{-1}$

   E 1300 $\text{km} \cdot \text{h}^{-1}$

   F 1450 $\text{km} \cdot \text{h}^{-1}$

   G 1760 $\text{km} \cdot \text{h}^{-1}$

9. Determine the aircraft speed, with a speed of sound of 340 $\text{m} \cdot \text{s}^{-1}$, for the following Mach Cone Angles:

   A 58.21 degrees

   B 49.07 degrees

   C 45.1 degrees

   D 39.46 degrees

   E 31.54 degrees

Chapter 27

Wave Nature of Matter - Grade 12

27.1 Introduction

In chapters 30 and 31 the so called wave-particle duality if light is described. This duality states that light displays properties of both waves and of particles, depending on the experiment performed. For example, interference and diffraction of light are properties of its wave nature, while the photoelectric effect is a property of its particle nature. In fact we call a particle of light a photon.

Hopefully you have realised that nature loves symmetry. So, if light which was originally believed to be a wave also has a particle nature then perhaps particles, also display a wave nature. In other words matter which which we originally thought of as particles may also display a wave-particle duality.

27.2 de Broglie Wavelength

Einstein showed that for a photon, its momentum, p, is equal to its energy, E divided the speed of light, c:

$$p = \frac{E}{c}.$$

The energy of the photon can also be expressed in terms of the wavelength of the light, λ:

$$E = \frac{hc}{\lambda},$$

where h is Planck's constant. Combining these two equations we find that the the momentum of the photon is related to its wavelength

$$p = \frac{hc}{c\lambda} = \frac{h}{\lambda},$$

or equivalently

$$\lambda = \frac{h}{p}.$$

In 1923, Louis de Broglie proposed that this equation not only holds for photons, but also holds for particles of matter. This is known as the de Broglie hypothesis

Definition: De Broglie Hypothesis
A particle of mass m moving with velocity v has a wavelength λ related to is momentum $p = mv$ by

$$\lambda = \frac{h}{p} = \frac{h}{mv} \tag{27.1}$$

This wavelength, λ, is known as the de Broglie wavelength of the particle.

Since the value of Planck's constant is incredibly small $h = 6.63 \times 10^{-34}$ J · s, the wavelike nature of everyday objects is not really observable.

The de Broglie hypothesis was proposed by French physicist Louis de Broglie (15 August 1892 – 19 March 1987) in 1923 in his PhD thesis. He was awarded the Nobel Prize for Physics in 1929 for this work, which made him the first person to receive a Nobel Prize on a PhD thesis.

Worked Example 174: de Broglie Wavelength of a Cricket Ball

Question: A cricket ball has a mass of $0{,}150\,\text{kg}$ and is bowled towards a bowler at $40\,\text{m}\cdot\text{s}^{-1}$. Calculate the de Broglie wavelength of the cricket ball?

Answer

Step 1 : Determine what is required and how to approach the problem

We are required to calculate the de Broglie wavelength of a cricket ball given its mass and speed. We can do this by using:

$$\lambda = \frac{h}{mv}$$

Step 2 : Determine what is given

We are given:

- The mass of the cricket ball $m = 0{,}150\,\text{kg}$
- The velocity of the cricket ball $v = 40\,\text{m}\cdot\text{s}^{-1}$

and we know:

- Planck's constant $h = 6{,}63 \times 10^{-34}$ J · s

Step 3 : Calculate the de Broglie wavelength

$$\begin{aligned} \lambda &= \frac{h}{mv} \\ &= \frac{6{,}63 \times 10^{-34}\ \text{J}\cdot\text{s}}{(0{,}150\,\text{kg})(40\ \ \text{m}\cdot\text{s}^{-1})} \\ &= 1{,}10 \times 10^{-34}\,\text{m} \end{aligned}$$

This wavelength is considerably smaller than the diameter of a proton which is approximately 10^{-15} m. Hence the wave-like properties of this cricket ball are too small to be observed.

Worked Example 175: The de Broglie wavelength of an electron

Question: Calculate the de Broglie wavelength of an electron moving at 40 m·s^{-1}.

Answer

Step 1 : Determine what is required and how to approach the problem

We required to calculate the de Broglie wavelength of an electron given its speed. We can do this by using:

$$\lambda = \frac{h}{mv}$$

Step 2 : Determine what is given

We are given:

- The velocity of the electron $v = 40\,\mathrm{m \cdot s^{-1}}$

and we know:

- The mass of the electron $m = 9{,}11 \times 10^{-31}\,\mathrm{kg}$
- Planck's constant $h = 6{,}63 \times 10^{-34}\ \mathrm{J \cdot s}$

Step 3 : Calculate the de Broglie wavelength

$$\begin{aligned} \lambda &= \frac{h}{mv} \\ &= \frac{6{,}63 \times 10^{-34}\ \mathrm{J \cdot s}}{(9{,}11 \times 10^{-31}\,\mathrm{kg})(40\ \mathrm{m \cdot s^{-1}})} \\ &= 1{,}82 \times 10^{-5}\,\mathrm{m} \\ &= 0{,}0182\,\mathrm{mm} \end{aligned}$$

Although the electron and cricket ball in the two previous examples are travelling at the same velocity the de Broglie wavelength of the electron is much larger than that of the cricket ball. This is because the wavelength is inversely proportional to the mass of the particle.

Worked Example 176: The de Broglie wavelength of an electron

Question: Calculate the de Broglie wavelength of a electron moving at $3 \times 10^5\,\mathrm{m \cdot s^{-1}}$. ($\frac{1}{1000}$ of the speed of light.)

Answer

Step 1 : Determine what is required and how to approach the problem

We required to calculate the de Broglie wavelength of an electron given its speed. We can do this by using:

$$\lambda = \frac{h}{mv}$$

Step 2 : Determine what is given

We are given:

- The velocity of the electron $v = 3 \times 10^5\,\mathrm{m \cdot s^{-1}}$

and we know:

- The mass of the electron $m = 9{,}11 \times 10^{-31}\,\mathrm{kg}$
- Planck's constant $h = 6{,}63 \times 10^{-34}\ \mathrm{J \cdot s}$

Step 3 : Calculate the de Broglie wavelength

$$\begin{aligned} \lambda &= \frac{h}{mv} \\ &= \frac{6{,}63 \times 10^{-34}\ \mathrm{J \cdot s}}{(9{,}11 \times 10^{-31}\,\mathrm{kg})(3 \times 10^5\,\mathrm{m \cdot s^{-1}})} \\ &= 2{,}43 \times 10^{-9}\,\mathrm{m} \end{aligned}$$

This is the size of an atom. For this reason, electrons moving at high velocities can be used to "probe" the structure of atoms. This is discussed in more detail at the end of this chapter. Figure 27.1 compares the wavelengths of fast moving electrons to the wavelengths of visible light.

Since the de Broglie wavelength of a particle is inversely proportional to its velocity, the wavelength decreases as the velocity increases. This is confirmed in the last two examples with the electrons. De Broglie's hypothesis was confirmed by Davisson and Germer in 1927 when they observed a beam of electrons being diffracted off a nickel surface. The diffraction means that the moving electrons have a wave nature. They were also able to determine the wavelength of the electrons from the diffraction. To measure a wavelength one needs two or more diffacting centres such as pinholes, slits or atoms. For diffraction to occur the centres must be separated by a distance about the same size as the wavelength. Theoretically, all objects, not just sub-atomic particles, exhibit wave properties according to the de Broglie hypothesis.

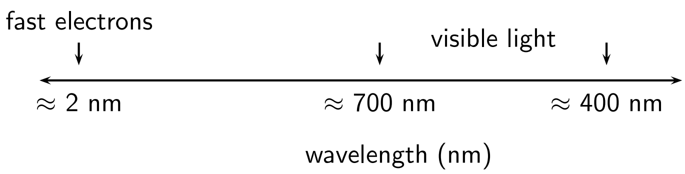


Figure 27.1: The wavelengths of the fast electrons are much smaller than that of visible light.

27.3 The Electron Microscope

We have seen that under certain circumstances particles behave like waves. This idea is used in the electron microscope which is a type of microscope that uses electrons to create an image of the target. It has much higher magnification or resolving power than a normal light microscope, up to two million times, allowing it to see smaller objects and details.

Let's first review how a regular optical microscope works. A beam of light is shone through a thin target and the image is then magnified and focused using objective and ocular lenses. The amount of light which passes through the target depends on the densities of the target since the less dens regions allow more light to pass through than the denser regions. This means that the beam of light which is partially transmitted through the target carries information about the inner structure of the target.

The original form of the electron microscopy, the transmission electron microscopy, works in a similar manner using electrons. In the electron microscope, electrons which are emitted by a cathode are formed into a beam using magnetic lenses. This electron beam is then passed through a very thin target. Again, the regions in the target with higher densities stop the electrons more easily. So, the amount of electrons which pass through the different regions of the target depend their densities. This means that the partially transmitted beam of electrons carries information about the densities of the inner structure of the target. The spatial variation in this information (the "image") is then magnified by a series of magnetic lenses and it is recorded by hitting a fluorescent screen, photographic plate, or light sensitive sensor such as a CCD (charge-coupled device) camera. The image detected by the CCD may be displayed in real time on a monitor or computer. In figure **??** is an image of the polio virus obtained with a transmission electron microscope.

The structure of an optical and electron microscope are compared in figure 27.3. While the optical microscope uses light and focuses using lenses, the electron microscope uses electrons and focuses using electromagnets.

The first electron microscope prototype was built in 1931 by the German engineers Ernst Ruska and Maximillion Knoll. It was based on the ideas and discoveries of Louis de Broglie. Although it was primitive and was not ideal for practical use, the instrument was still capable of magnifying objects by four hundred times. The first practical electron microscope was built at the

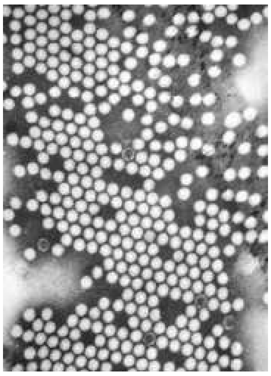


Figure 27.2: The image of the polio virus using a transmission electron microscope.

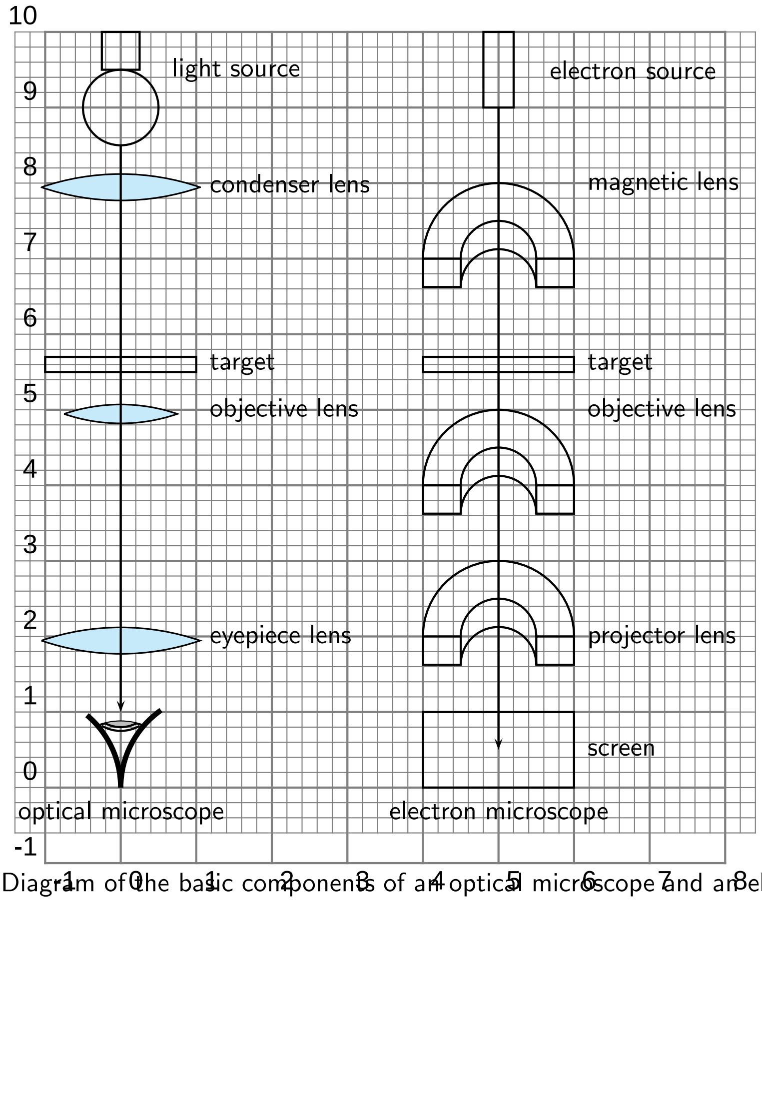


Figure 27.3: Diagram of the basic components of an optical microscope and an electron microscope.

Table 27.1: Comparison of Light and Electron Microscopes

	Light microscope	**Electron microscope**
Source	Bright lamp or laser	Electron gun
Radiation	U.V. or visible light	Electron beam produced by heating metal surface (e.g. tungsten)
Lenses	Curved glass surfaces	Electromagnets
Receiver	Eye; photographic emulsion or digital image	Fluorescent screen (for location and focusing image); photographic emulsion or digital image
Focus	Axial movement of lenses (up and down)	Adjustment of magnetic field in the electromagnets by changing the current
Operating Pressure	Atmospheric	High vacuum

University of Toronto in 1938, by Eli Franklin Burton and students Cecil Hall, James Hillier and Albert Prebus.

Although modern electron microscopes can magnify objects up to two million times, they are still based upon Ruska's prototype and his correlation between wavelength and resolution. The electron microscope is an integral part of many laboratories. Researchers use it to examine biological materials (such as microorganisms and cells), a variety of large molecules, medical biopsy samples, metals and crystalline structures, and the characteristics of various surfaces.

Electron microscopes are very useful as they are able to magnify objects to a much higher resolution. This is because their de Broglie wavelengths are so much smaller than that of visible light. You hopefully remember that light is diffracted by objects which are separated by a distance of about the same size as the wavelength of the light. This diffraction then prevents you from being able to focus the transmitted light into an image. So the sizes at which diffraction occurs for a beam of electrons is much smaller than those for visible light. This is why you can magnify targets to a much higher order of magnification using electrons rather than visible light.

Extension: High-Resolution Transmission Electron Microscope (HRTEM)

There are high-resolution TEM (HRTEM) which have been built. However their resolution is limited by spherical and chromatic aberration. Fortunately though, software correction of the spherical aberration has allowed the production of images with very high resolution. In fact the resolution is sufficient to show carbon atoms in diamond separated by only 89 picometers and atoms in silicon at 78 picometers. This is at magnifications of 50 million times. The ability to determine the positions of atoms within materials has made the HRTEM a very useful tool for nano-technologies research. It is also very important for the development of semiconductor devices for electronics and photonics.

Transmission electron microscopes produce two-dimensional images.

Extension: Scanning Electron Microscope (SEM)

The Scanning Electron Microscope (SEM) produces images by hitting the target with a primary electron beam which then excites the surface of the target. This causes secondary electrons to be emitted from the surface which are then detected. So the the electron beam in the SEM is moved across the sample, while detectors build an image from the secondary electrons.

Generally, the transmission electron microscope's resolution is about an order of magnitude better than the SEM resolution, however, because the SEM image relies on surface processes rather than transmission it is able to image bulk samples and has a much greater depth of view, and so can produce images that are a good representation of the 3D structure of the sample.

27.3.1 Disadvantages of an Electron Microscope

Electron microscopes are expensive to buy and maintain. they are also very sensitive to vibration and external magnetic fields. This means that special facilities are required to house microscopes aimed at achieving high resolutions. Also the targets have to be viewed in vacuum, as the electrons would scatter with the molecules that make up air.

Extension: Scanning Electron Microscope (SEM)

Scanning electron microscopes usually image conductive or semi-conductive materials best. A common preparation technique is to coat the target with a several-nanometer layer of conductive material, such as gold, from a sputtering machine; however this process has the potential to disturb delicate samples.

The targets have to be prepared in many ways to give proper detail, which may result in artifacts purely the result of treatment. This gives the problem of distinguishing artifacts from material, particularly in biological samples. Scientists maintain that the results from various preparation techniques have been compared, and as there is no reason that they should all produce similar artifacts, it is therefore reasonable to believe that electron microscopy features correlate with living cells.

Interesting Fact

The first electron microscope prototype was built in 1931 by the German engineers Ernst Ruska and Maximillion Knoll. It was based on the ideas and discoveries of Louis de Broglie. Although it was primitive and was not ideal for practical use, the instrument was still capable of magnifying objects by four hundred times. The first practical electron microscope was built at the University of Toronto in 1938, by Eli Franklin Burton and students Cecil Hall, James Hillier and Albert Prebus.

Although modern electron microscopes can magnify objects up to two million times, they are still based upon Ruska's prototype and his correlation between wavelength and resolution. The electron microscope is an integral part of many laboratories. Researchers use it to examine biological materials (such as microorganisms and cells), a variety of large molecules, medical biopsy samples, metals and crystalline structures, and the characteristics of various surfaces.

27.3.2 Uses of Electron Microscopes

Electron microscopes can be used to study:

- the topography of an object – how its surface looks.
- the morphology of particles making up an object – its shape and size.
- the composition of an object – the elements and compounds that the object is composed of and the relative amounts of them.
- the crystallographic information of the object – how the atoms are arranged in the object.

27.4 End of Chapter Exercises

1. If the following particles have the same velocity, which has the shortest wavelength: electron, hydrogen atom, lead atom?
2. A bullet weighing 30 g is fired at a velocity of $500\,\text{m}\cdot\text{s}^{-1}$. What is its wavelength?
3. Calculate the wavelength of an electron which has a kinetic energy of 1.602×10^{-19} J.
4. If the wavelength of an electron is 10^{-9} m what is its velocity?
5. Considering how one calculates wavelength using slits, try to explain why we would not be able to physically observe diffraction of the cricket ball in first worked example.

Chapter 28

Electrodynamics - Grade 12

28.1 Introduction

In Grade 11 you learnt how a magnetic field is generated around a current carrying conductor. You also learnt how a current is generated in a conductor that moves in a magnetic field. This chapter describes how conductors moved in a magnetic field are applied in the real-world.

28.2 Electrical machines - generators and motors

We have seen that when a conductor is moved in a magnetic field or when a magnet is moved near a conductor, such that the magnetic field is not parallel to the conductor, a current flows in the conductor. The amount of current depends on the speed at which the conductor experiences a changing magnetic field, the number of turns of the conductor and the orientation of the plane of the conductor with respect to the magnetic field. The effect of the orientation of the conductor with respect to the magnetic field is shown in Figure 28.1.

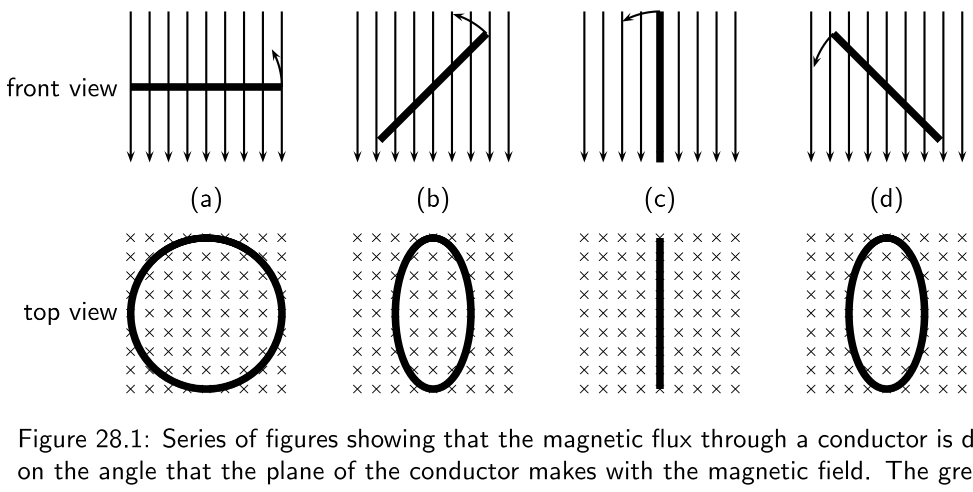


Figure 28.1: Series of figures showing that the magnetic flux through a conductor is dependent on the angle that the plane of the conductor makes with the magnetic field. The greatest flux passes through the conductor when the plane of the conductor is perpendicular to the magnetic field lines as in (a). The number of field lines passing through the conductor decreases, as the conductor rotates until it is parallel to the magnetic field.

If the current flowing in the conductor were plotted as a function of the angle between the plane of the conductor and the magnetic field, then the current would vary as shown in Figure 28.2. The current alternates about the zero value and is also known as an *alternating current* (abbreviated AC).

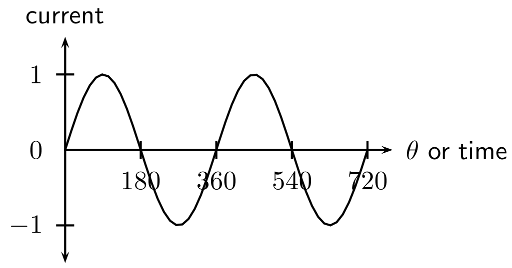
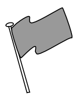


Figure 28.2: Variation of current as angle of plane of conductor with the magnetic field changes.

28.2.1 Electrical generators

AC generator

The principle of rotating a conductor in a magnetic field is used in electricity generators. A generator converts mechanical energy into electrical energy.



Definition: Generator
A generator converts mechanical energy into electrical energy.

The layout of an AC generator is shown in Figure 28.3. The conductor in the shape of a coil is connected to a ring. The conductor is then manually rotated in the magnetic field generating an alternating emf. The slip rings are connected to the load via brushes.

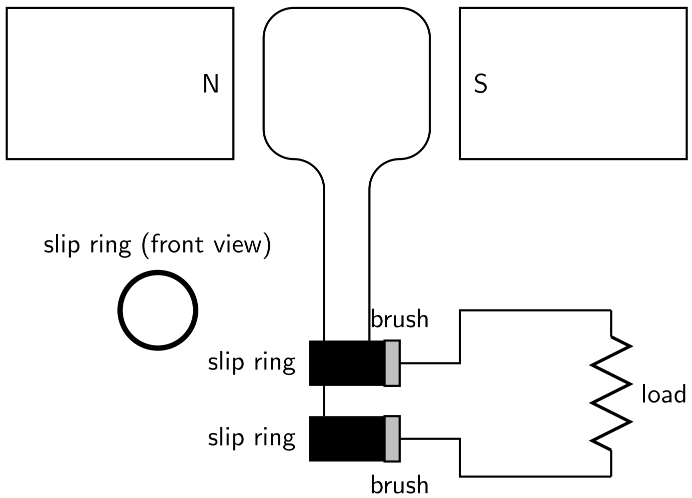
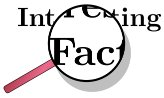


Figure 28.3: Layout of an alternating current generator.

If a machine is constructed to rotate a magnetic field around a set of stationary wire coils with the turning of a shaft, AC voltage will be produced across the wire coils as that shaft is rotated, in accordance with Faraday's Law of electromagnetic induction. This is the basic operating principle of an AC generator.
In an AC generator the two ends of the coil are each attached to a slip ring that makes contact with brushes as the coil turns. The direction of the current changes with every half turn of the coil. As one side of the loop moves to the other pole of the magnetic field, the current in it changes direction. The two slip rings of the AC generator allow the current to change directions and become alternating current.



AC generators are also known as alternators. They are found in motor cars to charge the car battery.

DC generator

A DC generator is constructed the same way as an AC generator except that there is one slip ring which is split into two pieces, called a commutator, so the current in the external circuit does not change direction. The layout of a DC generator is shown in Figure 28.4. The split-ring commutator accommodates for the change in direction of the current in the loop, thus creating DC current going through the brushes and out to the circuit.

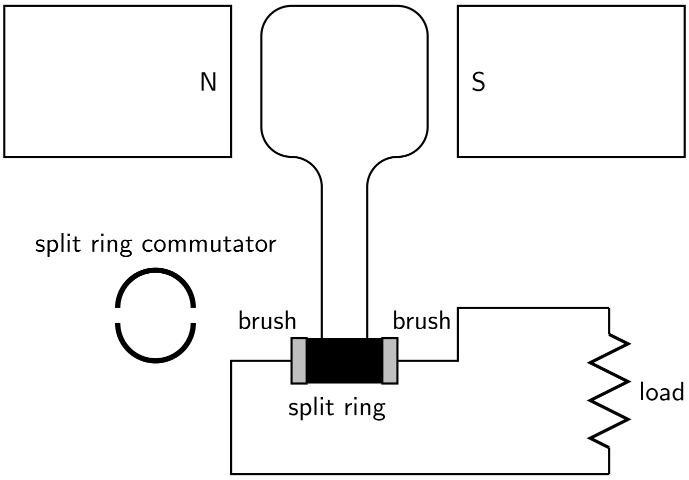


Figure 28.4: Layout of a direct current generator.

The shape of the emf from a DC generator is shown in Figure 28.5. The emf is not steady but is more or less the positive halves of a sine wave.

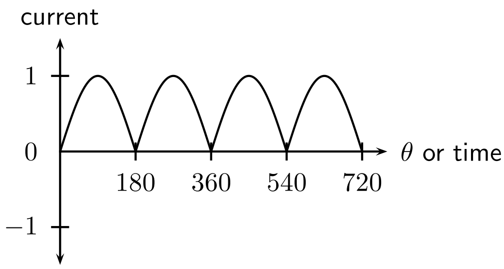


Figure 28.5: Variation of emf in a DC generator.

AC versus DC generators

The problems involved with making and breaking electrical contact with a moving coil should be obvious (sparking and heat), especially if the shaft of the generator is revolving at high speed. If the atmosphere surrounding the machine contains flammable or explosive vapors, the practical problems of spark-producing brush contacts are even greater.
An AC generator (alternator) does not require brushes and commutators to work, and so is immune to these problems experienced by DC generators. The benefits of AC over DC with regard to generator design is also reflected in electric motors. While DC motors require the use

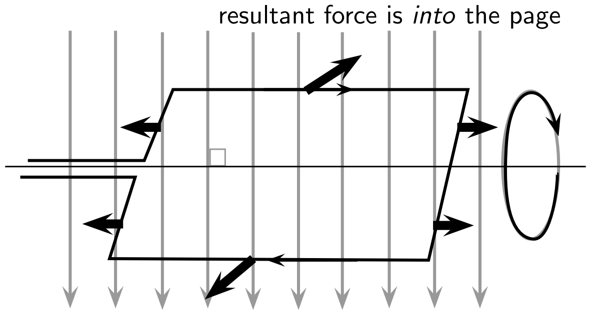
of brushes to make electrical contact with moving coils of wire, AC motors do not. In fact, AC and DC motor designs are very similar to their generator counterparts. The AC motor being dependent upon the reversing magnetic field produced by alternating current through its stationary coils of wire to rotate the rotating magnet around on its shaft, and the DC motor being dependent on the brush contacts making and breaking connections to reverse current through the rotating coil every 1/2 rotation (180 degrees).

28.2.2 Electric motors

The basic principles of operation for a motor are the same as that of a generator, except that a motor converts electrical energy into mechanical energy.

Definition: Motor
An electric motor converts electrical energy into mechanical energy.

Both motors and generators can be explained in terms of a coil that rotates in a magnetic field. In a generator the coil is attached to an external circuit and it is mechanically turned, resulting in a changing flux that induces an emf. In a motor, a current-carrying coil in a magnetic field experiences a force on both sides of the coil, creating a torque which makes it turn.
Any coil carrying current can feel a force in a magnetic field, the force is the Lorentz force on the moving charges in the conductor. We know that if the coil is parallel to the magnetic field then the Lorentz force will be zero. The charge of opposite sides of the coil will be in opposite directions because the charges are moving in opposite directions. This means the coil will rotate.



Instead of rotating the loops through a magnetic field to create electricity, a current is sent through the wires, creating electromagnets. The outer magnets will then repel the electromagnets and rotate the shaft as an electric motor. If the current is AC, the two slip rings are required to create an AC motor. An AC motor is shown in Figure 28.6

If the current is DC, split-ring commutators are required to create a DC motor. This is shown in Figure 28.7.

28.2.3 Real-life applications

Cars

A car contains an alternator that charges up its battery power the car's electric system when its engine is running. Alternators have the great advantage over direct-current generators of not using a commutator, which makes them simpler, lighter, less costly, and more rugged than a DC generator.

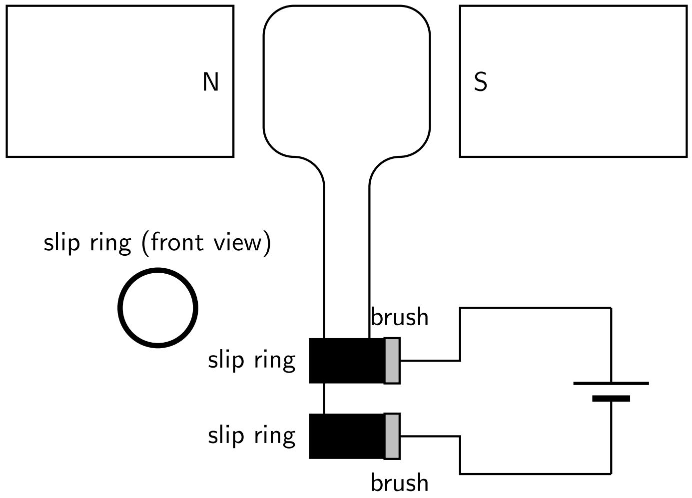


Figure 28.6: Layout of an alternating current motor.

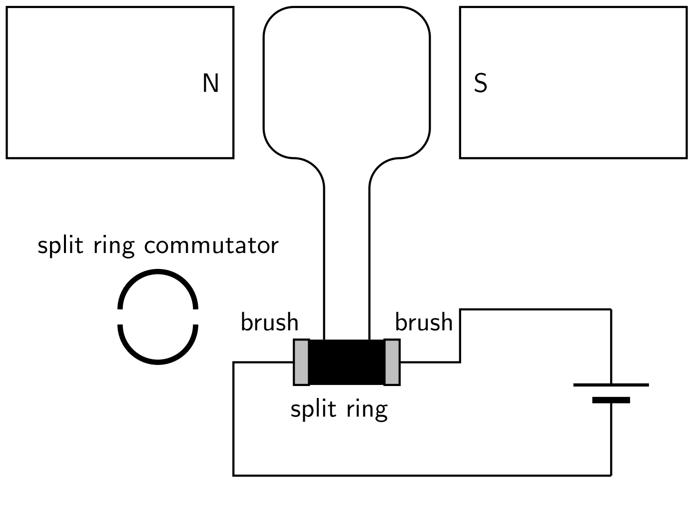


Figure 28.7: Layout of a direct current motor.

Activity :: Research Topic : Alternators

Try to find out the different ampere values produced by alternators for different types of machines. Compare these to understand what numbers make sense in the real world. You will find different numbers for cars, trucks, buses, boats etc. Try to find out what other machines might have alternators.

A car also contains a DC electric motor, the starter motor, to turn over the engine to start it. A starter consists of the very powerful DC electric motor and starter solenoid that is attached to the motor. A starter motor requires very high current to crank the engine, that's why it's connected to the battery with large cables.

Electricity Generation

AC generators are mainly used in the real-world to generate electricity.

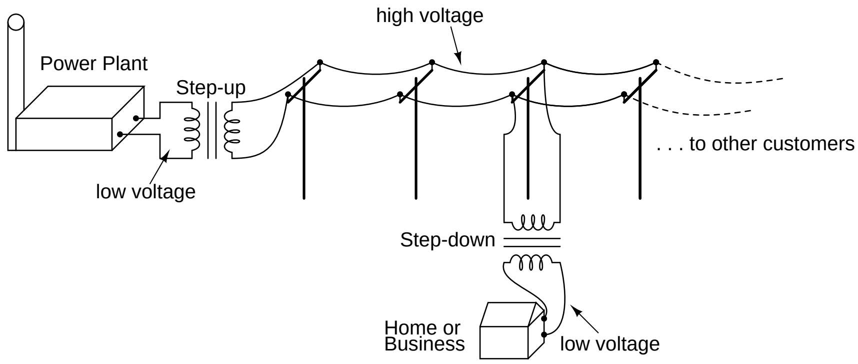


Figure 28.8: AC generators are used at the power plant to generate electricity.

28.2.4 Exercise - generators and motors

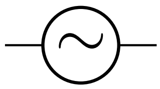
1. State the difference between a generator and a motor.
2. Use Faraday's Law to explain why a current is induced in a coil that is rotated in a magnetic field.
3. Explain the basic principle of an AC generator in which a coil is mechanically rotated in a magnetic field. Draw a diagram to support your answer.
4. Explain how a DC generator works. Draw a diagram to support your answer. Also, describe how a DC generator differs from an AC generator.
5. Explain why a current-carrying coil placed in a magnetic field (but not parallel to the field) will turn. Refer to the force exerted on moving charges by a magnetic field and the torque on the coil.
6. Explain the basic principle of an electric motor. Draw a diagram to support your answer.
7. Give examples of the use of AC and DC generators.
8. Give examples of the use of motors

28.3 Alternating Current

Most students of electricity begin their study with what is known as *direct current* (DC), which is electricity flowing in a constant direction. DC is the kind of electricity made by a battery, with definite positive and negative terminals).
However, we have seen that the electricity produced by a generator alternates and is therefore known as *alternating current*(AC). The main advantage to AC is that the voltage can be changed using transformers. That means that the voltage can be stepped up at power stations to a very high voltage so that electrical energy can be transmitted along power lines at low current and therefore experience low energy loss due to heating. The voltage can then be stepped down for use in buildings and street lights.

In South Africa alternating current is generated at a frequency of 50 Hz.

The circuit symbol for alternating current is:



Graphs of voltage against time and current against time for an AC circuit are shown in Figure 28.9

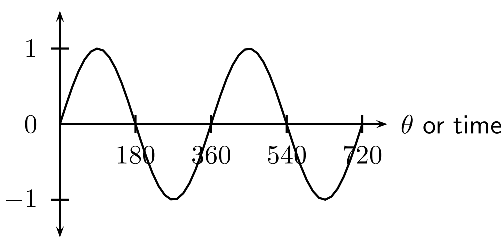


Figure 28.9: Graph of current or voltage in an AC circuit.

In a DC circuit the current and voltage are constant. In an AC circuit the current and voltage vary with time. The value of the current or voltage at any specific moment in time is called the *instantaneous current or voltage* and is calculated as follows:

$$\begin{aligned} i &= I_{max} sin(2\pi f t) \\ v &= V_{max} sin(2\pi f t) \end{aligned}$$

i is the instantaneous current. I_{max} is the maximum current. v is the instantaneous voltage. V_{max} is the maximum voltage. f is the frequency of the AC and t is the time at which the instantaneous current or voltage is being calculated.
This average value we use for AC is known as the root mean square (rms) average. This is defined as:

$$\begin{aligned} I_{rms} &= \frac{I_{max}}{\sqrt{2}} \\ V_{rms} &= \frac{V_{max}}{\sqrt{2}} \end{aligned}$$

Since AC varies sinusoidally, with as much positive as negative, doing a straight average would get you zero for the average voltage. The rms value by-passes this problem.

28.3.1 Exercise - alternating current

1. Explain the advantages of alternating current.
2. Write expressions for the current and voltage in an AC circuit.
3. Define the rms (root mean square) values for current and voltage for AC.
4. What is the period of the AC generated in South Africa?
5. If the mains supply is 200 V AC, calculate rms voltage.
6. Draw a graph of voltage vs time and current vs time for an AC circuit.

28.4 Capacitance and inductance

Capacitors and inductors are found in many circuits. Capacitors store an electric field, and are used as temporary power sources as well as minimize power fluctuations in major circuits. Inductors work in conjunction with capacitors for electrical signal processing. Here we explain the physics and applications of both.

28.4.1 Capacitance

You have learnt about capacitance and capacitors in Grade 11. Please read through section 17.5 to recap what you learnt about capacitance in a DC circuit.
In this section you will learn about capacitance in an AC circuit. A capacitor in an AC circuit has *reactance*. Reactance in an AC circuit plays a similar role to resistance in a DC circuit.
The reactance of a capacitor X_C is defined as:
$X_C = \frac{1}{2\pi fC}$
where C is the capacitance and f is the AC frequency.
If we examine the equation for the reactance of a capacitor, we see that the frequency is in the denominator. Therefore, when the frequency is low, the capacitive reactance is very high. This is why a capacitor blocks the flow of DC and low frequency AC because its reactance increases with decreasing frequency.
When the frequency is high, the capacitive reactance is low. This is why a capacitor allows the flow of high frequency AC because its reactance decreases with increasing frequency.

28.4.2 Inductance

An inductor is a passive electrical device used in electrical circuits for its property of inductance. An inductor is usually made as a coil (or solenoid) of conducting material, typically copper wire, wrapped around a core either of air or of ferromagnetic material.
Electrical current through the conductor creates a magnetic flux proportional to the current. A change in this current creates a change in magnetic flux that, in turn, generates an emf that acts to oppose this change in current.
Inductance (measured in henries, symbol H) is a measure of the generated emf for a unit change in current. For example, an inductor with an inductance of 1 H produces an emf of 1 V when the current through the inductor changes at the rate of 1 A·s^{-1}.
The inductance of an inductor is determined by several factors:

- the shape of the coil; a short, fat coil has a higher inductance than one that is thin and tall.
- the material that conductor is wrapped around.
- how the conductor is wound; winding in opposite directions will cancel out the inductance effect, and you will have only a resistor.

The inductance of a solenoid is defined by:
$L = \frac{\mu_0 A N^2}{l}$
where μ_0 is the permeability of the core material (in this case air), A is the cross-sectional area of the solenoid, N is the number of turns and l is the length of the solenoid.

Definition: Permeability
Permeability is the property of a material which describes the magnetisation developed in that material when excited by a source.

The permeability of free space is $4\pi \mathrm{x} 10^{-7}$ henry per metre.

Worked Example 177: Inductance I

Question:
Answer
Determine the inductance of a coil with a core material of air. A cross-sectional area of $0{,}3\mathrm{m}_2$, with 1000 turns and a length of 0,1 m
Step 1 : Determine how to approach the problem
We are calculating inductance, so we use the equation:

$$L = \frac{\mu_0 A N^2}{l}$$

The permeability is that for free space:$4\pi \mathrm{x} 10^{-7}$ henry per metre.
Step 2 : Solve the problem

$$\begin{aligned} L &= \frac{\mu_0 A N^2}{l} \\ &= \frac{(4\pi textrmx10\_7)(0{,}3)(1000)}{0{,}1} \\ &= 3{,}8\mathrm{x}10\_3 \; H/m \end{aligned}$$

Step 3 : Write the final answer
The inductance of the coil is $3{,}8x10^{-3}$ H/m.

Worked Example 178: Inductance II

Question: Calculate the inductance of a 5 cm long solenoid with a diameter of 4 mm and 2000 turns.
Answer
Again this is an inductance problem, so we use the same formula as the worked example above.

$$\mathrm{r} = \frac{4\ \mathrm{mm}}{2} = 2\ \mathrm{mm} = 0{,}002\ \mathrm{m}$$
$$\mathrm{A} = \pi \mathrm{r}^2 = \pi \times 0{,}002^2$$

$$\begin{aligned}
\mathrm{L} &= \frac{\mu_0 \mathrm{AN}^2}{l} \\
&= \frac{4\pi \times 10^{-7} \times 0{,}002^2 \times \pi \times 2000^2}{0{,}05} \\
&= 0{,}00126\ \mathrm{H} \\
&= 1{,}26\ \mathrm{mH}
\end{aligned}$$

An inductor in an AC circuit also has a reactance, X_L that is defined by:
$X_L = 2\pi f L$
where L is the inductance and f is the frequency of the AC.
If we examine the equation for the reactance of an inductor, we see that inductive reactance increases with increasing frequency. Therefore, when the frequency is low, the inductive reactance is very low. This is why an inductor allows the flow of DC and low frequency AC because its reactance decreases with decreasing frequency.
When the frequency is high, the inductive reactance is high. This is why an inductor blocks the flow of high frequency AC because its reactance increases with increasing frequency.

28.4.3 Exercise - capacitance and inductance

1. Describe what is meant by reactance.
2. Define the reactance of a capacitor.
3. Explain how a capacitor blocks the flow of DC and low frequency AC but allows the flow of high frequency AC.
4. Describe what is an inductor
5. Describe what is inductance
6. What is the unit of inductance?
7. Define the reactance of an inductor.
8. Write the equation describing the inductance of a solenoid.
9. Explain that how an inductor blocks high frequency AC, but allows low frequency AC and DC to pass.

28.5 Summary

1. Electrical generators convert mechanical energy into electrical energy.
2. Electric motors convert electrical energy into mechanical energy.
3. There are two types of generators - AC and DC. An AC generator is also called an alternator.
4. There are two types of motors - AC and DC.
5. Alternating current (AC) has many advantages over direct current (DC).
6. Capacitors and inductors are important components in an AC circuit.
7. The reactance of a capacitor or inductor is affected by the frequency of the AC.

28.6 End of chapter exercise

1. [SC 2003/11] Explain the difference between alternating current (AC) and direct current (DC).
2. Explain how an AC generator works. You may use sketches to support your answer.
3. What are the advantages of using an AC motor rather than a DC motor.
4. Explain how a DC motor works.
5. At what frequency is AC generated by Eskom in South Africa?

IEB 2001/11 HG1 - **Work, Energy and Power in Electric Circuits**
Mr. Smith read through the agreement with Eskom (the electricity provider). He found out that alternating current is supplied to his house at a frequency of 50 Hz. He then consulted a book on electric current, and discovered that alternating current moves to and fro in the conductor. So he refused to pay his Eskom bill on the grounds that every electron that entered his house would leave his house again, so therefore Eskom had supplied him with nothing!
Was Mr. smith correct? Or has he misunderstood something about what he is paying for? Explain your answer briefly.

6. What do we mean by the following terms in electrodynamics?
   A inductance
   B reactance
   C solenoid
   D permeability

Chapter 29

Electronics - Grade 12

29.1 Introduction

Electronics and electrical devices surround us in daily life. From the street lights and water pumps to computers and digital phones, electronics have enabled the digital revolution to occur. All electronics are built on a backbone of simple circuits, and so an understanding of circuits is vital in understanding more complex devices.

This chapter will explain the basic physics principles of many of the components of electronic devices. We will begin with an explanation of capacitors and inductors. We see how these are used in tuning a radio. Next, we look at **active** components such as transistors and operational amplifiers. Lastly, the chapter will finish with an explanation of digital electronics, including logic gates and counting circuits.

Before studying this chapter, you will want to remind yourself of:

- The meaning of voltage (V), current (I) and resistance (R), as covered in Grade 10 (see chapter 10), and Grade 11 (see chapter 19).
- Capacitors in electric circuits, as covered in Grade 11 (see section 17.6).
- Semiconductors, as covered in Grade 11 (see chapter 20).
- The meaning of an alternating current (see section 28.3).
- Capacitance (C) and Inductance (L) (see section 28.4).

29.2 Capacitive and Inductive Circuits

Earlier in Grade 12, you were shown alternating currents (a.c.) and you saw that the voltage and the current varied with time. If the a.c. supply is connected to a resistor, then the current and voltage will be proportional to each other. This means that the current and voltage will 'peak' at the same time. We say that the current and voltage are **in phase**. This is shown in Figure 29.1.

When a capacitor is connected to an alternating voltage, the maximum voltage is proportional to the maximum current, but the maximum voltage does not occur at the same time as the maximum current. The current has its maximum (it peaks) one quarter of a cycle before the voltage peaks. Engineers say that the 'current leads the voltage by $90°$'. This is shown in Figure 29.2.

For a circuit with a capacitor, the instantaneous value of $\frac{V}{I}$ is not constant. However the value of $\frac{V_{\max}}{I_{\max}}$ is useful, and is called the **capacitive reactance** (X_C) of the component. Because it is still a voltage divided by a current (like resistance), its unit is the ohm. The value of X_C (C

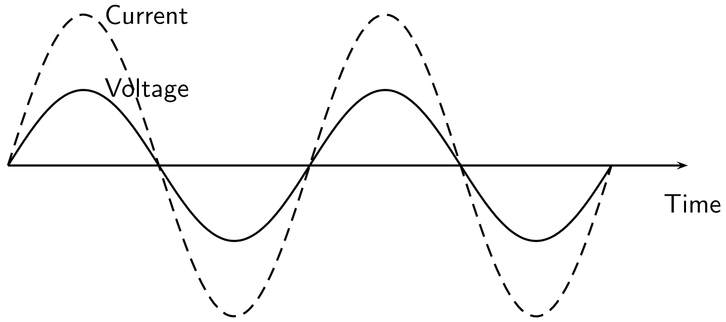


Figure 29.1: The voltage and current are in phase when a resistor is connected to an alternating voltage.

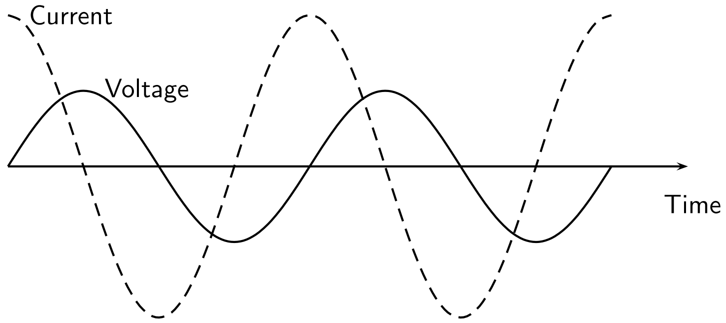


Figure 29.2: The current peaks (has its maximum) one quarter of a wave before the voltage when a capacitor is connected to an alternating voltage.

standing for capacitor) depends on its capacitance (C) and the frequency (f) of the alternating current (in South Africa 50 Hz).

$$X_C = \frac{V_{\max}}{I_{\max}} = \frac{1}{2\pi f C} \tag{29.1}$$

Inductors are very similar, but the current peaks $90°$ after the voltage. This is shown in Figure 29.3. Engineers say that the 'current lags the voltage'. Again, the ratio of maximum voltage to maximum current is called the reactance — this time inductive reactance (X_L). The value of the reactance depends on its inductance (L).

$$X_L = \frac{V_{\max}}{I_{\max}} = 2\pi f L \tag{29.2}$$

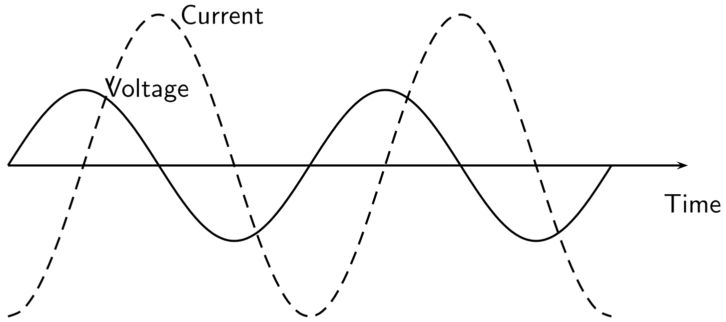


Figure 29.3: The current peaks (has its maximum) one quarter of a wave after the voltage when an inductor is connected to an alternating voltage.

Definition: Reactance
The ratio of the maximum voltage to the maximum current when a capacitor or inductor is connected to an alternating voltage. The unit of reactance is the ohm.

While inductive and capacitive reactances are similar, in one sense they are opposites. For an inductor, the current peaks $90°$ **after** the voltage. For a capacitor the current peaks $90°$ **ahead** of the voltage. When we work out the total reactance for an inductor and a capacitor in series, we use the formula

$$X_{\text{total}} = X_L - X_C \tag{29.3}$$

to take this into account. This formula can also be used when there is more than one inductor or more than one capacitor in the circuit. The total reactance is the sum of all of the inductive reactances minus the sum of all the capacitive reactances. The magnitude (number) in the final result gives the ratio of maximum voltage to maximum current in the circuit as a whole. The sign of the final result tells you its phase. If it is positive, the current peaks $90°$ after the voltage, if it is negative, the current peaks $90°$ before the voltage.

If a series circuit contains resistors as well, then the situation is more complicated. The maximum current is still proportional to the maximum voltage, but the phase difference between them won't be $90°$. The ratio between the maximum voltage and maximum current is called the **impedance** (Z), and its unit is also the ohm. Impedances are calculated using this formula:

$$Z = \sqrt{X^2 + R^2} \tag{29.4}$$

where X is the total reactance of the inductors and capacitors in the circuit, and R is the total resistance of the resistors in the circuit.

It is easier to understand this formula by thinking of a right angled triangle. Resistances are drawn horizontally, reactances are drawn vertically. The hypotenuse of the triangle gives the impedance. This is shown in Figure 29.4.

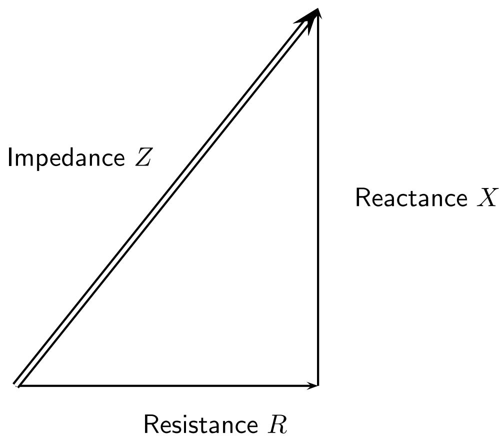


Figure 29.4: Visualizing the relationship between reactance, resistance and impedance.

Definition: Impedance
The maximum voltage divided by the maximum current for any circuit. The unit of impedance is the ohm.

It is important to remember that when resistors and inductances (or capacitors) are in a circuit, the current will not be in phase with the voltage, so the impedance is not a resistance. Similarly the current won't be exactly $90°$ out of phase with the voltage so the impedance isn't a reactance either.

Worked Example 179: The impedance of a coil

Question: Calculate the maximum current in a coil in a South African motor which has a resistance of 5 Ω and an inductance of 3 mH. The maximum voltage across the coil is 6 V. You can assume that the resistance and inductance are in series.
Answer

1. Calculate the reactance of the coil $X_L = 2\pi f L = 2\pi \times 50 \times 0{,}003 = 0{,}942\ \Omega$
2. Calculate the impedance of the coil
$Z = \sqrt{X^2 + R^2} = \sqrt{0{,}942^2 + 5^2} = 5{,}09\ \Omega$

3. Calculate the maximum current $I_{\text{max}} = V_{\text{max}}/Z = 6/5{,}09 = 1{,}18$ A.

Worked Example 180: An RC circuit

Question: Part of a radio contains a 30 Ω resistor in series with a 3 μF capacitor. What is its impedance at a frequency of 1 kHz?

Answer

1. Calculate the reactance of the capacitor

$$X_C = \frac{1}{2\pi fC} = \frac{1}{2\pi \times 10^3 \times 3 \times 10^{-6}} = 53{,}05\ \Omega$$

2. Calculate the impedance $Z = \sqrt{X^2 + R^2} = \sqrt{53{,}05^2 + 30^2} = 60{,}9\ \Omega$

Exercise: Capacitive and Inductive Circuits

1. Why is the instantaneous value of $\frac{V}{I}$ of little use in an a.c. circuit containing an inductor or capacitor?
2. How is the reactance of an inductor different to the reactance of a capacitor?
3. Why can the ratio of the maximum voltage to the maximum current in a circuit with a resistor and an inductor not be called a reactance?
4. An engineer can describe a motor as equivalent to a 30 Ω resistor in series with a 30 mH inductor. If the maximum value of the supply voltage is 350 V, what is the maximum current? Assume that the frequency is 50 Hz.
5. A timer circuit in a factory contains a 200 μF capacitor in series with a 10 kΩ resistor. What is its impedance? Assume that the frequency is 50 Hz.
6. A 3 mH inductor is connected in series with a 100 μF capacitor. The reactance of the combination is zero. What is the frequency of the alternating current?

Interesting Fact

Most factories containing heavy duty electrical equipment (e.g. large motors) have to pay extra money to their electricity supply company because the inductance of the motor coils causes the current and voltage to get out of phase. As this makes the electricity distribution network less efficient, a financial penalty is incurred. The factory engineer can prevent this by connecting capacitors into the circuit to reduce the reactance to zero, as in the last question above. The current and voltage are then in phase again. We can't calculate the capacitance needed in this chapter, because the capacitors are usually connected in parallel, and we have only covered the reactances and impedances of series circuits.

29.3 Filters and Signal Tuning

29.3.1 Capacitors and Inductors as Filters

We have already seen how capacitors and inductors pass current more easily at certain frequencies than others. To recap: if we examine the equation for the reactance of a capacitor, we see that the frequency is in the denominator. Therefore, when the frequency is low, the capacitive reactance is very high. This is why a capacitor blocks the flow of DC and low frequency AC because its reactance increases with decreasing frequency.

When the frequency is high, the capacitive reactance is low. This is why a capacitor allows the flow of high frequency AC because its reactance decreases with increasing frequency.

Therefore putting a capacitor in a circuit blocks the low frequencies and allows the high frequencies to pass. This is called a high pass filter. A filter like this can be used in the 'treble' setting of a sound mixer or music player which controls the amount of high frequency signal reaching the speaker. The more high frequency signal there is, the 'tinnier' the sound. A simple high pass filter circuit is shown in Figure 29.5.

Similarly, if we examine the equation for the reactance of an inductor, we see that inductive reactance increases with increasing frequency. Therefore, when the frequency is low, the inductive reactance is very low. This is why an inductor allows the flow of DC and low frequency AC because its reactance decreases with decreasing frequency.

When the frequency is high, the inductive reactance is high. This is why an inductor blocks the flow of high frequency AC because its reactance increases with increasing frequency.

Therefore putting an inductor in a circuit blocks the high frequencies and allows the low frequencies to pass. This is called a low pass filter. A filter like this can be used in the 'bass' setting of a sound mixer or music player which controls the amount of low frequency signal reaching the speaker. The more low frequency signal there is, the more the sound 'booms'. A simple low pass filter circuit is shown in Figure 29.6.

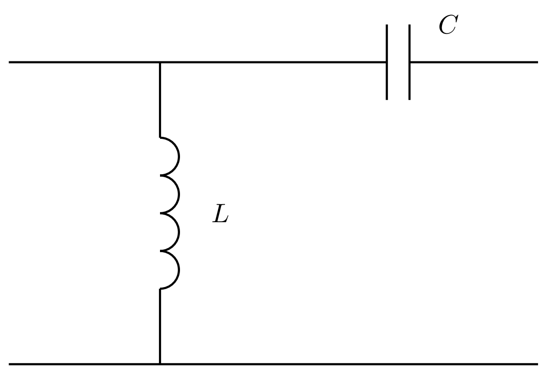


Figure 29.5: A high pass filter. High frequencies easily pass through the capacitor and into the next part of the circuit, while low frequencies pass through the inductor straight to ground.

29.3.2 LRC Circuits, Resonance and Signal Tuning

A circuit containing a resistor, a capacitor and an inductor all in series is called an LRC circuit. Because the components are in series, the current through each component at a particular time will be the same as the current through the others. The voltage across the resistor will be in phase with the current. The voltage across the inductor will be 90° ahead of the current (the current always follows or lags the voltage in an inductor). The voltage across the capacitor will be 90° behind the current (the current leads the voltage for a capacitor). The phases of the three voltages are shown in Figure 29.7.

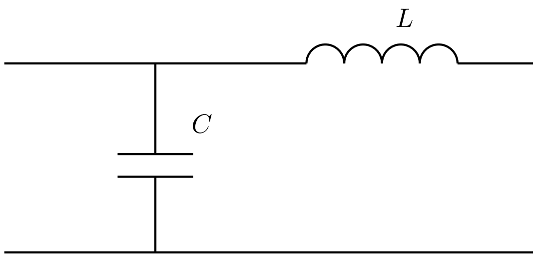


Figure 29.6: A low pass filter. Low frequencies pass through the inductor and into the next part of the circuit, while high frequencies pass through the capacitor straight to ground.

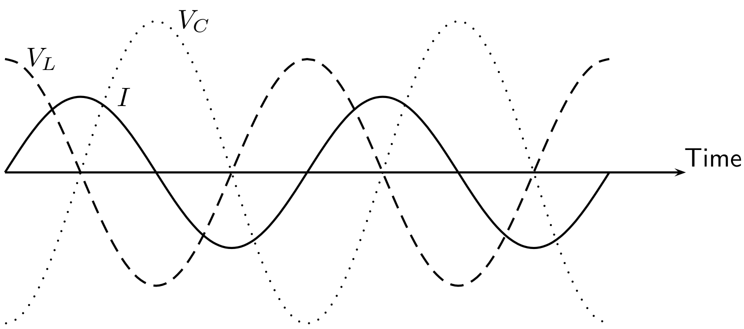


Figure 29.7: The voltages across the separate components of an LRC circuit. Looking at the peaks, you see that the voltage across the inductor V_L 'peaks' first, followed $90°$ later by the current I, followed $90°$ later by the voltage across the capacitor V_C. The voltage across the resistor is not shown — it is in phase with the current and peaks at the same time as the current.

The reactance of the inductor is $2\pi fL$, and the reactance of the capacitor is $1/2\pi fC$ but with the opposite phase. So the total reactance of the LRC circuit is

$$X = X_L - X_C = 2\pi fL - \frac{1}{2\pi fC}$$

The impedance of the circuit as a whole is given by

$$Z = \sqrt{X^2 + R^2} = \sqrt{\left(2\pi fL - \frac{1}{2\pi fC}\right)^2 + R^2}$$

At different frequencies, the impedance will take different values. The impedance will have its smallest value when the positive inductive reactance cancels out the negative capacitive reactance. This occurs when

$$2\pi fL = \frac{1}{2\pi fC}$$

so the frequency of minimum impedance must be

$$f = \frac{1}{2\pi\sqrt{LC}}$$

This is called the **resonant frequency** of the circuit. This is the frequency at which you can get the largest current for a particular supply voltage. It is also called the **natural frequency** of the circuit. This means the frequency at which the circuit would oscillate by itself.

Definition: Resonance
Resonance occurs when a circuit is connected to an alternating voltage at its natural frequency. A very large current in the circuit can build up, even with minimal power input.

An LRC circuit is very useful when we have a signal containing many different frequencies, and we only want one of them. If a signal like this is connected to an LRC circuit, then only the resonant frequency (and other frequencies close to it) will drive measureable currents. This means that an LRC circuit can select one frequency from a range. This process is called **signal tuning**.

Interesting Fact

When you set up a radio antenna, and amplify the radio signal it receives, you find many different bands of frequencies — one from each radio station. When you listen to the radio, you only want to listen to one station. An LRC circuit in the radio (the tuning circuit) is set so that its resonant frequency is the frequency of the station you want to listen to. This means that of the many radio stations broadcasting, you only hear one. When you adjust the tuning dial on the radio, it changes the capacitance of the capacitor in the tuning circuit. This changes the resonant frequency, and this changes the radio station that you will hear.

Exercise: Filters and Signal Tuning

1. Which component would you use if you wanted to block low frequencies?

2. Which component would you use if you wanted to block high frequencies?
3. Calculate the impedance of a series circuit containing a 50 Ω resistor, a 30 μF capacitor and a 3 mH inductor for frequencies of (a) 50 Hz, (b) 500 Hz, and (c) 5 000 Hz.
4. Calculate the resonant frequency of the circuit in the last question.
5. A radio station broadcasts at a frequency of 150 kHz. The tuning circuit in the radio contains a 0.3 mH inductor. What is the capacitance of the capacitor needed in the tuning circuit if you want to listen to this radio station?
6. State the relationship between the phase of the voltages across an inductor, a resistor and a capacitor in an LRC circuit.
7. Explain what is meant by resonance.
8. Explain how LRC circuits are used for signal tuning, for example in a radio.

29.4 Active Circuit Elements

The components you have been learning about so far — resistors, capacitors and inductors — are called **passive** components. They do not change their behaviour or physics in response to changes in voltage or current. **Active** components are quite different. Their response to changes in input allows them to make amplifiers, calculators and computers.

29.4.1 The Diode

A diode is an electronic device that allows current to flow in one direction only.

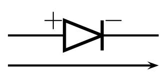


Figure 29.8: Diode circuit symbol and direction of flow of current.

A diode consists of two doped semi-conductors joined together so that the resistance is low when connected one way and very high the other way.

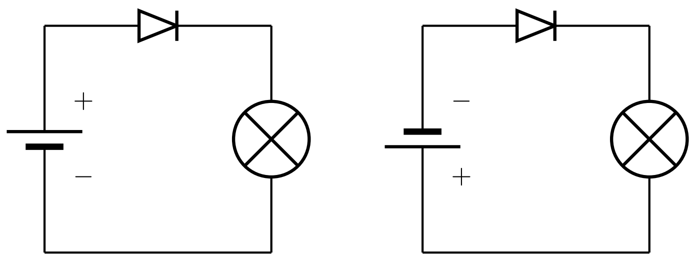


Figure 29.9: Operation of a diode. (a) The diode is forward biased and current is permitted. The negative terminal of the battery is connected to the negative terminal of the diode. (b) The diode is reverse biased and current flow is not allowed. The negative terminal of the battery is connected to the positive terminal of the diode.

A full explanation of diode operation is complex. Here is a simplified description. The diode consists of two semiconductor blocks attached together. Neither block is made of pure silicon — they are both **doped**. Doping was described in more detail in Section 10.3.

In short, p-type semiconductor has fewer free electrons than normal semiconductor. 'P' stands for 'positive', meaning a lack of electrons, although the material is actually neutral. The locations where electrons are missing are called **holes**. This material can conduct electricity well, because electrons can move into the holes, making a new hole somewhere else in the material. Any extra electrons introduced into this region by the circuit will fill some or all of the holes.

In n-type semiconductor, the situation is reversed. The material has an more free electrons than normal semiconductor. 'N' stands for 'negative', meaning an excess of electrons, although the material is actually neutral.

When a p-type semiconductor is attached to an n-type semiconductor, some of the free electrons in the n-type move across to the p-type semiconductor. They fill the available holes near the junction. This means that the region of the n-type semiconductor nearest the junction has no free electrons (they've moved across to fill the holes). This makes this n-type semiconductor positively charged. It used to be electrically neutral, but has since lost electrons.

The region of p-type semiconductor nearest the junction now has no holes (they've been filled in by the migrating electrons). This makes the p-type semiconductor negatively charged. It used to be electrically neutral, but has since gained electrons.

Without free electrons or holes, the central region can not conduct electricity well. It has a high resistance, and is called the **depletion band**. This is shown in Figure 29.10.

You can explain the high resistance in a different way. A free electron in the n-type semiconductor will be repelled from the p-type semiconductor because of its negative charge. The electron will not go into the depletion band, and certainly won't cross the band to the p-type semiconductor. You may ask, "But won't a free electron in the p-type semiconductor be attracted across the band, carrying a current?" But there are no free electrons in p-type semiconductor, so no current of this kind can flow.

If the diode is reverse-biased, the + terminal of the battery is connected to the n-type semiconductor. This makes it even more negatively charged. It also removes even more of the free electrons near the depletion band. At the same time, the − terminal of the battery is connected to the p-type silicon. This will supply free electrons and fill in more of the holes next to the depletion band. Both processes cause the depletion band to get wider. The resistance of the diode (which was already high) increases. This is why a reverse-biased diode does not conduct.

Another explanation for the increased resistance is that the battery has made the p-type semiconductor *more negative* than it used to be, making it repel any electrons from the n-type semiconductor which attempt to cross the depletion band.

On the other hand, if the diode is forward biased, the depletion band is made narrower. The negative charge on the p-type silicon is cancelled out by the battery. The greater the voltage used, the narrower the depletion band becomes. Eventually, when the voltage is about 0,6 V (for silicon) the depletion band disappears. Once this has occurred, the diode conducts very well.

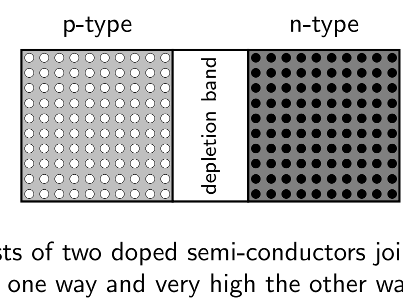


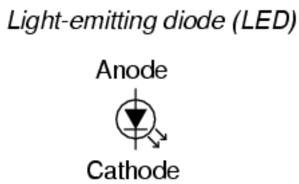
Figure 29.10: A diode consists of two doped semi-conductors joined together so that the resistance is low when connected one way and very high the other way.

Exercise: The Diode

1. What is a diode?
2. What is a diode made of?
3. What is the term which means that a diode is connected the 'wrong way' and little current is flowing?
4. Why is a diode able to conduct electricity in one direction much more easily than the other?

29.4.2 The Light Emitting Diode (LED)

A light-emitting diode (LED) is a diode device that emits light when charge flows in the correct direction through it. If you apply a voltage to force current to flow in the direction the LED allows it will light up.



Extension: Circuit Symbols

This notation of having two small arrows pointing away from the device is common to the schematic symbols of all light-emitting semiconductor devices. Conversely, if a device is light-activated (meaning that incoming light stimulates it), then the symbol will have two small arrows pointing toward it. It is interesting to note, though, that LEDs are capable of acting as light-sensing devices: they will generate a small voltage when exposed to light, much like a solar cell on a small scale. This property can be gainfully applied in a variety of light-sensing circuits.

The color depends on the semiconducting material used to construct the LED, and can be in the near-ultraviolet, visible or infrared part of the electromagnetic spectrum.

Interesting Fact

Nick Holonyak Jr. (1928) of the University of Illinois at Urbana-Champaign developed the first practical visible-spectrum LED in 1962.

Light emission

The wavelength of the light emitted, and therefore its color, depends on the materials forming the p-n junction. A normal diode, typically made of silicon or germanium, emits invisible far-infrared light (so it can't be seen), but the materials used for an LED have emit light corresponding to near-infrared, visible or near-ultraviolet frequencies.

LED applications

LEDs have many uses. Some of these are given here.

- thin, lightweight message displays, e.g. in public information signs (at airports and railway stations, among other places)
- status indicators, e.g. on/off lights on professional instruments and consumers audio/video equipment
- infrared LEDs in remote controls (for TVs, VCRs, etc)
- clusters of LEDs are used in traffic signals, replacing ordinary bulbs behind colored glass
- car indicator lights and bicycle lighting
- calculator and measurement instrument displays (seven segment displays), although now mostly replaced by LCDs
- red or yellow LEDs are used in indicator and [alpha]numeric displays in environments where night vision must be retained: aircraft cockpits, submarine and ship bridges, astronomy observatories, and in the field, e.g. night time animal watching and military field use
- red or yellow LEDs are also used in photographic darkrooms, for providing lighting which does not lead to unwanted exposure of the film
- illumination, e.g. flashlights (a.k.a. torches, UK), and backlighting for LCD screens
- signaling/emergency beacons and strobes
- movement sensors, e.g. in mechanical and optical computer mice and trackballs
- in LED printers, e.g. high-end color printers

LEDs offer benefits in terms of maintenance and safety.

- The typical working lifetime of a device, including the bulb, is ten years, which is much longer than the lifetimes of most other light sources.
- LEDs fail by dimming over time, rather than the abrupt burn-out of incandescent bulbs.
- LEDs give off less heat than incandescent light bulbs and are less fragile than fluorescent lamps.
- Since an individual device is smaller than a centimetre in length, LED-based light sources used for illumination and outdoor signals are built using clusters of tens of devices.

Because they are monochromatic, LED lights have great power advantages over white lights where a specific color is required. Unlike the white lights, the LED does not need a filter that absorbs most of the emitted white light. Colored fluorescent lights are made, but they are not widely available. LED lights are inherently colored, and are available in a wide range of colors. One of the most recently introduced colors is the emerald green (bluish green, about 500 nm) that meets the legal requirements for traffic signals and navigation lights.

Interesting Fact

The largest LED display in the world is 36 m high, at Times Square, New York, U.S.A.

There are applications that specifically require light that does not contain any blue component. Examples are photographic darkroom safe lights, illumination in laboratories where certain photo-sensitive chemicals are used, and situations where dark adaptation (night vision) must be preserved, such as cockpit and bridge illumination, observatories, etc. Yellow LED lights are a good choice to meet these special requirements because the human eye is more sensitive to yellow light.

Exercise: The Light Emitting Diode

1. What is an LED?
2. List 5 applications of LEDs.

29.4.3 Transistor

The diode is the simplest semiconductor device, made up of a p-type semiconductor and an n-type semiconductor in contact. It can conduct in only one direction, but it cannot control the size of an electric current. Transistors are more complicated electronic components which can control the size of the electric current flowing through them.

This enables them to be used in amplifiers. A small signal from a microphone or a radio antenna can be used to control the transistor. In response, the transistor will then increase and decrease a much larger current which flows through the speakers.

One of the earliest popular uses of transistors was in cheap and portable radios. Before that, radios were much more expensive and contained glass valves which were fragile and needed replacing. In some parts of the world you can still hear people talking about their 'transistor' — they mean their portable radio.

You can also use a small current to turn the transistor on and off. The transistor then controls a more complicated or powerful current through other components. When a transistor is used in this way it is said to be in **switching** mode as it is acting as a remotely controlled switch. As we shall see in the final sections of this chapter, switching circuits can be used in a computer to process and store digital information. A computer would not work without the millions (or billions) of transistors in it.

There are two main types of transistor - bipolar transistors (NPN or PNP), and field effect transistors (FETs). Both use doped semiconductors, but in different ways. You are mainly required to know about field effect transistors (FETs), however we have to give a brief description of bipolar transistors so that you see the difference.

Bipolar Transistors

Bipolar transistors are made of a doped semiconductor 'sandwich'. In an NPN transistor, a very thin layer of p-type semiconductor is in between two thicker layers of n-type

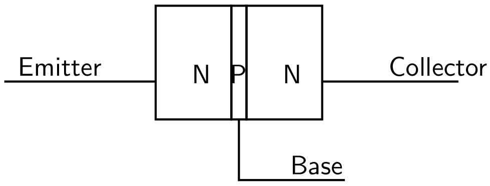


Figure 29.11: An NPN transistor. This is a type of bipolar transistor.

semiconductor. This is shown in Figure 29.11. Similarly an PNP transistor consists of a very thin n-type layer in between two thicker layers of p-type semiconductor.

In an NPN transistor a small current of electrons flows from the emitter (E) to the base (B). Simultaneously, a much larger current of electrons flows from the emitter (E) to the collector (C). If you lower the number of electrons able to leave the transistor at the base (B), the transistor automatically reduces the number of electrons flowing from emitter (E) to collector (C). Similarly, if you increase the current of electrons flowing out of the base (B), the transistor automatically also increases the current of electrons flowing from emitter (E) to collector (C). The transistor is designed so that the current of electrons from emitter to collector (I_{EC}) is proportional to the current of electrons from emitter to base (I_{EB}). The constant of proportionality is known as the **current gain** β. So $I_{EC} = \beta I_{EB}$.

How does it do it? The answer comes from our work with diodes. Electrons arriving at the emitter (n-type semiconductor) will naturally flow through into the central p-type since the base-emitter junction is forward biased. However if none of these electrons are removed from the base, the electrons flowing into the base from the emitter will fill all of the available 'holes'. Accordingly, a large depletion band will be set up. This will act as an insulator preventing current flow into the collector as well. On the other hand, if the base is connected to a positive voltage, a small number of electrons will be removed by the base connection. This will prevent the 'holes' in the base becoming filled up, and no depletion band will form. While some electrons from the emitter leave via the base connection, the bulk of them flow straight on to the collector. You may wonder how the electrons get from the base into the collector (it seems to be reverse biased). The answer is complicated, but the important fact is that the p-type layer is extremely thin. As long as there is no depletion layer, the bulk of the electrons will have no difficulty passing straight from the n-type emitter into the n-type collector. A more satisfactory answer can be given to a university student once band theory has been explained.

Summing up, in an NPN transistor, a small flow of electrons from emitter (E) to base (B) allows a much larger flow of electrons from emitter (E) to collector (C). Given that conventional current (flowing from $+$ to $-$) is in the opposite direction to electron flow, we say that a small conventional current from base to emitter allows a large current to flow from collector to emitter.

A PNP transistor works the other way. A small conventional current from emitter to base allows a much larger conventional current to flow from emitter to collector. The operation is more complicated to explain since the principal charge carrier in a PNP transistor is not the electron but the 'hole'.

The operation of NPN and PNP transistors (in terms of conventional currents) is summarized in Figure 29.12.

The transistor is considered by many to be one of the greatest discoveries or inventions in modern history, ranking with banking and the printing press. Key to the importance of the transistor in modern society is its ability to be produced in huge numbers using simple techniques, resulting in vanishingly small prices. Computer "chips" consist of millions of transistors and sell for

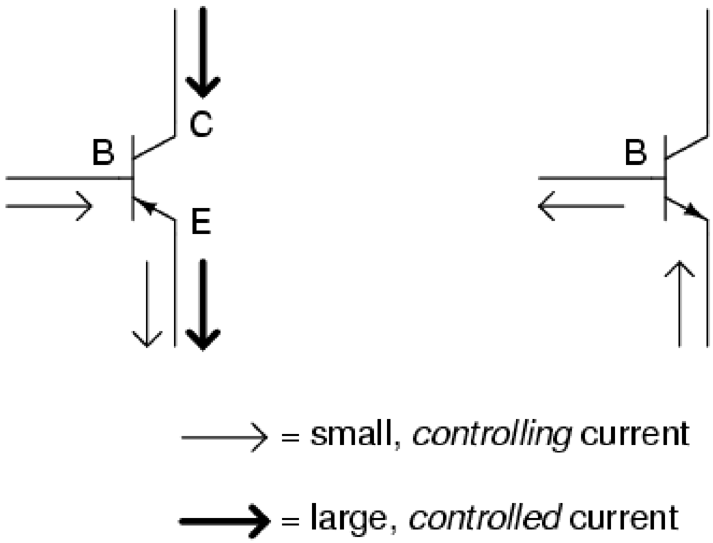


Figure 29.12: An overview of bipolar transistors as current amplifiers.

Rands, with per-transistor costs in the thousandths-of-cents. The low cost has meant that the transistor has become an almost universal tool for non-mechanical tasks. Whereas a common device, say a refrigerator, would have used a mechanical device for control, today it is often less expensive to simply use a few million transistors and the appropriate computer program to carry out the same task through "brute force". Today transistors have replaced almost all electromechanical devices, most simple feedback systems, and appear in huge numbers in everything from computers to cars.

The transistor was invented at Bell Laboratories in December 1947 (first demonstrated on December 23) by John Bardeen, Walter Houser Brattain, and William Bradford Shockley, who were awarded the Nobel Prize in physics in 1956.

The Field Effect Transistor (FET)

To control a bipolar transistors, you control the **current** flowing into or out of its base. The other type of transistor is the field effect transistor (FET). FETs work using control **voltages** instead. Accordingly they can be controlled with much smaller currents and are much more economic to use.

No-one would build a computer with billions of bipolar transistors — the current in each transistor's base might be small, but when you add up all of the base currents in the millions of transistors, the computer as a whole would be consuming a great deal of electricity and making a great deal of heat. Not only is this wasteful, it would prevent manufacturers making a computer of convenient size. If the transistors were too close together, they would overheat.

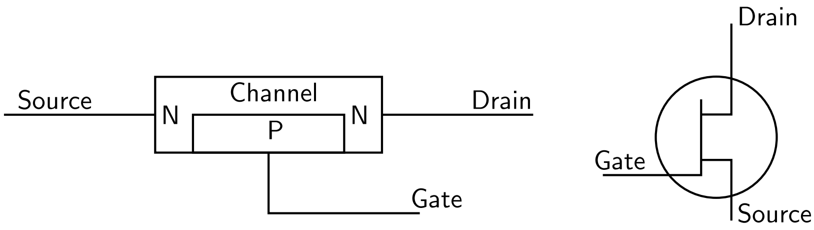


Figure 29.13: A field effect transistor (FET). The diagram on the left shows the semiconductor structure. The diagram on the right shows its circuit symbol.

The three terminals of the FET are called the *source* (S), *drain* (D) and *gate* (G), as shown in Figure 29.13. When the gate is not connected, a current of electrons can flow from source (S) to drain (D) easily along the channel. The source is, accordingly, the negative terminal of the transistor. The drain, where the electrons come out, is the positive terminal of the transistor. A few electrons will flow from the n-type channel into the p-type semiconductor of the gate when the device is manufactured. However, as these electrons are not removed (the gate is not connected), a depletion band is set up which prevents further flow into the gate.

In operation, the gate is connected to negative voltages relative to the source. This makes the p-n junction between gate and channel reverse-biased. Accordingly no current flows from the source into the gate. When the voltage of the gate is lowered (made more negative), the depletion band becomes wider. This enlarged depletion band takes up some of the space of the channel. So the lower the voltage of the gate (the more negative it is relative to the source), the larger the depletion band. The larger the depletion band, the narrower the channel. The narrower the channel, the harder it is for electrons to flow from source to drain.

The voltage of the gate is not the only factor affecting the current of electrons between the source and the drain. If the external circuit has a low resistance, electrons are able to leave the drain easily. If the external circuit has a high resistance, electrons leave the drain slowly. This creates a kind of 'traffic jam' which slows the passage of further electrons. In this way, the voltage of the drain regulates itself, and is more or less independent of the current demanded from the drain.

Once these two factors have been taken into account, it is fair to say that the positive output voltage (the voltage of the drain relative to the source) is proportional to the negative input voltage (the voltage of the gate relative to the source).

For this reason, the field effect transistor is known as a voltage amplifier. This contrasts with the bipolar transistor which is a current amplifier.

?

Exercise: Field Effect Transistors

1. What are the two types of bipolar transistor? How does their construction differ?
2. What are the three connections to a bipolar transistor called?
3. Why are very few electrons able to flow from emitter to collector in an NPN transistor if the base is not connected?

4. Why do you think a bipolar transistor would not work if the base layer were too thick?
5. "The bipolar transistor is a current amplifier." What does this statement mean?
6. Describe the structure of a FET.
7. Define what is meant by the source, drain and gate. During normal operation, what will the voltages of drain and gate be with respect to the source?
8. Describe how a depletion layer forms when the gate voltage is made more negative. What controls the width of the depletion layer?
9. "The field effect transistor is a voltage amplifier." What does this statement mean?
10. The amplifier in a cheap radio will probably contain bipolar transistors. A computer contains many field effect transistors. Bipolar transistors are more rugged and less sensitive to interference than field effect transistors, which makes them more suitable for a simple radio. Why are FETs preferred for the computer?

29.4.4 The Operational Amplifier

The operational amplifier is a special kind of voltage amplifier which is made from a handful of bipolar or field effect transistors. Operational amplifiers are usually called **op-amps** for short. They are used extensively in all kinds of audio equipment (amplifiers, mixers and so on) and in instrumentation. They also have many other uses in other circuit - for example comparing voltages from sensors.

Operational amplifiers are supplied on Integrated Circuits (I.C.s). The most famous operational amplifier I.C. is numbered 741 and contains a single operational amplifier on an integrated circuit ('chip') with eight terminals. Other varieties can be bought, and you can get a single integrated circuit with two or four '741'-type operational amplifiers on it.

The symbol for an op-amp is shown in Figure 29.14. The operational amplifier has two input terminals and one output terminal. The voltage of the output terminal is proportional to the difference in voltage between the two input terminals. The output terminal is on the right (at the sharp point of the triangle). The two input terminals are drawn on the left. One input terminal (labelled with a $+$ on diagrams) is called the **non-inverting input**. The other input terminal (labelled $-$) is called the **inverting input**. The labels $+$ and $-$ have nothing to do with the way in which the operational amplifier is connected to the power supply. Operational amplifiers must be connected to the power supply, but this is taken for granted when circuit diagrams are drawn, and these connections are not shown on circuit diagrams. Usually, when drawing electronic circuits, '0V' is taken to mean the negative terminal of the power supply. This is not the case with op-amps. For an op-amp, '0V' refers to the voltage midway between the $+$ and $-$ of the supply.

The output voltage of the amplifier V_{out} is given by the formula

$$V_{out} = A\left(V_{+} - V_{-}\right) \tag{29.5}$$

here A is a constant called the **open loop gain**, and V_{+} and V_{-} are the voltages of the two input terminals. That said, the output voltage can not be less than the voltage of the negative terminal of the battery supplying it or higher than the positive terminal of the battery supplying it. You will notice that V_{out} is positive if $V_{+} > V_{-}$ and negative if $V_{+} < V_{-}$. This is why the $-$ input is called the inverting input: raising its voltage causes the output voltage to *drop*.

The input resistance of an operational amplifier is very high. This means that very little current flows into the input terminals during operation.

If all of the transistors in the operational amplifier were identical then the output voltage would be zero if the two inputs were at equal voltages. In practice this is not quite the case, and for sensitive work a **trimming potentiometer** is connected. This is adjusted until the op-amp is zeroed correctly.

Simple operational amplifiers require the trimming potentiometer to be built into the circuit containing them, and an example is shown in Figure 29.15. Other operational amplifier designs incorporate separate terminals for the trimming potentiometer. These special terminals are labelled **offset** on the manufacturer's diagram. The exact method of connecting the potentiometer to the offset terminals can depend on the design of the operational amplifier, and you need to refer to the manufacturer's data sheet for details of which potentiometer to use and how to connect it.

For most commercially produced operational amplifiers (known as op-amps for short), the open loop gain A is very large and does not stay constant. Values of 100 000 are typical. Usually a designer would want an amplifier with a stable gain of smaller value, and builds the operational amplifier into a circuit like the one in Figure 29.15.

Extension: Calculating the gain of the amplifier in Figure 29.15.

1. The input resistance of the operational amplifier is very high. This means that very little current flows into the inverting input of the op-amp. Accordingly, the current through resistor R_1 must be almost the same as the current through resistor R_2. This means that the ratio of the voltage across R_1 to the voltage across R_2 is the same as the ratio of the two resistances.
2. The open loop gain A of the op-amp is very high. Assuming that the output voltage is less than a few volts, this means that the two input terminals must be at very similar voltages. We shall assume that they are at the same voltage.
3. We want the output voltage to be zero if the input voltage is zero. Assuming that the transistors within the op-amp are very similar, the output voltage will only be zero for zero input voltage if V_+ is very close to zero. We shall assume that $V_+ = 0$ when the trimming potentiometer is correctly adjusted.
4. It follows from the last two statements that $V_- \approx 0$, and we shall assume that it is zero.
5. With these assumptions, the voltage across R_2 is the same as V_{out}, and the voltage across R_1 is the same as V_{in}. Since both resistors carry the same current (as noted in point 1), we may say that the magnitude of $V_{out}/V_{in} = R_2/R_1$. However, if V_{in} is negative, then V_{out} will be positive. Therefore it is customary to write the gain of this circuit as $V_{out}/V_{in} = -R_2/R_1$.

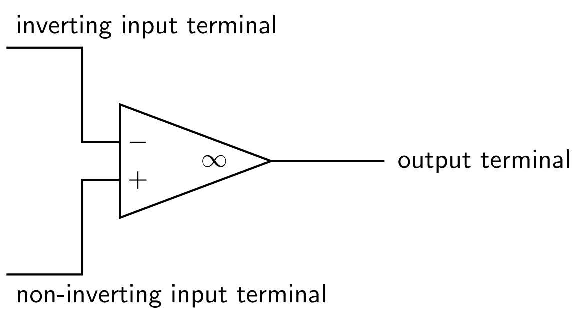


Figure 29.14: Circuit symbol for an operational amplifier. The amplifier must also be connectd to the + and − terminals of the power supply. These connections are taken for granted and not shown.

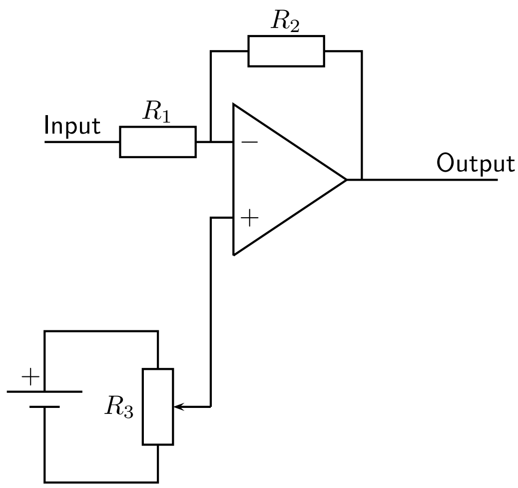


Figure 29.15: An inverting amplifier built using an operational amplifier. The connections from battery to operational amplifier are not shown. The output voltage $V_{out} = -R_2 V_{in}/R_1$, as explained in the text. The potentiometer R_3 is a trimming potentiometer. To set it, the input is connected to zero volts. The trimming potentiometer is then adjusted until $V_{out} = 0$. In all operational amplifier circuits, zero volts is midway between the $+$ and $-$ of the supply.

?

Exercise: Operational Amplifiers

1. What are operational amplifiers used for?
2. Draw a simple diagram of an operational amplifier and label its terminals.
3. Why is a trimming potentiometer is needed when using an op-amp?

29.5 The Principles of Digital Electronics

The circuits and components we have discussed are very useful. You can build a radio or television with them. You can make a telephone. Even if that was all there was to electronics, it would still be very useful. However, the great breakthrough in the last fifty years or so has been in **digital** electronics. This is the subject which gave us the computer. The computer has revolutionized the way business, engineering and science are done. Small computers programmed to do a specific job (called microprocessors) are now used in almost every electronic machine from cars to washing machines. Computers have also changed the way we communicate. We used to have telegraph or telephone wires passing up and down a country — each one carrying one telephone call or signal. We now have optic fibres each capable of carrying tens of thousands of telephone calls using **digital** signals.

So, what is a digital signal? Look at Figure 29.16. A normal signal, called an **analogue** signal, carries a smooth wave. At any time, the voltage of the signal could take any value. It could be 2,00 V or 3,53 V or anything else. A digital signal can only take certain voltages. The simplest case is shown in the figure — the voltage takes one of two values. It is either **high**, or it is **low**. It never has any other value.

These two special voltages are given symbols. The low voltage level is written 0, while the high voltage level is written as 1. When you send a digital signal, you set the voltage you want (0 or 1), then keep this fixed for a fixed amount of time (for example 0.01 μs), then you send the next 0 or 1. The digital signal in Figure 29.16 could be written 01100101.

Why are digital signals so good?

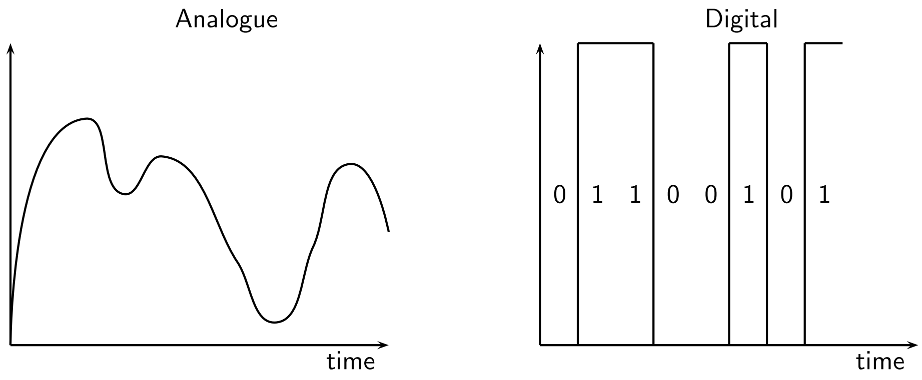


Figure 29.16: The difference between normal (analogue) signals and digital signals. The analogue signal is on the left.

1. Using a computer, any information can be turned into a pattern of 0s and 1s. Pictures, recorded music, text and motion pictures can all be turned into a string of 0s and 1s and transmitted or stored in the same way. The computer receiving the signal at the other end converts it back again. A Compact Disc (CD) for example, can store music or text or pictures, and all can be read using a computer.
2. The 0 and the 1 look very different. You can immediately tell if a 0 or a 1 is being sent. Even if there is interference, you can still tell whether the sender sent a 0 or a 1. This means that fewer mistakes are made when reading a digital signal. This is why the best music recording technologies, and the most modern cameras, for example, all use digital technology.
3. Using the 0s and 1s you can count, and do all kinds of mathematics. This will be explained in more detail in the next section.

The simplest digital circuits are called **logic gates**. Each logic gate makes a decision based on the information it receives. Different logic gates are set up to make the decisions in different ways. Each logic gate will be made of many microscopic transistors connected together within a thin wafer of silicon. This tiny circuit is called an Integrated Circuit or I.C. - all the parts are in one place (integrated) on the silicon wafer.

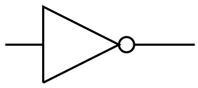
29.5.1 Logic Gates

There are five main types of logic gate: NOT, AND, OR, NAND and NOR. Each one makes its decision in a different way.

The NOT Gate

Problem: You want an automatic circuit in your office to turn on the heating in the winter. You already have a digital electronic temperature sensor. When the temperature is high, it sends out a 1. When the office is cold, it sends out a 0. If this signal were sent straight to the heater, the heater would turn on (1) when it was already hot, and would stay off when it was cold. This is wrong! To make the heater work, we need a circuit which will change a 0 (from the sensor) into a 1 (to send to the heater). This will make the heater come on when it is cold. You also want it to change a 1 (from the sensor) into a 0 (to send to the heater). This will turn the heater off when the room is hot. This circuit is called an **inverter** or **NOT gate**. It changes 0 into 1 (1 is NOT 0). It changes 1 into 0 (0 is NOT 1). It changes a signal into what it is NOT.

The symbol for the NOT gate is:



The action of the NOT gate is written in a table called a **truth table**. The left column shows the possible inputs on different rows. The right column shows what the output (decision) of the circuit will be for that input. The truth table for the NOT gate is shown below.

Input	Output
0	1
1	0

When you read the truth table, the top row says, "If the input is 0, the output will be 1." For our heater, this means, "If the room is cold, the heater will turn on." The bottom row says, "If the input is 1, the output will be 0." For our heater, this means, "If the room is hot, the heater will switch off."

The AND Gate

Problem: An airliner has two toilets. Passengers get annoyed if they get up from their seat only to find that both toilets are being used and they have to go back to their seat and wait. You want to fit an automatic circuit to light up a display if both toilets are in use. Then passengers know that if the light is off, there will be a free toilet for them to use. There is a sensor in each toilet. It gives out a 0 of the toilet is free, and a 1 if it is in use. You want to send a 1 to the display unit if **both** sensors are sending 1s. To do this, you use an AND gate.

The symbol for the AND gate is:

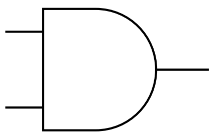


Figure 29.17: Symbol for the AND logic gate.

The truth table for the AND gate is shown below. An AND gate has two inputs (the NOT gate only had one). This means we need four rows in the truth table, one for each possible set of inputs. The first row, for example, tells us what the AND gate will do if both inputs are 0. In our airliner, this means that both toilets are free. The right column has a 0 showing that the output will be 0, so the display will not light up. The second row has inputs of 0 and 1 (the first toilet is free, the other is in use). Again the output is 0. The third row tells us what will happen if the inputs are 1 and 0 (the first toilet is in use, and the second is free). Finally, the last line tells us what will happen if both inputs are 1 (both toilets are in use). It is only in this case that the output is 1 and the display lights up.

Inputs		Output
A	B	
0	0	0
0	1	0
1	0	0
1	1	1

This device is called an AND gate, because the output is only 1 if one input AND the other input are both 1.

Extension: Using 0 and 1 to mean True and False

When we use logic gates we use the low voltage state 0 to represent 'false'. The high voltage state 1 represents 'true'. This is why the word AND is so appropriate. A AND B is true (1) if, and only if, A is true (1) AND B is true (1).

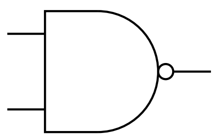
Extension: AND and multiplication

Sometimes, the AND operation is written as multiplication. A AND B is written AB. If either A or B are 0, then AB will also be 0. For AB to be 1, we need A and B to both be 1. Multiplication of the numbers 0 and 1 does exactly the same job as an AND gate.

The NAND Gate

Problem: You build the circuit for the airliner toilets using an AND gate. Your customer is pleased, but she says that it would be better if the display lit up when there **was** a free toilet. In other words, the display should light up unless both toilets are in use. To do this we want a circuit which does the opposite of an AND gate. We want a circuit which would give the output 0 if an AND gate would give 1. We want a circuit which would give the output 1 if an AND gate would give 0. This circuit is called a NAND gate.

The symbol for the NAND gate is:



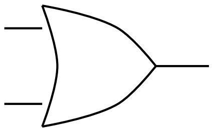
The truth table for the NAND gate is shown below.

Inputs		Output
A	B	
0	0	1
0	1	1
1	0	1
1	1	0

You may have noticed that we could have done this job on the airliner by using our earlier circuit, with a NOT gate added between the original AND gate and the display. This is where the word NAND comes from — it is short for NotAND.

The OR Gate

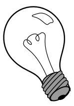
Problem: A long, dark corridor has two light switches — one at each end of the corridor. The switches each send an output of 0 to the control unit if no-one has pressed the switch. If someone presses the switch, its output is 1. The lights in the corridor should come on if either switch is pressed. To do this job, the control unit needs an OR gate. The symbol for the OR gate is:



The truth table for the OR gate is shown.

Inputs		Output
A	B	
0	0	0
0	1	1
1	0	1
1	1	1

You can see that the output is 1 (and the lights come on in the corridor) if either one switch OR the other is pressed. Pressing both switches also turns on the lights, as the last row in the table shows.



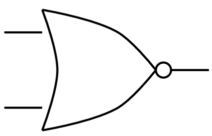
Extension: OR and addition

Sometimes you will see A OR B written mathematically as A+B. This makes sense, since if A=0 and B=0, then A OR B = A+B = 0. Similarly, if A=0 and B=1, then A OR B = A+B = 1. If A=1 and B=0, then A OR B = A+B = 1 once again. The only case where the OR function differs from normal addition is when A=1 and B=1. Here A OR B = 1 in logic, but A+B=2 in arithmetic. However, there is no such thing as '2' in logic, so we define + to mean 'OR', and write 1+1=1 with impunity!

If you wish, you can prove that the normal rules of algebra still work using this notation: A+(B+C) = (A+B)+C, A(BC) = (AB)C, and A(B+C) = AB + AC. This special kind of algebra where variables can only be 0 (representing false) or 1 (representing true) is called Boolean algebra.

The NOR Gate

The last gate you need to know is the NOR gate. This is opposite to the OR gate. The output is 1 if both inputs are 0. In other words, the output switches on if neither the first NOR the second input is 1. The symbol for the NOR gate is:



The truth table for the NOR gate is shown below.

Inputs		Output
A	B	
0	0	1
0	1	0
1	0	0
1	1	0

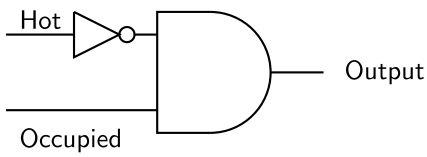
The examples given were easy. Each job only needed one logic gate. However any 'decision making' circuit can be built with logic gates, no matter how complicated the decision. Here is an example.

Worked Example 181: An Economic Heating Control

Question: A sensor in a building detects whether a room is being used. If it is empty, the output is 0, if it is in use, the output is 1. Another sensor measures the temperature of the room. If it is cold, the output is 0. If it is hot, the output is 1. The heating comes on if it receives a 1. Design a control circuit so that the heating only comes on if the room is in use and it is cold.

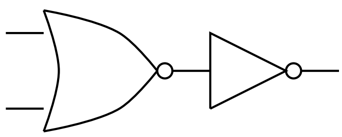
Answer

The first sensor tells us whether the room is occupied. The second sensor tells us whether the room is hot. The heating must come on if the room is occupied AND cold. This means that the heating should come on if the room is occupied AND (NOT hot). To build the circuit, we first attach a NOT gate to the output of the temperature sensor. This output of the NOT gate will be 1 only if the room is cold. We then attach this output to an AND gate, together with the output from the other sensor. The output of the AND gate will only be 1 if the room is occupied AND the output of the NOT gate is also 1. So the heating will only come on if the room is in use and is cold. The circuit is shown below.



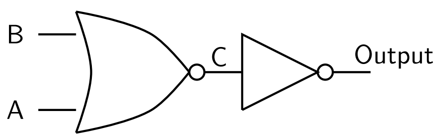
Worked Example 182: Solving a circuit with two logic gates

Question: Compile the truth table for the circuit below.



Answer

Firstly, we label the inputs A and B. We also label the point where the two gates are connected C.



Next we prepare a truth table. There is a column for each of the inputs, for the intermediate point C and also for the output. The truth table has four rows, since there are four possible inputs — 00, 01, 10 and 11.

A	B	C	Output
0	0		
0	1		
1	0		
1	1		

Next we fill in the C column given that we know what a NOR gate does.

A	B	C	Output
0	0	1	
0	1	0	
1	0	0	
1	1	0	

Next, we can fill in the output, since it will always be the opposite of C (because of the NOT gate).

A	B	C	Output
0	0	1	0
0	1	0	1
1	0	0	1
1	1	0	1

Finally we see that this combination of gates does the same job as an OR gate.

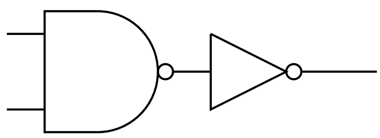
Each logic gate is manufactured from two or more transistors. Other circuits can be made using logic gates, as we shall see in the next section. We shall show you how to count and store numbers using logic gates. This means that if you have enough transistors, and you connect them correctly to make the right logic gates, you can make circuits which count and store numbers.

In practice, the cheapest gate to manufacture is usually the NAND gate. Additionally, Charles Peirce showed that NAND gates alone (as well as NOR gates alone) can be used to reproduce all the other logic gates.

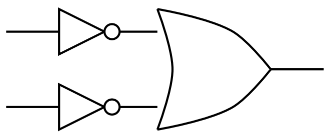
?

Exercise: The Principles of Digital Electronics

1. Why is digital electronics important to modern technology and information processing?
2. What two symbols are used in digital electronics, to represent a “high” and a “low”? What is this system known as?
3. What is a logic gate?
4. What are the five main types of logic gates? Draw the symbol for each logic gate.
5. Write out the truth tables for each of the five logic gates.
6. Write out the truth table for the following circuit. Which single gate is this circuit equivalent to?



7. Write out the truth table for the following circuit. Which single gate is this circuit equivalent to?



29.6 Using and Storing Binary Numbers

In the previous section, we saw how the numbers 0 and 1 could represent 'false' and 'true' and could be used in decision making. Often we want to program a computer to count with numbers. To do this we need a way of writing any number using nothing other than 0 and 1. When written in this way, numbers are called **binary numbers**.

Definition: Binary
A way of writing any number using only the digits 0 and 1.

29.6.1 Binary numbers

In normal (denary) numbers, we write 9+1 as 10. The fact that the '1' in 10 is the second digit from the right tells us that it actually means 10 and not 1. Similarly, the '3' in 365 represents 300 because it is the third digit from the right. You could write 365 as $3 \times 100 + 6 \times 10 + 5$. You will notice the pattern that the nth digit from the right represents 10^{n-1}. In binary, we use the nth digit from the right to represent 2^{n-1}. Thus 2 is written as 10 in binary. Similarly $2^2 = 4$ is written as 100 in binary, and $2^3 = 8$ is written as 1000 in binary.

Worked Example 183: Conversion of Binary Numbers to Denary Numbers

Question: Convert the binary number 10101 to its denary equivalent.
Answer
We start on the right. The '1' on the right does indeed represent one. The next '1' is in the third place from the right, and represents $2^2 = 4$. The next '1' is in the fifth place from the right and represents $2^4 = 16$. Accordingly, the binary number 10101 represents 16+4+1 = 21 in denary notation.

Worked Example 184: Conversion of Denary Numbers to Binary Numbers

Question: Convert the decimal number 12 to its binary equivalent.
Answer
Firstly we write 12 as a sum of powers of 2, so 12 = 8+4. In binary, eight is 1000, and four is 100. This means that twelve = eight + four must be 1000+100 = 1100 in binary. You could also write 12 as $1 \times 8 + 1 \times 4 + 0 \times 2 + 0 \times 1 = 1100$ in binary.

How do you write numbers as a sum of powers of two? The first power of two (the largest) is the largest power of two which is not larger than the number you are working with. In our last example, where we wanted to know what twelve was in binary, the largest power of two which is not larger than 12 is 8. Thus $12 = 8 +$ something. By arithmetic, the 'something' must be 4, and the largest power of two not larger than this is 4 exactly. Thus $12 = 8 + 4$, and we have finished.

A more complicated example would be to write one hundred in binary. The largest power of two not larger than 100 is 64 (1000000 in binary). Subtracting 64 from 100 leaves 36. The largest power of two not larger than 36 is 32 (100000 in binary). Removing this leaves a remainder of 4, which is a power of two itself (100 in binary). Thus one hundred is $64 + 32 + 4$, or in binary $1000000 + 100000 + 100 = 1100100$.

Once a number is written in binary, it can be represented using the low and high voltage levels of digital electronics. We demonstrate how this is done by showing you how an electronic counter works.

29.6.2 Counting circuits

To make a counter you need several 'T flip flops', sometimes called 'divide by two' circuits. A T flip flop is a digital circuit which swaps its output (from 0 to 1 or from 1 to 0) whenever the input changes from 1 to 0. When the input changes from 0 to 1 it doesn't change its output. It is called a **flip flop** because it changes (flips or flops) each time it receives a pulse.

If you put a series of pulses 10101010 into a T flip flop, the result is 01100110. Figure 29.18 makes this clearer.

As you can see from Figure 29.18, there are half as many pulses in the output. This is why it is called a 'divide by two' circuit.

If we connect T flip flops in a chain, then we make a counter which can count pulses. As an example, we connect three T flip flops in a chain. This is shown in Figure 29.19.

When this circuit is fed with a stream of pulses, the outputs of the different stages change. The table below shows how this happens. Each row shows a different stage, with the first stage at the top. We assume that all of the flip flops have 0 as their output to start with.

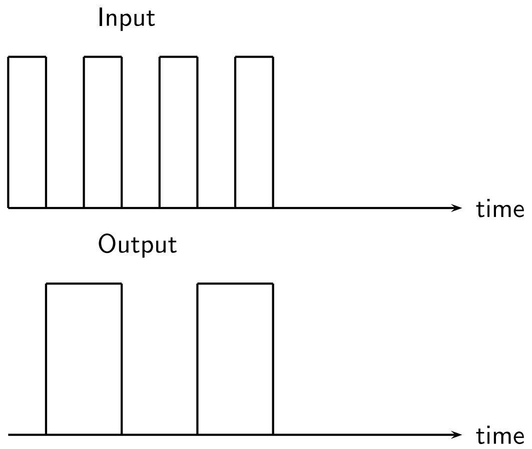


Figure 29.18: The output of a T flip flop, or 'divide by two' circuit when a square wave is connected to the input. The output changes state when the input goes from 1 to 0.

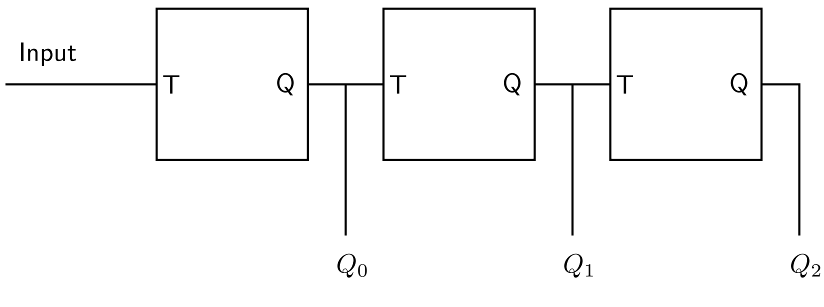


Figure 29.19: Three T flip flops connected together in a chain to make a counter. The input of each flip flop is labelled T, while each output is labelled Q. The pulses are connected to the input on the left. The outputs Q_0, Q_1 and Q_2 give the three digits of the binary number as the pulses are counted. This is explained in the text and in the next table.

Input	Output 1	Output 2	Output 3	Number of pulse	Number in binary
1	0	0	0	0	000
0	1	0	0	1	001
1	1	0	0	1	001
0	0	1	0	2	010
1	0	1	0	2	010
0	1	1	0	3	011
1	1	1	0	3	011
0	0	0	1	4	100
1	0	0	1	4	100
0	1	0	1	5	101
1	1	0	1	5	101
0	0	1	1	6	110
1	0	1	1	6	110
0	1	1	1	7	111
1	1	1	1	7	111
0	0	0	0	8	1000
1	0	0	0	8	1000
0	1	0	0	9	1101
1	1	0	0	9	1101

The binary numbers in the right hand column count the pulses arriving at the input. You will notice that the output of the first flip flop gives the right most digit of the pulse count (in binary). The output of the second flip flop gives the second digit from the right (the 'twos' digit) of the pulse count. The output of the third flip flop gives the third digit from the right (the 'fours' digit) of the pulse count. As there are only three flip flops, there is nothing to provide the next digit (the 'eights' digit), and so the eighth pulse is recorded as 000, not 1000.

This device is called a **modulo 8** counter because it can count in eight stages from 000 to 111 before it goes back to 000. If you put four flip flops in the counter, it will count in sixteen stages from 0000 to 1111, and it is called a modulo 16 counter because it counts in sixteen stages before going back to 0000.

Definition: Modulo
The modulo of a counter tells you how many stages (or pulses) it receives before going back to 0 as its output. Thus a modulo 8 counter counts in eight stages 000, 001, 010, 011, 100, 101, 110, 111, then returns to 000 again.

Interesting Fact

If a counter contains n flip flops, it will be a modulo 2^n counter. It will count from 0 to $2^n - 1$.

29.6.3 Storing binary numbers

Counting is important. However, it is equally important to be able to remember the numbers. Computers can convert almost anything to a string of 0s and 1s, and therefore to a binary number. Unless this number can be stored in the computer's memory, the computer would be useless.

The memory in the computer contains many parts. Each part is able to store a single 0 or 1. Since 0 and 1 are the two binary digits, we say that each part of the memory stores one **bit**.

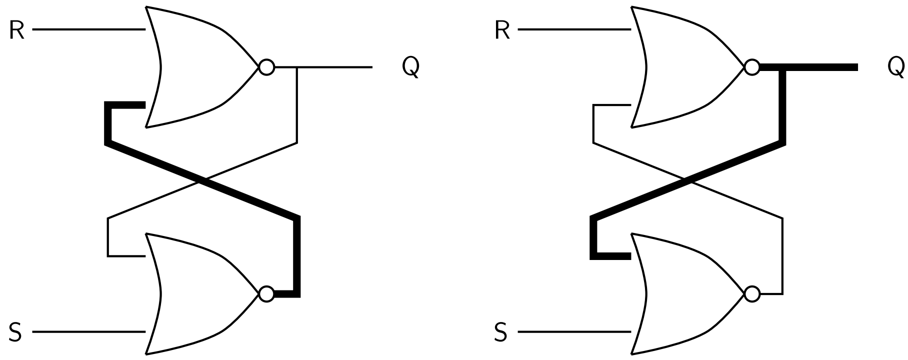


Figure 29.20: A bistable circuit made from two NOR gates. This circuit is able to store one bit of digital information. With the two inputs set to 0, you can see that the output could be (and will remain) either 0 or 1. The circuit on the left shows an output of 0, the circuit on the right shows an output of 1. Wires carrying high logic levels (1) are drawn thicker. The output of the bistable is labelled Q.

Definition: Bit
One bit is a short way of saying one 'binary digit'. It is a single 0 or 1.

If you have eight bits, you can store a binary number from 00000000 to 11111111 (0 to 255 in denary). This gives you enough permutations of 0s and 1s to have one for each letter of the alphabet (in upper and lower case), each digit from 0 to 9, each punctuation mark and each control code used by a computer in storing a document. When you type text into a word processor, each character is stored as a set of eight bits. Each set of eight bits is called a **byte**. Computer memories are graded according to how many bytes they store. There are 1024 bytes in a kilobyte (kB), 1024×1024 bytes in a megabyte (MB), and $1024 \times 1024 \times 1024$ bytes in a gigabyte (GB).

To store a bit we need a circuit which can 'remember' a 0 or a 1. This is called a **bistable** circuit because it has two stable states. It can stay indefinitely either as a 0 or a 1. An example of a bistable circuit is shown in Figure 29.20. It is made from two NOR gates.

To store the 0 or the 1 in the bistable circuit, you set one of the inputs to 1, then put it back to 0 again. If the input labelled 'S' (set) is raised, the output will immediately become 1. This is shown in Figure 29.21.

To store a 0, you raise the 'R' (reset) input to 1. This is shown in Figure 29.22.

Once you have used the S or R inputs to set or reset the bistable circuit, you then bring both inputs back to 0. The bistable 'remembers' the state. Because of the ease with which the circuit can be Reset and Set it is also called a **RS flip flop** circuit.

A computer memory will be able to store millions or billions of bits. If it used our circuit above, it would need millions or billions of NOR gates, each of which is made from several transistors. The computer memory is made of many millions of transistors.

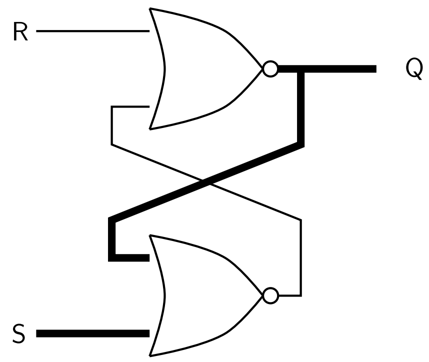


Figure 29.21: The output of a bistable circuit is **set** (made 1) by raising the 'S' input to 1. Wires carrying high logic levels (1) are shown with thicker lines.

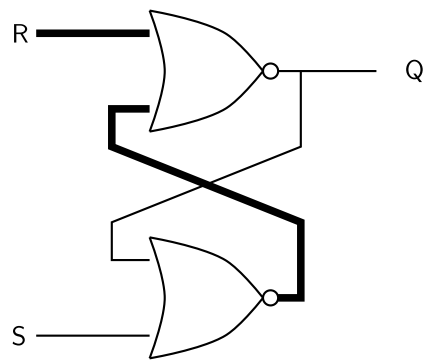


Figure 29.22: The output of a bistable circuit is **reset** (made 0) by raising the 'R' input to 1. Wires carrying high logic levels (1) are shown with thicker lines.

The bistable circuits drawn here don't remember 0s or 1s for ever — they lose the information if the power is turned off. The same is true for the RAM (Random Access Memory) used to store working and temporary data in a computer. Some modern circuits contain special memory which can remember its state even if the power is turned off. This is used in FLASH drives, commonly found in USB data sticks and on the memory cards used with digital cameras. These bistable circuits are much more complex.

You can also make T flip flops out of logic gates, however these are more complicated to design.

Exercise: Counting Circuits

1. What is the term *bit* short for?
2. What is 43 in binary?
3. What is 1100101 in denary?
4. What is the highest number a modulo 64 counter can count to? How many T flip flops does it contain?
5. What is the difference between an RS flip flop and a T flip flop?
6. Draw a circuit diagram for a bistable circuit (RS flip flop). Make three extra copies of your diagram. On the first diagram, colour in the wires which will carry high voltage levels (digital 1) if the R input is low, and the S input is high. On the second diagram, colour in the wires which carry high voltage levels if the S input of the first circuit is now made low. On the third diagram, colour in the wires which carry high voltage levels if the R input is now made high. On the final diagram, colour in the wires carrying high voltage levels if the R input is now made low again.
7. Justify the statement: a modern computer contains millions of transistors.

Exercise: End of Chapter Exercises

1. Calculate the reactance of a 3 mH inductor at a frequency of 50 Hz.
2. Calculate the reactance of a 30 μF capacitor at a frequency of 1 kHz.
3. Calculate the impedance of a series circuit containing a 5 mH inductor, a 400 μF capacitor and a 2 kΩ resistor at a frequency of 50 kHz.
4. Calculate the frequency at which the impedance of the circuit in the last question will be the smallest.
5. Which component can be used to block low frequencies?
6. Draw a circuit diagram with a battery, diode and resistor in series. Make sure that the diode is forward biased so that a current will flow through it.

7. When building a complex electronic circuit which is going to be powered by a battery, it is always a good idea to put a diode in series with the battery. Explain how this will protect the circuit if the user puts the battery in the wrong way round.
8. Summarize the differences betwen a bipolar and field effect transistor.
9. What does an operational amplifier (op-amp) do?
10. What is the difference between a digital signal and an analogue signal?
11. What are the advantages of digital signals over analogue signals?
12. Draw the symbols for the five logic gates, and write down their truth tables.
13. Draw a circuit diagram with an AND gate. Each input should be connected to the output of a separate NOT gate. By writing truth tables show that this whole circuit behaves as a NOR gate.
14. Convert the denary number 99 into binary.
15. Convert the binary number 11100111 into denary.
16. Explain how three T flip flops can be connected together to make a modulo 8 counter. What is the highest number it can count up to?
17. Draw the circuit diagram for an RS flip flop (bistable) using two NOR gates.
18. Show how the circuit you have just drawn can have a stable output of 0 or 1 when both inputs are 0.
19. Operational (and other) amplifiers, logic gates, and flip flops all contain transistors, and would not work without them. Write a short newspaper article for an intelligent reader who knows nothing about electronics. Explain how important transistors are in modern society.

Chapter 30

Electromagnetic Radiation

30.1 Introduction

This chapter will focus on the electromagnetic (EM) radiation. Electromagnetic radiation is a self-propagating wave in space with electric and magnetic components. These components oscillate at right angles to each other and to the direction of propagation, and are in phase with each other. Electromagnetic radiation is classified into types according to the frequency of the wave: these types include, in order of increasing frequency, radio waves, microwaves, infrared radiation, visible light, ultraviolet radiation, X-rays and gamma rays.

30.2 Particle/wave nature of electromagnetic radiation

If you watch a colony of ants walking up the wall, they look like a thin continuous black line. But as you look closer, you see that the line is made up of thousands of separated black ants.

Light and all other types of electromagnetic radiation seems like a continuous wave at first, but when one performs experiments on the light, one can notice that the light can have both wave and particle like properties. Just like the individual ants, the light can also be made up of individual bundles of energy, or quanta of light.

Light has both wave-like and particle-like properties (wave–particle duality), but only shows one or the other, depending on the kind of experiment we perform. A wave-type experiment shows the wave nature, and a particle-type experiment shows particle nature. One cannot test the wave and the particle nature at the same time. A particle of light is called a photon.

Definition: Photon
A photon is a quantum (energy packet) of light.

The particle nature of light can be demonstrated by the interaction of photons with matter. One way in which light interacts with matter is via the photoelectric effect, which will be studied in detail in Chapter 31.

?

Exercise: Particle/wave nature of electromagnetic radiation

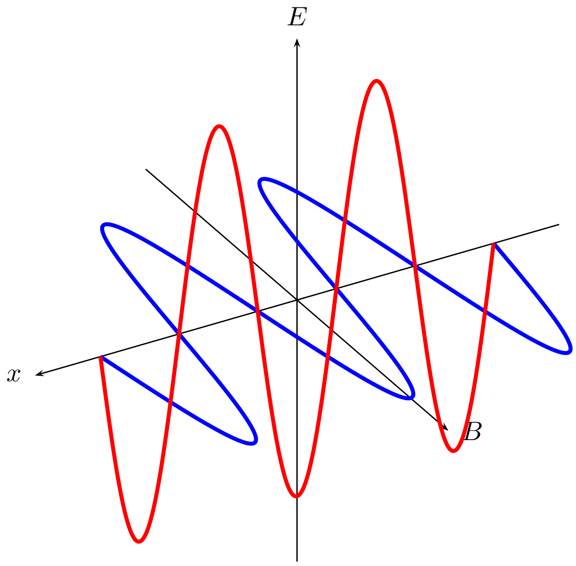
1. Give examples of the behaviour of EM radiation which can best be explained using a wave model.

2. Give examples of the behaviour of EM radiation which can best be explained using a particle model.

30.3 The wave nature of electromagnetic radiation

Accelerating charges emit electromagnetic waves. We have seen that a changing electric field generates a magnetic field and a changing magnetic field generates an electric field. This is the principle behind the propagation of electromagnetic waves, because electromagnetic waves, unlike sound waves, do not need a medium to travel through. EM waves propagate when an electric field oscillating in one plane produces a magnetic field oscillating in a plane at right angles to it, which produces an oscillating electric field, and so on. The propagation of electromagnetic waves can be described as *mutual induction*.

These mutually regenerating fields travel through space at a constant speed of $3 \times 10^8\,\mathrm{m \cdot s^{-1}}$, represented by c.



30.4 Electromagnetic spectrum

Observe the things around you, your friend sitting next to you, a large tree across the field. How is it that you are able to see these things? What is it that is leaving your friend's arm and entering your eye so that you can see his arm? It is light. The light originally comes from the sun, or possibly a light bulb or burning fire. In physics, light is given the more technical term electromagnetic radiation, which includes all forms of light, not just the form which you can see with your eyes.

Electromagnetic radiation allows us to observe the world around us. It is this radiation which reflects off of the objects around you and into your eye. The radiation your eye is sensitive to is only a small fraction of the total radiation emitted in the physical universe. All of the different fractions taped together make up the electromagnetic spectrum.

Extension: Dispersion

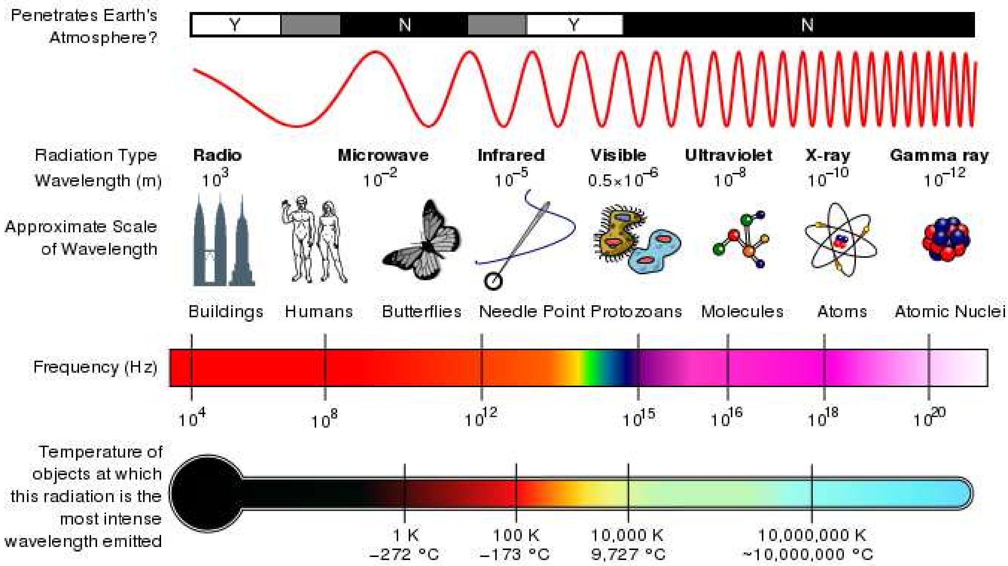


Figure 30.1: The electromagnetic spectrum as a function of frequency. The different types according to wavelength are shown as well as everyday comparisons.

When white light is split into its component colours by a prism, you are looking at a portion of the electromagnetic spectrum.

The wavelength of a particular electromagnetic radiation will depend on how it was created.

?

Exercise: Wave Nature of EM Radiation

1. List one source of electromagnetic waves. Hint: consider the spectrum diagram and look at the names we give to different wavelengths.
2. Explain how an EM wave propagates, with the aid of a diagram.
3. What is the speed of light? What symbol is used to refer to the speed of light? Does the speed of light change?
4. Do EM waves need a medium to travel through?

The radiation can take on any wavelength, which means that the spectrum is continuous. Physicists broke down this continuous band into sections. Each section is defined by how the radiation is created, not the radiations wavelength. But each category is continuous within the min and max wavelength of that category, meaning there are no wavelengths excluded within some range.

The spectrum is in order of wavelength, with the shortest wavelength at one end and the longest wavelength at the other. The spectrum is then broken down into categories as detailed in Table 30.1.

Since an electromagnetic wave is still a wave, the following equation that you learnt in Grade 10 still applies:

$$c = f \cdot \lambda$$

Table 30.1: Electromagnetic spectrum

Category	Range of Wavelengths (nm)	Range of Frequencies (Hz)
gamma rays	<1	$> 3 \times 10^{19}$
X-rays	1-10	3×10^{17}-3×10^{19}
ultraviolet light	10-400	$7{,}5 \times 10^{14}$-3×10^{17}
visible light	400-700	$4{,}3 \times 10^{14}$-$7{,}5 \times 10^{14}$
infrared	700-10^5	3×10^{12}-$4{,}3 \times 10^{19}$
microwave	$10^5 - 10^8$	3×10^{9}-3×10^{12}
radio waves	$> 10^8$	$< 3 \times 10^{9}$

Worked Example 185: EM spectrum I

Question: Calculate the frequency of red light with a wavelength of $4{,}2 \times 10^{-7}$ m
Answer
We use the formula: $c = f\lambda$ to calculate frequency. The speed of light is a constant 3×10^8m/s.

$$\begin{aligned} c &= f\lambda \\ 3 \times 10^8 &= f \times 4{,}2 \times 10^{-7} \\ f &= 7{,}14 \times 10^{14}\text{Hz} \end{aligned}$$

Worked Example 186: EM spectrum II

Question: Ultraviolet radiation has a wavelength of 200 nm. What is the frequency of the radiation?
Answer
Step 1 : To calculate the frequency we need to identify the wavelength and the velocity of the radiation.
Recall that all radiation travels at the speed of light (c) in vacuum. Since the question does not specify through what type of material the Ultraviolet radiation is traveling, one can assume that it is traveling through a vacuum. We can identify two properties of the radiation - $wavelength$ (200 nm) and speed (c). From previous chapters, we know that the period of the wave is the time it takes for a wave to complete one cycle or one wavelength.
Step 2 : Since we know the wavelength and we know the speed, lets first calculate the Period (T).

$$\begin{aligned} T &= \frac{distance}{speed} \\ &= \frac{\lambda}{c} \\ distance &= 200 \text{ nm} \\ &= 200 \times 10^{-9} \text{ m} \\ speed &= 3.0 \times 10^8 \text{ m} \cdot \text{s}^{-1} \\ T &= \frac{200 \times 10^{-9} \text{ m}}{3.0 \times 10^8 \text{ m} \cdot \text{s}^{-1}} \\ &= 6.67 \times 10^{-16} \text{ s} \end{aligned}$$

Step 3 : From the Period (T), we can calculate the frequency (f).

$$\begin{aligned} f &= \frac{1}{T} \\ &= \frac{1}{6.67 \times 10^{-16}\ \text{s}} \\ &= 1.5 \times 10^{15}\ \text{Hz} \end{aligned}$$

Examples of some uses of electromagnetic waves are shown in Table 30.2.

Table 30.2: Uses of EM waves

Category	**Uses**
gamma rays	used to kill the bacteria in marshmallows
X-rays	used to image bone structures
ultraviolet light	bees can see into the ultraviolet because flowers stand out more clearly at this frequency
visible light	used by humans to observe the world
infrared	night vision, heat sensors, laser metal cutting
microwave	microwave ovens, radar
radio waves	radio, television broadcasts

In theory the spectrum is infinite, although realistically we can only observe wavelengths from a few hundred kilometers to those of gamma rays due to experimental limitations.

Humans experience electromagnetic waves differently depending on their wavelength. Our eyes are sensitive to visible light while our skin is sensitive to infrared, and many wavelengths we do not detect at all.

?

Exercise: EM Radiation

1. Arrange the following types of EM radiation in order of increasing frequency: infrared, X-rays, ultraviolet, visible, gamma.
2. Calculate the frequency of an EM wave with a wavelength of 400 nm.
3. Give an example of the use of each type of EM radiation, i.e. gamma rays, X-rays, ultraviolet light, visible light, infrared, microwave and radio and TV waves.

30.5 The particle nature of electromagnetic radiation

When we talk of electromagnetic radiation as a particle, we refer to photons, which are packets of energy. The energy of the photon is related to the wavelength of electromagnetic radiation according to: h is called Planck's constant.

Definition: Planck's constant
Planck's constant is a physical constant named after Max Planck.

$$h = 6{,}626 \times 10^{-34} \text{ J} \cdot \text{s}$$

The energy of a photon can be calculated using the formula: $E = hf$ or $E = h\frac{c}{\lambda}$. Where E is the energy of the photon in joules (J), h is planck's constant, c is the speed of light, f is the frequency in hertz (Hz) and λ is the wavelength in metres (m).

Worked Example 187: Calculating the energy of a photon I

Question: Calculate the energy of a photon with a frequency of 3×10^{18} Hz
Answer
We use the formula: $E = hf$

$$\begin{aligned} E &= hf \\ &= 6{,}6 \times 10^{-34} \times 3 \times 10^{18} \\ &= 2 \times 10^{-15}\,\text{J} \end{aligned}$$

Worked Example 188: Calculating the energy of a photon II

Question: What is the energy of an ultraviolet photon with a wavelength of 200 nm?
Answer
Step 1 : Determine what is required and how to approach the problem.
We are required to calculate the energy associated with a photon of ultraviolet light with a wavelength of 200 nm.
We can use:

$$E = h\frac{c}{\lambda}$$

Step 2 : Solve the problem

$$\begin{aligned} E &= h\frac{c}{\lambda} \\ &= (6{,}626 \times 10^{-34})\frac{3 \times 10^{8}}{200 \times 10^{-9}} \\ &= 9{,}939 \times 10^{-10}\,\text{J} \end{aligned}$$

30.5.1 Exercise - particle nature of EM waves

1. How is the energy of a photon related to its frequency and wavelength?
2. Calculate the energy of a photon of EM radiation with a frequency of 10^{12} Hz.
3. Determine the energy of a photon of EM radiation with a wavelength of 600 nm.

30.6 Penetrating ability of electromagnetic radiation

Different kinds of electromagnetic radiation have different penetrabilities. For example, if we take the human body as the object. Infrared light is emitted by the human body. Visible light is reflected off the surface of the human body, ultra-violet light (from sunlight) damages the skin, but X-rays are able to penetrate the skin and bone and allows for pictures of the inside of the human body to be taken.

If we compare the energy of visible light to the energy of X-rays, we find that X-rays have a much higher energy. Usually, kinds of electromagnetic radiation with higher energy have higher penetrabilities than those with low energies.

Certain kinds of electromagnetic radiation such as ultra-violet radiation, X-rays and gamma rays are very dangerous. Radiation such as these are called ionising radiation. Ionising radiation loses energy as it passes through matter, breaking molecular bonds and creating ions.

Excessive exposure to radiation, including sunlight, X-rays and all nuclear radiations, can cause destruction of biological tissue.

30.6.1 Ultraviolet(UV) radiation and the skin

UVA and UVB are different ranges of frequencies for ultraviolet (UV) light. UVA and UVB can damage collagen fibres which results in the speeding up skin aging. In general, UVA is the least harmful, but it can contribute to the aging of skin, DNA damage and possibly skin cancer. It penetrates deeply and does not cause sunburn. Because it does not cause reddening of the skin (erythema) it cannot be measured in the SPF testing. There is no good clinical measurement of the blocking of UVA radiation, but it is important that sunscreen block both UVA and UVB.

UVB light can cause skin cancer. The radiation excites DNA molecules in skin cells, resulting in possible mutations, which can cause cancer. This cancer connection is one reason for concern about ozone depletion and the ozone hole.

As a defense against UV radiation, the body tans when exposed to moderate (depending on skin type) levels of radiation by releasing the brown pigment melanin. This helps to block UV penetration and prevent damage to the vulnerable skin tissues deeper down. Suntan lotion, often referred to as sunblock or sunscreen, partly blocks UV and is widely available. Most of these products contain an SPF rating that describes the amount of protection given. This protection, however, applies only to UVB rays responsible for sunburn and not to UVA rays that penetrate more deeply into the skin and may also be responsible for causing cancer and wrinkles. Some sunscreen lotion now includes compounds such as titanium dioxide which helps protect against UVA rays. Other UVA blocking compounds found in sunscreen include zinc oxide and avobenzone.

Extension: What makes a good sunscreen?

- UVB protection: Padimate O, Homosalate, Octisalate (octyl salicylate), Octinoxate (octyl methoxycinnamate)
- UVA protection: Avobenzone
- UVA/UVB protection: Octocrylene, titanium dioxide, zinc oxide, Mexoryl (ecamsule)

Another means to block UV is by wearing sun protective clothing. This is clothing that has a UPF rating that describes the protection given against both UVA and UVB.

30.6.2 Ultraviolet radiation and the eyes

High intensities of UVB light are hazardous to the eyes, and exposure can cause welder's flash (photo keratitis or arc eye) and may lead to cataracts, pterygium and pinguecula formation.

Protective eyewear is beneficial to those who are working with or those who might be exposed to ultraviolet radiation, particularly short wave UV. Given that light may reach the eye from the sides, full coverage eye protection is usually warranted if there is an increased risk of exposure, as in high altitude mountaineering. Mountaineers are exposed to higher than ordinary levels of UV radiation, both because there is less atmospheric filtering and because of reflection from snow and ice.

Ordinary, untreated eyeglasses give some protection. Most plastic lenses give more protection than glass lenses. Some plastic lens materials, such as polycarbonate, block most UV. There are protective treatments available for eyeglass lenses that need it which will give better protection. But even a treatment that completely blocks UV will not protect the eye from light that arrives around the lens. To convince yourself of the potential dangers of stray UV light, cover your lenses with something opaque, like aluminum foil, stand next to a bright light, and consider how much light you see, despite the complete blockage of the lenses. Most contact lenses help to protect the retina by absorbing UV radiation.

30.6.3 X-rays

While x-rays are used significantly in medicine, prolonged exposure to X-rays can lead to cell damage and cancer.

For example, a mammogram is an x-ray of the human breast to detect breast cancer, but if a woman starts having regular mammograms when she is too young, her chances of getting breast cancer increases.

30.6.4 Gamma-rays

Due to the high energy of gamma-rays, they are able to cause serious damage when absorbed by living cells.

Gamma-rays are not stopped by the skin and can induce DNA alteration by interfering with the genetic material of the cell. DNA double-strand breaks are generally accepted to be the most biologically significant lesion by which ionising radiation causes cancer and hereditary disease.

A study done on Russian nuclear workers exposed to external whole-body gamma-radiation at high cumulative doses shows a link between radiation exposure and death from leukaemia, lung, liver, skeletal and other solid cancers.

Extension: Cellphones and electromagnetic radiation

Cellphone radiation and health concerns have been raised, especially following the enormous increase in the use of wireless mobile telephony throughout the world. This is because mobile phones use electromagnetic waves in the microwave range. These concerns have induced a large body of research. Concerns about effects on health have also been raised regarding other digital wireless systems, such as data communication networks.

The World Health Organization has officially ruled out adverse health effects from cellular base stations and wireless data networks, and expects to make recommendations about mobile phones in 2007-08.

Cellphone users are recommended to minimise radiation, by for example:

1. Use hands-free to decrease the radiation to the head.
2. Keep the mobile phone away from the body.
3. Do not telephone in a car without an external antenna.

30.6.5 Exercise - Penetrating ability of EM radiation

1. Indicate the penetrating ability of the different kinds of EM radiation and relate it to energy of the radiation.

2. Describe the dangers of gamma rays, X-rays and the damaging effect of ultra-violet radiation on skin

30.7 Summary

1. Electromagnetic radiation has both a wave and particle nature.

2. Electromagnetic waves travel at a speed of $3 \times 10^8\ m \cdot s^{-1}$ in a vaccum.

3. The Electromagnetic consists of the follwing types of radiation: radio waves, microwaves, infrared, visible, ultraviolet, X-rays, gamma-rays.

4. Gamma-rays have the most energy and are the most penetrating, while radio waves have the lowest energy and are the least penetrating.

30.8 End of chapter exercise

1. What is the energy of a photon of EM radiation with a frequency of 3×10^8 Hz?

2. What is the energy of a photon of light with a wavelength of 660 nm?

3. List the main types of electromagnetic radiation in order of increasing wavelength.

4. List the main uses of:

   A radio waves

   B infrared

   C gamma rays

   D X-rays

5. Explain why we need to protect ourselves from ultraviolet radiation from the Sun.

6. List some advantages and disadvantages of using X-rays.

7. What precautions should we take when using cell phones?

8. Write a short essay on a type of electromagnetic waves. You should look at uses, advantages and disadvantages of your chosen radiation.

9. Explain why some types of electromagnetic radiation are more penetrating than others.

Chapter 31

Optical Phenomena and Properties of Matter - Grade 12

31.1 Introduction

For centuries physicists argued about whether light was a particle or a wave. It was assumed that light could only be one or the other, but not both.

In earlier chapters on waves (Chapters 6, 24, 25, 26) and optics (Chapters 7 and 13), you studied how light or other electromagnetic radiation propagates like a wave. The wave nature of light of was demonstrated by the propagation of light in examples such as diffraction, interference, and polarisation of light.

You also saw in Chapter 30 on electromagnetic radiation how light sometimes behaves as a particle. This chapter looks at evidence supporting the *particle model of light*. The idea that light can have both wave and particle properties was one of the most important discoveries of the twentieth century.

31.2 The transmission and scattering of light

31.2.1 Energy levels of an electron

We have seen that the electrons in an atom have different energy levels. When the electron receives enough energy, it can jump up to a higher energy level. This is called *'exciting'* the electron. When the electron in a high energy level sheds some energy, it drops to a lower energy level.

We have also seen that the energy associated with light at a specific wavelength is given by:

$$E = \frac{hc}{\lambda}.$$

In the particle model of light, this means that each packet of light (photon) at a wavelength λ has energy:

$$E = \frac{hc}{\lambda}.$$

For the electron to receive energy, it *absorbs* a photon and gets its energy. When an electron loses energy to drop to a lower level, it *emits* (gives off) a photon with that energy.

31.2.2 Interaction of light with metals

When light encounters or passes through a material, the photons of the light interact with the atoms or molecules of the material. Depending on the strength of the interactions and how often they happen, the light will pass through the material or be scattered in some other direction.

Each wavelength of light relates to a particular energy, and the closer that energy is to the energy difference between two of the levels of the atom, the likelier the photon is to interact with the atom.

When visible or ultraviolet (UV) radiation shines on a metal, the photons are *absorbed* by the electrons in the metal. The electrons are then excited up to a higher energy level. When an electron returns to a lower energy level, another photon is emitted. This is how light is **reflected** off a metal surface.

In previous chapters, you have studied geometrical optics, which tells us what happens to rays of light when they are reflected off a surface or refracted through a lens. That tells us what happens to light rays, made up of many photons, on a large scale. If you look at a smaller level, i.e. on a microscopic scale, then reflection and refraction happen by all the photons interacting with the atoms of the lens or mirror. The photons get absorbed and re-emitted many times before emerging as the finals rays of light that we see.

Scattering of light is responsible for many effects in everyday life. We see that certain materials are red or blue, for example, since they contain materials that have energy level differences that correspond to the energies of the photons that make up red or blue light. These materials then reflect the red or blue light and absorb the other wavelengths in the visible spectrum. White objects reflect photons of all wavelengths in the visual spectrum, while black objects absorb these photons.

Interesting Fact

Because a truly black object absorbs all the visual wavelengths of light, and does not re-emit photons at visual wavelengths, we can say that 'black' is not a colour itself, but rather a *lack of* colour! Also, since black objects absorb visual light, they heat up more than objects of other colours which reflect light at certain wavelengths.

Activity :: Investigation : Reflection and absorption

Aim:
Investigate the interaction of light with differently coloured metal objects

Apparatus:
Find some differently coloured metal objects (at least 5) which will not be damaged if left in the sun for 15 minutes. Make sure to include at least one white item and one black item.

Method:
At the start of your lesson set out your objects in direct sunlight. Leave them there for around 15 minutes.

Alternatively, if it is a sunny day, you can use your teachers' cars for this experiment - as long as there are some cars of different colours and they have been standing in direct sunlight for the same length of time!

After 15 minutes is up, touch each of the items/cars (be careful not to burn yourself!) and compare their temperatures (rate them from 1 to 5 with 1 being cold and 5 being very hot) in a table such as the example table below:

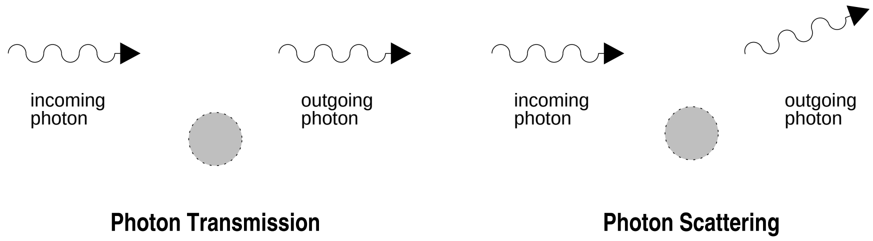


Figure 31.1: Diagrams of photon transmission (left) and scattering (right).

Object	**Colour**	**Temperature rating**
e.g. car 1	e.g. red	e.g. 3

Questions:

1. Which object was the hottest and what was it made of?
2. Which object was the coolest and what was it made of?
3. How did the temperatures of the black and white objects compare to each other? (which one was hotter and which was cooler?)
4. Try to explain the reasons for the different temperatures of the objects with respect to their colours and the materials of which they are made.

Metals generally reflect most wavelengths of visible light, but they will reflect the light in a certain direction, given by the laws of reflection in geometrical optics. This is different to most materials, like wood or fabric, which reflect light in all directions. Metals have this property since they have electrons that are not bound to atoms and can move freely through the metal. This is unlike most other materials that have their electrons bound closely to the atoms. These free electrons in metals can then absorb and reflect photons of a wide range of energies.

Ultraviolet light (which has shorter wavelengths than visible light) will pass through some substances, such as many plastics, because they do not have the right energy levels to absorb it and re-emit it. *X-rays* (also short wavelengths) will also pass through most materials, since the energies of X-rays correspond to the energy levels of atomic nuclei. Such nuclei are much smaller than atoms, so it is much less likely for an X-ray to hit a nucleus instead of the whole atom.

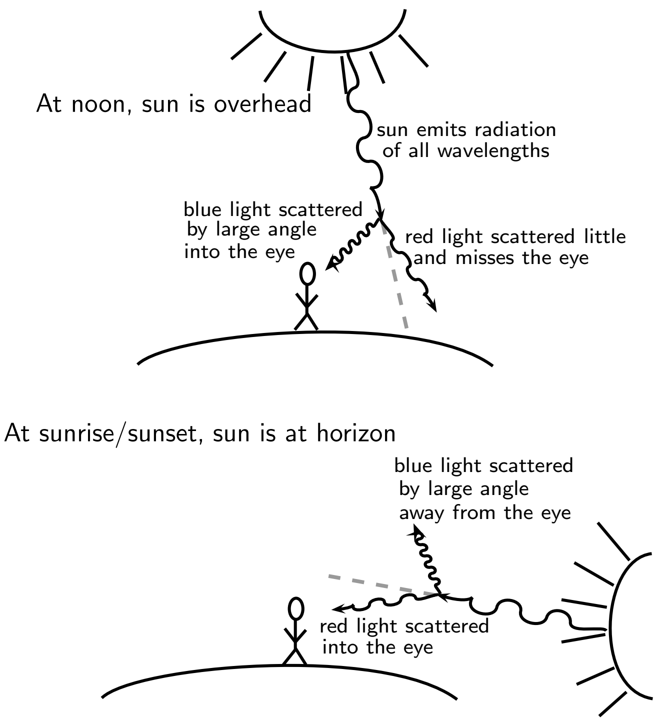
Most materials will absorb *infrared radiation* (longer wavelengths than visible light), since the energies of that radiation often correspond to rotational or vibrational energy levels of molecules.

31.2.3 Why is the sky blue?

The sun emits light in many different wavelengths, including all of the visible wavelengths. Light which is made up of all the visible wavelengths appears white. So what causes the sky to look blue?

The air is not just full of nitrogen and oxygen gases. It is also full of tiny dust grains. The light from the sun scatters off these many dust grains.

The chance that the light will scatter off one of these dust grains is *bigger* for *shorter* wavelengths. The *short wavelength* **blue light** is therefore scattered much more than the other colors. At noon, when the light from the sun is coming straight down (see the picture below), the scattered blue light reaches your eyes from all directions and so the sky appears blue. The other wavelengths do not get scattered much and therefore miss your line of sight. At sunrise or sunset, the direction of the light coming from the sun is now straight towards your eyes (see picture below). Therefore the scattered blue light can't be seen because it is scattered out of your line of sight. The redder colours (oranges and reds) can now be seen because they are not scattered as much and still fall in your line of sight.



?

Exercise: Transmission and scattering of light

1. Explain how visible light is reflected from metals.
2. Explain why the sky is blue.

31.3 The photoelectric effect

Around the turn of the twentieth century, it was observed by a number of physicists (including Hertz, Thomson and Von Lenard) that when light was shone on a metal, electrons were emitted by the metal. This is called the photoelectric effect. (*photo*- for light, *electric*- for the electron.)

Definition: The photoelectric effect
The photoelectric effect is the process whereby an electron is emitted by a metal when light shines on it.

At that time, light was thought to be purely a wave. Therefore, physicists thought that if a more intense (i.e. brighter) light was shone on a metal, then the electrons would be knocked out with greater kinetic energies than if a faint light was shone on them. However, Von Lenard observed that this did not happen at all. The *intensity* of the light made *no difference* to the kinetic energy of the emitted electrons! Also, it was observed that the electrons were emitted immediately when light was shone on the metal - there was no time delay.

Einstein solved this problem by proposing that light is made up of packets of energy called *quanta* (now called photons) which interacted with the electrons in the metal like particles instead of waves. Each incident photon would transfer all its energy to one electron in the metal. For a specific colour of light (i.e. a certain wavelength or frequency), the energy of the photons is given by $E = hf = hc/\lambda$, where h is Planck's constant. The energy needed to knock an electron out of the metal is called the **work function** (symbol ϕ) of the metal. Therefore, the amount of energy left over as the kinetic energy (E_k) of the emitted electron would be the difference between the incoming photon's energy and the energy needed to knock out the electron (work function of the metal):

$$E_k = hf - \phi$$

Increasing the intensity of the light (i.e. making it brighter) did not change the wavelength of the light and therefore the electrons would be emitted with *the same* kinetic energy as before! This solved the paradox and showed that light has *both* a **wave nature** and a **particle nature**. Einstein won the Nobel prize for this quantum theory and his explanation of the photoelectric effect.

Increasing the intensity of the light actually means increasing the *number* of incident photons. Therefore, since each photon only gives energy to one electron, more incident photons means*more* electrons would be knocked out of the metal, but their kinetic energies would be *the same* as before.

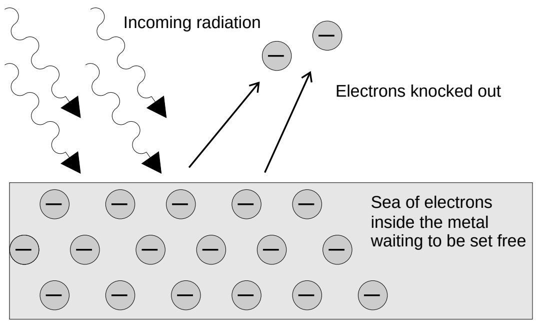


Figure 31.2: The photoelectric effect: Incoming photons on the left hit the electrons inside the metal surface. The electrons absorb the energy from the photons, and are ejected from the metal surface.

Interesting Fact

The photoelectric effect was first observed in the experiments of Heinrich Hertz in 1887. In 1899 J.J. Thomson proved that it was electrons that were emitted. The photoelectric effect was theoretically explained by Albert Einstein in 1905.

The discovery and understanding of the photoelectric effect was one of the major breakthroughs in science in the twentieth century as it provided concrete evidence of the particle nature of light. It overturned previously held views that light was composed purely of a continuous transverse wave. On the one hand, the wave nature is a good description of phenomena such as diffraction and interference for light, and on the other hand, the photoelectric effect demonstrates the particle nature of light. This is now known as the 'dual-nature' of light. (dual means two)

While solving problems we need to decide for ourselves whether we should consider the wave property or the particle property of light. For example, when dealing with interference and diffraction, light should be treated as a wave, whereas when dealing with photoelectric effect we consider the particle nature.

31.3.1 Applications of the photoelectric effect

We have learnt that a metal contains electrons that are free to move between the valence and conduction bands. When a photon strikes the surface of a metal, it gives *all* its energy to one electron in the metal.

- If the photon energy is *equal* to the energy between two energy levels then the electron is excited to the higher energy level.
- If the photon energy is *greater* than or *equal to* the work function (energy needed to escape from the metal), then the electron is emitted from the surface of the metal (the photoelectric effect).

The work function is different for different elements. The smaller the work function, the easier it is for electrons to be emitted from the metal. Metals with low work functions make good conductors. This is because the electrons are attached less strongly to their surroundings and can move more easily through these materials. This reduces the resistance of the material to the flow of current i.e. it conducts well. Table 31.1 shows the work functions for a range of elements.

Element	Work Function (J)
Aluminium	$6,9 \times 10^{-19}$
Beryllium	$8,0 \times 10^{-19}$
Calcium	$4,6 \times 10^{-19}$
Copper	$7,5 \times 10^{-19}$
Gold	$8,2 \times 10^{-19}$
Lead	$6,9 \times 10^{-19}$
Silicon	$1,8 \times 10^{-19}$
Silver	$6,9 \times 10^{-19}$
Sodium	$3,7 \times 10^{-19}$

Table 31.1: Work functions of selected elements determined from the photoelectric effect. (From the *Handbook of Chemistry and Physics.)*

Interesting Fact

The **electron volt** is the kinetic energy gained by an electron passing through a potential difference of one volt (1 V). A **volt** is not a measure of energy, but the **electron volt** is a unit of energy. When you connect a 1.5 Volt battery to a circuit, you can give 1.5 eV of energy to every electron.

Worked Example 189: The photoelectric effect - I

Question: Ultraviolet radiation with a wavelength of 250 nm is incident on a silver foil (work function $\phi = 6{,}9 \times 10^{-19}$). What is the maximum kinetic energy of the emitted electrons?

Answer

Step 1 : Determine what is required and how to approach the problem

We need to determine the maximum kinetic energy of an electron ejected from a silver foil by ultraviolet radiation.

The photoelectric effect tells us that:

$$\begin{aligned} E_k &= E_{photon} - \phi \\ E_k &= h\frac{c}{\lambda} - \phi \end{aligned}$$

We also have:

Work function of silver: $\phi = 6{,}9 \times 10^{-19}$ J

UV radiation wavelength = 250 nm = 250×10^{-9} m

Planck's constant: $h = 6{,}63 \times 10^{-34}\ \mathrm{m^2kgs^{-1}}$

speed of light: $c = 3 \times 10^8\ \mathrm{ms^{-1}}$

Step 2 : Solve the problem

$$\begin{aligned} E_k &= \frac{hc}{\lambda} - \phi \\ &= [6{,}63 \times 10^{-34} \times \frac{3 \times 10^8}{250 \times 10^{-9}}] - 6{,}9 \times 10^{-19} \\ &= 1{,}06 \times 10^{-19}\ \mathrm{J} \end{aligned}$$

The maximum kinetic energy of the emitted electron will be $1{,}06 \times 10^{-19}$ J.

Worked Example 190: The photoelectric effect - II

Question: If we were to shine the same ultraviolet radiation ($f = 1{,}2 \times 10^{15}$ Hz), on a gold foil (work function = $8{,}2 \times 10^{-19}$ J), would any electrons be emitted from the surface of the gold foil?

Answer

For the electrons to be emitted from the surface, the energy of each photon needs to be *greater* than the work function of the material.

Step 1 : Calculate the energy of the incident photons

$$\begin{aligned} E_{photon} &= hf \\ &= 6{,}63 \times 10^{-34} \times 1{,}2 \times 10^{15} \\ &= 7{,}96 \times 10^{-19}\ \mathrm{J} \end{aligned}$$

Therefore each photon of ultraviolet light has an energy of $7{,}96 \times 10^{-19}$ J.

Step 2 : Write down the work function for gold.

$$\phi_{gold} = 8{,}2 \times 10^{-19}\ \mathrm{J}$$

Step 3 : Is the energy of the photons greater or smaller than the work function?

$$\begin{aligned} 7{,}96 \times 10^{-19} \text{ J} &< 8{,}2 \times 10^{-19} \text{ J} \\ E_{photons} &< \phi_{gold} \end{aligned}$$

The energy of each photon is less than the work function of gold. Therefore, the photons do not have enough energy to knock electrons out of the gold. No electrons would be emitted from the gold foil.

Extension: Units of energy

When dealing with calculations at a small scale (like at the level of electrons) it is more convenient to use different units for energy rather than the Joule (J). We define a unit called the electron-Volt (eV) as the kinetic energy gained by an electron passing through a potential difference of one volt.

$$E = q \times V$$

where q is the charge of the electron and V is the potential difference applied. The charge of 1 electron is $1{,}6 \times 10^{-19}$ C, so 1 eV is calculated to be:

$$1 \text{ eV} = (1{,}610^{-19} \text{ C} \times 1 \text{ V}) = 1{,}6 \times 10^{-19} \text{ J}$$

You can see that $1{,}6 \times 10^{-19}$ J is a very small amount of energy and so using electron-Volts (eV) at this level is easier.
Hence, 1eV $= 1.6 \times 10^{-}19$ J which means that 1 J $= 6.241 \times 10^{18}$ eV

31.3.2 Real-life applications

Solar Cells

The photo-electric effect may seem like a very easy way to produce electricity from the sun. This is why people choose to make solar panels out of materials like silicon, to generate electricity. In real-life however, the amount of electricity generated is less than expected. This is because not every photon knocks out an electron. Other processes such as reflection or scattering also happen. This means that only a fraction $\approx 10\%$ (depends on the material) of the photons produce photoelectrons. This drop in efficiency results in a lower measured current. Much work is being done in industry to improve this efficiency so that the panels can generate as high a current as possible, and create as much electricity as possible form the sun. But even these smaller electrical currents are useful in applications like solar-powered calculators.

Exercise: The photoelectric effect

1. Describe the photoelectric effect.
2. List two reasons why the observation of the photoelectric effect was significant.
3. Refer to Table 31.1: If I shine ultraviolet light with a wavelength of 288 nm onto some aluminium foil, what would be the kinetic energy of the emitted electrons?

4. I shine a light of an unknown wavelength onto some silver foil. The light has only enough energy to eject electrons from the silver foil but not enough to give them kinetic energy. (Refer to Table 31.1 when answering the questions below:)
   A If I shine the same light onto some copper foil, would electrons be ejected?
   B If I shine the same light onto some silicon, would electrons be ejected?
   C If I increase the intensity of the light shining on the silver foil, what happens?
   D If I increase the frequency of the light shining on the silver foil, what happens?

31.4 Emission and absorption spectra

31.4.1 Emission Spectra

You have learnt previously about the structure of an atom. The electrons surrounding the atomic nucleus are arranged in a series of levels of increasing energy. Each element has its own distinct set of energy levels. This arrangement of energy levels serves as the atom's unique fingerprint.

In the early 1900s, scientists found that a liquid or solid heated to high temperatures would give off a broad range of colours of light. However, a gas heated to similar temperatures would emit light only at certain specific colours (wavelengths). The reason for this observation was not understood at the time.
Scientists studied this effect using a discharge tube.

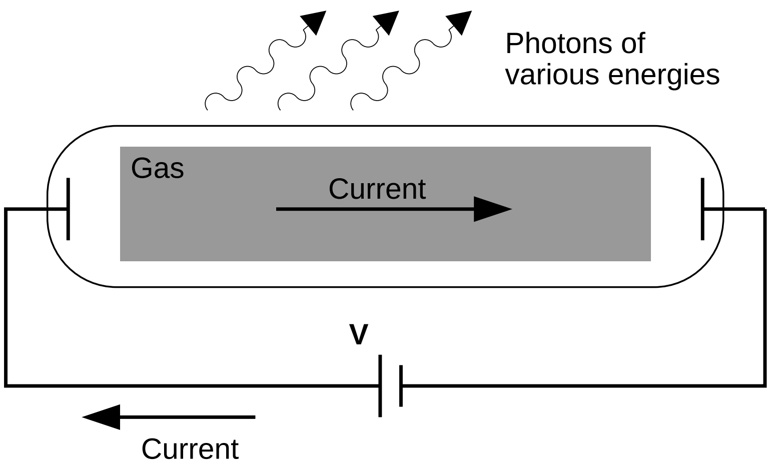


Figure 31.3: Diagram of a discharge tube. The tube is filled with a gas. When a high enough voltage is applied to both ends of the tube, the gas ionises and acts like a conductor, allowing a current to flow through the circuit. The current excites the atoms of the ionised gas. When the atoms fall back to their ground state, they emit photons to carry off the excess energy.

A discharge tube (Figure 31.3) is a glass gas-filled tube with a metal plate at both ends. If a large enough voltage difference is applied between the two metal plates, the gas atoms inside the tube will absorb enough energy to make some of their electrons come off i.e. the gas atoms are ionised. These electrons start moving through the gas and create a current, which raises some electrons in other atoms to higher energy levels. Then as the electrons in the atoms fall back down, they emit electromagnetic radiation. The amount of light emitted at different wavelengths, called the **emission spectrum**, is shown for a discharge tube filled with hydrogen gas in figure 31.4 below. Only certain wavelengths (i.e. colours) of light are seen as shown by the thick black lines in the picture.

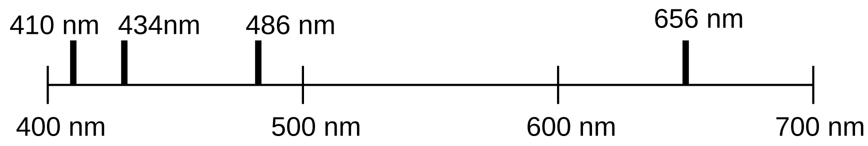


Figure 31.4: Diagram of the emission spectrum of hydrogen in the visible spectrum. Four lines are visible, and are labeled with their wavelengths. The three lines in the 400–500 nm range are in the blue part of the spectrum, while the higher line (656 nm) is in the red/orange part.

Eventually, scientists realized that these lines come from photons of a specific energy, emitted by electrons making transitions between specific energy levels of the atom. Figure **??** shows an example of this happening. When an electron in an atom falls from a higher energy level to a lower energy level, it emits a photon to carry off the extra energy. This photon's energy is equal to the energy difference between the two energy levels. As we previously discussed, the frequency of a photon is related to its energy through the equation $E = hf$. Since a specific photon frequency (or wavelength) gives us a specific colour, we can see how each coloured line is associated with a specific transition.

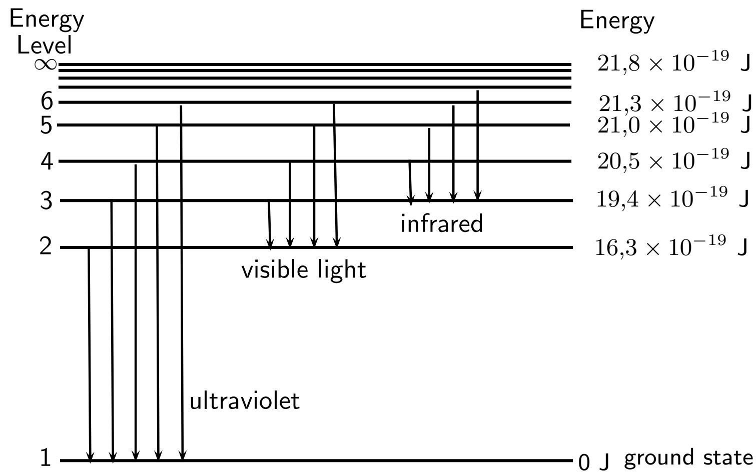


Figure 31.5: In this diagram are shown some of the electron energy levels for the hydrogen atom. The arrows show the electron transitions from higher energy levels to lower energy levels. The energies of the emitted photons are the same as the energy difference between two energy levels. You can think of absorption as the opposite process. The arrows would point upwards and the electrons would jump up to higher levels when they absorp a photon of the right energy.

Visible light is not the only kind of electromagnetic radiation emitted. More energetic or less energetic transitions can produce ultraviolet or infrared radiation. However, because each atom has its own distinct set of energy levels (its fingerprint!), each atom has its own distinct emission spectrum.

31.4.2 Absorption spectra

As you know, atoms do not only emit photons; they also absorb photons. If a photon hits an atom and the energy of the photon is the same as the gap between two electron energy levels in the atom, then the electron can absorb the photon and jump up to the higher energy level.

If the atom has no energy level differences that equal the incoming photon's energy, it cannot absorb the photon, and can only scatter it.

Using this effect, if we have a source of photons of various energies we can obtain the **absorption spectra** for different materials. To get an absorption spectrum, just shine white light on a sample of the material that you are interested in. White light is made up of all the different wavelengths of visible light put together. In the absorption spectrum, the energy levels corresponding to the absorbed photons show up as black lines because the photons of these wavelengths have been absorbed and don't show up. Because of this, the absorption spectrum is the exact *inverse* of the emission spectrum. Look at the two figures below. In figure 31.6 you can see the emission lines of hyrodrogen. Figure 31.7 shows the absorption spectrum. It is the exact opposite of the emission spectrum! Both emission and absorption techniques can be used to get the same information about the energy levels of an atom.

Figure 31.6: *Emission* spectrum of Hydrogen.

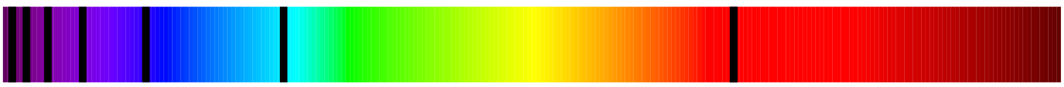


Figure 31.7: *Absorption* spectrum of Hydrogen.

Worked Example 191: Absorption

Question: I have an unknown gas in a glass container. I shine a bright white light through one side of the container and measure the spectrum of transmitted light. I notice that there is a black line (*absorption* line) in the middle of the visible red band at 642 nm. I have a hunch that the gas might be hydrogen. If I am correct, between which 2 energy levels does this transition occur? (Hint: look at figure 31.5 and the transitions which are in the visible part of the spectrum.)

Answer

Step 1 : What is given and what needs to be done?

We have an absorption line at 642 nm. This means that the substance in the glass container absorbed photons with a wavelength of 642 nm. We need to calculate which 2 energy levels of hydrogen this transition would correspond to. Therefore we need to know what energy the absorbed photons had.

Step 2 : Calculate the energy of the absorbed photons

$$\begin{aligned} E &= \frac{hc}{\lambda} \\ &= \frac{(6{,}63 \times 10^{-34}) \times (3 \times 10^{8})}{642 \times 10^{-9}} \\ &= 3{,}1 \times 10^{-19} \text{ J} \end{aligned}$$

The absorbed photons had energy of $3{,}1 \times 10^{-19}$

Step 3 : Find the energy of the transitions resulting in radiation at visible wavelengths

Figure 31.5 shows various energy level transitions. The transitions related to visible wavelengths are marked as the transitions beginning or ending on Energy Level 2. Let's find the energy of those transitions and compare with the energy of the absorbed photons we've just calculated.

Energy of transition (absorption) from Energy Level 2 to Energy Level 3:

$$\begin{aligned} E_{2_3} &= E_2 - E_3 \\ &= 16{,}3 \times 10^{-19} \text{ J} - 19{,}4 \times 10^{-19} \text{ J} \\ &= -3{,}1 \times 10^{-19} \text{ J} \end{aligned}$$

Therefore the energy of the photon that an electron must absorb to jump from Energy Level 2 to Energy Level 3 is $3{,}1 \times 10^{-19}$ J. (NOTE: The minus sign means that *absorption* is occurring.)
This is the same energy as the photons which were absorbed by the gas in the container! Therefore, since the transitions of all elements are unique, we can say that the gas in the container is hydrogen. The transition is absorption of a photon between Energy Level 2 and Energy Level 3.

31.4.3 Colours and energies of electromagnetic radiation

We saw in the explanation of why the sky is blue that different *wavelengths* or *frequencies* of light correspond to different *colours* of light. The table below gives the wavelengths and colours for light in the visible spectrum:

Colour	Wavelength range (nm)
violet	390 - 455
blue	455 - 492
green	492 - 577
yellow	577 - 597
orange	597 - 622
red	622 - 780

Table 31.2: Colours and wavelengths of light in the visible spectrum.

We also know that the energy of a photon of light can be found from:

$$E = hf = \frac{hc}{\lambda}$$

Therefore if we know the frequency or wavelength of light, we can calculate the photon's energy and vice versa.

Activity :: Investigation : Frequency,wavelength and energy relation
Refer to table 31.2: Copy the table into your workbook and add two additional columns.

1. In the first new column write down the lower and upper frequencies for each colour of light.
2. In the second column write down the energy range (in Joules) for each colour of light.

Questions

1. Which colour of visible light has the highest energy photons?
2. Which colour of visible light has the lowest energy photons?

Worked Example 192: Colours of light

Question: A photon of wavelength 500 nm is emitted by a traffic light.

1. What is the energy of the photon?
2. What is the frequency of the photon?
3. Use table 31.2 to determine the colour of the light.

Answer

Step 1 : What information is given and what do we need to do find?

We are given $\lambda = 500 \times 10^{-9}$ m and we need to find the photon's *energy*, *frequency* and *colour*.

Step 2 : Use the equation $E = \frac{hc}{\lambda}$ to find the photon's energy

$$\begin{aligned} E &= \frac{hc}{\lambda} \\ &= \frac{(6{,}63 \times 10^{-34}) \times (3 \times 10^{8})}{500 \times 10^{9}} \\ &= 3{,}98 \times 10^{-19} \text{ J} \end{aligned}$$

The energy of the photon is $3{,}98 \times 10^{-19}$ J.

Step 3 : We know the energy of the photon, now we can use $E = hf$ to solve for the frequency

$$\begin{aligned} E &= hf \\ f &= \frac{E}{h} \\ &= \frac{3.98 \times 10^{-19}}{6{,}63 \times 10^{-34}} \\ &= 6 \times 10^{14} \text{ Hz} \end{aligned}$$

The frequency of the photon is 6 $\times 10^{14}$ Hz

Step 4 : Use the table to find the colour of light

The wavelength given in the question is 500 nm. We can see in the table that green light has wavelengths between 492 - 577 nm. Therefore 500 nm is in this range and the colour of the light is **green**.

Worked Example 193: Colours and energies of light

Question: I have some sources which emit light of the following wavelengths:

1. 400 nm,
2. 580 nm,
3. 650 nm,
4. 300 nm.

What are the colours of light emitted by the sources (see table 31.2) and which source emits photons with the highest energy and which with the lowest energy?

Answer

Step 1 : What information is given, and what do we need to do?

4 wavelengths of light are given and we need to find their *colours*.

We also need to find which colour light photons have the highest energy and which one has the lowest energy.

Step 2 : To find the colours of light, we can compare the wavelengths to those given in table 31.2

1. 400 nm falls into the range for *violet* light (390 - 455 nm).
2. 580 nm falls into the range for *yellow* light (577 - 597 nm).
3. 650 nm falls into the range for *red* light (622 - 780 nm).
4. 300 nm is not shown in the table. However, this wavelength is just a little shorter than the shortest wavelength in the violet range. Therefore 300 nm is *ultraviolet*.

Step 3 : To find the colour of the light whose photons have the highest and lowest energies respectively, we need to calculate the energies of all the photons

We know $E = \frac{hc}{\lambda}$

For 400 nm:

$$\begin{aligned} E &= \frac{hc}{\lambda} \\ &= \frac{(6{,}63 \times 10^{-34}) \times (3 \times 10^{8})}{400 \times 10^{-9}} \\ &= 4{,}97 \times 10^{-19} \text{ J} \end{aligned}$$

For 580 nm:

$$\begin{aligned} E &= \frac{hc}{\lambda} \\ &= \frac{(6{,}63 \times 10^{-34}) \times (3 \times 10^{8})}{580 \times 10^{-9}} \\ &= 3{,}43 \times 10^{-19} \text{ J} \end{aligned}$$

For 650 nm:

$$\begin{aligned} E &= \frac{hc}{\lambda} \\ &= \frac{(6{,}63 \times 10^{-34}) \times (3 \times 10^{8})}{650 \times 10^{-9}} \\ &= 3{,}06 \times 10^{-19} \text{ J} \end{aligned}$$

For 300 nm:

$$\begin{aligned} E &= \frac{hc}{\lambda} \\ &= \frac{(6{,}63 \times 10^{-34}) \times (3 \times 10^{8})}{300 \times 10^{-9}} \\ &= 6{,}63 \times 10^{-19} \text{ J} \end{aligned}$$

Therefore, the photons with the highest energy are the **ultraviolet** photons.
The photons with the lowest energy are from light which is **red**.

31.4.4 Applications of emission and absorption spectra

The study of spectra from stars and galaxies in astronomy is called *spectroscopy*. Spectroscopy is a tool widely used in astronomy to learn different things about astronomical objects.

Identifying elements in astronomical objects using their spectra

Measuring the spectrum of light from a star can tell astronomers what the star is made of! Since each element emits or absorbs light only at particular wavelengths, astronomers can identify what elements are in the stars from the lines in their spectra. From studying the spectra of many stars we know that there are many different types of stars which contain different elements and in different amounts.

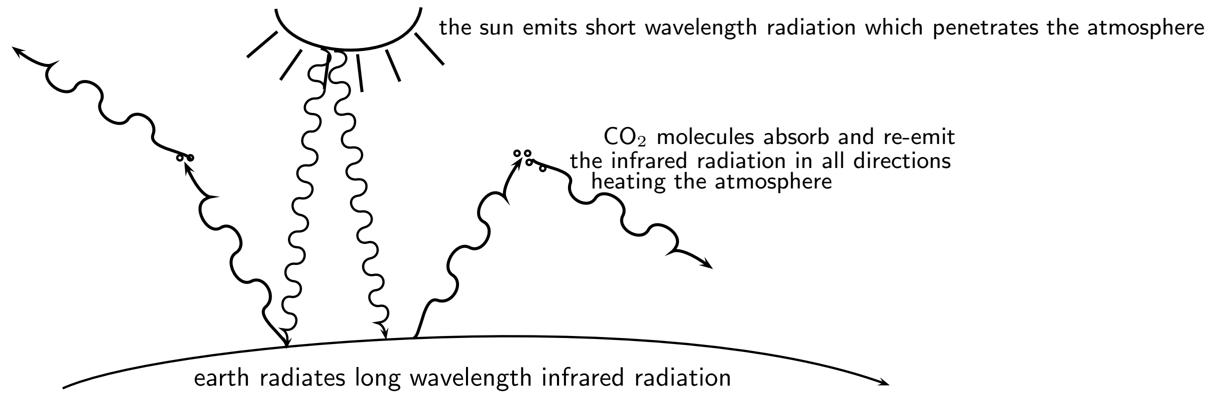
Determining velocities of galaxies using spectroscopy

You have already learned in Chapter 24 about the Doppler effect and how the frequency (and wavelength) of sound waves changes depending on whether the object emitting the sound is moving *towards* or *away* from you. The same thing happens to electromagnetic radiation. If the object emitting the light is moving *towards* us, then the wavelength of the light appears shorter (called **blue-shifted**). If the object is moving *away* from us, then the wavelength of its light appears stretched out (called **red-shifted**).

The Doppler effect affects the spectra of objects in space depending on their motion relative to us on the earth. For example, the light from a distant galaxy, which is moving away from us at some velocity, will appear red-shifted. This means that the emission and absorption lines in the galaxy's spectrum will be shifted to a longer wavelength (lower frequency). Knowing where each line in the spectrum would normally be if the galaxy was not moving, and comparing to their red-shifted positions, allows astronomers to precisely measure the velocity of the galaxy relative to the earth!

Global warming and greenhouse gases

The sun emits radiation (light) over a range of wavelengths which are mainly in the visible part of the spectrum. Radiation at these wavelengths passes through the gases of the atmosphere to warm the land and the oceans below. The warm earth then radiates this heat at longer infrared wavelengths. Carbon-dioxide (one of the main greenhouse gases) in the atmosphere has energy levels which correspond to the infrared wavelengths which allow it to absorb the infrared radiation. It then also emits at infrared wavelengths in all directions. This effect stops a large amount of the infrared radiation getting out of the atmosphere, which causes the atmosphere and the earth to heat up. More radiation is coming in than is getting back out.



Therefore increasing the amount of greenhouse gases in the atmosphere increases the amount of trapped infrared radiation and therefore the overall temperature of the earth. The earth is a very sensitive and complicated system upon which life depends and changing the delicate balances of temperature and atmospheric gas content may have disastrous consequences if we are not careful.

Activity :: Investigation : The greenhouse effect
In pairs try to find the following information (e.g. in books, on the internet) and report back to the class in a 5 minute presentation which includes the following:

1. What other gases besides carbon dioxide are responsible for the greenhouse effect?
2. Where do greenhouse gases come from? (are they human-made or natural?)
3. Investigate one serious side-effect which could arise if the earth's temperature were to go up significantly. Present some ways in which this effect could be avoided.

?

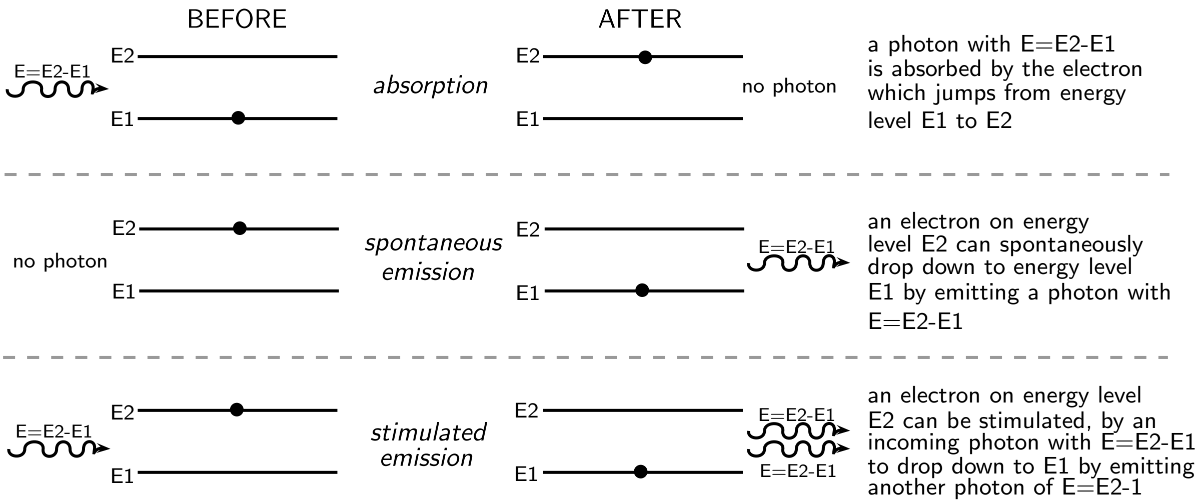
Exercise: Emission and absorption spectra

1. Explain how atomic emission spectra arise and how they relate to each element on the periodic table.
2. How do the lines on the atomic spectrum relate to electron transitions between energy levels?
3. Explain the difference between of atomic absorption and emission spectra.
4. Describe how the absorption and emission spectra of the gases in the atmosphere give rise to the Greenhouse Effect.
5. Using table 31.2 calculate the frequency range for yellow light.
6. What colour is the light emitted by hydrogen when an electron makes the transition from energy level 5 down to energy level 2? (Use figure 31.5 to find the energy of the released photon.)
7. I have a glass tube filled with hydrogen gas. I shine white light onto the tube. The spectrum I then measure has an absorption line at a wavelength of 474 nm. Between which two energy levels did the transition occur? (Use figure 31.5 in solving the problem.)

31.5 Lasers

A laser is a device that produces a special type of light: all the laser photons are identical! They all have the *same* wavelength (and frequency), amplitude and phase. Since they all have the same wavelength, this means they all have the same *colour* and the light is called **monochromatic**. (*Note: mono* means "one" or "single" and *chromatic* means "colour".) This is very different to most other light sources which produce light with a range of wavelengths (e.g. white light from the sun consists of all the visible wavelengths.)

Laser light is highly directional and can be focused very well. This focus allows laser beams to be used over long distances, and to pack a lot of energy into the beam while still requiring reasonably small amounts of energy to be generated. Each centimetre of a typical laser beam contains many billions of photons. These special properties of laser light come from the way in which the laser photons are created and the energy levels of the material that makes up the laser. These properties make laser light extremely useful in many applications from CD players to eye surgery.
The term **LASER** stands for **L**ight **A**mplification by the **S**timulated **E**mission of **R**adiation. This **stimulated emission** is different to the **spontaneous emission** already discussed earlier. Let's review the absorption and emission processes which can occur in atoms.



- **Absorption**: As you can see in the picture above, **absorption** happens when an electron jumps up to a higher energy level by *absorbing* a photon which has an energy equal to the energy difference between the two energy levels.
- **Spontaneous emission**: **Spontaneous emission** is when an electron in a higher energy level drops down to a lower energy level and a photon is emitted with an energy equal to the energy difference between the two levels. There is no interference in this process from outside factors. Usually spontaneous emission happens very quickly after an electron gets into an excited state. In other words, the lifetime of the excited state is very short. However, there are some excited states where an electron can remain in the higher energy level for a longer time than usual before dropping down to a lower level. These excited states are called **metastable states.**
- **Stimulated emission**: As the picture above shows, **stimulated emission** happens when a photon with an energy equal to the energy difference between two levels interacts with an electron in the higher level. This *stimulates* the electron to emit an identical photon and drop down to the lower energy level. This process results in two photons at the end.

Definition: Spontaneous Emission
Spontaneous emission occurs when an atom is in an unstable excited state and randomly decays to a less energetic state, emitting a photon to carry off the excess energy. The unstable state decays in a characteristic time, called the lifetime.

Definition: Meta-stable state
A meta-stable state is an excited atomic state that has an unusually long lifetime, compared to the lifetimes of other excited states of that atom. While most excited states have lifetimes measured in microseconds and nanoseconds (10^{-6} s and 10^{-9} s), meta-stable states can have lifetimes of milliseconds (10^{-3} s) or even seconds.

Definition: Stimulated emission
Stimulated emission occurs when a photon interacts with an excited atom, causing the atom to decay and emit another identical photon.

31.5.1 How a laser works

The important process involved in how a laser works is *stimulated emission* - as you can tell from what 'laser' stands for! You can imagine that stimulated emission can lead to more and more identical photons being released in the following way: Imagine we have an electron in an excited metastable state and it drops down to the ground state by emitting a photon. If this photon then travels through the material and meets another electron in the metastable excited state this will cause the electron to drop down to the lower energy level and another photon to be emitted. Now there are two photons of the same energy. If these photons then both move through the material and *each* interacts with another electron in a metastable state, this will result in them *each* causing an additional photon to be released. i.e. from 2 photons we then get 4, and so on! This is how laser light is produced.

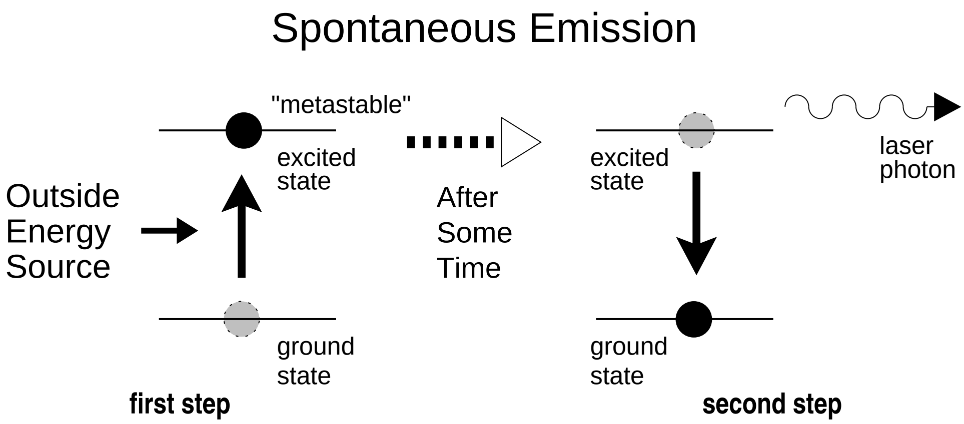


Figure 31.8: Spontaneous emission is a two step process, as shown here. First, energy from an external source is applied to an atom in the laser medium, raising its energy to an excited (metastable) state. After some time, it will decay back down to its ground state and emit the excess energy in the form of a photon. This is the first stage in the formation of a laser beam.

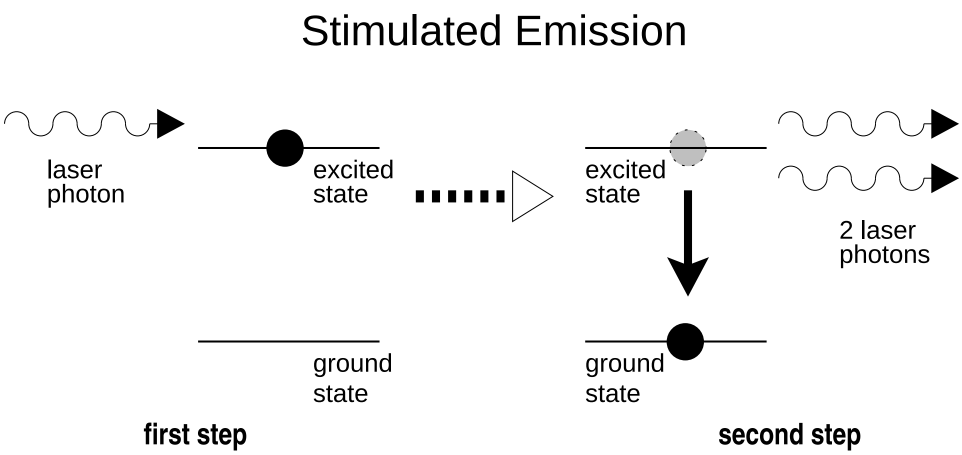
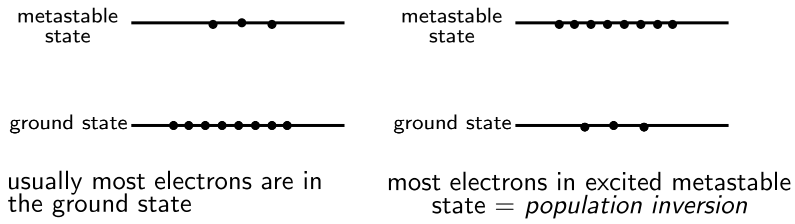


Figure 31.9: Stimulated emission is also a two step process, as shown here. First, a laser photon encounters an atom that has been raised to an excited state, just like in the case of spontaneous emission. The photon then causes the atom to decay to its ground state and emit another photon identical to the incoming photon. This is the second step in the creation of a laser beam. It happens many, many times as the laser photons pass through the optical cavity until the laser beam builds up to full strength.

This can only happen if there are many electrons in a metastable state. If most of the electrons are in the ground state, then they will just *absorb* the photons and no extra photons will be emitted. However, if more electrons are in the excited metastable state than in the ground state, then the process of stimulated emission will be able to continue. Usually in atoms, most of the electrons are in the lower energy levels and only a few are in excited states. When most of the electrons are in the excited metastable state and only a few are in the ground state, this is called **population inversion** (the populations are *swapped* around) and this is when stimulated emission can occur. To start off the process, the electrons first have to be excited up into the metastable state. This is done using an external energy source.

Definition: Population inversion
Population inversion is when more atoms are in an excited state than in their ground state. It is a necessary condition to sustain a laser beam, so that there are enough excited atoms that can be stimulated to emit more photons.



Therefore, materials used to make laser light *must* must have metastable states which can allow population inversion to occur when an external energy source is applied. Some substances which are used to make lasers are listed in table 31.3. You can see that gases (such as Helium-Neon mixture), liquids (such as dyes), and solids (such as the precious stone ruby) are all used to make lasers.

Material	Type	Wavelength	Uses
Helium–Neon	gas	632,8 nm	scientific research, holography
Argon ion	gas	488,0 nm	medicine,
Carbon dioxide	gas	10,6 μm	industry (cutting, welding), surgery
Helium–Cadmium	vapor	325 nm	printing, scientific research
Ruby	solid–state	694,3 nm	holography
Neodymium YAG (Yttrium Aluminium Garnet)	solid–state	1,064 μm	industry, surgery, research
Titanium–Sapphire	solid–state	650–1100 nm	research
Laser diode	semiconductor	375–1080 nm	telecommunications, industry, printing, CD players, laser pointers

Table 31.3: A selection of different lasers. The laser material and general type of each laser is given, along with typical wavelengths of the laser light they create. Examples of the real-world applications it is used for are also given. All these materials allow a population inversion to be set up.

The first working laser, using synthetic ruby as the laser material, was made by Theodore H. Maiman at Hughes Research Laboratories in Malibu, California. Later in the same year the Iranian physicist Ali Javan, together with William Bennet and Donald Herriot, made the first gas laser using helium and neon. Javan received the Albert Einstein Award in 1993.

31.5.2 A simple laser

A laser consists of a number of different parts that work together to create the laser beam. Figure 31.10 shows the different parts of the laser, while Figure 31.11 shows how they create the laser beam.

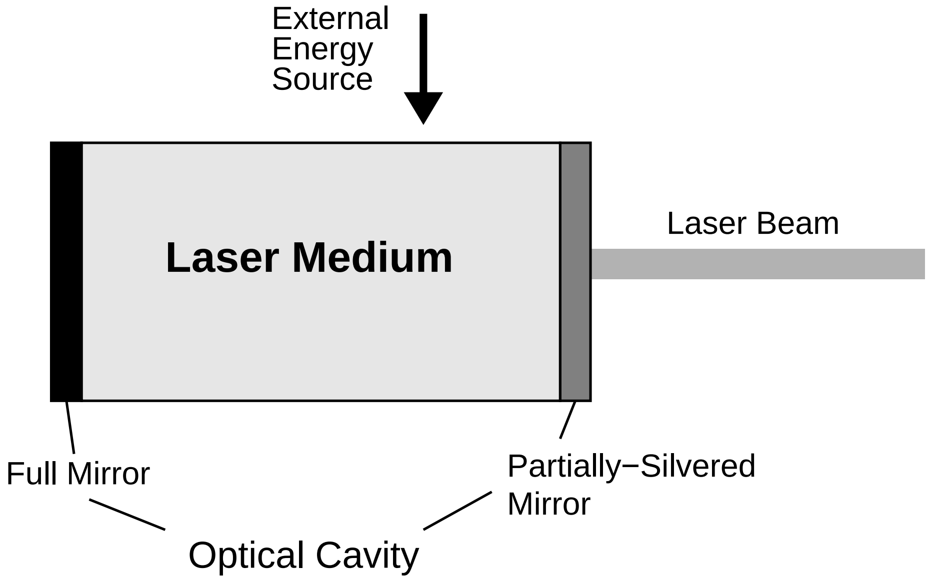


Figure 31.10: Diagram of a laser showing the main components.

The basis of the laser is the laser material. The laser material consists of the atoms that are used to create the laser beam. Many different materials can be used as laser material, and their energy levels determine the characteristics of the laser. Some examples of different lasers are shown in Table 31.3. The laser material is contained in the optical cavity.
Before the laser is turned on, all the atoms in the laser material are in their ground state. The first step in creating a laser beam is to add energy to the laser material to raise most of the electrons into an excited metastable state. This is called *pumping* the laser.
The creation of the laser beam starts through the process of *spontaneous emission*, shown in Figure 31.8. An electron drops down to the ground state and emits a photon with energy equal to the energy difference of the two energy levels. This laser photon is the beginning of the laser

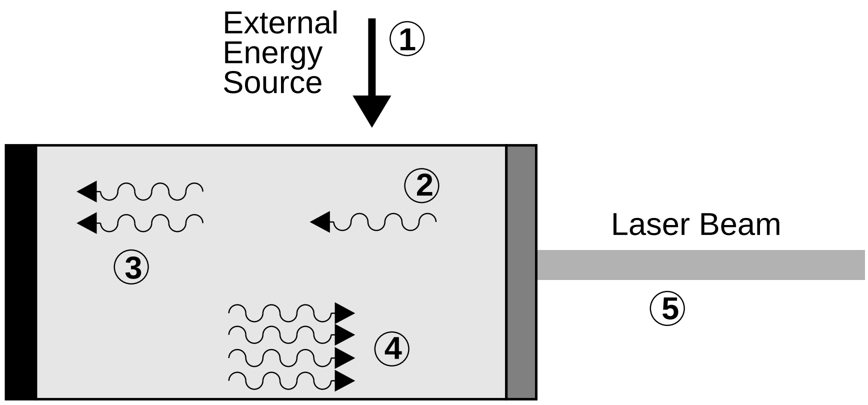


Figure 31.11: Diagram of a laser showing the process of creating a laser beam. (1) A source of external energy is applied to the laser medium, raising the atoms to an excited state. (2) An excited atom decays though spontaneous emission, emitting a photon. (3) The photon encounters another excited atom and causes it to decay through stimulated emission, creating another photon. (4) The photons bounce back and forth through the laser medium between the mirrors, building up more and more photons. (5) A small percentage of the photons pass through the partially-silvered mirror to become the laser beam we see.

beam.
Sometimes a laser photon runs into another excited electron. Then stimulated emission occurs and the electron drops down to the ground state and emits an additional identical photon as shown in Figure 31.9. Since the laser material typically has a large number of atoms, one laser photon passing through this material will rapidly cause a large number of photons just like it to be emitted.
The optical cavity keeps the laser photons inside the laser cavity so they can build up the laser beam. At each end is a concave mirror; one is a full mirror and one is a partial mirror. The full mirror is totally reflective. The partial mirror transmits a small amount of the light that hits it(less than 1%). The mirrors are carefully aligned so that photons that reflect off one mirror become "trapped", and bounce back and forth between the mirrors many times causing more and more stimulated emission. The photons that eventually escape through the partially-silvered mirror become the laser beam that we see.
As the photons bounce between mirrors, they continually pass through the laser material, stimulating those atoms to emit more photons. This creates an ever increasing beam of photons, all with the same characteristics, all traveling in the same direction. In this way, the optical cavity helps to amplify the original laser photons into a concentrated, intense beam of photons.
The laser cavity also helps to narrow the frequency range of laser light emitted. The distance between the two mirrors defines the cavity *mode* which only allows light of a narrow range of frequencies to continue being reflected back and forth. Light of other frequencies damped out. (This is just like in the chapter on the physics of music where a pipe of a certain length corresponds to a particular wavelength of sound.) Therefore only a narrow frequency of light can be emitted.

Interesting Fact

In 1953, Charles H. Townes and graduate students James P. Gordon and Herbert J. Zeiger produced the first maser, a device operating on similar principles to the laser, but producing microwave rather than optical radiation. Townes's maser was incapable of making a continuous beam. Nikolay Basov and Aleksandr Prokhorov of the former Soviet Union worked independently and developed a method of making a continuous beam using more than two energy levels. Townes, Basov and Prokhorov shared the Nobel Prize in Physics in 1964.

31.5.3 Laser applications and safety

Although the first working laser was only produced in 1958, lasers are now found in many household items. For example, lasers are well-known through their use as cheap laser pointers. However, lasers can be very dangerous to the human eye since a large amount of energy is focused into a very narrow beam. **NEVER POINT A LASER POINTER INTO SOMEBODY'S EYES - it can blind them.**

Other uses include:

- Semiconductor lasers which are small, efficient and cheap to make are found in CD players.
- He-Ne Lasers are used in most grocery shops to read in the price of items using their barcodes. This makes the cashiers' job much quicker and easier.
- High energy lasers are used in medicine as a cutting and welding tool. Eye surgery in particular make use of the precision of lasers to reattach the retinas of patients' eyes. The heat from cutting lasers also helps to stop the bleeding on a wound by burning the edges (called cauterising).

Activity :: Case Study : Uses of lasers
Do research in a library or on the Internet on one application of laser technology. Explain how the technology works by using a laser.

You will need to present your findings to the class in the form of a poster. You can think of any useful application, but to give you some ideas of where to start, some applications are listed below:

- laser printers
- laser communication and fibre optics
- optical storage
- using lasers as precision measurement tools
- your own ideas...

?

Exercise: Lasers

1. Explain what is meant by *spontaneous emission of radiation*.
2. Explain what is meant by *stimulated emission of radiation*.
3. List the similarities and differences between spontaneous emission of radiation and stimulated emission of radiation.
4. How is the light emitted by a laser different from the light emitted by a light bulb?
5. Describe using a simple diagram, how a laser works. Your description should include the following concepts: metastable state and population inversion.
6. Give examples of some materials that have been used for lasers. What do all these materials have in common?
7. Describe how the laser cavity affects:
   - increasing amplification
   - concentrating beam intensity
   - narrowing the frequency of the beam
8. List some applications of lasers.

31.6 Summary

1. Light of the correct frequency can emit electrons from a metal. This is called the photoelectric effect.
2. A metal has a work function which is the minimum energy needed to emit an electron from the metal.
3. Emission spectra are formed by glowing gases. The pattern of the spectra is characteristic of the specific gas.
4. Absorption spectra are formed when certain frequencies of light is absorbed by a material.
5. Lasers are devices that produce a special type of light that has many uses.
6. Lasers have many uses, including being used in CD and DVD players, to cut material, in surgery, in printing, in telecommunications and as laser pointers.

31.7 End of chapter exercise

1. What is the photoelectric effect?
2. Calculate the energy of a photon of red light with a wavelength of 400 nm.
3. Will ultraviolet light with a wavelenth of 990 nm of be able to emit electrons from a sheet of calcium with a work function of 2,9 eV?
4. What does the acronym LASER stand for?
5. Name three types of lasers and their uses.
6. Write a short essay on the benefits lasers have had on modern society.

Appendix A

GNU Free Documentation License

PREAMBLE

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or non-commercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "Document", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "you". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "Modified Version" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "Secondary Section" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a

Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "Invariant Sections" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "Cover Texts" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A "Transparent" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "Opaque".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The "Title Page" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

A section "Entitled XYZ" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as "Acknowledgements", "Dedications", "Endorsements", or "History".) To "Preserve the Title" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or non-commercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section A.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections A and A above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

1. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.
2. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.
3. State on the Title page the name of the publisher of the Modified Version, as the publisher.
4. Preserve all the copyright notices of the Document.
5. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.

6. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.
7. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.
8. Include an unaltered copy of this License.
9. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.
10. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.
11. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.
12. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.
13. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.
14. Do not re-title any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.
15. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties–for example, statements of peer review or that the text has been approved by an organisation as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section A above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled “History” in the various original documents, forming one section Entitled “History”; likewise combine any sections Entitled “Acknowledgements”, and any sections Entitled “Dedications”. You must delete all sections Entitled “Endorsements”.

COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an “aggregate” if the copyright resulting from the compilation is not used to limit the legal rights of the compilation’s users beyond what the individual works permit. When the Document is included an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section A is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document’s Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section A. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled “Acknowledgements”, “Dedications”, or “History”, the requirement (section A) to Preserve its Title (section A) will typically require changing the actual title.

TERMINATION

You may not copy, modify, sub-license, or distribute the Document except as expressly provided for under this License. Any other attempt to copy, modify, sub-license or distribute the Document is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See http://www.gnu.org/copyleft/.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.

ADDENDUM: How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

> Copyright © YEAR YOUR NAME. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License".

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the "with...Texts." line with this:

with the Invariant Sections being LIST THEIR TITLES, with the Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.