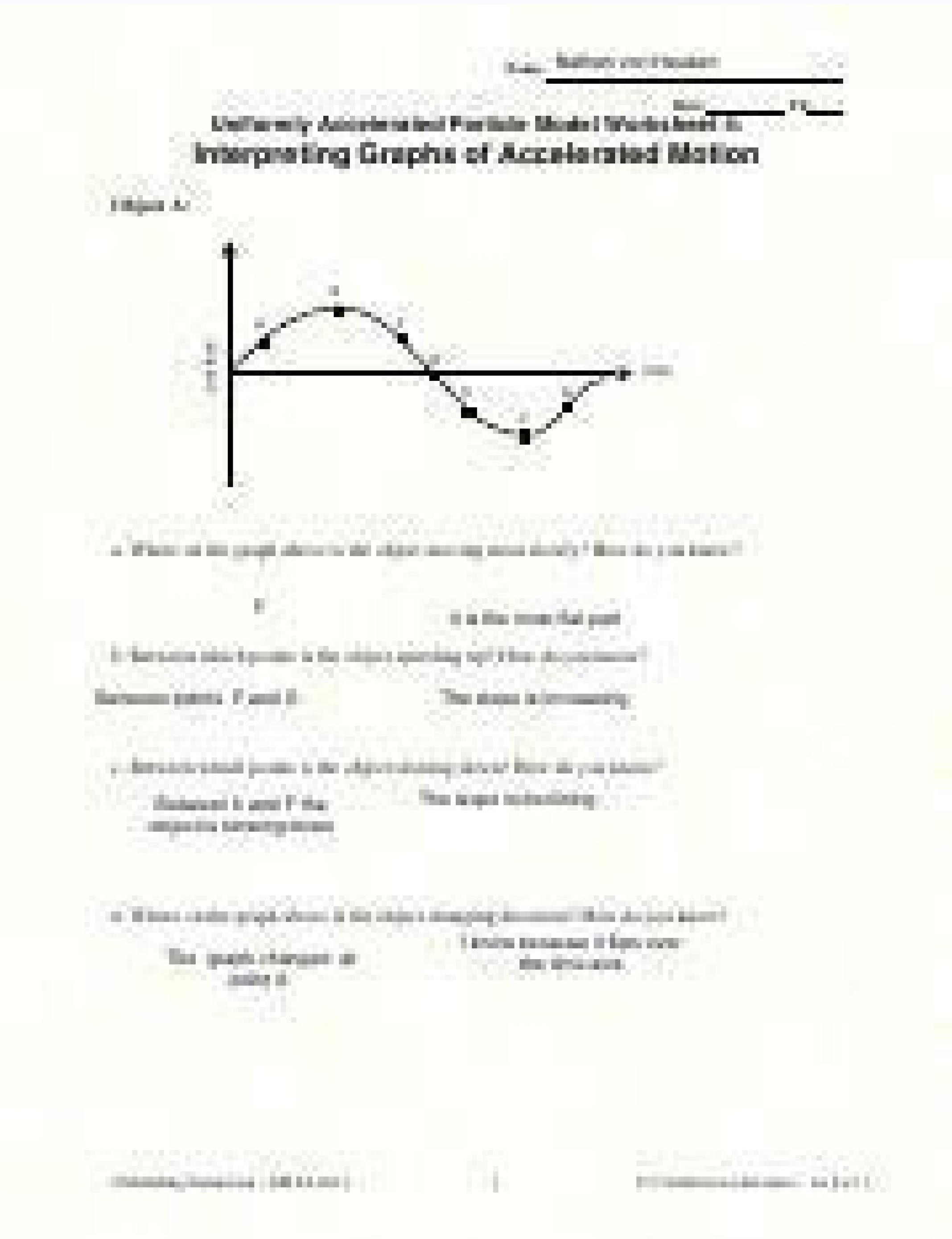
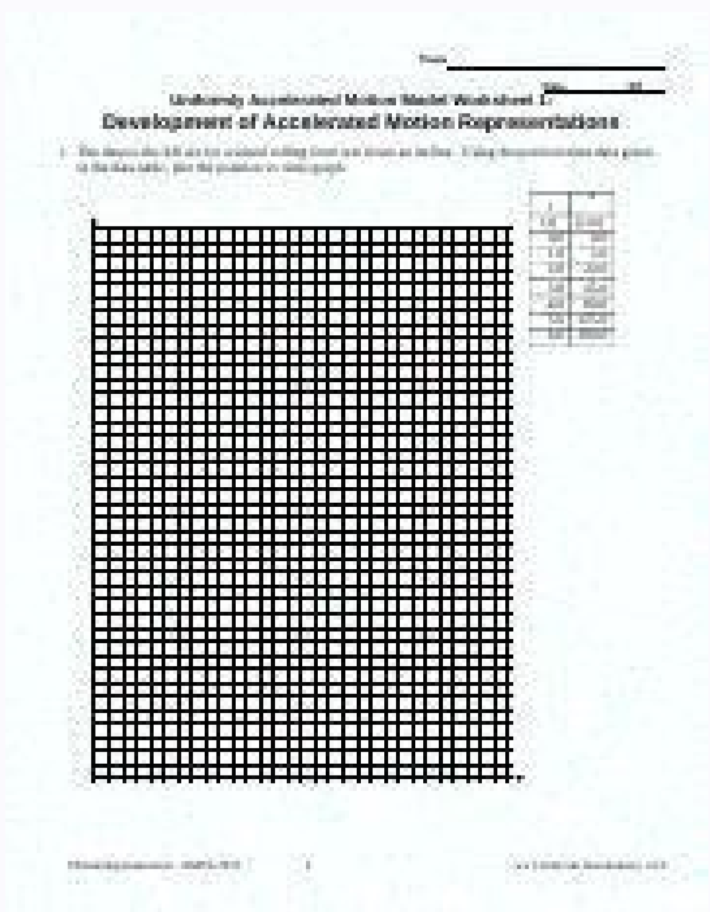
Uniformly accelerated motion model worksheet

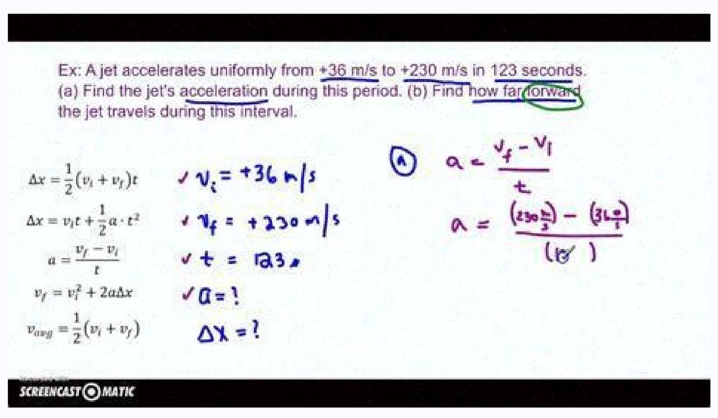
Graphs of Flim versus FN and Ffrk versus FN are shown in Figure 3.10. The slope of each graph is equal to the respective coefficient of friction. The graphs show that μk In Figure 3.9a, a horizontal force, FA, gradually increasing from 0 N, is applied to a box resting on a horizontal surface. At first, the box does not move horizontally. This is because the static frictional force, Ffrs, keeps increasing such that it is equal and opposite to the applied force, FA. The box just begins to move when the maximum static frictional force between the surfaces is reached. This maximum static friction is called the limiting friction, Flim. As the box moves, there is kinetic friction, Ffrk, between the two surfaces. But this kinetic friction does not vary with the applied force but rather stays fairly constant. The kinetic frictional force is also less than the limiting frictional force, Flim. Worked example 3.6: Acceleration of a crate on a rough surface Q A 50 kg crate is pulled along a horizontal floor with a 300 N horizontal force (Figure 3.11). Determine the acceleration of the box if the coefficient of kinetic friction between the crate and the floor is 0.60. FN Relationship between frictional force and normal contact force Experiment shows that both limiting and kinetic frictional forces are directly proportional to the normal contact force, FN. Therefore FA Ffrk mg Figure 3.11 Free-body diagram showing forces acting on the crate. Flim ∝ FN, in which case Flim = μs FN...............(3.11) and Ffrk ∝ FN, in which case Ffrk = μk FN...............(3.12) The constant of proportionality, μ, is called the coefficient of friction between the two surfaces.



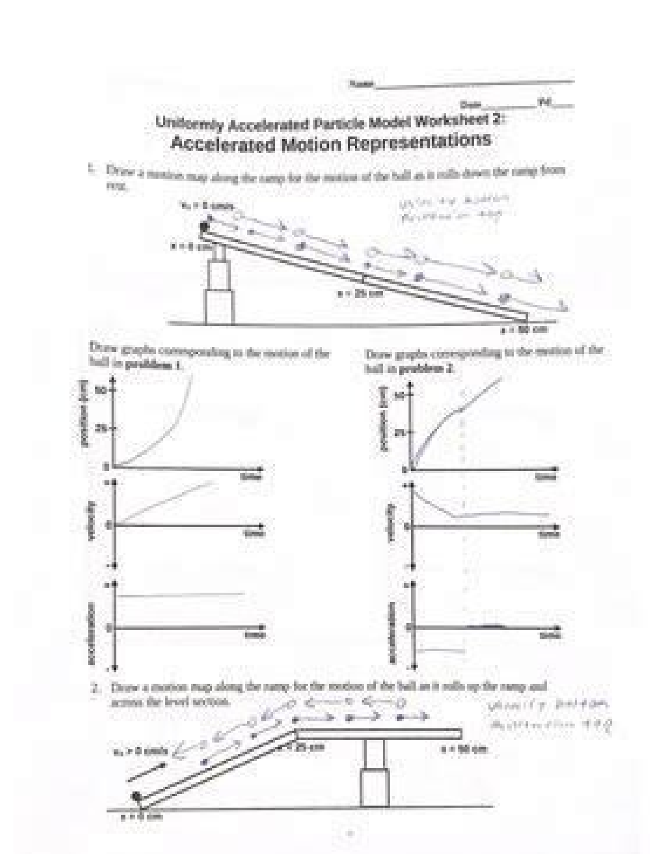
This maximum static friction is called the limiting friction, Flim. As the box moves, there is kinetic friction, Ffrk, between the two surfaces. But this kinetic friction does not vary with the applied force but rather stays fairly constant. The kinetic frictional force is also less than the limiting frictional force, Flim. Worked example 3.6: Acceleration of a crate on a rough surface Q A 50 kg crate is pulled along a horizontal floor with a 300 N horizontal force (Figure 3.11). Determine the acceleration of the box if the coefficient of kinetic friction between the crate and the floor is 0.60. FN Relationship between frictional force and normal contact force Experiment shows that both limiting and kinetic frictional forces are directly proportional to the normal contact force, FN. Therefore FA Ffrk mg Figure 3.11 Free-body diagram showing forces acting on the crate. Flim ∝ FN, in which case Flim = μs FN...............(3.11) and Ffrk ∝ FN, in which case Ffrk = μk FN...............(3.12) The constant of proportionality, μ, is called the coefficient of friction between the two surfaces. The larger this coefficient, the greater the frictional force between surfaces, since the coefficient represents the ratio of force of friction to normal contact force. ■ The constant μs is called the coefficient of static friction. ■ The constant μk is called the coefficient of kinetic (or sliding) friction. ITQ 3 In Figure 3.9, is the pair of forces FA and Ffrs a Newton's third law pair? Explain your answer. A The crate does not accelerate vertically. Referring to Figure 3.11, we get, by application of Newton's second law, FN – mg = 0 FN = mg = 50 kg × 9.8 m s−2 = 490 N Applying Newton's second law in the horizontal direction, we get, for an acceleration, a FA − Ffrk = ma FA − μkFN = ma 300 N − (0.60 × 490 N) = 50 kg × a 300 N – 294 N = 50 kg × a a = 0.12 m s−2 ITQ 4 What would be the S.I. unit for coefficient of friction? Chapter 3   Forces and motion Determining the coefficient of static friction Consider a 20 kg child just about to slide from rest down a plane inclined at an angle of 30° to the horizontal (Figure 3.12). How can we estimate the coefficient of static friction between the child and the plane, making use of the relationship between frictional force and the normal contact force? A little geometry of triangles will show that x = θ (since b + x = 90° and b + θ = 90°). The weight, mg, of the child has two components: mg sin θ along the plane and mg cos θ perpendicular (or normal) to the plane. (See Appendix 3 for a discussion of components.) Applying Newton's second law along the plane, taking down the plane as positive, we get mg sin θ − Flim = m × 0 (since there is no acceleration along the plane) mg sin θ = Flim...............(3.13) Perpendicular to the plane, we get FN − mg cos θ = m × 0 (since there is no acceleration perpendicular to the plane) mg cos θ = FN...............(3.14) Dividing equation 3.13 by equation 3.14, we get F mg sin θ = lim FN mg cos θ Flim Since = μs, according to equation 3.11, we get, FN tan θ = μs...............(3.15) We obtain the rather interesting result that the coefficient of static friction does not depend of the mass of the child, but only on the nature of the two surfaces in contact. The coefficient of static friction is equal to the tan of the angle the inclined plane makes with the horizontal when the object (the child) is just about to slip. For this particular case, we get μs = tan 30° = 0.58. Newton's laws and upthrust forces in fluids Upthrust forces Consider a cuboid of cross-section area A held submerged in a fluid (liquid or gas; Figure 3.13). The pressure of a fluid acts perpendicular to any surface within it. Therefore, the pressure of the fluid acts perpendicularly to the surfaces of the cuboid. The pressure, P, acting perpendicularly on a surface due to a head, h, of fluid of density ρ is given by the equation P = hρg...............(3.16) We derive this equation as follows: weight of fluid above A = mass of fluid above A × g = volume of fluid above A × density of fluid × g = volume of fluid × ρ × g = hAρg force Since pressure = area hAρg = hρg Then, P = A By Pascal's principle, the pressure of a fluid acts equally in all directions at a given point in the fluid. Fluid pressure also acts perpendicularly to any surface placed within the fluid. In Figure 3.13, the forces (= pressure × area) on the top and the bottom of the cuboid are given, respectively, by F1 = P1A = h1ρgA F2 = P2A = h2ρgA Since h2 > h1, the net force on the cuboid in the vertical direction is upward and is given by F2 − F1 = (h2 − h1)ρgA...............(3.17) This net force, F2 − F1, is called the upthrust, U, on the cuboid. If the weight, Wc, of the cuboid is less than the upthrust, then, by Newton's second law (equation 3.3), the cuboid will accelerate upwards when released. FN Flim mg sin θ U − Wc = mcac...............(3.18) mg cos θ b where mc = mass of cuboid; ac = acceleration of cuboid. x mg θ Figure 3.12 Free-body diagram of a child just about to overcome static friction on an inclined plane. We can determine the acceleration, ac, of the cuboid by combining equation 3.17 and equation 3.18: (h2 − h1)ρgA = mcac 45 46 Unit 1   Module 1   Mechanics a b P1 h1 U P h2 P P P P2 Figure 3.13 (a) Pressures on the surfaces of a cuboid submerged in a fluid of density ρ. (b) Free-body diagram showing vertical forces acting on the cuboid.



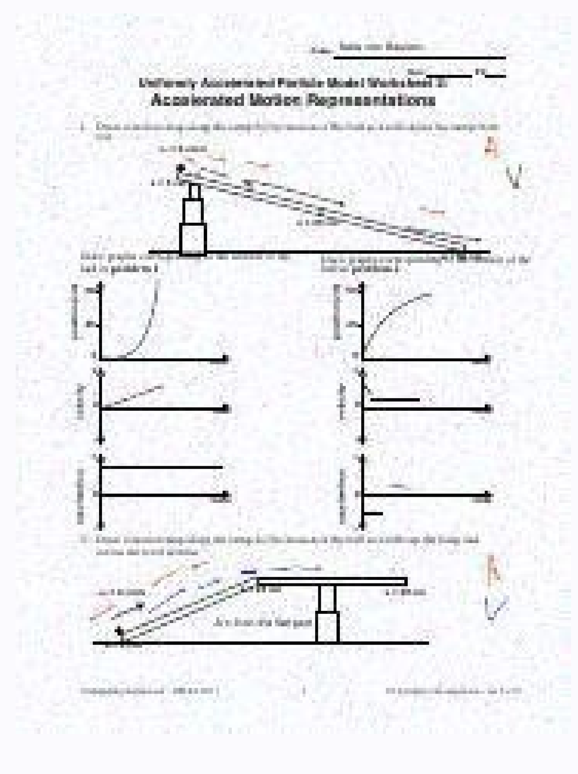
The box just begins to move when the maximum static frictional force between the surfaces is reached. This maximum static friction is called the limiting friction, Flim. As the box moves, there is kinetic friction, Ffrk, between the two surfaces. But this kinetic friction does not vary with the applied force but rather stays fairly constant. The kinetic frictional force is also less than the limiting frictional force, Flim. Worked example 3.6: Acceleration of a crate on a rough surface Q A 50 kg crate is pulled along a horizontal floor with a 300 N horizontal force (Figure 3.11). Determine the acceleration of the box if the coefficient of kinetic friction between the crate and the floor is 0.60. FN Relationship between frictional force and normal contact force Experiment shows that both limiting and kinetic frictional forces are directly proportional to the normal contact force, FN. Therefore FA Ffrk mg Figure 3.11 Free-body diagram showing forces acting on the crate. Flim ∝ FN, in which case Flim = μs FN...............(3.11) and Ffrk ∝ FN, in which case Ffrk = μk FN...............(3.12) The constant of proportionality, μ, is called the coefficient of friction between the two surfaces.



As the box moves, there is kinetic friction, Ffrk, between the two surfaces. But this kinetic friction does not vary with the applied force but rather stays fairly constant. The kinetic frictional force is also less than the limiting frictional force, Flim. Worked example 3.6: Acceleration of a crate on a rough surface Q A 50 kg crate is pulled along a horizontal floor with a 300 N horizontal force (Figure 3.11). Determine the acceleration of the box if the coefficient of kinetic friction between the crate and the floor is 0.60. FN Relationship between frictional force and normal contact force Experiment shows that both limiting and kinetic frictional forces are directly proportional to the normal contact force, FN.



As the box moves, there is kinetic friction, F_{frk}, between the two surfaces. But this kinetic friction does not vary with the applied force but rather stays fairly constant. The kinetic frictional force is also less than the limiting frictional force, F_{lim}. Worked example 3.6: Acceleration of a crate on a rough surface Q A 50 kg crate is pulled along a horizontal floor with a 300 N horizontal force (Figure 3.11). Determine the acceleration of the box if the coefficient of kinetic friction between the crate and the floor is 0.60. FN Relationship between frictional force and normal contact force Experiment shows that both limiting and kinetic frictional forces are directly proportional to the normal contact force, F_N. Therefore FA Ffrk mg Figure 3.11 Free-body diagram showing forces acting on the crate. $F_{lim} \propto F_N$, in which case $F_{lim} = \mu_s F_N$..............(3.11) and $F_{frk} \propto F_N$, in which case $F_{frk} = \mu_k F_N$..............(3.12) The constant of proportionality, μ, is called the coefficient of friction between the two surfaces. The larger this coefficient, the greater the frictional force between surfaces, since the coefficient represents the ratio of force of friction to normal contact force. ■ The constant μ_s is called the coefficient of static friction. ■ The constant μ_k is called the coefficient of kinetic (or sliding) friction.



As the box moves, there is kinetic friction, F_{frk}, between the two surfaces.
But this kinetic friction does not vary with the applied force but rather stays fairly constant. The kinetic frictional force is also less than the limiting frictional force, F_{lim}. Worked example 3.6: Acceleration of a crate on a rough surface Q A 50 kg crate is pulled along a horizontal floor with a 300 N horizontal force (Figure 3.11). Determine the acceleration of the box if the coefficient of kinetic friction between the crate and the floor is 0.60. FN Relationship between frictional force and normal contact force Experiment shows that both limiting and kinetic frictional forces are directly proportional to the normal contact force, F_N. Therefore FA Ffrk mg Figure 3.11 Free-body diagram showing forces acting on the crate. $F_{lim} \propto F_N$, in which case $F_{lim} = \mu_s F_N$..............(3.11) and $F_{frk} \propto F_N$, in which case $F_{frk} = \mu_k F_N$..............(3.12) The constant of proportionality, μ, is called the coefficient of friction between the two surfaces. The larger this coefficient, the greater the frictional force between surfaces, since the coefficient represents the ratio of force of friction to normal contact force. ■ The constant μ_s is called the coefficient of static friction. ■ The constant μ_k is called the coefficient of kinetic (or sliding) friction. ITQ 3 In Figure 3.9, is the pair of forces FA and Ffrs a Newton's third law pair? Explain your answer. A The crate does not accelerate vertically.

Referring to Figure 3.11, we get, by application of Newton's second law, $F_N - mg = 0$ $F_N = mg = 50 \text{ kg} \times 9.8 \text{ m s}^{-2} = 490$ N Applying Newton's second law in the horizontal direction, we get, for an acceleration, a $F_A - F_{frk} = ma$ $F_A - \mu_k F_N = ma$ 300 N − (0.60 × 490 N) = 50 kg × a 300 N − 294 N = 50 kg × a a = 0.12 m s^{-2} ITQ 4 What would be the S.I. unit for coefficient of friction? Chapter 3 Forces and motion Determining the coefficient of static friction Consider a 20 kg child just about to slide from rest down a plane inclined at an angle of 30° to the horizontal (Figure 3.12). How can we estimate the coefficient of static friction between the child and the plane, making use of the relationship between frictional force and the normal contact force? A little geometry of triangles will show that x = θ (since b + x = 90° and b + θ = 90°). The weight, mg, of the child has two components: mg sin θ along the plane and mg cos θ perpendicular (or normal) to the plane. (See Appendix 3 for a discussion of components.) Applying Newton's second law along the plane, taking down the plane as positive, we get $mg \sin\theta - F_{lim} = m \times 0$ (since there is no acceleration along the plane) $mg \sin\theta = F_{lim}$..............(3.13) Perpendicular to the plane, we get $F_N - mg \cos\theta = m \times 0$ (since there is no acceleration perpendicular to the plane) $mg \cos\theta = F_N$..............(3.14) Dividing equation 3.13 by equation 3.14, we get F mg sin θ = lim FN mg cos θ Flim Since = μs, according to equation 3.11, we get, FN tan θ = μs..............(3.15) We obtain the rather interesting result that the coefficient of static friction does not depend of the mass of the child, but only on the nature of the two surfaces in contact. The coefficient of static friction is equal to the tan of the angle the inclined plane makes with the horizontal when the object (the child) is just about to slip. For this particular case, we get $\mu_s = \tan 30° = 0.58$.

Newton's laws and upthrust forces in fluids Upthrust forces Consider a cuboid of cross-section area A held submerged in a fluid (liquid or gas; Figure 3.13). The pressure of a fluid acts perpendicular to any surface within it. Therefore, the pressure of the fluid acts perpendicularly to the surfaces of the cuboid. The pressure, P, acting perpendicularly on a surface due to a head, h, of fluid of density ρ is given by the equation $P = h\rho g$..............(3.16) We derive this equation as follows: weight of fluid above A = mass of fluid above A × g = volume of fluid above A × density of fluid × g = volume of fluid × ρ × g = hAρg force Since pressure = area hAρg = hρg Then, P = A By Pascal's principle, the pressure of a fluid acts equally in all directions at a given point in the fluid. Fluid pressure also acts perpendicularly to any surface placed within the fluid. In Figure 3.13, the forces (= pressure × area) on the top and the bottom of the cuboid are given, respectively, by $F_1 = P_1A = h_1\rho gA$ $F_2 = P_2A = h_2\rho gA$ Since $h_2 > h_1$, the net force on the cuboid in the vertical direction is upward and is given by $F_2 - F_1 = (h_2 - h_1)\rho gA$..............(3.17) This net force, $F_2 - F_1$, is called the upthrust, U, on the cuboid. If the weight, W_c, of the cuboid is less than the upthrust, then, by Newton's second law (equation 3.3), the cuboid will accelerate upwards when released. FN Flim mg sin θ $U - W_c = m_ca_c$..............(3.18) mg cos θ b where m_c = mass of cuboid; a_c = acceleration of cuboid. x mg θ Figure 3.12 Free-body diagram of a child just about to overcome static friction on an inclined plane. We can determine the acceleration, a_c, of the cuboid by combining equation 3.17 and equation 3.18: $(h_2 - h_1)\rho gA = m_ca_c$ 45 46 Unit 1 Module 1 Mechanics a b P1 h1 U P h2 P P P P2 Figure 3.13 (a) Pressures on the surfaces of a cuboid submerged in a fluid of density ρ. (b) Free-body diagram showing vertical forces acting on the cuboid.

Wc Archimedes' principle Since $(h_2 - h_1)A$ is the volume of the cuboid submerged, this is the same as the volume V of fluid of mass m displaced by the cuboid. moving plate of area A F v stationary bottom of vessel d liquid Equation 3.17 therefore becomes $F_2 - F_1 = U = (h_2 - h_1)\rho gA = V\rho g = mg$ That is, U = mg..............(3.19) Equation 3.19 tells us that: The upthrust on a solid submerged in a fluid is equal to the weight of the fluid displaced. This relationship holds whether the solid is wholly or partly submerged in the fluid. The statement above is known as Archimedes' principle and holds for any shape of solid immersed in a fluid. One consequence of Archimedes' principle arises when we apply Newton's second law to a floating object. We obtain the law of flotation which states that: When an object floats, the weight of fluid displaced is equal to the weight of the object. Newton's laws and resistive forces in fluids Viscous forces in fluids When there is relative motion between a liquid and a solid, forces of friction arise within the liquid as layers of liquid slide along each other if the liquid flow is orderly. This type of friction within a liquid is called viscosity. Figure 3.14 shows a solid plate of surface area A being forced along the surface of a liquid of depth d in a vessel. If the solid plate is moved at a speed v the layer of liquid in contact with the surface moves with the same speed, v. Figure 3.14 Velocity gradient in a liquid. The layer at the bottom of the vessel is stationary as it rests on a stationary surface. A velocity gradient, (v − 0), is set up across the layers of liquid. Experiments d show that the force, F, required to keep the plate moving at speed v is proportional to the area of the surface and the velocity gradient: F∝A F = ηA v d v d ....(3.20) The constant, η, is called the coefficient of viscosity of the liquid.

The applied force, F, is used to overcome the viscous drag force of the liquid on the plate. In liquids, the viscous forces are due to attractive forces of cohesion between molecules of the liquid (see Chapter 16 for forces between molecules). In gases, the forces are due to change in momenta of molecules as they bombard each other across adjacent layers. ITQ 5 A 5000 tonne ship floats first in salt water (density 1025 kg m^{-3}) and then in fresh water (density 1000 kg m^{-3}). (1 tonne = 1000 kg) (a) What weight of salt water does the ship displace? (b) What weight of fresh water does the ship displace? Chapter 3 Forces and motion Viscous drag on a sphere at low speeds At low speeds, the flow of fluid around a sphere is laminar (orderly) – and the drag on a sphere is then given by application of Stokes' law Discussion example 3.7: Object falling through the air Q A small sphere of mass m and radius r is dropped through the air (density ρ). (a) Draw a labelled free-body diagram to show the forces acting on the sphere. (b) Write the equation for the buoyant force (upthrust) acting on the sphere. (c) Write the equation that enables the calculation of the acceleration, a, of the sphere. A (a) $F_d = 6\pi\eta rv$..............(3.21) where η is the viscosity coefficient and r is the sphere radius. Drag at high speeds It can be shown that the drag force, F_d, on a sphere moving at high speed through a fluid is given by Fd drag U upthrust $F_d = DA\rho v^2$ ....(3.22) where D is called the drag coefficient (which has a value of about 0.5 for a sphere and may be as large as 2, depending on the shape of the object), A is the crosssection area of the object perpendicular to the object's velocity, v, and ρ is the density of the fluid.

At high speeds the drag force becomes proportional to v^2, but the flow is no longer orderly – it is turbulent. In this type of situation, Stokes' law does not apply.

The resistance (drag) does not depend on the viscosity of the fluid, but on its density – so the resistance force might be better referred to simply as 'drag' or 'air resistance' rather than viscous drag. Equation 3.22 shows that the drag force is proportional to the square of the speed, v. Terminal velocity Equation 3.23 shows that as the speed, v, increases from zero, there comes a point when the net force (F_d − mg), assuming negligible upthrust, on an object of mass m falling through a fluid = 0. Hence the acceleration, a, = 0, which means that the velocity is constant. This constant velocity is called terminal velocity. A graph of the magnitude of the velocity versus time, showing the magnitude of the terminal velocity, v_T, is shown in Figure 3.16. v vT t Figure 3.16 Graph of magnitude of velocity versus time for an object falling through a fluid and achieving terminal velocity. ITQ 6 By what factor will the drag force increase if the speed of an object is tripled? W weight Figure 3.15 Forces acting on a sphere falling through air. (b) upthrust = weight of air displaced by sphere = (volume of air × density of air) × g = (⁴⁄₃)πr 3ρg (c) Applying Newton's second law to the motion of the sphere (Figure 3.15) and the equation for viscous drag on a sphere (equation 3.22), assuming non-laminar flow of the air around the sphere, we get $U + F_d - W = ma$ (⁴⁄₃)πr 3ρg + 0.5Aρv2 − mg = ma Therefore, (⁴⁄₃)πr 3ρg + 0.5πr2ρv2 − mg = ma..............(3.23) Static and dynamic equilibrium Translational equilibrium Translational equilibrium occurs when the net force on an object is zero.

By Newton's second law, the acceleration of the object is zero. The object is either stationary (static equilibrium), as in the case of the box resting on the table (Figure 3.6), or moving at constant velocity (dynamic equilibrium), as in the case of terminal velocity, v_T (Figure 3.16). In Figure 3.17, for translational equilibrium to occur, the net force on the beam must be zero: $F_2 - F_1 - F_3 = 0$ Rotational equilibrium For a beam pivoted as shown in Figure 3.17, F_3 provides a clockwise turning effect about the pivot (also called a fulcrum) whereas F_1 provides an anticlockwise turning effect. F_2 provides no turning effect since this force acts 47 48 Unit 1 Module 1 Mechanics Clockwise torques are designated positive and anticlockwise torques, negative. F2 beam d12 pivot The net torque, ΣT, produced about the pivot is given by $\Sigma T = F_3 \times d_{23} + F_2 \times 0 - F_1 \times d_{12}$..............(3.25) d23 F1 F3 Figure 3.17 A light, extended body in equilibrium. where d_{23} represents the perpendicular distance between the line of action of F_3 and the pivot (the latter coinciding with the point of application of F_2 on the axis of the beam) and d_{12} represents the perpendicular distance between the line of action of F_1 and the pivot. The symbol 'Σ' (Greek capital sigma) is used to mean 'the sum of'. T = F × d..............(3.24) When the net torque is zero, the beam either remains stationary (static, rotational equilibrium) or rotates about the pivot at a constant frequency (dynamic, rotational equilibrium). where T is the moment (or torque), F is the force and d is the perpendicular distance from the line of action of the force to the pivot.

ITQ 7 A uniform metre rule of mass 400 g is pivoted freely at the 20.0 cm mark. Calculate the torque about the pivot. (The weight of a uniform rod acts at the centre of the rod.) at the pivot.

Turning effect, called moment or torque, is calculated by the equation Summary ✓ Momentum, p, of an object of mass m moving ✓ According to Archimedes' principle, the upthrust on a body, wholly or partly immersed in a fluid, is equal to the weight of the fluid displaced. with a velocity v is defined as mass times velocity: p = mv..............(3.1) ✓ According to Newton's first law of motion, every object continues in its state of rest or of uniform motion in a straight line unless acted on by a net external force. ✓ Upthrust on an object immersed in a fluid originates from the fact that the fluid pressure at the top of the object is less than the fluid pressure at the bottom, and that the fluid pressure at any given level is equal on all sides of the object. ✓ According to Newton's second law of motion, the net force, F, applied on a mass, m, is directly proportional to the rate of change of momentum of the mass and is in the same direction as the change in momentum. For an initial velocity v_0 of the mass changed to a final velocity v in a time t, the following equations arise out of Newton's second law: F = ma..............(3.3) $Ft = mv - mv_0$..............(3.4) The product Ft is called the impulse of a force on an object and is equal to the change in momentum produced on the object. ✓ According to Newton's third law of motion, if an object A exerts a force on another object B then B exerts an equal and opposite force on object A. In equation form, the force of A on B is equal to the negative of the force of B on A: $F_{AB} = -F_{BA}$ ✓ It is believed that friction between solids is due to forces of attraction between molecules and also to molecular protrusions on the surfaces of two solids rubbing against each other. ✓ Friction within fluids (called viscosity) is due, in the case of liquids, to forces of cohesion between adjacent layers of molecules. For orderly flow of layers, the viscous force, F, on a sphere moving with velocity v through a liquid is given by the Stoke's law equation: $F_d = 6\pi\eta rv$..............(3.21) where η is the viscosity coefficient of the liquid and r is the sphere radius. ✓ In the case of gases, viscous forces arise from molecules moving from one layer to the next and colliding with other molecules.

Chapter 3 Forces and motion ✓ As an object moves quickly through a fluid of density ρ, the flow of the layers of fluid becomes disorderly; the force of drag, F_d, on the object is directly proportional to the square of the velocity, v, of the object through the fluid. For a sphere radius r ✓ For translational equilibrium of a body, the net force on the body is equal to zero. The body either remains at rest or moves with uniform velocity. ✓ For rotational equilibrium of a body about an axis, the net moment (or torque) on the body about any given axis is equal to zero. The body either remains at rest or rotates with constant frequency. $F_d = DA\rho v^2$..............(3.22) where D is called the drag coefficient (which has a value of about 0.5 for a sphere, to 2, depending on the shape of the object), A is the cross-section area of the object perpendicular to the object's velocity, v, and ρ is the density of the fluid. Review questions 5 Newton's first law of motion; Newton's second law of motion; Newton's third law of motion; Newton's laws and frictional forces MA = 450 g MA pulley FT 1 (a) Draw a free-body diagram of a box resting on a horizontal floor, labelling the forces involved. (b) Which of Newton's laws explains why the two forces are equal and opposite? Explain your answer. 2 A 50 kg box rests on a rough ramp inclined at an angle of 40° to the horizontal. (a) Draw a labelled free-body diagram to show the inclined plane and forces acting on the box. (b) Calculate the components of the box's weight along the plane and perpendicular to the plane. (c) Determine the normal contact force (magnitude and direction) acting on the box.

3 A 1.5 kg box is given an initial velocity of 2.0 m s^{-1} up a plane, from the bottom of the plane. The plane is inclined at 30° to the horizontal.

The coefficient of kinetic friction between the box and the plane is 0.44. Determine: (a) the acceleration of the box as it travels up the plane (b) the distance the box travels up the plane before coming to a stop. (Hint: it will be useful to draw a diagram.) 4 (a) Draw free-body diagrams showing two masses (M_A = 40 kg and M_B = 35 kg) in an Atwood machine arrangement. (b) (i) Calculate the magnitude of the acceleration of the masses. (ii) Calculate the magnitude of the force of tension in the string.

(c) Suggest one everyday use of the Atwood machine arrangement. Explain why it is used in this application. MB MB = 600 g MB g Figure 3.18 Figure 3.18 shows two masses connected by a light string that rises over a frictionless pulley. The coefficient of kinetic friction between mass A and the surface shown is 0.42.

(a) Draw and label a free-body diagram showing all the forces acting on mass A. (b) Determine the acceleration of the system. 6 In order to move a heavy truck, a person ties a rope to a firm tree and the truck as shown (Figure 3.19). When the person pulls on the rope in the direction of the arrow, with a force of 480 N, as shown, the truck just begins to move. Estimate the force of tension in the rope. F = 480 N 20˚ Figure 3.19 20˚ 49 50 Unit 1 Module 1 Mechanics 7 Calculate, using components, the resultant of two forces. F_1 = 1800 N in a direction 40° east of north and F_2 = 2500 N in a direction 10° north of west. (Hint: a diagram will help.) 8 A passenger pulls a 50 kg suitcase horizontally with the strap at an angle of 60° to the horizontal. If the force in the strap is 100 N and the suitcase moves at a constant velocity of 0.15 m s^{-1}, determine the coefficient of kinetic friction between the suitcase and the floor. (Hint: a free-body diagram will help.) 9 A car accelerates from rest to 40 km/h along a distance of 12 m. During this time, a pendant with a small ball attached remained inclined at an angle θ to the vertical. Determine the angle θ: (a) during the acceleration (b) when the car achieves a constant velocity of 40 km/h. Newton's laws and upthrust forces in fluids; Newton's laws and resistive forces in fluids 10 (a) Derive a formula for the upthrust on a cuboid of height h and cross-sectional area A immersed in a fluid of density ρ. (b) A loaded 2000 tonne barge, of dimensions 40 m (length), 30 m (width) and 5.0 m (height), traverses the Panama Canal from the Atlantic side to the Pacific side. Assuming the average densities of sea water in the North Atlantic and Pacific Oceans are 1025 kg m^{-3} and 1040 kg m^{-3}, respectively, calculate: (i) the mass of sea water displaced by the ship in each of the waters (ii) the volume of sea water displaced by the floating ship in each ocean (iii) the change in the height of the waterline on the side of the ship, and whether the ship rises or drops as it moves from the North Atlantic side to the Pacific side. 11 (a) State the 'law of flotation'. (b) The density of sea water generally increases with depth.

The density of a submarine can be altered by pumping in or pumping out water from compartments called 'ballast' tanks. State and explain how, in terms of upthrust forces and density of ocean water, (i) a submarine can be made to sink deeper in the ocean (ii) a submarine can be made to rise from deep under the ocean. 12 (a) Explain the origin of viscous forces in: (i) liquids (ii) gases. (b) (i) What is the condition for terminal velocity, v_T, of a solid falling through a gas, in terms of upthrust, drag and weight? (ii) Write an equation that illustrates your answer in part (b)(i) for a sphere of mass m and radius r falling in a gas of density ρ. State, with a reason, the choice of the equation given in your answer. (Hint: see equation 3.22.) 13 The terminal velocity, v_T, of a sphere is related to the radius, r, of the sphere by the equation, $v_T = cr^n$, where c and n are constants. In an experiment, values of v_T and corresponding values of r are measured. How can the values of c and n be determined by plotting a suitable graph? (Hint: see Appendix 3 about logarithms and graphs.) Static and dynamic equilibrium 14 Two painters, Ian (mass 60 kg) and Sheila (mass 45 kg), are standing on a trestle (a thick board) of mass 10 kg, one storey above the ground. The trestle is 5.0 m long and rests on narrow supports at 0.50 m and 3.0 m, respectively, from the left end, L. Sheila stands 1.0 m from L. (a) How far from L can Ian stand on the trestle without causing it to overturn? (b) Determine the reaction forces at each of the supports when the trestle is just about to overturn as described in (a) above. 15 Forces of 120 N at an angle of 30° north of east and 100 N at an angle of 35° south of east are applied horizontally to a crate of mass 50 kg lying on a horizontal floor. If the forces cause the crate to move at a constant velocity, determine (a) the net force being applied to the crate (magnitude and direction) (b) the coefficient of kinetic friction between the crate and the floor. Chapter 3 Forces and motion Practical exercises and challenges 1 Atwood machine Set up an Atwood machine consisting of masses m and M (where M > m) connected by a light string that passes over a fixed, low-friction pulley (Figure 3.20). Let the mass M start falling from rest from a height h above M the floor. Release the mass M and measure the time taken for it to fall the distance h.

Do this several times, using various values of h. h Determine the acceleration, a, from the values of h and t by plotting a m graph of h versus t2 (according to 1 equation 2.8, h = 2 at2). Compare the experimental value of a with the theoretical value (equation 3.8). Make an estimate of the percentage error expected in your experiment, and compare this with the actual percentage error obtained. Figure 3.20 (a) Plot a graph of distance fallen, h, versus time, t. (b) Using slopes from the h versus t graph, plot a graph of velocity, v, versus time, t. Explain the shape of this graph and determine the terminal velocity, v_T. (c) What is the value of the acceleration of the sphere moving at terminal velocity? (d) Terminal velocity, v_T, can be calculated from equation 3.21: 4 (3)πr3ρg + 6πηrv − mg = ma Determine the theoretical and experimental percentage errors in v_T. Answers to ITQs 1 (a) Position A.

(Position B is approximately the normal position of the head.) (b) The force of the car seat on the torso when the car is suddenly hit from behind. 2 The gravitational pull of the box on the Earth. By the third law, the pull of the Earth on the box is equal and opposite to the pull of the box on the Earth. 2 Coefficient of static friction Investigate the coefficient of static friction between a shoe bottom and two different types of tiles (or wooden surfaces) and an inclined plane. Use trigonometry to determine the angle of incline of the plane to the horizontal. You can use an arrangement such as shown in Figure 3.12 (using a shoe to replace the child), along with equation 3.15. Design your experiment so that you can conclude, with a reasonable degree of confidence, whether the coefficients were the same or different. 3 No! FA is the pull of a person on the box, whereas Ffrs is not the pull of the box on the person but rather the force (friction) of the floor on the box. F 4 No unit. Since μ = Ffr , the unit of force in the N numerator and denominator cancel each other. 3 Viscous drag Drop a small sphere (or several identical spheres) of mass m and radius r several times through thick lubricating oil in a tall transparent tube. (You may wish to make marks at equal intervals on the tube corresponding to different heights, h, or place a suitable cm scale beside the tube. If you have a digital video camera, you can examine your video on a computer, frame by frame, and determine both h and t rather easily.) 7 We calculate torque in S.I. units: 5 When a ship floats, weight of ship = weight of water displaced, regardless of the kind of water.

W = mg = 5000 × 1000 kg × 9.8 m s^{-2} = 4.9 × 10^7 N in both cases (a) and (b) 6 By a factor of 9 (3^2)! T = 0.30 m × (0.400 × 9.8) N ≈ 1.2 N m Answers to Review questions 3 (a) −8.6 m s^{-2} (b) 0.23 m 5 3.8 m s^{-2} 7 2200 N, 40° W of N 9 (a) 27° (b) 0° 15 (a) 186 N, 0.81° N of E (b) 0.38 51 52 Chapter 4 Work, energy and power Learning objectives ■ Define and apply the concept of work as the product of force and displacement in the direction of the force (W = Fx). ■ Define energy as an ability to do work, and describe work as a process in which energy is converted. ■ List and give examples of different forms of energy. 1 ■ Derive and use the formula for kinetic energy: Ek = 2 mv2. ■ Distinguish between different types of potential energy, such as gravitational, electrical, elastic and strain energy. ■ Derive and apply the formula $\Delta E_p = mg\Delta h$ to potential energy changes near the Earth's surface. ■ State and apply the law of conservation of energy. ■ Apply the concept of power as the rate of doing work or rate of converting energy (P = W and P = F × v).

■ Apply the concept of efficiency to solving problems involving energy transfer. t ■ Describe examples of energy conversion in industry and in everyday life.

■ Apply the concept of energy conversion to the Caribbean situation in terms of renewable and non-renewable sources of energy. ■ Discuss critically mechanisms for the efficient use of energy in the Caribbean. ■ Discuss the need for sustainability in the use of energy for society today. Work and energy Work Work is defined as the product of force and distance moved in the direction of the force. Consider a force of 5.0 N used to pull a box. If the box does not move, then no work is done by the force on the box.

If the box moves a distance of 3.0 m in the direction of the force, then the work done by the force is 5.0 N × 3.0 m = 15 N m Worked example 4.1 shows that if a force F moves an object through a displacement x at an angle θ between F and x, then work done by the force is given by the equation W = Fx cos θ..............(4.1) Work done in moving an object round a circle Figure 4.2a shows a householder mowing a lawn by pushing a lawn mower in circles. The horizontal component, F_H, of the force exerted by the person always moves the lawn mower in the direction of F_H, even though the direction of F_H is always changing. Hence, work, W_c, done in moving the lawn mower round a circle of radius r is given by $W_c = F_H \times 2\pi r$ In Figure 4.2b, the work done by gravity in moving the Earth in one complete circle is zero, since, for circular motion, the force of gravity, F_G, is always at right angles to the direction the Earth is moving and so cos θ is always 0. The topic of forces involved in circular motion is discussed in Chapter 6. Strictly speaking though, the motion of the Earth around the Sun is not perfectly circular, but slightly elliptical. The angle between the direction of the force of gravity and the direction of motion of the Earth is 90° only when the two bodies are at their closest and farthest approaches (i.e. at perihelion and aphelion). There will be a small tangential component of the force of gravity that will do work, causing the Earth to be slowing down during one half of the orbit and speeding up in the other. The net work done by this tangential component during one complete orbit is zero. Chapter 4 Work, energy and power Energy Worked example 4.1: Traveller pulling a suitcase Q A traveller pulls a suitcase with a force of 6.0 N along a horizontal floor (Figure 4.1). The angle the strap makes with the horizontal is 40°, and the horizontal distance pulled to the check-in area is 9.0 m. How much work has the traveller done: (a) in moving the suitcase along the horizontal floor to check-in? (b) in lifting the suitcase as it moves the 9.0 m distance along the floor? θ Energy, the ability to do work, comes in many forms. Energy can be converted from one form to another by natural and artificial means. Here we outline some of the main forms. Figure 4.1 Traveller pulling a suitcase along a horizontal floor. A Let F = force in the strap, θ = angle between the horizontal and the strap and d_H = horizontal distance moved by the suitcase. (a) To find the work done by the traveller in moving the suitcase along the floor we need to know the horizontal component of the force, F_H, exerted by the traveller on the suitcase and the distance, d_H, moved by the suitcase along the floor. By simple trigonometry, $F_H = F \cos\theta$ (see Appendix 3). F_H is called the horizontal component of the force, F. Work, W_H, done in moving the suitcase along the horizontal floor is given by: $W_H = F_H \times d_H = F \cos\theta \times d_H = 6.0 \text{ N} \times \cos 40° \times 9.0 \text{ m} \approx 41$ N m (b) To find the work done by the traveller in lifting the suitcase by the force, F, in the strap, we need to know the force, F_V, exerted on the suitcase by the traveller in the vertical direction. F_V is called the vertical component of F and is equal to F sin θ. Since the suitcase moves horizontally only, then the vertical distance, d_V = 0.

Work, W_V, done in the vertical direction by force F is given by $W_V = F_V \times d_V = 0$ N m No work is done on the suitcase by the traveller in lifting the suitcase! a b If something is capable of doing a large amount of work we say it has much energy. If it is unable to do much work, we say it has little energy. As we shall see shortly, work represents an energy conversion process. That is why we equate 1 unit of work done with 1 unit of energy converted. Therefore, in the S.I., 1 N m of work done ≡ 1 J of energy converted..............(4.2) CHECK IN F Energy is defined as the ability to do work.

(a) Mechanical energy is energy that has to do with forces applied to masses. Kinetic energy and potential energy are each a subset of mechanical energy. These two will be discussed in the next two main sections. (b) Thermal energy is energy due to the random movements of molecules and forces between molecules, and thus is kinetic and potential energy at the molecular level (see Chapter 14). The flow of thermal energy from one substance to another, or within a substance, due to temperature differences, is called heat energy. (c) Chemical energy is energy due to the arrangement of electrons of an atom in discrete energy states on account of electrical forces. When elements react with each other, electrons are moving from one energy state to another as compounds are formed. Energy is taken in to the system if electrons end up in higher energy states than they were originally (an endothermic reaction). If electrons end up in lower energy states, energy is given off (an exothermic reaction). path taken by Earth F FH Sun's gravity Figure 4.2 Work done in moving an object in a circle in two different situations: (a) a householder mowing a lawn; (b) the Earth orbiting the Sun.

ITQ 1 Does the Sun's gravity do net work in maintaining the Earth's motion for 6 months from perihelion to aphelion? 53 54 Unit 1 Module 1 Mechanics (d) Electrical energy is energy due to charges present in matter. Charges in motion constitute an electric current. The energy of an electric current is due to the presence of the charge carriers in the electric field created by the electromotive force that drives the current – it is electrical kinetic energy.

Electric energy is also stored in electric fields. Molecules, made up of static electric charges, store electrical energy in electric fields within them when they undergo 'polarization' (discussed in Chapter 18). Energy is also stored in fields between two oppositely charged plates, even if there are no molecules between the plates. (e) Nuclear energy: this form of energy is obtained when a large nucleus (e.g. uranium-235) is split, or when small nuclei (e.g. hydrogen-2) fuse together. Some mass is 'lost' in the reaction, but appears as energy according to the equation $E = mc^2$ (E = energy, m = nuclear mass converted, c = speed of light).

Nuclear energy is discussed further in Chapter 27. (f) Electromagnetic wave energy (also called radiant energy): electromagnetic waves have electric and magnetic properties, have no mass and all travel at the speed of light in free space. These waves carry energy along with them.

The electromagnetic wave energy (mainly ultraviolet light, visible light and infrared light) from the Sun has been the main source of energy on the Earth. Kinetic energy is defined as energy due to the motion of a mass. A moving car, even with its engine switched off, can push a stationary car on a horizontal road as a result of a crash. Hence, the moving car can do work and, therefore possesses kinetic energy.

Work–energy principle We derive the formula for the kinetic energy, E_k, of a mass m travelling at a speed v as follows. We imagine a low-friction dynamics cart of mass m at rest on a smooth desk (i.e. the opposing force of friction between the cart and the desk surface can be neglected). We consider a spring balance F x0 Work done, $W = F \times (x - x_0)$ Since F = ma (by Newton's second law of motion), then $W = ma(x - x_0)$..............(4.3) Applying the linear kinematics equation (equation 2.9) $v^2 = v_0^2 + 2a(x - x_0)$ gives us v2 − v02 2a = x − x0 ....(4.4) Substituting equation 4.4 into equation 4.3, we get (v2 − v02) W = ma 2a mv2 − mv02 That is, W = 2 1 1 or W = 2 mv2 − 2 mv02.............. (4.5) Since the left-hand side of equation 4.5 represents work, then each term on the right-hand side represents energy. This is so because, as pointed out earlier, work represents a process of conversion of one form of energy to another. The 1 term 2 mv2 is called kinetic energy, since the moving cart itself can now do work after the force has been removed. Equation 4.5 is an expression of the work–energy principle which states that the net work done on an object is equal to its change in kinetic energy. Worked example 4.2: Caught! Kinetic energy cart constant, horizontal force, F, that moves the cart through a horizontal distance x − x0 (Figure 4.3). F x x - x0 Figure 4.3 A horizontal force, F, pulling a dynamics cart a horizontal distance, x − x0, along a smooth desk. Q A cricket ball, travelling approximately horizontally when caught, causes the fielder's hand to move a distance of 15 cm in the direction of the ball's motion in order to stop the ball. If the mass of the ball is 0.160 kg and it was travelling at 12 m s^{-1} when caught, estimate the average force of the ball on the fielder's hands. A ■ stopping force of ball on hand, F_{bh} = ? ■ mass of ball, m = 0.160 kg ■ initial velocity of ball, v_0 = 12 m s^{-1} ■ final velocity of ball, v = 0 m s^{-1} ■ stopping distance of ball, $(x - x_0)$ = 15 cm = 0.15 m ■ stopping force of hand on ball, F_{hb} = ? By the work–energy principle, work done in stopping ball = change in kinetic energy of ball Therefore, Fhb × (x − x0) = 12 mv2 − 12 mv02 Therefore, Fhb × 0.15 m = 0 − 12 × 0.160 kg × 122 (m s−1)2 Or, 0.15Fhb = −11.52 N Therefore, F_{hb} ≈ −77 N The stopping force of the hand on the ball is negative, i.e. opposite to the motion of the ball. Therefore, the force of the ball on the hand is equal and opposite to that, and is therefore equal to +77 N, i.e. in the direction of the ball's original motion. Chapter 4 Work, energy and power It follows that the kinetic energy (Ek or KE) of an object moving at a speed v is given by 1 Ek = 2 mv2..............(4.6) and is equal to the work done in moving the object from rest to a speed v. Potential energy Potential energy is energy in stored form. Table 4.1 provides a summary of various types of potential energy. Table 4.1 Some types of potential energy Type of potential energy Description Worked example 4.3: Should this driver be charged for speeding? Gravitational Energy stored due to an arrangement of masses. Elastic Energy stored in materials that have been stretched or compressed. This type of energy is also called strain energy. Q Electrical A A car of mass 1200 kg leaves skid marks during an emergency stop in a 40 km/h zone. An officer, noticing this, asks the driver to pull over to the side of the road. The measured length of the skid is 8.4 m. Should the driver be charged for speeding? Assume the coefficient of kinetic friction between the car and the road is 0.80 (see Chapter 3 for discussion of the equation for coefficient of friction).

■ velocity of car before skidding, v_0 = ? final velocity of car, v = 0 m s^{-1} ■ length of skid marks, $x - x_0$ = 8.4 m ■ mass of car, m = 1200 kg ■ coefficient of kinetic friction, μ_k = 0.80 The force of friction during stopping is given by $F_{fr} = \mu_k F_N$ where F_N is the normal contact force of the road on the car (see Chapter 3). i.e. $F_{fr} = \mu_k mg$ By the work–energy principle work done by friction in stopping car = change in kinetic energy of the car Therefore, −Ffr × (x − x0) = 12 mv2 − 12 mv02 Note: the minus sign on the left is for the fact that if (x − x0) is defined as positive, the force of friction must be designated negative, since friction opposes motion. Therefore, −μkmg × (x − x0) = 12 mv2 − 12 mv02 Or, −μkg × (x − x0) = 12 v2 − 12 v02 Note that the m's cancel, so the mass of the car does not matter! Substituting: −0.80 × 9.8 m s−2 × 8.4 m = 0 − 0.5v02 Therefore v_0^2 ≈ 131.7 m s^{-2} × m Or v_0 ≈ 11.5 m s^{-1} This speed needs to be converted back into km/h. Since 1 m = 10^{-3} km and 1 s = (1/3600) h, we get v_0 ≈ 11.5 m s^{-1} × 10^{-3} km ÷ (1/3600) h ≈ 41 km/h Should this driver should be charged for speeding in the 40 km/h speed limit zone? Can the driver be successfully convicted of the driving offence if charged? ■ ITQ 2 What law was being applied to arrive at the last statement in Worked example 4.2 concerning the force of the ball on the hand? Energy stored due to of electrons around nuclei of atoms (this form of potential energy is also called 'chemical' potential energy) ■ electric fields (within capacitors, for example) ■ arrangement of charges within electric fields ■ arrangements Nuclear Energy stored within the nucleus of the atom and is released when nuclear mass is converted into various forms of energy.

In the section on kinetic energy, we showed that an object gains kinetic energy on account of work done on it. Work done on an object, or on a system of objects, can also result in a gain in potential energy. We consider two types of potential energy in the next sections: gravitational and elastic. Gravitational potential energy gravitational field lines Figure 4.4 The Earth's gravitational field. Figure 4.4 shows the direction of the Earth's gravitational field. The direction of this field is defined as the direction a small mass would move under the influence of gravity if placed in the field. The field is radial and its strength weakens with increasing distance from the Earth. The gravitational field strength, g, at any point in the field is defined as the force of gravity per unit mass at that point in the field.
The force of gravity on a mass is called its weight, W.
Hence W g= m or W = mg ....(4.7) 55 56 Unit 1 Module 1 Mechanics For distances close to the Earth's surface, we can assume that g does not vary noticeably with distance from the surface. The average value of gravitational field strength, g, close to the Earth's surface is 9.8 N kg−1. The value of g varies slightly with latitude on the Earth.
The magnitude of g in N kg−1 is the same as the magnitude of g, the acceleration due to gravity, whose unit is m s−2. Getting it right! Although the value of g is the same in N kg−1 and m s−2, gravitational field strength and acceleration due to gravity are two different concepts. Hence, strictly speaking, it is incorrect to say that the weight of a stationary object is its mass times the acceleration due to gravity – rather the weight is the mass times the gravitational field strength. To raise a mass, m, at constant speed, from a height h1 to a height h2 close to the Earth's surface, an agent must apply a constant force, F, equal but opposite to the weight, mg, so that the mass is raised without accelerating (Figure 4.5). The work done by the agent is given by WA = F(h2 − h1) mass and the Earth, this energy is called gravitational potential energy. Equation 4.8 tells us that the change in gravitational potential energy, ΔEG, as a mass is raised through a distance, Δh = h2 − h1, in a gravitational field is given by ΔEG = mgΔh ....(4.9) If h1 = 0 (i.e. the mass was originally resting on the Earth's surface) and h2 = h, then ΔEG = mgh..............(4.10) We commonly refer to mgh as the potential energy of the mass in the Earth's gravitational field. Thus, the potential energy, PE, due to the mass at a height h close to the surface of the Earth is given by PE = mgh..............(4.11) Getting it right! Strictly speaking, mgh represents potential energy gained by the system of two masses, i.e. by the mass m and the Earth, which underwent separation by a force F. We should not strictly be referring to the potential energy of a single mass but of a system of more than one mass. i.e., WA = mg(h2 − h1) or WA = mgh2 − mgh1..............(4.8) The difference shown on the right of equation 4.8 represents energy, since work is an energy conversion process. The separated mass–Earth system can do work if allowed.
Hence, energy is stored in the arrangement; this energy is called potential energy. Since the work was done against gravity to achieve the separation between the Elastic potential energy (strain energy) An object is said to be elastic if it returns to its original shape and size after a force that deforms it is removed. Robert Hooke (1635–1703) investigated the elasticity of wires by hanging weights on them. He used weights to provide the stretching force (F = mg) is directly proportional to the resulting extension, ΔL, provided that the proportional limit (PL) was not reached (Figure 4.6a). This last statement is called Hooke's law, and can be expressed mathematically as F ∝ ΔL i.e. F = kΔL..............(4.12) F h2 mg h1 Figure 4.5 Raising a mass against gravity close to the Earth's surface.
The constant k is called the stiffness constant of the object. If the object is a spring, k is called the spring constant of the spring. Rearranging equation 4.12 to make k the subject, we find that the S.I. unit for k is N m−1. Just beyond the proportional limit is the elastic limit (EL). If the wire passes the elastic limit, it remains deformed, even after the deforming force is removed. The amount of the deformation is called the permanent set of the material. Since the proportional limit and the elastic Chapter 4 Work, energy and power ... the elastic limit. Beyond the elastic limit, a small increase in the stretching force causes the wire to elongate by large amounts, i.e., to 'yield'. The wire then expands irregularly with increasing applied force; passes its ultimate strength (US) (the maximum force it can withstand) and then breaks suddenly. Figure 4.6a shows a graph of force, F, versus extension, ΔL, for a wire. Experiments show that the value of the stiffness constant, k, depends both on the thickness of the wire as well as the material it is made of. The work done in stretching a wire elastically by a force, F that varies from 0 to F during the stretching process is equal to the average force multiplied by the extension produced.
Thus, (0 + F) ΔL WE = 2 1 i.e. WE = 2 FΔL..............(4.13) Since F = kΔL (assuming the proportional limit has not been exceeded), then equation 4.13 becomes 1 WE = 2 kΔL2 ....(4.14) The work done in stretching the wire is equal to the energy stored in the wire. The energy stored in the wire is called elastic potential energy (or strain energy, ITQ 3 How can the stiffness constant, k, of the wire be obtained from the graphs in Figures 4.6a and 4.6b? F/N Figure 4.6 (a) Force, F, versus extension, ΔL, for a wire. (b) Extension, ΔL, versus stretching force, F, for the same wire. since the wire is under strain). Hence, the elastic potential energy stored in the wire that is stretched a distance ΔL is given by 1 PEE = 2 kΔL2..............(4.15) Comparing equation 4.13 with the F versus ΔL graph in 1 Figure 4.6a, we see that 2 FΔL is the area of the shaded triangle of base ΔL and height F. Thus, the work done in stretching a wire a distance ΔL is equal to the area under the corresponding F versus ΔL graph. This area therefore also represents the potential energy stored in the wire. Young's modulus and strain energy Tensile stress on a wire is defined as the force applied per unit cross-section area of the wire. Tensile strain on a wire is defined as the extension per unit length produced in the wire on account of the tensile stress. If a force, F, results in an extension, ΔL, of a wire of original length L and cross-sectional area A, we obtain the following equations for tensile stress and tensile strain: F tensile stress = ...(4.16) A ΔL tensile strain = ...(4.17) L ITQ 5 What does the slope of a PEE versus ΔL2 graph represent? ITQ 4 Referring to Figure 4.6a, how much potential energy is stored in the wire if it is stretched 0.20 cm? What are the S.I. units for (a) tensile stress (b) tensile strain? 57 Unit 1 Module 1 Mechanics Discussion example 4.4: Energy stored in a compressed spring Q A (a) How much energy is stored in a spring of spring constant 200 N m−1 that is compressed a distance 5.0 cm? (b) How much force is needed to hold the spring in its compressed state? Note: the same equations involving force, extension and potential energy for a wire apply for compression of a spring as well, so long as Hooke's law as obeyed by the compressed spring. Information needed for calculation purposes: (a) energy stored, PE = ? spring constant, k = 200 N m−1 compression, ΔL = 5.0 cm = 0.050 m PE = 12 kΔL2 = 12 × 200 N m−1 × 0.0502 m2 = 0.25 J (b) force required, F = ? energy stored = average force × compression Therefore, PEE = 12 FΔL 2PEE ΔL 2 × 0.25 J = 0.050 m = 10 N Alternatively, F = kΔL = 200 N m−1 × 0.050 m = 10 N Therefore, F = Figure 4.7 shows a graph of tensile stress versus tensile strain for a wire. It has a similar shape to that in Figure 4.6a since A and L are constants for any given wire. For the linear region, stress is directly proportional to strain.
Hence, in this region, stress = E × strain where E is a constant. The equation for E is therefore given by F stress strain ....(4.18) E = F/A ΔL/L..............(4.19) or EAΔL L 5 4 3 2 1 0 0 1 2 3 4 Strain / x 10 −3 Figure 4.7 Graph of tensile stress versus tensile strain for a wire. of the wire. E is the called the stiffness constant of the material of which the object is made and is referred to as its elastic modulus or its Young's modulus. The Young's modulus of a material is defined as tensile stress divided by tensile strain, providing the material does not exceed its elastic limit. Work done in stretching a wire is given by 1 WE = 2 FΔL..............(4.13) Substituting equation 4.20 into equation 4.13 gives 1 EAΔL WE = 2 L 2..............(4.21) Equation 4.13 enables us to derive an equation for the strain energy stored per unit volume (WE) in a wire V stretched within the elastic region. Since the volume, V, of 1 the wire = AL, and WE = 2 FΔL, then WE = V i.e. 1 2 FΔL AL WE 12 F ΔL = V A L energy stored per unit volume = 2 stress × strain..............(4.22) The energy per unit volume stored in the linear region of the graph in Figure 4.7 is therefore equal to the area under the line. This is so because the area of a triangle 1 (the shaded area) is 2 base × height. In this case, base is the strain and stress is the height. So the area of the 1 triangle is 2 stress × strain, which is, in fact, the energy stored per unit volume. ITQ 7 What feature of the stress versus strain graph (Figure 4.7) represents the Young's modulus? Rearranging, we get, F= Stress / x 10 8 Pa 58 ....(4.20) Investigations show that the constant, E, does not depend on the dimensions of the wire, but only on the material ITQ 6 Using equations 4.20 and 4.21 with the Hooke's Law equation, F = kΔL, Express k in terms of the constants in these two equations. Chapter 4 Work, energy and power Materials under strain store strain energy. It is thought that some tsunamis result from underwater plates of the Earth pushing against and bending each other until they snap, releasing tremendous amounts of strain energy. When work is done against friction, an energy conversion from one form into heat takes place. This is shown by the fact that the bottom of the box feels warmer immediately at the end of the push than at the beginning of the push.
Conservation of energy Work done against gravity Non-conservative and conservative forces Work done against friction Now suppose we move a box of mass m at constant speed from ground level on to a platform, h meters above (Figure 4.9). The vertical force, F, required in Figure 4.9a must be equal in magnitude to the weight of the box, mg. The work done, W, against gravity, by this force is therefore given by W = F × h = mg × h A = mgh B Figure 4.8 Moving a box via two possible paths to a door. Consider pushing a box along the floor at a constant speed in one of two paths, A or B, toward a door. The length of path A is 5 m and of path B is 7 m. A constant force of 40 N is being applied (Figure 4.8). The work done by the mover against friction between the box and the floor is the applied force times the distance the box moves in the direction of the force. The box could be pushed along a smooth plank of length h to the same level, h. If the plank is inclined at an angle θ to the horizontal ground, a force, F1, equal to the component of mg along the plane (i.e. equal to mg sin θ) is required to push the box up the plane at constant speed. (Refer to Chapter 1 as well as Appendix 3 for a treatment of components of vectors.) Work done, W1, in pushing the box up the smooth plank to level, h, is therefore given by W1 = F1 × d = mg sin θ × d along path A: work done = 40 N × 5 m = 200 J along path B: work done = 40 N × 7 m = 280 N m Thus we see that the work done against friction in moving the box between the starting and ending positions depends on the path taken. We call such a force a non-conservative force. Hence, friction is non-conservative force. h d sin θ = h Therefore, W1 = mgh (same as before) If we had decided to move the box first along the horizontal direction AB, on a smooth, horizontal plank, and then lift the box vertically, the work done, against gravity is still mgh. AC = dB F C F1 mg F1 h h mg A B θ ITQ 9 Why would the work done when moving the box horizontally and then vertically also equal to mgh? 59 60 Unit 1 Module 1 Mechanics Discussion example 4.5: Spring-loaded toy foam ball gun Q A A spring of spring constant k is compressed a distance d within the barrel of a toy foam ball gun, in which there is a foam ball of mass m. When the trigger is pulled, would the speed of the ball leaving the gun be the same if the gun were pointed horizontally or vertically (Figure 4.10)? a of a foam ball of ball foam ball v v bb b v1 v1 v1 v d d Hence, work done against gravity between two levels is the same, irrespective of the path taken (e.g. A and C in Figure 4.9b) does not depend on the path taken. When work done between two points of a force does not depend on the path taken, we call such a force a conservative force. Gravity is therefore a conservative force. Unlike doing work against friction, which causes the production of one form of energy into heat energy, which usually dissipates into the environment, when work is done against gravity, an energy conversion into potential energy (stored energy) results. Conservation of mechanical energy Figure 4.10 A foam ball gun aimed in two ways: (a) horizontally, (b) vertically. A ■ Ball leaving gun horizontally Applying conservation of mechanical energy, we get mechanical energy before triggering = mechanical energy after triggering PE1 + KE1 = PE2 + KE2 Therefore, 12 kd 2 + 0 = 0 + 12 mv2 Therefore, speed of ball on leaving gun, kd 2 v = m ■ Ball leaving gun vertically Again applying conservation of mechanical energy, we get mechanical energy before triggering = mechanical energy after triggering PE1 + KE1 = PE2 + KE2 However, this situation involves first raising the ball a height d (i.e. giving the ball some gravitational potential energy) and then the ball leaves the gun with a speed v1. We therefore get, 12 kd2 + 0 = mgd + 12 mv12 Therefore speed on leaving the gun, v1 = 2(12 kd2 − mgd) m kd2 − 2mgd = m Hence, the ball leaves the gun with a lower speed if the gun is pointed upward than if it were pointed horizontally. v1 =
Potential energy and kinetic energy are called mechanical energy. Experiments have confirmed that, so long as only conservative forces are involved, the total mechanical energy in a system remains constant. This statement is referred to as the principle of conservation of mechanical energy. In equation form, PE1 + KE1 = PE2 + KE2..............(4.23) Here the subscript '1' indicates an initial state and the subscript '2' indicates a subsequent state. PE and KE represent potential energy and kinetic energy, respectively. The law of conservation of energy As every child who has been on a slide would know, their bottoms gets hot as they slide down. Heat energy is generated through friction between their bottoms and the slide. This brings us to a more general statement on conservation of energy called the law of conservation of energy. In an isolated system, the total energy remains constant. Energy is neither created nor destroyed, but may be converted from one form to another. In equation form: PE1 + KE1 + WNC = PE2 + KE2..............(4.25) where WNC is work done by non-conservative forces. Work, energy and power Discussion example 4.6: Child on a slide Discussion example 4.6: Child on a slide, revisited Q A child of mass m slides from rest at the top of a frictionless inclined plane (Figure 4.11). With what speed does the child leave the slide (where height h2 = 0)? Suppose there was a coefficient of kinetic friction, μk, between the child and the slide (Figure 4.12). With that speed would the child have reached the bottom of the slide? friction h1 h1 mg mg θ h2 Figure 4.11 Child sliding down a 'frictionless' slide. A Since friction is negligible, and assuming there are no other non-conservative forces involved, we can apply the conservation of mechanical energy. PE1 + KE1 = PE2 + KE2 mgh1 + 12 mv12 = mgh2 + 12 mv22 mgh1 + 0 = 0 + 12 mv22 v2 = 2gh1 v2 = √2gh1 ....(4.24) Note that the mass of the child does not matter. Rearranging equation 4.25, we get, WNC = PE2 − PE1 + KE2 − KE1 i.e. WNC = ΔPE + ΔKE..............(4.27) Equation 4.27 is another way of expressing mathematically the law of conservation of energy. This equation says that the change in potential energy plus the change in kinetic energy in a system is equal to the work done by the non-conservative forces in the system. Power Power is the rate of doing work or converting energy. Thus, average power, P, involving an energy conversion, E, in a time t is given by P= E t ....(4.28) Power is measured in joules per second (J s−1). ITQ 10 Using the delta (Δ) notation, re-write equation 4.23 to illustrate conservation of mechanical energy. h2 θ Figure 4.12 Child sliding down a plane where there is friction. A With reference to Figure 4.12, the magnitude of the force of friction, Ffrk, is given by Ffrk = μkFN..............(3.12) where FN is the normal contact force between the child and the slide.
But, FN = mg cos θ (since mg cos θ is the perpendicular component of mg to the plane – refer to Chapter 1 as well as Appendix 3 for a treatment of components of vectors.) Therefore, Ffrk = μkmg cos θ The work done on the child by friction, Ffrk, along the length of the plane, is negative, since Ffrk is opposite to the direction of motion the length, d, of the plane. Equation 4.25 becomes, mgh1 + 12 mv12 − (μkmg cos θ)d = mgh2 + 12 mv22 mgh1 + 0 − (μkmg cos θ)d = 0 + 12 mv22 i.e. v2 = 2gh1 − 2μkg cos θd Notice again, that the mass of the child does not matter, since the m cancels. Again, the mass of the child does not matter. v2 = √2gh1 − 2μkg cos θd ....(4.26) Comparing equations 4.24 and 4.26, we see that the speed of the child at the bottom of the plane is reduced on account of friction between the child and the slide. One J s−1 in the S.I. is called one watt (W). A power station converts and supplies energy at a particular rate, usually in megawatts (MW). Since P = Et and E = F × d (where F is the applied force and d the distance moved in the direction of the force), then Fd P= t Or, P= Fv ....(4.29) where v = t is the speed of the object to which the force is being applied. 61 62 Unit 1 Module 1 Mechanics Worked example 4.8: Operating power requirements of a car Q A (a) How much power is required by a car to keep it going straight at a steady 60 km/h on a level road if the total opposing frictional force is 800 N? (b) What must the engine power be if the motor is to be running at 80% efficiency? Power required by car to overcome friction (a) P = Fv = 800 N × 60 × 1000 m h−1 × 1 h 3600 s = 13 000 W = 13 kW (b) Since efficiency is 80%, power supplied to car 80% P = 100% × 13 kW = 16 kW ≈ Discussion example 4.9: Incandescent and fluorescent lamps Q A 15 W compact fluorescent lamp (CFL) is claimed by a manufacturer to deliver the same light output (intensity) as a 40 W incandescent (filament) lamp. How do the efficiencies of the CFL and the incandescent lamp compare? A assume the useful light output power in each case = P. Efficiency of incandescent lamp = Ei Efficiency of CFL = Ec useful light power output, P Ec = × 100% P × 100% power input P = × 100% 40 useful light power output, P 15 W power input Therefore, 15 E P Therefore, c = Ei 15 40 40 = 15 = 2.7 Ei 15 the CFL is 2.7 times as efficient as the incandescent lamp. 1 , or roughly 40% of. This means that the CFL needs only 2.7 power that the incandescent lamp needs to produce the same amount of lighting.
ITQ 11 Another type of lamp has been developed that is even more efficient than the CFL. What is it? Efficiency The efficiency of an energy conversion is defined as efficiency energy converted to useful form × 100% (Eff) total energy input ....(4.30) E = Euseful × 100% input Since energy = power, P, × time, t, then we may define efficiency in terms of power as follows: ×t E Eff = Puseful × t × 100% input P Thus, Eff = Puseful × 100% Pinput ....(4.31) Sources of energy Energy is not only the ability to do mechanical work but also to make things happen in the physical world. Thus, energy can cause a chemical change, such as takes place in a chemical reaction, or even a biochemical change, such as occurs in photosynthesis or respiration. Energy for living and non-living usage In the biotic (living) world, energy is vital for survival. Energy in the form of stored in food is needed for living organisms, including ourselves, to survive. This fact is often overlooked, especially in Physics courses, where energy is often thought of in terms of inanimate use only.
In the abiotic (non-living) world, energy is needed to operate devices such as cars, cookers, lamps and a whole host of others, so typical of modern society.
However, energy is an essential requirement in both worlds! Sometimes, a balancing act has to be decided on within a country when making choices concerning developing energy for inanimate use and energy for living things. In addition to getting things done, energy usage must not have too deleterious an effect so as to degrade either the living or the non-living environment. In other words, energy use must be sustainable to the environment, biotic and abiotic, both in terms of availability and quality. Renewable and non-renewable energy Energy sources fall into two broad categories – renewable and non-renewable. A renewable energy source is one that replenishes itself during an ordinary lifetime. Examples of renewable energy sources are solar, wind, wave, ocean-thermal and biomass. Some view nuclear fusion involving heavy hydrogen as a potentially renewable energy source since there is, for all practical Chapter 4 Work, energy and power purposes, an unlimited supply of heavy hydrogen in our oceans and this hydrogen is recycled through the natural water cycles occurring on the Earth. A non-renewable energy source is not replenished in one's lifetime. For example, fossil fuels such as oil, natural gas and coal took millions of years to form; these fuels will therefore not be replenished in our lifetime.
Fissile nuclear material, such as uranium, is another type of non-renewable energy source (nuclear energy is discussed in Chapter 27). However, efforts have been made in 'breeder' reactors to convert products of nuclear fission into fissionable material, such as plutonium, since naturally occurring fissile material is not believed to be in abundance.
Geothermal energy is also thought by some to be non-renewable.
Most of the energy utilized in the Caribbean and the rest of the world today comes from fossil fuels. Because of several factors, such as increase in population worldwide, and hence increase in demand for energy, and the pace of 'development', where people are wanting more and more modern amenities, the quantity of fossil fuel reserves is depleting rapidly. According to some estimates, at the rate at which we are consuming energy, fossil fuel oil is expected to last only 100 years or so. Further, the use of fossil fuels are implicated in pollution and global warming. The present thinking is therefore to make more use of the renewable and 'cleaner' energy sources and to be more efficient in the use of energy (i.e. to 'waste' less energy in getting work done). With these factors in mind, we outline here some possible prospects for energy, especially for the Caribbean. Sources of energy within and outside the Caribbean The student is advised to use internet search engines to obtain fairly up-to-date information on these energy sources and the issues (local, regional and international) arising out of developing them. Here, we will list a few countries and pose some questions concerning prospects for development of energy sources nationally. Guyana Guyana has many mountains. Are there natural, dependable, waterfalls which can be made to generate electrical energy continuously? Would there be a need to empolder large areas so as to form storage for water to be used in hydropower? Would this stored water act as a haven for malaria-carrying mosquitoes? Would people or animals have to be displaced so as to accomplish the empoldering? Guyana is also mostly forest. Can the forest be used as a source of biomass energy or should it be preserved? The Iwokrama Rain Forest Reserve in Guyana has received international recognition.
Is the international community obligated to give Guyana something in return for preserving its forest? In 2014, China was given access to logging in Guyana. Is this a good thing for Guyana, for China, or for the world? How can adverse environmental effects due to the cutting down of large numbers of trees be reduced? Incidentally, in China and India, cook stoves using wood and wood products as fuel were developed to replace the more traditional 'fireside' which burns wood. It is claimed that these stoves have helped to reduce both pollution and respiratory ailments, but also have reduced the need for importation of fossil fuel for cooking purposes. Guyana is close to the equator. Guyana has tried using solar-voltaic energy on a small scale to power a hospital at the hinterland location of Orealla, and also in the capital, Georgetown, to power traffic lights. How successful are these projects? Should there be larger scale use of photo-voltaics, thus making use of solar energy directly? What about disposal problems when the life of these photo-voltaic cells come to an end (photo-voltaics contain chemicals that are very toxic to the environment)? Trinidad and Tobago Trinidad and Tobago has reserves of fossil fuels. Natural gas can be seen bubbling in streamlets present on the surface of its famous Pitch Lake in La Brea. Natural gas burns much more cleanly than oil (less pollution) – but how much natural gas potential does this twin-island republic have? Like Guyana, Trinidad and Tobago is close to the equator. Should there be an increase in direct use of solar energy, for example, using photo-voltaics (solar-electric cells), at this stage of the country's development?
Should the country wait until its gas reserves are nearly exhausted? The rich, volcanic soil of Trinidad and Tobago lends itself to food production. For every square metre of soil devoted to direct solar energy capture, e.g. by photo-voltaics, this means one square metre less for food production. How should Trinidadians and Tobagonians do a balancing act here? Should they make more use of solar energy (in Guyana, both fish and paddi have been successfully dried by the Sun). Should crop residues be considered as a significant possible source of energy? 63 64 Unit 1 Module 1 Mechanics Figure 4.14 Wind turbines similar to these are being tried out in Jamaica.
The volcanic soil in Jamaica is rich. Should developments in food energy be accelerated there?
Figure 4.13 Darrieus rotor, wind-powered generators. A 120 kW Darrieus rotor, wind-powered generator has been tried in Antigua. Barbados Barbados is a little further north of the equator than either Trinidad and Tobago or Guyana. Yet it is in a tropical zone. Barbadians make much use of solar thermal energy (direct heating by the Sun). Solar thermal panels are placed on roofs to capture the heat of the Sun for heating water for household and hotel use. Barbados, like Guyana and Trinidad and Tobago, has had a history of growing sugar cane to produce food (sugar) and alcohol. Should this industry, as in Brazil, be expanded to produce fuel alcohol to power vehicles? What are the possibilities? What are some constraints, given the rather small size of this island? Should Barbados at this point be increasing its local food production and thus decreasing its food energy import bill? The Windward & Leeward Islands Like Trinidad and Tobago, there is rich volcanic soil in these islands. But many of the islands are small and their populations are increasing. Should land area be used for direct capture of solar energy, or for growing more food energy? Should land be used to produce biomass energy such as alcohol or combustible wood products? What are the developments and possibilities of wind power in these islands (Figure 4.13)?
What are the drawbacks? Jamaica Jamaica has been embarking on development of wind power (Figure 4.14). Are there good prospects for wave power? Does the mountainous topography of Jamaica open possibilities for hydropower?
Is there swift running water whose kinetic energy can be converted into electrical energy? If so, how much? United Kingdom and the United States The UK and the USA have embarked, especially in recent times, on a multi-pronged approach to deal with their energy supplies. The UK, although oil-producing, has invested in renewable energies (especially wind) and in nuclear energy. The UK has also worked on improvements in efficiency in electrical power generation, distribution and consumption. In their combined heat and power (CHP) approach, they use the 'waste' heat from power stations to provide hot water to towns. Laws have been passed to limit the amount of pollution a power-generating system can emit, and incentives have been offered as well. Efficiency and (polluting) emission standards have been set in the manufacture of vehicles and appliances.
A similar approach has been taken by the USA. The UK, although having only about half the solar intensity as the Caribbean, has also invested heavily in solar-voltaic systems.
Is there value in Caribbean territories adopting such a multi-pronged approach to deal with their national energy supplies? Electrical power generation, distribution and consumption By far, the most widely used source of power today is electrical. There are many reasons for this. ■ Electrical energy is easily generated mechanically (using magnets rotating relative to coils of wire) or electrochemically (as in batteries and photo-voltaic cells). Chapter 4 Work, energy and power ■ Electrical energy is easily inter-convertible. A lamp, for example, easily converts electrical energy to light. A microphone easily converts sound into electrical energy.
A battery easily stores electrical energy by chemical means. ■ Electrical energy can travel very fast (almost at the speed of light along wires), at the flick of a switch. ■ The use of wires enables electrical power to reach remote areas easily. One approach to conserving energy, especially from fossil fuels and other non-renewable energy resources (e.g. nuclear resources), is by improvements in efficient energy usage.
Efficiencies in the usage of electrical energy should be continually improved in areas such as: ■ generation – less energy wasted (mainly as heat) during the generation process ■ distribution – less energy wasted as heat as the electrical energy is sent along wires ■ consumption – appliances used by consumers need to be made more efficient, so that less energy is needed to accomplish the same amount of work; consumption habits and lifestyles must also improve, such as turning off electrical appliances and equipment (e.g. lights and air-conditioning) when they are not needed to function.
Due care must also be taken that the effects on the environment associated with energy usage are, at the least, not very harmful or are beneficial at most. Thus, sustainability will be achieved, not only in terms of the amount of energy available but also in preserving the quality of life and the environment. Nobel Prize for energy-saving invention The Nobel Prize in Physics for 2014 was awarded jointly to Isamu Akasaki, Hiroshi Amano and Shuji Nakamura who invented the efficient and environmentally friendly LED (light-emitting diode). As about one-fourth of the world's electricity consumption is used for lighting purposes, this very energy-efficient invention will significantly contribute to saving the Earth's fossil fuel and other energy resources. LEDs are discussed in more detail on page 559. 65 66 Unit 1 Module 1 Mechanics Summary ✓ Work, W, is defined as the product of force, F, and distance, x, moved in the direction of force. If θ is the angle between the direction of the force applied to an object and the direction the object is moved, W = Fx cos θ..............(4.1) ✓ Energy is the ability to do work. ✓ According to the work–energy principle, the net work, W, done on a mass, m, is equal to its change in kinetic energy: W = 12 mv2 − 12 mv02..............(4.5) ✓ The change in gravitational potential energy of a mass raised a height, Δh, above, but close to, the Earth, is given by ΔEG = mgΔh..............(4.9) ✓ Work done in stretching a linear object of stiffness constant k and extension ΔL by a force F (provided the elastic limit is not reached), are given by: 1 WE = 2 FΔL..............(4.13) WE = 2 kΔL2..............(4.14) ✓ Through the work done in stretching the object, elastic potential energy (or strain energy) is stored in the object.
The expression for this elastic potential energy is the same as the righthand side of equations 4.13 and 4.14, providing the object does not reach its elastic limit. ✓ Tensile stress and tensile strain are defined by the equations: tensile stress = F..............(4.16) A ....(4.17) tensile strain = ΔL L ✓ The Young's modulus of elasticity, E, of a material is defined as: E = stress ....(4.18) strain providing that the material has not reached its elastic limit.
✓ Potential energy and kinetic energy are called mechanical energy. Experiments have confirmed that as long as only conservative forces are involved, the total mechanical energy in a system remains constant. This is a statement of conservation of mechanical energy. In equation form: PE1 + KE1 = PE2 + KE2..............(4.23) ✓ The law of conservation of energy can take many forms. However, in any isolated system, the total energy remains constant. Energy is neither created nor destroyed, but may be converted from one form to another (the law of conservation of energy). In equation form: PE1 + KE1 + WNC = PE2 + KE2..............(4.25) where WNC is work done by non-conservative forces. ✓ Power is the rate of doing work or converting energy. If E, converted in a time t the average power, P, is given by P = E t ....(4.28) ✓ If a force, F, that causes a body to move a distance d in the direction of the force in a time t, average power is given by: P = Fd t or P = Fv..............(4.29) where v is the average velocity of the body during the energy conversion. ✓ The efficiency of an energy conversion is defined as efficiency energy converted to useful form = × 100% (Eff) total energy input ....(4.30) Chapter 4 Work, energy and power Review questions Work and energy; Kinetic energy; Potential energy 1 (a) Define: (i) work (ii) energy (b) How much work is done in lifting a 30 kg box from a point A on the ground to a point B on a platform 1.5 m directly above A? (c) Show that the same amount of work would be done if the box were carried from ground level by sliding the box along a smooth 3.0 m long plank whose one end rests on the ground away from A and the other end rests on point B. 2 (a) State the work–energy principle. (b) Derive the work–energy principle equation for a mass m moving initially with velocity v0 and ending up with a velocity v. State any assumptions made. (c) Using the work–energy principle equation from part (b) above, show that the kinetic energy, EK, of a body of mass m and travelling at a speed v is given by the formula: 1 EK = 2 mv2 3 (a) Derive the formula for elastic potential energy, PEE, in terms of applied force F, stiffness constant k, and extension ΔL of a wire. State any assumptions made. (b) Figure 4.6a shows a F versus ΔL graph for a wire of length 2.0 m and diameter 1.2 mm. (i) Determine the stiffness constant of the wire from the graph. (ii) Determine the Young's modulus for the material of the wire. (iii) What is a practical advantage of knowing the Young's modulus as compared with the stiffness constant of the wire?
4 Figure 4.7 is a graph of tensile stress versus tensile strain for a wire 1.40 m long and 0.36 mm in diameter.
(a) What feature of the graph represents strain energy per unit volume? (b) Using the graph, determine the strain energy stored in the wire when it has been stretched 3.0 mm. Conservation of energy 5 A skier, starting from rest, skis down a slope from a height of 300 m to the horizontal ground level. (a) Show that the speed with which the skier reaches the ground level does not depend on either the mass of the skier or the angle of incline of the slope. State any assumptions made.
(b) Calculate the speed with which the skier reaches the ground level. 6 Figure 4.15 shows a roller-coaster. When it is at the highest point, H, it moves with a speed of 0.25 m s−1. The mass of the roller-coaster and its occupants is 600 kg.
H 10 m P Figure 4.15 67 68 Unit 1 Module 1 Mechanics (a) Assuming no mechanical power is being supplied to the roller-coaster at H, estimate: (i) the gravitational potential energy converted during the ride from H to a point P a vertical distance of 10 m below the level of H (ii) the speed of the roller-coaster at point P. (b) In practice, the speed at P is less than the calculated value.
Explain this. 7 A child of mass 18 kg slides from rest from a vertical height h1 = 3.8 m at the top of a smooth inclined plane (see Figure 4.11). With what speed does the child reach the ground (where height h2 = 0)? 8 Two identical balls, A and B, are released from rest from the same height, 2.8 m, along two smooth inclined planes that touch the horizontal ground. Ball A is released on a plane inclined at 30° to the horizontal. The plane on which ball B is released is inclined at 60° to the horizontal. Ignore energy due to rotation of the balls. (a) What is the speed of each ball at the bottom of their respective inclined planes? (b) Which ball would reach the bottom first? Explain your answer. 9 A spring of spring constant 40 N m−1 is compressed a distance 20 cm within the barrel of a toy foam ball gun, in which there is a foam ball of mass 15 g. When the trigger is pulled, what would the speed of the ball be leaving the gun: (a) if the gun was aimed horizontally? (b) if the gun was aimed vertically upward? (Hint: see Discussion example 4.5.)
10 A box slides from rest from a height of 2.4 m down a plank inclined at 40° to the horizontal. If the coefficient of kinetic friction between the box and the plank is 0.22, (a) derive a formula for the speed with which the box reaches the ground at the point where the bottom end of the plank is resting (b) calculate the speed with which the box reaches the ground at the bottom end of the plank. 11 (a) Differentiate between the principle of conservation of mechanical energy and the law of conservation of energy. (b) A box of mass 2.0 kg slides down a plank at constant speed. If the plank is inclined at an angle of 35° to the horizontal, estimate the force of kinetic friction between the box and the plank. Power 12 A 1200 kg car ascends a road inclined at 20° to the horizontal at 25 km/h. Neglecting frictional opposition, estimate the power required to keep going at this speed. 13 A power station delivers 25 MW of electrical power. How many tonnes of fuel oil would be needed by this station to deliver the 25 MW of power for 1 month of 30 days, assuming a conversion efficiency of 35%. (1 tonne of oil equivalent in energy = 4.2 × 1010 J; 1 tonne = 1000 kg) Sources of energy within the Caribbean; Electrical power generation, distribution and consumption 14 Describe two ways EACH in which fossil fuel energy can be conserved in Caribbean homes through (a) measures involving reduced wastage of energy (b) sensible choice of equipment and appliances (c) making use of renewable energy.
15 (a) Outline problems involved in the large-scale use of photo-voltaic electricity in terms of availability of solar energy and that a large area is required to capture adequate amounts of solar energy. (b) It is planned to replace a 20 MW output fossil fuel power station with a 20 MW output solar-voltaic power station. If the average solar insolation (power incident normally per square meter on the surface of the Earth) at a given location is 350 W m−2 and the efficiency of the solar-voltaic cells is 20%, estimate the area of land surface required for generating the 20 MW of electrical power using solar-voltaic panels. 16 Discuss three reasons why the Caribbean should be making more use of renewable energy today.
17 Should the Caribbean be spending more on importation of food or importation of fuel? Justify your point of view. Practical exercises and challenges All experiments involve some degree of danger. The practical investigations described below should be carried out only under the direct supervision of suitably qualified and experienced personnel.
1 Light bulb efficiencies Devise and carry out a method for testing a manufacturer's claim that their '15 W CFL' gives as much light as a '40 W incandescent' lamp. (Make sure you pay strict attention to control of variables.) Chapter 4 Work, energy and power 2 Energy audit Design and carry out an energy audit in a home. Attempt to quantify energy used monthly as energy from (a) food, (b) non-renewable energy (e.g. for appliances in the home and for transportation), and (c) renewable sources of energy used. As an example of a use of renewable energy, consider if the home uses the Sun's heat for drying clothes. You can estimate the energy used per wash load as equivalent to that used per load in an average drier and the equation energy = power × time (equation 4.28) can be used in estimating electrical energy. A similar energy audit could be attempted regionally or even nationally. Such audits will give comparative amounts of expenditure on food, non-renewable and renewable energy sources. The audit could even be extended to include: ■ importation figures ■ estimates of energy resources present nationally and possible duration ■ estimates of energy resources that are currently in use. 3 Solar insolation (power per square metre) Devise and carry out an experiment to determine the solar power per square metre incident normally at your location. (An apparatus similar to an insulated solar water-heater with a glass heat-trap can be used. The energy trapped could be estimated using the equation: heat energy absorbed = mass × specific heat capacity × temperature change (see Chapter 12 Thermal properties of matter.) Other ideas and equipment can be obtained through an internet search. Answers to ITQs 1 Yes – but only a small amount, since the Earth's orbit is almost circular. The slight tangential component of gravity will do work on the Earth as it moves along its elliptical path, causing a slight changing of its speed (and therefore kinetic energy). The distance between the Sun and the Earth will also be changing (and therefore altering the potential energy of the system). However, the total energy of the system at any point will be constant, on account of conservation of energy. 2 Newton's third law of motion. (The force of A on B is equal and opposite to the force of B on A.) 3 (a) k = slope, since F = kΔL and F is plotted against ΔL. (b) 1/k = slope, since F = kΔL and ΔL is plotted against F. 1 1 4 PE = 2 FΔL = 2 × 500 × 0.002 = 0.50 J 1 1 5 2 k (since PE = 2 kΔL2) 6 (a) stress (N m−2 or pascals, Pa) (b) strain (no units – just a ratio) 7 The slope of the graph, since stress = E × strain 8 k = EA L 9 Work done against gravity along AB = 0 N m. Then work done in lifting the box vertically from B = mgh. Therefore total work = mgh.
10 ΔPE + ΔKE = 0 for conservation of mechanical energy (i.e. only conservative forces are involved). 11 The LED (light-emitting diode) lamp. Answers to Review questions 1 (b) 440 N m (or J) 3 (b) (i) 2.5 × 105 N m−1 (ii) 4.4 × 1011 N m−2 5 (b) 77 m s−1 7 8.6 m s−1 9 (a) 7.3 m s−1 (b) 7.0 m s−1 11 (b) 11 N 13 4400 tonnes 15 (b) 2.9 × 105 m2 69 70 Chapter 5 Momentum and collisions Learning objectives ■ Define 'linear momentum'. ■ State Newton's second law of motion in terms of momentum.
■ Define 'impulse' of a force in terms of Newton's second law. ■ Draw and interpret F–t graphs. ■ Apply the concept of impulse to solve problems. ■ State the principle of conservation of momentum. ■ Apply the principle of conservation of momentum in one and two dimensions. ■ Distinguish between inelastic and perfectly elastic collisions. ■ Solve problems involving elastic and inelastic collisions. Linear momentum System in which mass is constant In 1932, James Chadwick (an English physicist, 1891–1974) achieved one of the greatest breakthroughs in understanding the structure of the atom – showing the existence of the neutron. He did this by applying conservation of linear momentum and energy to collisions involving nuclei of atoms. We will return to this breakthrough at the end of this chapter (page 79). Motion of all objects, from the tiniest particle to the largest mass, involves a quantity called linear momentum. Linear momentum is defined as the product of the mass, m, of an object and its velocity, v.
Thus, a defining equation for linear momentum, p, is given by p = mv ....(5.1) Since velocity, v, is a vector quantity and mass, m, is a scalar quantity, it follows that linear momentum, p, is a vector quantity and its direction is that of v. According to Newton's second law of motion, the net force, F, applied to an object of mass m is directly proportional to the object's rate of change of momentum. Thus if the momentum change in time Δt is Δ(mv) then the net force is given by F∝ Δ(mv) Δt ....(5.2) ITQ 1 Given that equation 5.1 defines momentum as the product of mass and velocity, what is the SI unit of momentum? For a system consisting of a constant mass, m, the relationship shown by equation 5.2 becomes F∝m Δv Δt or F = km Δv Δt where k is a constant. In Chapter 3, equations 3.2 and 3.3, we saw that the constant k was set to be equal to 1 so that 1 N of force gives a mass of 1 kg an acceleration, a, of 1 m s−2 (where a = Δv).
Δt Δv Thus, F = m ....(5.3) Δt or, F = ma..............(5.4) since a = Δv Δt For constant acceleration a during a time t, (v − v0) a= t Hence, F = mv − mv0 t ....(5.5) where v = final velocity of the object, v0 = initial velocity of the object and t = time interval during the action of the force on the object. ITQ 2 According to equation 5.6, what are two ways of producing the same change in momentum of a mass? Chapter 5 Momentum and collisions System in which mass is changing If the mass of an object is changing with time, then, since momentum is the product of mass and velocity, its momentum is also changing with time. Therefore, by equation 5.2, there must also be a net force on that object. This kind of situation is characteristic of rocket motion and is discussed in a later section on conservation of momentum. Although the product, Ft, results in a change in momentum of an object, we usually reserve the term 'impulse' for forces acting for a very short duration. Such impulsive forces can sometimes be very large, as illustrated in Worked example 5.1. Sometimes, impulsive forces are so large, they can result in unintended damage, as in Worked example 5.2 – taken from cricket. Linear momentum and impulse Rearranging equation 5.5, we get Ft = mv − mv0..............(5.6) The right-hand side of equation 5.6 represents change in momentum (final momentum − initial momentum). The product, Ft, on the left-hand side is equal to the change in momentum due to the force F acting for time t. We define this product of force and time as the impulse (J) of the force on the mass. Thus we have the result that the impulse of a force on an object equals the change in momentum of the object. Impulse, J, is defined as follows: J = Ft assuming the force is constant during the time t. Or, since 'impulsive forces' act during a very short time, Δt, J = FΔt It follows from equation 5.6 that the S.I. unit for impulse is the newton second (N s) and that 1 N s of impulse = 1 kg m s−1 of change in momentum. Worked example 5.1: Penalty shot Q Consider a penalty shot at a goal.
A soccer ball, starting from rest, leaves the player's foot with a velocity of 25 m s−1. If the mass of the ball is 0.44 kg and the ball is in contact with the player's boot for 10 ms, estimate the average force of the shoe on the ball. A ■ initial velocity of ball, v0 = 0 m s−1 ■ final velocity of ball, v = 25 m s−1 ■ time of contact, t = 10 ms = 0.010 s ■ average force of boot on ball, F = ? ■ mass of ball, m = 0.44 kg impulse imparted c hange in momentum Use equation 5.6, = by the force of the ball Thus, Ft = mv − mv0 m(v − v0) F = t 0.44 kg × 25 m s−1 = 0.010 s = 1100 N (a rather large force!) Worked example 5.2: Force of a cricket ball on a bat Q A A cricket ball of mass 156 g, moving horizontally at 140 km/h, is hit along the same line of its incidence. If the ball leaves the bat with a velocity of 120 km/h, and is in contact with the bat for 5.0 ms, estimate the average force of the ball on the bat. Getting it right! We need to be careful here. We are asked to find the force of the ball on the bat, which will change the momentum of the bat. But we do not know the mass of the bat. Since we have data concerning the ball, we will first find the force of the bat on the ball (which will produce a change in momentum of the ball). Then we will apply Newton's third law to find the force of the ball on the bat. ■ initial velocity of ball, v0 = 140 km/h 140 × 103 m = ≈ 39 m s−1 3600 s ■ final velocity of ball, v = −120 km/h −120 × 103 m ≈ −33 m s−1 = 3600 s (The final velocity is in the negative direction to the initial velocity.) ■ mass of ball, m = 156 g = 0.156 kg ■ time of contact, t = 5.0 ms = 0.0050 s ■ average force of bat on ball, F = ? Ft = mv − mv0 F= m(v − v0) t 0.156 kg × (−33 − 39) m s−1 0.0050 s = −2246 N ≈ −2250 N Hence the average force of the ball on the bat, by application of Newton's third law, is +2250 N, i.e. in the direction of the original velocity of the ball, which direction, of course, we expected.
Note that, since the maximum force of the ball on the bat will be larger than this average force, this maximum force is sometimes able to break the bat! = 71 0.3 72 Unit 1 Module 1 Mechanics Force versus time graphs crumple zone crumple zone Consider the graph of a constant 30 N force acting on an object for 0.30 s. Since the product, Ft, represents impulse, the area under the F versus t graph in Figure 5.1a is equal to the impulse due to the force F. In practical situations, the graph of impulsive force, F, against time, t, shows a somewhat smoothly changing force (Figure 5.1b) rather than an abruptly changing force (as in Figure 5.1a). Yet, by considering this curve as a sequence of impulses, each acting for a very short time during which the force is constant, we can show that the total impulse due to a varying force is always equal to the area under the F versus t graph. a Figure 5.2 The crumple zones of vehicles fold during a crash, lengthening the time taken to come to a stop. Worked example 5.3 Estimating impulse from a graph Q A b F/N F/N 40 40 30 30 20 20 10 10 0 0 0.1 0.2 0.3 0.4 0.5 t/s b 0 0 0.1 0.2 0.3 0.4 0.5 t/s 0 0 0.1 0.2 0.3 0.4 0.5 t/s Figure 5.1 Graphs of impulsive force, F, versus time, t. The impulse represented by each small square on the graph: = 2 N × 0.02 s = 0.04 N s We estimate the total area under the graph by counting squares, half or greater, as 1 square and less than half as 0 square. Here we have counted the number of squares within each of the larger blocks in Figure 5.1b. total number of tiny squares: ≈ [17 + 7 + 2] + [(3 × 25) + 21 + 11] + [17 + 7 + 2] ≈ 159 impulse ≈ 159 × 0.04 N s ≈ 6.36 N s ≈6Ns 0.2 0.3 0.4 0.5 Worked example 5.4: t / s Estimating the velocity of a hit ball Q If Figure 5.1b represents the F versus t graph for the force of a bat on a stationary cricket ball of mass 156 g, estimate the velocity with which the ball leaves the bat. A ■ F/N 40 Estimate the impulse in Figure 5.1b. mass of ball, m = 156 g = 0.156 kg ■ initial velocity of ball, v0 = 0 m s−1 ■ final velocity of ball on leaving bat, v = ? Using the graph in Figure 5.1b, ■ impulse imparted to ball, Ft ≈ 6.36 N s Since Ft = mv − mv0 And, v0 = 0 m s−1 Ft Then, = v m 6.36 N s v≈ 0.156 kg ≈ 40.8 m s−1 ≈ 41 m s−1 ITQ 3 What do we normally mean when we say that a person is acting 'impulsively'? ITQ 4 Calculate the impulse due to the force F in Figure 5.1a. ITQ 5 In the high jump at an athletics competition, what safety feature is present to protect the high jumper from large, impulsive forces?
Explain how the feature works. Chapter 5 Momentum and collisions Practical applications of the impulse concept Conservation of linear momentum To drive a nail into wood requires a very large force. A hammer head is able to deliver a large enough force by making use of impulse. The hammer head has a large mass. The carpenter gives the head a fairly large initial velocity – hence a large initial momentum. The hammer head comes to quick stop upon contact with the nail due to the friction between the nail and the wood.
According to equation 5.6, the change in momentum of the hammer head during such a short contact time results in a very large impulsive force. Cricket bats, tennis racquets and golf clubs make use of a similar type of impulsive force. Hence these types of sports equipment are made very stiff, and some have relative large masses as well. Where small impulsive forces are required, a soft or weakly elastic material is used.
For example, when a speeding car makes a sudden stop, and a passenger crashes into an airbag, the time taken for the passenger to come to a stop is lengthened, as compared to crashing into a hard metal steering wheel. Hence the impulsive force on the passenger is reduced, and so is the injury to the passenger. Seat belts have some elasticity for the same reason.
Today's cars are also made with relatively soft crumple zones at the front and the rear (Figure 5.2). These zones fold and lengthen the time taken for the car to come to a stop during a crash – resulting in less injury to the occupants who would have been travelling at the same initial speed as the car. Packing materials for delicate items such as eggs, glassware and electronic equipment are usually made of Styrofoam (polystyrene impregnated with air) so that when dropped, the items come to a lengthened stop on the polystyrene rather than an abrupt stop on a hard surface, thus reducing the force on the item. Experiments have confirmed that, as long as no external forces are acting on a system of bodies, the total momentum of the system remains constant. This statement is known as the principle of conservation of linear momentum. A few examples below will illustrate the principle. Consider the following: ■ mA = mass of object A ■ mB = mass of object B ■ vA = velocity of object A just before a collision with B ■ vA' = velocity of object A just after the collision with B ■ vB = velocity of object B just before the collision with A ■ vB' = velocity of object B just after the collision with A If we assume that there are no external forces acting on the system of objects A and B, then, according to the principle of conservation of momentum, total momentum of system before collision = total momentum of system after collision That is, mAvA + mBvB = mAvA' + mBvB'..............(5.7) Equation 5.7 is a vector equation since momentum is a vector quantity. The equation therefore holds, not only along a straight line, but also in two, and even three, dimensions. Momentum conservation in one dimension where masses do not change Worked example 5.5: A car colliding with a stationary truck Q A 1200 kg car, freely rolling along a straight horizontal road at 30 km/h, collides with a stationary truck of mass 10 000 kg. If, as a result of the collision, the truck moves off with a velocity of 5.0 km/h, find the final velocity of the car just after the collision. A It usually simplifies calculation if before-collision and after-collision diagrams are drawn with the data placed on the diagram (see Figure 5.3). Note, however, that the diagrams do not need to be artistic representations. Circular balls, labelled A and B, will work just as well. a before collision after collision b ? mA = 1200 kg mB = 10 000 kg vA = 30 km/h (8.3 m s –1) vB = 0 m s –1 vA ' = ? vB ' = 5.0 km/h (1.4 m s –1) Figure 5.3 A before-and-after diagram involving a collision between a car and a truck. Continued >> 73 74 Unit 1 Module 1 Mechanics Worked example 5.5: A car colliding with a stationary truck (continued) For this collision, 30 000 m ≈ 8.3 m s−1 ■ vA = 30 km/h = 3600 s ■ vA' = ? ■ vB = 0 m s−1 5000 m ≈ 1.4 m s−1 ■ vB' = 5.0 km/h = 3600 s ■ mA = 1200 kg ■ mB = 10 000 kg Assuming there are no external forces (e.g. friction) acting on the system, then, momentum of system before collision = momentum of system after collision mAvA + mBvB = mAvA' + mBvB' 1200 kg × 8.3 m s−1 + 0 kg m s−1 = 1200 kg × vA' + 10 000 kg × 1.4 m s−1 (9960 − 14 000) kg m s−1 = 1200 kg × vA' vA' = −3.4 m s−1 The negative sign tells us that after the collision, the car moves in a direction opposite to its original direction. Worked example 5.6: Gun and bullet Q A A bullet of mass 100 g is fired from a gun of mass 2.4 kg, originally at rest (Figure 5.4). If the bullet leaves the gun at 90 m s−1, calculate the velocity of the gun just after firing. a bullet (B) b gun (A) before firing mA = 2.4 kg mB = 100 g = 0.100 kg vA = 0 m s –1 vB = 0 m s –1 after firing vA ' = ? vB ' = 90 m s –1 Figure 5.4 A before-and-after diagram involving an internal explosion within a gun. Assuming there are no external forces acting on the system, we apply the principle of conservation of momentum.
There was an internal force of an explosion, acting within the system, that propelled the bullet forward and the gun backwards. But this is an action–reaction pair of forces that together have zero net effect on the total momentum of the system of the gun and bullet, since the forces are equal and opposite to each other (according to Newton's third law). Note: internal forces do not alter the total momentum but they can change other properties of the system – its kinetic energy for example, as in this case.
mAvA + mBvB = mAvA' + mBvB' 2.4 kg × 0 m s−1 + 0.100 kg × 0 m s−1 = 2.4 kg × vA' + 0.100 kg × 90 m s−1 −9.0 kg m s−1 = vA' 2.4 kg −1 vA' ≈ −3.8 m s The negative sign in the answer tells us that the gun has recoiled (moved in the opposite direction to the bullet). Chapter 5 Momentum and collisions a b particle of gas (Δm) rocket (M) v v + Δv v – ve Momentum conservation involving varying mass Consider a rocket and fuel of total mass M + Δm moving with velocity v at a certain instant of time (Figure 5.5a). In a short time, Δt, the rocket burns fuel and expels gas of mass Δm with a speed of ve relative to the rocket (Figure 5.5b). Relative to an observer at rest on the ground, a tiny mass of gas, Δm, is expelled with a velocity of v − ve. Worked example 5.7: Thrust on a rocket Q A rocket ejects burnt fuel at a rate of 40 kg s−1, with a speed of 3.0 × 103 m s−1 (relative to the rocket), in a direction opposite to which the rocket is travelling. Estimate the thrust on the rocket. A ■ The rocket velocity changes to v + Δv. Applying conservation of momentum to the system, we get: total momentum of the system before expulsion of gases = total momentum of the system after expulsion of gases Therefore, Figure 5.5 Rocket before-andafter diagrams, expelling a small quantity of gas. speed of expulsion of burnt fuel, ve = 3.0 × 103 m s−1 Δm = 40 kg s−1 ■ rate of ejection of burnt fuel, Δt Δm ■ thrust on rocket, F = ve Δt = 3.0 × 103 m s−1 × 40 kg s−1 = 1.2 × 105 N Discussion example 5.8: Rocket with space shuttle at vertical lift-off Q (M + Δm)v = M(v + Δv) + Δm(v − ve) from which we get Δmv = MΔv + Δmv − Δmve which gives us MΔv = Δmve Dividing both sides by Δt, we get M Δm Δv = ve Δt Δt ....(5.8) ■ The left-hand side of equation 5.8 is mass times acceleration of the rocket, in other words it is the force, F, on the rocket. ■ The right-hand side of equation 5.8 is ejection speed Figure 5.6 Rocket with space shuttle at vertical lift-off. of the gases (relative to the rocket) times the rate at which the rocket is losing mass. Hence, the force exerted on the rocket (called the thrust on the rocket) by the mass of gas being expelled is given by: F = ve Δm Δt ....(5.9) At what point does a rocket achieve vertical lift-off during launching of a space shuttle? A Vertical lift-off from the ground is achieved when the thrust provided by the expelled gases is slightly greater than the total weight of the rocket plus fuel plus payload. That is why, in a launch of a space shuttle, for example, the rockets are seen firing for a few seconds before lift-off actually takes place. 75 76 Unit 1 Module 1 Mechanics Momentum conservation in two dimensions We mentioned earlier that the momentum conservation equation (equation 5.7) is a vector equation.
Let us see how this equation works in two dimensions. Worked example 5.9: Oblique collision involving two cars Q Figure 5.7 shows the path taken by two cars on a very slippery road just before and after colliding. Data is given in the diagram. Calculate the speed and direction with which car B emerges from the collision. vA ' = 3.0 m s –1 A mA = 1200 kg θ A ' = 30˚ mB = 1000 kg A B vA = 4.0 m s –1 θB ' = ? vB = 0 m s –1 B vB ' = ? Figure 5.7 Oblique collision between two cars. A This is a two-dimensional situation. Since the total momentum is conserved, the component of the momentum in any direction must also be conserved. Along the broken line (forward direction) momentum must be conserved. Along a direction perpendicular to the forward direction, momentum must also be conserved. The velocities of each of the cars can each be expressed as a component in the forward direction and a component perpendicular to the forward direction. (See Chapter 1 and Appendix 3 for more about resolving of vectors into orthogonal components.) Assuming no external forces are acting on the system of the two cars, we can apply the principle of conservation of momentum in the forward direction and in the direction perpendicular to the forward direction.
■ Along the forward direction, we get mAvA + mBvB = mAvA' cos 30° + mBvB' cos θB' 1200 kg × 4.0 m s−1 + 0 kg m s−1 = 1200 kg × 3.0 m s−1 × cos 30° + 1000 kg × vB' cos θB' 4800 kg m s−1 = 3118 kg m s−1 + 1000 kg × vB' cos θB' or 1.682 m s−1 = vB' cos θB'..............(i) ■ Along the direction perpendicular to the forward direction, there is initially no vertical component of velocity and hence no net momentum. We get 0 = mAvA' sin 30° − mBvB' sin θB' 0 = 1200 kg × 3.00 m s−1 × sin 30° − 1000 kg × vB' sin θB' 0 = 1800 kg m s−1 − 1000 kg × vB' sin θB' or 1.800 m s−1 = vB' sin θB' ....(ii) Dividing equation (ii) by equation (i) gives 1.800 m s−1 vB' sin θB' ≈ 1.683 m s−1 vB' cos θB' 1.07 ≈ tan θB' θB' ≈ 46.9° θB' ≈ 47° Substituting, θB' = 46.9° into equation (ii) gives 1.800 m s−1 = vB' sin 46.9°, from which we get 1.800 m s−1 = vB' 0.730 That is, vB' ≈ 2.5 m s−1 The final velocity of B is therefore 2.5 m s−1 at an angle of 47° clockwise from the original direction of car A. Chapter 5 Momentum and collisions Elastic and inelastic collisions Tell-tale signs of inelastic collisions An elastic collision is one in which kinetic energy is conserved. An inelastic collision is one in which kinetic energy is not conserved. In equation form, and using our usual notation, for a system of two masses, A and B, colliding elastically we have Worked example 5.10 shows us that when objects become joined together on account of a collision, the collision is inelastic. Other tell-tale signs of an inelastic collision are: 1 2 2 mAvA + 1 2 2 mBvB = 1 2 2 mAvA' + 1 2 2 mBvB' ....(5.10) ■ objects become permanently deformed or crushed by the collision (work is done to cause the damage and heat energy is liberated) ■ sound, heat or some other form of energy is produced and leaves the system of colliding objects. Worked example 5.10: Is this collision elastic? Tell-tale signs of elastic collisions Q A 1100 kg car moving at 30 km/h collides with a stationary 100 kg motorbike. Immediately after the collision, they are joined and move together in the direction of the car. (a) Determine the common velocity after the collision. (b) Was the collision elastic? Tell-tale signs that a collision is elastic include: We first draw a diagram with the data (Figure 5.8). ■ objects do not join together or become permanently A is converted to potential energy and then back to kinetic energy, enabling the objects to move off from each other deformed after collision before collision bike (B) car (A) ■ objects tend to bounce off each other – kinetic energy ■ sound, heat or other forms of non-mechanical energy are not produced by the collision. mA = 1100 kg mB = 100 kg vA = 30 km h–1 (8.3 m s –1) vB = 0 km h –1 (0 m s –1) after collision B) g h –1 (0 m s –1) v'=? Figure 5.8 Collision between a car and a motorbike. 30 000 m ≈ 8.3 m s−1 3600 s Assuming no external forces acting on the system, we get, from the principle of the conservation of momentum: mAvA + mBvB = (mA + mB) × v' 1100 kg × 8.3 m s−1 + 0 kg m s−1 = (1100 kg + 100 kg) × v' Therefore, v' = 7.6 m s−1 (≈ 27 km h−1) (b) Kinetic energy before collision = 12 × 1100 kg × (8.3

m s−1)2 + 12 × 100 kg × (0 m s−1)2 = 37 889.5 J ≈ 38 000 J Kinetic energy after collision = 12 × 1200 kg × (7.6 m s−1)2 = 34 656 J ≈ 35 000 J Since KE before collision ≠ KE after collision, this collision is inelastic. (a) 30 km/h = It should be noted that the kinetic theory of gases (Chapter 14) assumes that molecules of a gas collide v'=? elastically with each other and with the walls of the container. In a perfectly elastic collision, the sum of the kinetic energies of the system of colliding particles before collision is equal to the sum of the kinetic energies after the collision. Head-on elastic collision equation A very simple, but useful, equation can tell us whether the head-on collision between two balls is elastic. (In a 'head-on collision' both objects move along the same straight line before and after the collision. This analysis applies only to a collision that takes place 'head-on', i.e. along a straight line.) We employ our usual notation.

By the conservation of momentum (which is true in all collisions as long as there are no external forces acting on the system): mAvA + mBvB = mAvA' + mBvB' or, mA(vA − vA') = mB(vB' − vB)...............(i) ITQ 6 When pool balls of the same mass collide and bounce away from each other, is the collision perfectly elastic? Explain your answer. 77 78 Unit 1 Module 1 Mechanics By conservation of kinetic energy in an elastic collision: 1 2 2 mAvA1 1 1 + 2 mBvB2 = 2 mAvA'2 + 2 mBvB'2...............(5.10) i.e. mAvA2 + mBvB2 = mAvA'2 + mBvB'2 mAvA2 − vA2 = mBvB'2 − vB2)...............(ii) Discussion example 5.12: Elastic collision of pool balls of equal masses, case 1 Q A Applying the difference of two squares algebra formula (see Appendix 3) to equation (ii) gives mA(vA − vA')(vA + vA') = mB(vB' − vB)(vB' + vB) This simplifies greatly by (vA + vA') = (vB' + vB) ....(5.11) Equation 5.11 is simple, yet very powerful. A close inspection of the equation tells us that in an elastic collision that takes place along a straight line: ■ the difference between the initial velocities if the colliding objects is equal to the negative of the difference of the velocities after collision, and ■ the relationship between the velocity differences is independent of the masses of the objects. Discussion example 5.11: Objects moving apart during an 'explosion' Q Two dynamics trolleys are held together by a string with a compressed spring between them.

If the string is cut, the trolleys are released and they move apart. Is this moving apart (or 'explosion') perfectly elastic? A Since the initial velocity of each trolley was zero, then the total initial kinetic energy of the system = 0 J Since both trolleys move apart after the 'explosion', then there is final kinetic energy in the system (kinetic energy is proportional to the square of speed and is therefore always positive). The initial kinetic energy of the system is not equal to the final kinetic energy of the system and therefore the 'explosion' is inelastic. ITQ 7 In a bumper to bumper collision, car A (rolling at 12 km/h) undergoes an almost perfectly elastic collision with a small truck B (rolling at 5 km/h in the same direction).

If, after the collision, truck B moves in the same original direction at 8 km/h, what is the final velocity of the car A? vA mA mB vB = 0 vB = 0 vA' = vA' = ? vB' = ? vB' = ? mB mB mA mA mB mB before collision after collision ....(iii) Dividing equation (iii) by equation (i) gives mA(vA − vA')(vA + vA') mBvB' − vB) (vB' + vB) = mA(vA − vA') mBvB' − vB) Consider a head-on collision, with ball A originally moving and ball B originally at rest (Figure 5.9). afterafter collision collision Figure 5.9 Head-on elastic collision of pool balls of equal masses. A If the motion of each ball is along the same straight line (e.g. along the x-axis), then, by conservation of momentum, mAvA = mAvA' + mBvB' i.e. vA = vA' + vB' (since mA = mB)...............(i) Using the elastic collision equation gives vA − vB = vA − 0 = −(vA' − vB') Using equation (i) gives vA' + vB' = −(vA' − vB') Rearranging gives 2vA' = 0 Therefore, vA' = 0 ....(ii) And so from equation (i), vA = vB' This is the behaviour of the moving ball coming to a stop and the stationary ball acquiring almost the same velocity as the incident ball sometimes occurs during collisions of pool balls of equal masses. This shows that such collisions are almost perfectly elastic.

(The reader is encouraged to check on the internet (e.g. on YouTube) to see how professional pool players can achieve such 'trick' shots where the incoming ball becomes stationary and the ball that is hit moves off.) Chapter 5 Momentum and collisions Discussion example 5.13: Elastic collision of pool balls of equal masses, case 2 Q Consider a glancing collision of pool balls of equal masses, with ball B originally at rest (Figure 5.10). vA ° before collision vA vB = 0 θA ° after collision θB ° vB Figure 5.10 Glancing collision of pool balls of equal masses. m. A If the collision is perfectly elastic, then 1 1 2 2 2 mAvA = 2 mAvA'2 + 2 mBvB'2 i.e. vA2 = vA'2 + vB'2 By Pythagoras' theorem, this means that vA' and vB' must be at right angles to each other, i.e. θA + θB = 90° Note: this does not necessarily mean that θA = θB! Discovery of the neutron by James Chadwick When a beryllium target is bombarded by alpha particles (Figure 5.11), protons were emitted, each of which was detected as having a certain amount of energy. The maximum amount of energy detected from one such proton was far greater than that expected if the radiation from the beryllium target was a gamma ray photon (which is electromagnetic in nature). Chadwick therefore suggested that the radiation consisted of particles that had mass. Chadwick assumed that the mass of a particle of the radiation was nearly equal to the mass of a proton (the latter was well known). Applying principles of conservation of momentum and energy, he found that the maximum energy detected (from one proton) was consistent with the assumption that the mass of the unknown particle was nearly the same as the mass of a proton. (Note: the actual calculations used by Chadwick to calculate the mass of a particle of the unknown radiation were very involved. The energies taken into account were the kinetic energies of all the particles involved at each stage of the collisions and their 'mass-energies' as well. This was because Einstein had established that energy changes, ΔE, had a 'mass equivalent', Δm, related by the equation, ΔE = Δmc2, where c = the speed of light in free space.

The concept of mass-energy is discussed in more detail in Chapter 27.) One of the greatest breakthroughs in understanding the structure of the atom was achieved experimentally by James Chadwick through the application of theory concerning collisions.

Ernest Rutherford had suggested that, apart from protons, the nucleus also contained particles which had mass but no net charge (see Chapter 25). However, there was no experimental confirmation of Rutherford's hypothesis. Chadwick was able to deduce the existence and characteristics of the neutron by applying the principles of momentum and energy conservation to collisions between the previously unidentified radiation and protons. Chadwick found that when the radiation from polonium (alpha source) beryllium target plate collects ions paraffin wax alpha neutrons particles protons to amplifier Figure 5.11 Chadwick's apparatus used in confirming the existence of the neutron. 79 80 Unit 1 Module 1 Mechanics Summary ✓ Linear momentum is defined as the product of ✓ According to the principle of conservation of momentum, as long as there is no net force acting on a system of masses, the total momentum of the system remains constant. For a two-mass system, an object's mass and its velocity (p = mv). Linear momentum is a vector quantity. ✓ According to Newton's second law, the net applied force on an object is directly proportional to the object's rate of change Δp of momentum (F ∝ Δt) and the change in momentum takes place along the direction of the net force. ■ If the mass of the object does not change, then Δv ....(5.3) At ■ If the mass changes (as in the case of rockets, but the velocity of the exhaust gases with respect to the rocket is constant), then F= m Δm ....(5.5) F = vc At ■■ If both the mass and the velocity are changing then F=m Δv Δm +v dt dt ✓ The impulse, J, of a force on an object is defined as the product of the force and the short duration of time during which the force acts (J = FΔt). mAvA + mBvB = mAvA' + mBvB'...............(5.7) ✓ A perfectly elastic collision, by definition, is one in which kinetic energy is conserved. For a twomass system, 1 2 2 mAvA1 1 1 + 2 mBvB2 = 2 mAvA'2 + 2 mBvB'2...............(5.10) ✓ For a perfectly elastic head-on collision (i.e. in one dimension only) vA − vB = −(vA' − vB')...............(5.11) ✓ Linear momentum of a system is conserved in all collisions (providing, of course, there is no net force acting on the system). Kinetic energy is conserved only in perfectly elastic collisions. ✓ Collisions between molecules are assumed to be perfectly elastic (unless there is a change in molecular structure caused by a collision or energy is given to or taken in during the collision). ✓ It follows from Newton's second law that impulse = change in momentum of an object: FΔt = mv − mv0...............(5.6) Review questions Linear momentum 1 (a) Define linear momentum. (b) State Newton's second law of motion in momentum form. (c) Using the equation from part (b), show that we can obtain the equation F = ma, where a is the acceleration produced by F on the mass. (d) Define the S.I. unit of force (the Newton) using the equation F = ma. 2 A motor car of mass 1200 kg is moving at 15 km/h along a straight, horizontal road when the engine cuts off.

The car comes to a stop in 20 s. Estimate: (a) the initial momentum of the car (b) the average frictional force bringing the car to rest. 3 A chunk of rock, estimated to be 100 tonnes in mass (1 tonne = 1000 kg), is heading from space toward the Earth at a speed of 12 km per second. (Such rocks have been thought to have brought tremendous destruction on Earth upon impact.) If missiles of mass 500 kg and speed 8.0 km per second are to intercept the rock head-on in rapid succession, how many such missiles would be needed to reduce the speed of the rock almost to zero at the time of the interception? State any assumptions made. Chapter 5 Momentum and collisions 4 (a) Starting with the equation that represents Newton's second law of motion in momentum form, derive the equation for the thrust, F, produced on a mass, M, by a varying mass, m, which is being ejected at a constant speed, ve, relative to the mass M. (b) Water is sprayed from a hose horizontally on to a vertical side of a motor car. If the speed of the water is 2.0 m s−1 and the cross-sectional area of the water exit from the nozzle is 2.5 cm2, estimate the force of the water on the car, assuming the water does not rebound (density of water = 1000 kg m−3).

5 During lift-off of a rocket, a commentator mentioned that 2500 kg of gas was being expelled every second at a speed of 3000 m s−1 relative to the rocket. Use this information to estimate the mass of the rocket and its contents at lift-off. Linear momentum and impulse, Force versus time graphs, Practical applications of impulse concept 6 (a) Define impulse of a force. (b) Starting with the equation F = mΔv/Δt (equation 5.3), show that impulse is equal to a change in momentum. (c) Figure 5.12 shows a graph of force, F, on a 0.42 kg soccer ball versus time, t. If the ball was approaching the player horizontally at 8.2 m s−1 just before being kicked horizontally in the opposite direction, estimate: (i) the impulse imparted to the ball (ii) the velocity of the ball upon leaving the player's foot. F/N 1200 1000 800 600 400 200 0 0 0.01 0.02 0.03 0.04 0.05 0.06 0.07 0.08 t/s Figure 5.12 7 (a) Just before lift-off, the mass of a rocket plus fuel is 2000 kg. If the engines can eject burnt fuel at an average speed of 1600 m s−1 relative to the rocket, estimate the rate of ejection of burnt fuel (in kg s−1) required to achieve lift-off. (b) Figure 5.13 shows graphs of total weight (W) of a rocket plus fuel versus time (t), and thrust (T) on the rocket produced by gases being expelled versus time. 4 F / x 10 N T 8 7 6 5 4 W 3 2 1 0 0 1 2 3 4 5 6 7 8 9 10 t/s Figure 5.13 (i) From the graph of W versus t, estimate the rate (in kg s−1) at which fuel is being expelled during the first 5.0 s. Explain your answer. (ii) At what time was lift-off actually achieved?

Explain your answer. (iii) Calculate the average speed with which the gases were being expelled by the rocket at lift-off.

(iv) How can change in momentum of an object be obtained from a force versus time graph? (v) Find the change in momentum of the rocket between take off and t = 7 s. 8 A 1300 kg motor car, travelling at 80 km/h, collides with a sturdy concrete wall and is brought to a stop. On account of a certain design feature, the car does not stop immediately but comes to a total stop in 0.40 s. (a) What is the name given to the design feature, and how does it help reduce the impulsive force on the car? (b) Estimate the impulsive force on the car in bringing it to a stop? Where does this force come from? (c) If a driver of mass 70 kg was strapped firmly to the car frame by a rigid seat belt, estimate the impulsive force of the seat belt on the driver.

81 82 Unit 1 Module 1 Mechanics (d) What property of the seat belt helps reduce the impulsive force on the driver in such a collision? Explain your answer. 9 A 0.16 kg billiard ball travelling at 12 m s−1 is incident on an edge of the billiard table at an angle of 60° to the normal.

The collision lasts for 0.018 s, and the ball rebounds with a velocity of 10 m s−1 at an angle of 60° to the normal. Determine, giving magnitude and direction: (a) the change in momentum of the ball (b) the average force of the table's edge on the ball (c) the average force of the ball on the table's edge. 13 (a) Prove that in an oblique, elastic collision involving two equal masses, the masses move off at 90° to each other just after the collision. (b) Figure 5.14 shows tracks made by an alpha particle colliding with a nucleus in liquid helium (an alpha particle is a helium nucleus). Explain the feature of the photograph that suggests an elastic collision had taken place. Conservation of linear momentum, Elastic and inelastic collisions, Discovery of the neutron 10 (a) State the principle of conservation of momentum. (b) A 12 000 kg truck, moving freely at 1.4 m s−1 along a straight, horizontal road, collides with a stationary 1100 kg car that is free to move. Both vehicles become coupled and move off together.

Determine the common velocity with which they move off. Figure 5.14 (c) Explain why the length of an alpha particle track is directly proportional to the square of the original speed of the alpha particle, given that the length of the incident alpha particle track (part of which has not been shown) up to the point of the collision, in terms of the length, ls, of the short track shown branching to the left. (d) If the velocity of the incident alpha particle at the time of the collision was 2.8 × 104 m s−1 and the other particle was at rest, determine the speeds (by calculation and directions (by measurement) of each particle after the collision. 11 (a) Define perfectly elastic collision.

(b) Using the definition of a perfectly elastic collision, show that the collision in question 10b is not elastic. (c) Explain the discrepancy between the total kinetic energy of the two vehicles before and after the collision. 12 (a) Derive the equation, vA − vB = −(vA' − vB') (equation 5.11) for a perfectly elastic head-on collision involving two bodies, A and B, with initial velocities, vA and vB, and final velocities, vA' and vB', respectively. (b) Ball A of mass mA and initial velocity vA in the x direction, collides head-on elastically with ball B of mass mB and initially moving with velocity vB = 0. Applying conservation of momentum, show that m vA' = vA − mB vB' A ( ) (ii) By considering the equation in part (c) only, describe the direction of motion of A just after the collision in each of the three cases.

Give an everyday example of your answer in each of the cases: (i) mA > mB (ii) mA < mB (iii) mA = mB 14 A 0.60 kg ball moving to the right with a speed of 3.6 m s−1 collides head-on, elastically, with a 0.30 kg ball moving in the opposite direction with a speed of 2.0 m s−1. (a) Determine the velocity of each ball just after the collision. 15 Ball A, of mass 0.80 kg, travelling at 2.4 m s−1, collides elastically with a stationary ball, B, of mass 0.60 kg, as shown in Figure 5.15. Calculate the speed and direction of ball, B, just after the collision. vA' = 1.6 m s−1 vA' Before collision vA = 2.4 m s−1 vB = 0 m s−1 after collision θA' = 40° θB' vB' Figure 5.15 Chapter 5 Momentum and collisions Mixed exercises Practical activities and challenges 16 Two pendula, of length 1.00 m, are suspended freely with bob A and B touching. Bob A has a mass of 50 g and bob B a mass of 75 g. When bob A is displaced through an angle of 90° and then released, bob A collides elastically with bob B. (a) Determine the speed of bob A just before the collision. (Hint: use conservation of energy.) (b) Determine the speed of bob B just after the collision. (c) Determine the vertical height to which bob B rises after the collision. (Hint: use conservation of energy.) (d) Determine the angle through which bob B is displaced. Momentum and kinetic energy conservation can be investigated using apparatus arrangements like those shown in Figures 5.16 and 5.17. 17 Use the same data as in problem 16 above, except that this time the collision is completely inelastic (ball A sticks on to ball B).

(a) Determine the speed of bob A just before the collision. (Hint: Use conservation of energy.) (b) Determine the common speed of bobs A and B just after the collision. (c) Determine the vertical height to which the bobs rise after the collision. (Hint: use conservation of energy.) (d) Determine the angle from the vertical through which the bobs are displaced. photogate #1 v2i = 0 photogate #1 m2 = m1 before collision v1i air track m1 Figure 5.16 Air-track apparatus for investigating collisions. The photogate timers in Figure 5.16 can measure the time the black cards mounted on the masses take to pass the respective photogates, from which velocities of the masses can be determined. 18 A horizontal conveyor belt transports and off-loads boxes of mass 8.0 kg at an average speed of 1.2 m s−1. The boxes are loaded vertically on to the belt at the rate of 50 per minute. (a) Estimate: (i) the total force of the belt required to keep the boxes moving at the stated speed (ii) the output power of the conveyor system. (b) Why would the conveyor system require a power greater than that calculated in part (a)(ii) above? Figure 5.17 Low-friction dynamics carts for investigating collisions. The speeds of the carts can be determined using photogate timers (as in Figure 5.16, in conjunction with black cards mounted on each cart). Alternatively, a digital video camera can be used to photograph the events and speeds determined by checking the stills for displacements, as determined from the metre rule, occurring from frame to frame.

83 84 Unit 1 Module 1 Mechanics 1 Conservation of momentum Using the apparatus in Figure 5.16 or Figure 5.17, or otherwise, and other necessary equipment, investigate conservation of linear momentum and conservation of kinetic energy. Consider these situations: (a) when objects move apart after a collision (m1 > m2, m1 < m2 and m1 = m2) (b) when objects stick together during a collision.

■ To what extent is momentum conserved in each of the collisions above? ■ To what extent is kinetic energy conserved in each collision? ■ How would you classify each collision in terms of being elastic or inelastic? 2 Impulsive force during a collision Using a digital video camera (or a motion sensor connected to a data logger that can measure and plot speed as a function of time), investigate the average impulsive force experienced during collision of two dynamics carts: (a) during an inelastic collision where they stick together upon colliding (b) during a collision where they move apart upon colliding.

(Hint: the main experimental problem is to determine Δt.

Then the equation FavΔt = mv − mv0 can be applied in order to determine Fav.) 3 Ballistics Design and carry out activities to check your theoretical calculations made in Review questions 16 and 17. Answers to ITQs 1 Since momentum = mass × velocity, the unit would be kg m s−1. 2 By using a large force for a short time, or a smaller force for a longer time, since mv − mv0 = Ft 3 A person who acts 'impulsively' acts without much thinking, i.e. reacts very quickly. 4 Impulse, J = Ft = 30 N × 0.3 s = 9 N s 5 The landing mat is soft and elastic so that the high jumper comes to a gentle stop. The impact with the mat reduces the jumper's momentum to zero. The change in momentum = Ft so if t is lengthened F is reduced.

6 As long as the sound of the balls colliding is heard, the collision cannot be perfectly elastic. However, pool balls can be made to engage in collisions that are almost perfectly elastic.

7 Use the equation for a perfectly elastic collision: vA − vB = −(vA' − vB')...............(5.11) 12 km/h − 5 km/h = −(vA' − 8 km/h) This gives vA' = 1 km/h Car A almost comes to a stop! Answers to Review questions 3 300 missiles 5 7.6 × 105 kg 7 (a) 12.25 kg s−1 (b) (i) 2.4 × 102 kg s−1 (ii) 3.2 s (iii) 1.3 × 102 m s−1 (iv) area under F versus t graph (v) ≈ 1.1 × 105 kg m s−1 9 (a) 1.78 kg m s−1 at 8.7° clockwise from the −x direction (c) 9.9 N at 8.7° clockwise from the +x direction 13 (b) The 90° angle produced by the collision between identical masses (alpha particle and helium nucleus) (c) 3.9 ls (d) vshort = 0.7 × 104 m s−1 (direction 60° anticlockwise from incident particle direction) vlong = 2.1 × 104 m s−1 (direction 30° clockwise from incident particle direction 15 (a) 2.07 m s−1 (b) 42° 17 (a) 4.4 m s−1 (b) 1.8 m s−1 (c) 0.16 m (d) 33° 85 Chapter 6 Circular motion and gravitation Learning objectives ■ Define angular displacement in radians.

■ Define angular velocity. ■ Use the concept of angular velocity to solve problems involving circular motion. ■ Apply the equation v = rω to solve problems involving circular motion. ■ Apply the equations for centripetal force (F = mrω2 ; F = mv2/r). ■ Solve problems involving horizontal circles, vertical circles, conical pendulums and banking.

Use Newton's law of universal gravitation to solve problems involving attraction between masses. Solve problems involving gravitational fields near the Earth's surface or above.

Derive equations and conditions for circular motion for near-Earth and geostationary orbits. Discuss applications of near-Earth and geostationary satellites. ■ Use equations for centripetal acceleration (a = rω2; a = ■ ■ ■ ■ ■ ■ ■ ■ Kinematics of uniform circular motion Uniform circular motion Period (T) and frequency (f) Kinematics refers to describing motion, usually by means of equations. Uniform circular motion is one such motion. The period of a particle under going uniform circular motion is defined as the time taken for one complete revolution. The frequency of the rotation is the number of rotations made per second. Thus, if a particle makes n revolutions in t seconds, its period, T, is given by the t n equation T = ; its frequency, f, is given by f = t .

Note that period is measured in seconds, since n is just a number. Frequency is measured in number of revolutions (or cycles) per second. One revolution per second is called a hertz (Hz) in the S.I. Uniform circular motion of an object is motion in a circle at constant speed. In Figure 6.1, a particle, P, moves with a constant speed, v, in a circle of radius r. Note that although the speed is constant, the velocity is not constant, since velocity encompasses both magnitude (speed) and direction. In circular motion the direction of motion of P is a tangent to the circle at the location of P and is therefore constantly changing. Hence the velocity of an object under going uniform circular motion is always changing. t n 1 t n 1 = T = n 1 Or = 1 × T 1 = But n f 1 t Therefore = f T Then y = P θ O Since T = x x Figure 6.1 A particle, P, moving in a circle at constant speed v. ITQ 1 In Figure 6.1, the rotation of P takes place about an axis. Describe this axis. ....(6.1) It is left as an exercise to the reader to show that 1 = T f ....(6.2) Thus period, T, and frequency, f, are reciprocals of each other. 86 Unit 1 Module 1 Mechanics Angular displacement (θ) The angle, θ (Figure 6.1), measured anticlockwise from the x-axis, is called the angular displacement of the particle, assuming timing of the motion of the particle began when it was crossing the x-axis. Anticlockwise measures of angular displacement from the x-axis are designated positive; clockwise are negative.

The angle θ is defined as the ratio of arc-length swept out (in a time, t) to the radius of the circular path. In equation form, for an arc length s swept out by the particle, s θ= r For one complete revolution in circular motion, s = 2πr = r i.e. θ = 2π rad ....(6.4) Angular velocity (ω) Angular velocity is defined as the rate of change of angular displacement with time. Hence, for a tiny angular displacement, Δθ, taking place in a tiny time, Δt, ω= Δθ Δt ....(6.5) But since the speed of the particle is constant, it follows that the angle swept out per unit time is constant. Hence angular velocity is constant, and we can therefore express equation 6.5 as ω= Q Calculate the angular velocity of the Earth about the Sun. Assume the path of the Earth is a circle and that it takes the Earth 365 days to make one complete orbit of the Sun. For a given period, T (the angular velocity, ω, is given by 2π ω = T or 2πf) T = 365 d × 24 h/d × 3600 s/h = 1.99 × 10−7 rad s−1 ....(6.3) Since both s and r have units of length, the units cancel. The angle, thus defined, is therefore only a number. The unit name given for the value of an angle, θ, is called the radian (abbreviated rad). θ= Worked example 6.1: Angular velocity of the Earth θ t ....(6.6) Centripetal acceleration, a Acceleration, a, is defined as the rate of change of linear velocity, i.e. a = Δv Δt . Figure 6.2a shows velocity vectors, of magnitude v, for a particle undergoing uniform circular motion. By extending the directions of the vectors, so that v1 and v2 intersect, and applying a little geometry, it can be seen that the angle ΔΘ, between the vectors, is equal to the angle Δθ. We also consider Figure 6.2b. The small angle Δθ, since speed v is constant, Δv, is approximately equal to the arc length subtended by Δθ for a circle of radius v. Hence, Δv = vΔθ (since s = Δv) But, since a = Δv/Δt, then a = ω... Hence, a = vω (equation 6.10), then 2π ω= T and ω = 2πf (since 1 T = f) ....(6.7) Relationship between linear and angular velocities (v and ω) Since, for one complete revolution, the area of a circle is swept out by the radius, the time taken is the period, T. We have... Hence, v = rω ....(6.11) ....(6.10) Figure 6.2b shows that the direction of Δv, and hence of the acceleration a, is perpendicular to the velocity vector in the limit as Δθ approaches zero. Hence, the acceleration is directed towards the centre of the circle, along the radius. It is for this reason that this acceleration is called a 'centre-seeking' acceleration or a centripetal acceleration. Chapter 6 Circular motion and gravitation a b v2 Δθ Worked example 6.2: Angular velocity, speed and centripetal acceleration of a CD bit Δv v2 − v1 v1 Δθ r v2 Q A CD of diameter 12.0 cm is spinning at 200 rpm (revolutions per minute). For a 'bit' of data, located near the edge of the CD, determine: (a) the angular velocity (b) the linear speed (c) the centripetal acceleration. A ■ θ Figure 6.2 (a) Velocity vectors v1 and v2, of same magnitude, v, for a particle undergoing uniform circular motion.

(b) The change in velocity, Δv, associated with angular displacement Δθ. Dynamics of uniform circular motion Dynamics is the study of the action of forces on masses. According to Newton's second law of motion, the net force, F, on an object is equal to its mass times its acceleration (equation 3.3).

Since, in circular motion at constant speed v, the magnitude of the centripetal acceleration, a, is given by a= v 2 r (equation 6.13) Then, using F = ma (equation 3.3), F= angular velocity, ω = ? ....(6.15) ....(6.14) Or, applying equation 6.12, F = mω2r ....(6.15) Further, since the direction of motion is tangential, i.e. directed toward the centre of the circle, the force is therefore also directed towards the centre of the circle, and are therefore perpendicular to the circular path taken by the object at the point of reference and the instantaneous direction of motion of the moving particle. Ball on string Figure 6.3 shows a ball of mass m being whirled at constant speed v in a vertical circle (Figure 6.3a) and in a horizontal circle (Figure 6.3b). C a FTC b mg v v D FT FT B mg mg r FT mg FTA A Conditions for uniform circular motion Equations 6.14 and 6.15 summarize the two conditions that must be met for uniform circular motion to occur. First, the particle must be moving at a constant speed, and second, there must be a constant 'centripetal' force, i.e. a force acting perpendicularly to the direction of motion of the particle at all times.

We now consider a few cases which illustrate the dynamics of circular motion. mg Figure 6.3 A ball of mass m being whirled at constant speed v: (a) in a vertical circle and (b) in a horizontal circle. ITQ 2 How does the force of tension in the string in Figure 6.3a compare: (a) between the points A and B? (b) between the points B and C? 87 88 Unit 1 Module 1 Mechanics Discussion example 6.3: Ball whirled in a vertical circle Q A ball of mass 64 g connected to a string is whirled in a vertical circle of radius 30 cm with gradually increasing speed (Figure 6.4).

C mg FTB D (a) Discuss the motion of the ball at low speeds. (b) If the maximum strength (tension) of the string is 3.4 N, (i) where, and at what speed, is the string most likely to break?

(ii) where will the ball land after the string breaks if at the point of breakage of the string, the ball was 40 cm above the ground? FTC r B FTA v A mg y −40 cm A Figure 6.4 Ball being whirled in a vertical circle and string breaking at lowest point. x −x0 (a) At position A, mv2 FTA − mg = r mv2 + mg r At position C, mv2 FTC + mg = r Or FTA = mv2 − mg r ....(i) ....(ii) Equations (i) and (ii) show that at a given speed, v, the tension in the string is greatest when the ball is at the lowest point, A, since the tension in the string is greatest there. We get, for this tension, mv 2 FTA + = mg + 0.064 kg × v2 = 3.4 N − 0.064 kg × 9.8 N kg−1 0.30 m (ii) For circular motion to be maintained, the net centripetal 2 force must be positive, since mv r is always positive. According to equation (ii), if v is too small, then mv 2 − mg < 0. Hence, circular motion r r cannot be maintained (the particle will not be able to reach C). For circular motion to be maintained, mv 2 − mg ≥ 0 at least. In this case v2 = rg. Or, 3.4 N = 0.213 v2 kg m−1 0.627 N 3.4 − 0.627 Therefore, v2 = (m s−1)2 0.213 Hence, v = 3.61 m s−1 ≈ 3.6 m s−1 (b) (i) We find the time the ball takes to fall to the ground when the ball is undergoing a projectile motion: x = v0 t + 12 at2, i.e. y = y0 + v0 t + 12 at2 ....(2.8) Taking the up direction as positive, we get −0.40 m − 0 = 0 + 12 × −9.8 m s−2 × t2 Therefore, t = 0.286 s = 0.29 s Using this time, t = 0.29 s, we calculate the horizontal displacement, x − x0, of the ball as follows, using the fact that the horizontal velocity is constant in projectile motion: x − x0 = vt Therefore, x − x0 = 3.61 m s−1 × 0.286 s = 1.0 m (approx.) Chapter 6 Circular motion and gravitation Note: For the ball to be moving in a vertical circle, the string would have to be inclined slightly upward to the horizontal so that there is an upward component of tension in the string that serves to keep the ball moving along a horizontal plane. For calculation purposes, we assume that the string and the ball in Figure 6.3b are moving in the same plane. We apply Newton's second law of motion as follows, taking the direction of the force of tension in each case. Ball moving in a vertical circle (Figure 6.4) ■ At: the two forces acting on the mass are the force of tension, FTA, in the string and the weight, mg, of the mass. These forces are acting in opposite directions on the ball. Therefore mv 2 FTA − mg = r mv 2 + mg ....(6.16) Or, FTA = r ■ At D and at B: the weight is acting in a vertical direction, and hence is at right angles to the force of tension, FT, in the string.

Hence the weight has no effect on the centripetal force. We get mv 2 FT = r ....(6.17) ■ At C: both the weight and the force of tension are acting in the same direction. Hence, mv2 FTC + mg = r Or, FTC = mv2 − mg ....(6.18) r mv 2 Since r is always positive, equations 6.16 to 6.18 tell us that the force of tension in the string is a maximum at the lowest point A and a minimum at the uppermost point C. See also Discussion example 6.3. Ball moving in a horizontal circle (Figure 6.3b) In Figure 6.3b, the only force acting on the object is the force of tension, FT, in the string. If the object is moving at constant speed, equation 6.14 tells us that this force of tension remains constant throughout the orbital path. Conical pendulum Our next case (Worked example 6.4) involves motion of a conical pendulum (Figure 6.5). Students performing 'simple pendulum' experiments sometimes encounter this motion, but in error ... Getting it right! To find the period of a conical pendulum, we cannot use the simple pendulum formula, T = 2π Lg , because a conical pendulum is not a simple pendulum. First of all, the angle of swing of a simple pendulum has to be made small if the simple pendulum formula, just stated, is to be true, whereas there is no restriction on the size of the angle between the vertical for a conical pendulum. Secondly, the motion of the bob in a simple pendulum is approximately a straight line (for small angles of swing), whereas the motion of a conical pendulum bob is circular. Period of a conical pendulum Using Figure 6.6 (page 90), we get FT cos θ = mg...............(i) since there is no vertical motion of the bob and FT sin θ = mv 2 r ....(ii) for circular motion of radius r. Combining equations i and ii, we get mg mv 2 sin θ = r cos θ Hence, g 2π r 2ω 2 sin θ = (since v = rω, where ω =) r T cos θ g sin θ = rω2 cos θ g sin θ = L sin θ ω2 cos θ g 4π2 sin θ = L sin θ × 2 T cos θ From which we get, T2 = L cos θ × 4π2 g Therefore, T = 2π L cos θ g which is the same as that of a simple pendulum when θ is small (cos θ = 1). As θ increases, the period of the conical pendulum decreases, but the period of the simple pendulum increases.

The simple pendulum formula is derived in Chapter 7: T = 2π L g ....(7.26) The derivation shows that for the simple pendulum, mg FT = −mg sin θ = mrω2 as compared with = mrω2 cos θ for the conical pendulum. 89 90 Unit 1 Module 1 Mechanics Worked example 6.4: Conical pendulum Q A Figure 6.5 shows a conical pendulum of length 1.0 m and mass 400 g.

The string is inclined at 40° to the vertical. The mass is moving at a constant speed, v, in a circle. (a) Draw a free-body diagram showing the forces acting on the pendulum bob.

(b) Calculate: (i) the tension in the string (ii) the period of the motion. (a) θ L FT FT cos θ FT sin θ mg Figure 6.6 Free-body diagram showing forces acting on a conical pendulum bob. 1.0 m 40˚ v 400 g Figure 6.5 Conical pendulum. (b) (i) The vertical component of the tension in the string is equal to the weight of the bob. Hence, FT cos 40° = mg mg Or, FT = cos 40° 0.400 kg × 9.8 N kg−1 0.766 i.e. FT = 5.1 N (ii) Let length of pendulum = L Then, radius, r, of the circular motion of the bob is given by r = L sin θ i.e. r = 1.0 sin 40° = 0.64 m 2 To find the period, T, we use the equation F = mv r , where F is the horizontal component of FT. We also use the fact that since v = s t (equation 6.9), then, for one complete revolution, the distance travelled is the circumference, and the time taken is the period of the motion. Thus, 2πr v= ....(6.19) T We therefore get m × 4π2r 2 FT sin 40° = rT2 = Or, T2 = m × 4π2r FT sin 40° 0.400 kg × 4π2 × 0.64 m 5.1 N × 0.64 i.e. T = 1.76 s ≈ 1.8 s = Chapter 6 Circular motion and gravitation Bucket of water Discussion example 6.5 is probably familiar to many students. If a bucket of water is whirled in a vertical circle at a certain speed, the water does not fall out of the bucket when the bucket is upside down at the top of the motion. Discussion example 6.5: Bucket whirled in a vertical circle Q (a) A cylindrical bucket containing water is whirled in a vertical circle. Draw a free-body diagram showing the forces acting on the water at the top and at the bottom of the circular motion. (b) What is the minimum (constant) speed at which the cylindrical bucket of water must be whirled in a vertical circle of radius r so that the water does not fall out when the bucket is upside down at the top of the motion?

(c) (i) What is the nature of the force that must be present at the lowest part of the motion in order for circular motion to take place? (ii) What is the magnitude of this force? A (a) B FNB mg v FNA bucket A water mg Figure 6.7 Free-body diagram showing the forces acting on water in a bucket at the top and at the bottom of circular motion. Getting it right! In Discussion example 6.5 and in all other cases of circular motion, we are using the convention that forces directed toward the centre of the circle are designated positive. (b) At the top of the motion, mv 2 FNB + mg = r where FNB is the normal contact force provided by the bottom of the bucket. The net force at the top must be a minimum in order for minimum speed to be achieved, so FNB must be zero. Hence, mv 2 mg = r which gives v 2 = rg or v = √rg (c) (i) There must be a net centripetal force (i.e. pointing toward the centre of the circle) if circular motion is to take place. The weight, mg, acts downward. Hence, at the lowest part of the motion, there must be an upward force acting on the water. This force is FNA and is the normal contact force of the bottom of the bucket on the water. (ii) For circular motion at speed, v, mv 2 FNA − mg = r Hence, FNA = mv 2 + mg r 91 92 Unit 1 Module 1 Mechanics Carnivals, fairs and circuses Bends, roundabouts and corners Carnivals, fairs and circuses employ circular motion in many of their activities. The merry-go-round, rollercoaster, rotor-ride (or gravitron) and the 'globe-of-death' are a few examples. We shall discuss the 'globe-of-death' at this point. Other examples will appear in the Review questions at the end of the chapter.

Curved sections of motorways are designed on the dynamics of circular motion. roundabout car Discussion example 6.6: 'Globe-of-death' Q Figure 6.8 shows two motorcyclists riding simultaneously inside a 'globe-of-death' of radius, r. (a) What is the minimum speed with which motorcyclists must ride so that the motorcycle does not fall off the top part of the sphere? (b) What provides the centripetal force at this speed at the top and at the bottom of the globe? (Assume the mass of each motorcyclist plus motorcycle is m.) (c) If the motorcycles are not to collide, what is one condition that must be met? Ffrs r Figure 6.9 Top view of a car on a roundabout. Figure 6.9 shows a car moving on a horizontal roundabout of radius r. The centripetal force is the force of static friction, Ffrs. If the car is not to skid on the roundabout, then its maximum safe speed, v, is given by Ffrs = mv 2 r ....(6.20) But, the maximum static friction, Ffrs, is related to the normal contact force, FN, and the coefficient of static friction, μs, by the equation Ffrs = μsFN (equation 3.11) Figure 6.8 'Globe-of-death'. A (a) As in the case of the bucket of water (Discussion example 6.5) the minimum speed, v, must be given by v = √rg (where r is the radius of the globe) if the motorcycle is to remain in contact with the globe at the top of the motion. (b) The centripetal force responsible for the circular motion of the motorcycle at the top of the globe at this speed is the weight of the motorcyclist plus motorcycle. At the bottom of the globe, the force, as in the case of the bucket and water, is the normal contact force provided by the globe (upward force). As well as providing the centripetal force the contact force must also balance the weight. (c) The period of motion of the cyclists must be the same. If speeds are constant, this means that v must be equal to 2πr . T Since, on a horizontal surface the normal contact force is equal to the weight resting on the surface, i.e. FN = mg (see Worked example 3.2a), Then, Ffrs = μsmg We therefore get, μsmg = mv2 r v2 = μsrg v = √μsrg...............(6.21) Equation 6.21 tells us that the maximum safe speed on a horizontal roundabout depends only on the radius of the roundabout, the coefficient of static friction between the tyres and the road and the acceleration due to gravity. Equation 6.21 applies, not only on horizontal roundabouts but also on horizontal bends or corners which vehicles are negotiating. Chapter 6 Circular motion and gravitation R2 a R1 R = mg b C C x2 x1 curvature of road h F2 h F1 F motion of car motion of car Toppling of vehicles Figure 6.10a shows a car rounding a bend at a low speed. Taking moments, T, about the centre of gravity, C, we get the following: T1 = R1x1 is an anticlockwise moment T2 = R2x2 is a clockwise moment These two moments are approximately equal in magnitude since we assume that the weight of the vehicle is distributed equally on each wheel and that the centre of gravity is at the centre of the vehicle.

T1 and T2 therefore result in no net toppling effect as their vector sum is zero. However, T3 = F1h, and T4 = F2h are clockwise moments whose effects are to tilt the car on the side wheels. At low speeds, F1 and F2 are small. Both F1 and F2 increase with 2 increasing speed (F = mv), hence the larger clockwise r moments produced cause the car to tilt clockwise. As the car tilts, the normal contact force, R, acts on one side only and is equal to the weight, mg, of the car (Figure 6.10b). From Figure 6.10b we get, ■ clockwise moment (tending to topple the car) mv 2 h r ■ anticlockwise moment (tending to keep the car horizontal = Fh = = Rx1 = mgx1 Thus, the car will topple if the clockwise moment is greater than the anticlockwise moment, i.e. if mv 2 r h > mgx1 gx1r v2 > h v> rgx1 h x1 curvature of road Figure 6.10 (a) Vehicle at low speed on a horizontal bend. (b) Vehicle at high speed on a bend. Normal contact forces are designated as R, static frictional forces as F. The maximum, safe speed (without toppling), then, is given by rgx1 v= ....(6.23) h To allow for large, safe speeds (without toppling) requires the following: ■ r must be large (the road has a very gentle bend) ■ x1 must be large (the width of the base of the car is large), and ■ h must be small (the centre of gravity of the car is low). Racing cars are built with wide bases and low centres of gravity for these reasons. Banking of roads The coefficient of static friction between car tyres and the road surface is reduced when water or oil films from the asphalt spread on the road surface. Hence, the maximum safe speed (i.e. without skidding) on such horizontal roads is reduced since v = √μsrg (equation 6.21). One way of designing a bend that does not require any frictional force at all is to 'bank' the road at a certain angle to the horizontal, called the banking angle, θ. Figure 6.11 shows a vertical section through one such banked road. In Figure 6.11a, a car of mass m is moving in a horizontal circle of radius r with a speed v. The normal contact force, FN, and the weight, mg, are shown in the free-body diagram of the car on the road. The horizontal component of FN provides the centripetal force. A little geometry will show that if the banking angle is θ, then the angle between FNx and the road is θ, and so is the angle between FNy and FN. For circular motion in the horizontal plane of r, ...(6.22) ITQ 3 Which law(s) of Newton, when applied to a mass resting on a horizontal surface, yields the result FN = mg? mv 2 = FN sin θ r ....(i) ITQ 4 Would the maximum safe speed on a horizontal, curved road be the same for all vehicles with the same type of tyres? 93 94 Unit 1 Module 1 Mechanics FN a FNy θ r v b car FNx Figure 6.11 (a) A vertical section showing a road that is banked at an angle θ to the horizontal. (b) A raceway showing increasing banking angles at a bend. θ mg But, applying Newton's second law in the vertical direction we get FNy = mg i.e. FN cos θ = mg or, mg = FN cos θ...............(ii) Dividing equation i by equation ii, we get mv 2 r F sin θ = N mg FN cos θ hence, v 2 = tan θ rg Newton's law of universal gravitation Isaac Newton (1643–1727) stated the following law, called the law of universal gravitation: All objects attract each other with the force of gravity. The force of attraction between two objects is directly proportional to the product of their masses and inversely proportional to the square of the distance between them. Figure 6.12 shows two objects of masses m1 and m2, separated by a distance d and attracting each other with a force F.

According to Newton's law of universal gravitation, ■ F is proportional to m1 and m2 ....(6.24) Equation 6.24 tells us two things: ■ for a given required speed, v, there is only one banking angle, and therefore ■ if higher speeds are required, then a steeper banking angle is necessary. Some race track bends have roadways with curved inclines, and so have smaller banking angles lower down and larger banking angles higher up (Figure 6.11b).

Drivers who wish to negotiate bends at very high speeds must use the steeper part of the curve. Dynamics of gravitational motion Gravity plays a very important role in circular motion today, especially concerning satellites. We shall first discuss the gravitational force before considering circular motion under gravity. ITQ 5 Masses m1 and m2 attract each other with a gravitational force of the same magnitude. According to which of Newton's laws of motion is the force of m1 on m2 equal (but opposite) to that of m2 on m1? ITQ 6 The force of gravity between two masses separated a distance d is F. What would the force of gravity be between the masses if they are separated a distance 3 × d? 1 d2 m1m2 And therefore, F is proportional to d2 Gm1m2 F= ....(6.25) d2 ■ F is proportional to The constant, G, is called the universal gravitation constant and has the value of 6.67 × 10−11 N m2 kg−2. Equation 6.25 shows us that the force of gravity decreases with separation distance and increases with mass.

m2 m1 F F Figure 6.12 Two masses separated by a distance d. d Gravitational field The question arises as to how a mass can exert a force on another mass without them touching each other. The concept of a gravitational field accounts for this. A gravitational field is believed to be associated with every mass. Table 6.1 Some useful geophysical data for this section Body Mass / kg Radius / km Surface gravitational Mean distance / km field strength / N kg−1 Earth 5.98 × 1024 6.38 × 103 9.8 1.50 × 106 (from Sun) Mars 6.42 × 1023 3.39 × 103 3.7 2.28 × 108 (from Sun) Moon 7.35 × 10 1.74 × 10 1.6 3.84 × 105 (from Earth) Sun 6.96 × 105 22 1.99 × 1030 3 274 – Chapter 6 Circular motion and gravitation We can also show that 1 N kg−1 (the unit for g) is the same as 1 m s−2 (the unit for acceleration), as follows: Since F = ma Then 1 N = 1 kg × 1 m s−2 X Dividing both sides by kg, we get 1 N kg−1 = 1 m s−2 Gravitational field strength at a distance h above the surface of a body Figure 6.13 Gravitational field of the Earth (shown by directed lines). A gravitational field is defined as a region in which a gravitational force acts.

The direction of the gravitational field is the direction a tiny mass would move if placed at any point in the field. Figure 6.13 shows the gravitational field direction (simply referred to as the gravitational field) above the Earth. The Earth is assumed to be a perfect sphere. Note that the field lines point toward the centre of the Earth. The gravitational field strength, g, at a point such as X in Figure 6.13 is defined as the gravitational force per unit mass at that point in the field. The gravitational force on a mass, m, is called its weight, W. It follows then, that the equation for the gravitational field strength, g, at X is given by g= W m ....(6.26) or, rearranging, W = mg ....(6.27) Equation 6.26 tells us that the unit of gravitational field strength, g, is N kg−1. If we apply Newton's second law of motion to a mass in free fall near the Earth's surface, we get F = ma Therefore, W = ma But, W = mg Let us now obtain an equation for the gravitational field strength g at a distance h from the surface of a body (e.g. the Earth). By Newton's law of universal gravitation (equation 6.25), the force of the Earth's gravity on a mass m a distance r from the Earth's centre (but above its surface) is given by mg = where M is the mass of the Earth. Therefore, g = GM r2 ....(6.29) Equation 6.29 tells us that g varies inversely with the square of the distance between the mass and the centre of the Earth. If R is the radius of the Earth and h is the height of the mass m above the Earth, equation 6.29 becomes g= GM (R + h)2 ....(6.30) Equation 6.30 can be used to determine the gravitational field strength (or the acceleration due to gravity) at or near the surface of any body, such as a moon or a planet. Discussion example 6.7: The value of g near the Earth's surface Q Using equation 6.30, determine the value of g at a height h above, but near to, the Earth's surface. A Near the Earth's surface, h ⋘ R. So for calculation purposes we can let h ≈ 0.

We get GM g= (R + h)2 ≈ GM R 2 ≈ 6.67 × 10−11 N m2 kg−2 × 5.98 × 1024 kg (6.38 × 106 m)2 Therefore, g = a...............(6.28) Thus, the acceleration due to gravity at a point in a gravitational field of strength g is numerically equal to g. GMm r2 ≈ 9.8 N kg−1 (or m s−2) 95 96 Unit 1 Module 1 Mechanics Gravitation and circular motion It can be shown that the orbit of a body moving around another under the influence of gravity only is an ellipse. However, unless otherwise stated, we shall be treating these orbits as circular since many of them are nearly so. Worked example 6.8: Orbital speed of near-Earth satellite Q (a) Determine the orbital speed, v, for a satellite orbiting near the Earth.

(b) How long will it take such a satellite to orbit the Earth? Note: an ellipse has two fixed points, A and B, called foci. The curve of an ellipse, as traced by a point C, is such that the sum of the distances, AC + CB, is always constant (Figure 6.14). A circle can be considered as a special case of an ellipse in which the two foci occur at the same point. In this latter case, AC = CB and forms the radius of the circle. ellipse A satellite v h R C B Figure 6.14 Illustrating an ellipse. The distance AC + BC is constant throughout the curved path. Near-Earth Satellite A satellite is a body that orbits another under the influence of gravity. The Moon is a natural satellite of the Earth, orbiting at a mean distance of 3.84 × 105 km from the Earth. Artificial satellites orbiting at about 200–500 km above the Earth's surface can be treated as being 'near Earth' since the Earth's mean radius is 6.38 × 10³ km. At these heights above the Earth, the atmosphere is very thin and offers little friction to the satellite. If m is the mass of a satellite and v is its orbital speed, and M is the mass of the Earth and R its radius, we can obtain the formula for the orbital speed for such a satellite. We apply the equations involving the force acting on the satellite to keep it in orbit (equations 6.14 and 6.25). mv2 GMm = (see equations 6.14 and 6.25) (R + h) (R + h)2 or, v2 = Figure 6.15 Near-Earth satellite circular motion (not drawn to scale). A (a) Let m be the mass of the satellite and v its orbital speed. Also, let M be the mass of the Earth and R its radius. Applying the near-Earth approximation (h ≈ 0) to equation 6.31, we get GM v = R 6.67 × 10−11 N m2 kg−2 × 5.98 × 1024 kg 6.38 × 106 m ≈ 7.9 km s−1 (b) The time taken to orbit the Earth is the period, T. 2πR Since v = (equation 6.19) T Then, T = 2πR It is left as an exercise to the reader to show that this period T ≈ 85 minutes.

= GM (R + h) Uses of near-Earth (and medium Earth) satellites and progression of severe weather disturbances. The reader is encouraged to research, using the internet, these and other uses. The proximity of such satellites to the Earth's surface (approximately 200 to 2000 km) and the short time in which they can orbit the Earth facilitate such uses. Near-Earth satellites (designated low Earth orbit – LEO – satellites) are used for photographing in some detail features present on the Earth's surface, such as crop growth over large areas, movement of armies and aircraft, A synchronized system of satellites orbiting at about 20 000 km above the Earth with a period of 12 hours forms a Global Positioning System (GPS). Such orbits are called medium Earth orbits – MEO. A GPS receiver needs v= GM (R + h) ....(6.31) Chapter 6 Circular motion and gravitation a motion of satellite A b D 1 day B C P equator path taken by geostationary satellite equator to lock on to the signal of at least three of these satellites to calculate the surface position (latitude and longitude) and track the movement of an object on the Earth's surface. Geostationary satellites A geostationary satellite is a satellite that has the same period as the Earth's rotation about its axis. Figure 6.16a shows such a satellite orbiting the Earth. When the satellite is at a point A it is directly above a point B on the Earth's surface. When the satellite reaches position D, the point B moves to a position C, since both the Earth's rotation about its axis at P and the satellite's orbit have the same period, i.e. will sweep out the same angle in the same time. To an observer on the Earth, the satellite therefore appears 'stationary' – hence the term 'geostationary'. To be geostationary, a satellite must orbit the Earth with the same period as the rotation of the Earth about its axis. The satellite must also follow a path that is directly above the equator, otherwise it will not appear to be over the same spot as the Earth rotates about its axis. Figure 6.16b shows the axis of the Earth's rotation through both of the poles. The Earth's equator is shown. A near-Earth satellite has too short a period (85 minutes) to be geostationary with the Earth. The Earth rotates about its axis in 24 hours (1 day). To be geostationary, a satellite must orbit high above the Earth and at a particular height to achieve a period of 24 hours, as Discussion example 6.9 shows. A geostationary Earth orbit is abbreviated as GEO. ITQ 7 What is the value of the period of the rotation of the Earth about the axis through its poles? Figure 6.16 Geostationary satellite (sometimes called geosynchronous satellite). (a) View from above one of the poles, P, of the Earth. (b) View showing the equator in relation to the orbital motion of the satellite. These diagrams are not to scale. The following is an alternative solution to that in Discussion example 6.8 for working out the height of a geostationary orbit. GMm = mω2r r2 Therefore r3 = Now ω = So r3 = r = 3 GM ω2 2π T GMT2 4π2 6.67 × 10−11 N m2 kg−2 × 5.98 × 1024 kg × (24 h × 3600 s/h)2 4π2 = 4.22 × 107 m Therefore, h = r − R = 4.22 × 107 m – 6.38 × 106 m = 3.58 × 107 m ≈ 36 000 km Uses of geostationary satellites Because geostationary satellites are in 'parked' orbits (relative to the Earth), they can be used to pin-point the exact location of objects and monitor specific areas continuously on the Earth's surface.

An international network of geostationary satellites is used primarily by meteorologists to track weather phenomena such as hurricanes and tsunamis, and by scientists studying movements of the Earth's plates and migration of animals. Weather-forecasting models make use of data collected by these satellites. 97 98 Unit 1 Module 1 Mechanics If a person wishes to place a phone call from D to E, on the other side of the Earth, the signal is relayed via satellites B and C. Of course, on account of the large distances involved, the signal takes some time to reach E in spite of the fact that it is travelling at the speed of light (3 × 108 m s−1). Persons making such international calls usually experience a delay (called 'latency') in hearing the response from the person at the other end for that reason. A P E D Weightlessness B C Figure 6.17 Geostationary satellite arrangement used for global communications, e.g. telephone and TV. Geostationary satellites are also used in global telecommunications. Because geostationary satellites remain in the same position relative to points on the Earth, satellite dishes sending signals to, or receiving signals from, such satellites could point to such 'parked' satellites and remain in fixed positions and orientations. Figure 6.17 shows three such satellites 'parked' so that they can relay telecommunications signals to cover the entire globe. Discussion example 6.9: Orbital height of a geostationary satellite Q At what height, h, above the surface of the Earth must a satellite orbit in order to be geostationary? A Taking M and R as the mass and radius of the Earth, respectively, and m as the mass of the satellite, we get GMm mv 2 = 2 (R + h) (R + h) (see a similar argument in Discussion example 6.7). GM Thus, = v 2 (R + h) But, v = Astronauts in space stations orbiting at a height h above the Earth often send back pictures to Earth showing apparent 'weightlessness' in their environment (Figure 6.18). Relative to the Earth, they and all other objects in the orbiter still possess weight since there exists a gravitational field due to the Earth at their location from the Earth. However, for circular motion, the orbiter, its material contents and the astronauts must be accelerating with the same centripetal acceleration, a, which is numerically equal to g, given by g= v2 (equation 6.13) (R + h) where, v = GM (equation 6.31) (R + h) Since the orbiting station and its contents, which include the astronauts, are at the same height from the Earth, the above equations tell us that relative to the station, the astronauts and other objects within the capsule would be stationary when released. In other words they appear 'weightless' since all will be moving with the same speed and with the same centripetal acceleration.

2π(R + h) (see equation 6.19) T Therefore, GM 4π2(R + h)2 = (R + h) T 2 GMT 2 ....(6.32) 4π2 By substituting the values for R, G, M and T, it can be shown that h ≈ 36 000 km, which is about 6 times the Earth's radius! Or, (R + h)3 = Figure 6.18 Apparent 'weightlessness' while working outside an orbiting spacecraft. Chapter 6 Circular motion and gravitation Summary ✓ For a particle moving in a circle of radius r, the angular displacement, θ, of the particle is defined as the ratio of the arc length, s, moved s by the particle to the radius (θ =).

The unit for r angular displacement is the radian (abbreviated rad). For one complete circular movement, the angular displacement is 2π rad. ✓ Newton's law of universal gravitation states ✓ Angular velocity, ω, is the rate of change of angular displacement with time (ω = Δθ). For Δt constant speed in circular motion, ω is related to period, T, and frequency, f, as follows: 2π ....(6.7) T ω = 2πf...............(6.8) ✓ A mass exerts a gravitational force by means ω= of a gravitational field. The gravitational field strength, g, at a point in a gravitational field is defined as the gravitational force per unit mass at that point in the field. Since the gravitational force on an object of mass, m, is called its weight, W, the equation for gravitational field strength, g, is given by ✓ Linear speed, v, of a particle moving in a circle that all objects in the universe attract each other with a force of gravity. The gravitational force, F, between two objects of masses m1 and m2 is directly proportional to the product of the masses and inversely proportional to the square of the separation distance, d, between their centres. In equation form, Gm1m2 F= ....(6.25) d2 The constant, G, is called the universal gravitation constant and has the value of 6.67 × 10−11 N m2 kg−2. is defined as the arc length, s, travelled per unit s time (v =). It can be shown that for a period, T, t 2πr v= ....(6.19) T v = rω...............(6.10) ✓ Objects undergoing circular motion experience a v2 ....(6.13) r a = rω2...............(6.12) a= ✓ By Newton's second law of motion, objects ....(6.26) W = mg...............(6.27) ✓ A satellite is a body that orbits another under the influence of gravity. The gravitational field strength at a height h above the surface of a body of mass M and radius R is given by the equation undergoing centripetal acceleration must also be experiencing a centripetal force F, given by mv2 ....(6.14) r F = mω2r...............(6.15) W m or, rearranging, g= centripetal acceleration, a, given by F= GM ....(6.30) (R + h)2 The orbital speed, v, of such a satellite is given by g= v= GM (R + h) ....(6.31) ✓ The period of a geostationary (or geosynchronous) satellite is 24 hours.

Such a satellite must follow a path directly overhead to the equator. 99 100 Unit 1 Module 1 Mechanics Review questions Note: values of geophysical constants (and other constants) may be found at the front of this book. Kinematics of circular motion 1 (a) Define: (i) uniform circular motion (ii) angular displacement, θ (in radians) (iii) angular velocity, ω. (b) Show that in uniform circular motion of radius r and of period T: (i) v = 2πr (where v is the linear speed) T (ii) v = rω 2 (a) Starting with the equation for centripetal 2 acceleration (a = vr) for a particle moving with a speed v in a circle of radius r, show that a = rω2, where ω is the angular velocity of the particle. (b) Estimate the centripetal acceleration of an object on the surface of the Earth: (i) at the North pole (ii) at the equator. 3 A circular grindstone of radius 15 cm rotates at 200 rpm (revolutions per minute). (a) What is its angular velocity, in rad s−1? (b) Estimate the speed with which sparks fly off the circumference of the grindstone assuming that they fly off at the tangential speed of the grindstone. 4 (a) Assuming that the Earth is a perfect sphere (which it is not quite) of radius 6.38 × 103 km, estimate the angular velocity of a person standing: (i) at the equator (ii) at the poles. (b) Draw a free-body diagram to show the forces (including the normal contact force) acting on a person standing on the rotating Earth: (i) at the equator (ii) at the poles. (c) Objects on the Earth have an apparent weight W = mg' due to the fact that the Earth is rotating. Estimate the percentage difference between their weights measured at the poles and at the equator, assuming the Earth to be a perfect sphere but taking rotation of the Earth into consideration. (Hint: weight as measured on a bathroom scale is the normal contact force of the scale on a person.) Dynamics of uniform circular motion 5 (a) Draw a free-body diagram to show the vertical forces acting on the person shown in Figure 6.19 sitting upside down in a roller-coaster seat while at the top of the motion. (Ignore the force of the safety seat-belt.) seat roller-coaster track person Figure 6.19 Person seated on a roller-coaster. (b) If the radius of the roller-coaster motion is 7.5 m, estimate the tangential speed of the roller-coaster for the person to feel 'weightless'. 6 In a 'rotor' ride a person stands on a floor against a vertical, padded wall.

The angular velocity of the wall and floor are gradually increased. The person 'feels' pressed against the wall. At a certain angular velocity, the floor is lowered and the person remains pressed against the wall, as shown in the diagram, and remains at the same place against the wall (Figure 6.20). F v N r mg Figure 6.20 A person suspended against a wall in a 'rotor' ride. (a) (i) Name the force, F, shown in the diagram. (ii) Explain in terms of the forces F and mg in Figure 6.20 why the person does not slide downward when the floor is lowered. (b) (i) Write an equation for F in terms of the force labelled N and the coefficient of static friction, μs, between the person and the padded wall. Chapter 6 Circular motion and gravitation (ii) Prove that the minimum tangential speed, v, of the rotation at which the person does not slide rg downwards is given by v = μ . s (iii) If the radius of the rotor cylinder is 4.0 m, what is the frequency of the rotor in rpm that corresponds to the minimum speed, v, in (b)(ii) above? 7 A person A is seated at the top of a

rotating Ferris wheel and person B is seated at the bottom (Figure 6.21).
A 8.4 m 9 (a) What is meant by the expression 'banking of a road?' (b) Roads are usually banked at bends so that friction is not necessary to maintain circular motion at given speeds. (i) What is the name of the horizontal non-friction force that enables circular motion on banked roads? (ii) Prove that for a given speed v on a road banked at an angle θ, the banking angle v2 where r is the equation is given by tan θ = rg radius of curvature of the bend. (iii) How can the same banking angle work for higher speeds? (c) Why are racing tracks usually banked with a curve of varying banking angles at bends on the track? 10 A photograph showed a motorcyclist leaning at an angle θ = 60° when negotiating a bend of radius 8.6 m. The coefficient of static friction between the tyres and the road is 0.80. The mass of the motorcyclist and motorcycle is 170 kg and the centre of gravity, C, of both combined is a distance 65 cm from the bottom of the motorcycle tyre. A sketch of the photograph of the motorcyclist on the motorcycle is shown in Figure 6.22. C B FN Figure 6.21 Two people seated on a Ferris wheel. (a) Draw a free-body diagram to show named forces acting on persons A and B.
(b) (i) At what tangential speed of the Ferris wheel will person A feel 'weightless'? (ii) Does this speed depend on the mass of the person? Explain. (c) If person B has a mass of 60 kg, what would be this person's apparent weight at the tangential speed of rotation obtained in part (b)? (Hint: the measured 'weight' of a person is the normal contact force of the person, e.g. on a spring bathroom scale.) 8 The coefficient of static friction between tyres and a horizontal, circular roundabout of radius 10.0 m is 0.82 on a dry day.
(a) Determine the maximum no-skid speed, in km/h, on a dry day. (b) Does the speed in (a) depend on the mass of the car? Explain your answer. θ P F Figure 6.22 A motorcyclist negotiating a bend.
(a) Determine the value of FN shown in the diagram. Explain your answer. (b) Consider moments about the centre of gravity, C. Which force produces a clockwise moment and which produces an anticlockwise moment on the motorcyclist-motorcycle system? Explain your answer. (c) Assuming no skidding was occurring in the photograph (as sketched), estimate the speed with which the motorcyclist was negotiating the bend. 101 102 Unit 1 Module 1 Mechanics 11 A passenger in a bus is holding on to a pole 60 cm from her body (Figure 6.23). The mass of the passenger is 65 kg and her centre of gravity is 90 cm above the floor. Shoulder height is 150 cm from the floor. (i) If the aircraft maintains the same speed during the banking, estimate the radius of the curve the aircraft will make. (ii) Will the aircraft remain at the same height above the ground while banking at the same speed? Explain your answer. Gravitation pole P A C 60 cm motion of bus B Figure 6.23 A passenger standing in a bus that is rounding a bend. F (a) (i) Copy the diagram and sketch the normal contact force FN of the floor of the bus on the passenger. (ii) Calculate the value of FN.
(b) The bus rounds a horizontal bend of radius 16 m at a speed of 30 km/h.
By considering moments about the centre of gravity, C, determine: (i) the moment produced by the centripetal force, F (ii) the force (magnitude and direction) that must be exerted on the passenger at the point A so as to keep her vertical during negotiating of the bend by the bus. 12 An aircraft of mass 5000 kg is flying horizontally at a speed of 300 km/h. (a) Estimate the force of lift on the aircraft as it maintains its horizontal orientation and level at the speed given. Explain your answer. (b) In order to make a turn, the aircraft banks at an angle θ = 20°, as shown in Figure 6.24. A θ B Figure 6.24 Banking of an aircraft. 13 (a) (i) State Newton's law of universal gravitation. (ii) Write an equation that illustrates the law in (a) (i) above. (b) Prove that the value of g at a height h above the Earth (mass M and radius R) is given by g = GM 2 (R + h) (c) At what height above the Earth's surface is the value of the gravitational field strength of the Earth half the value of that on the surface of the Earth? 14 (a) Calculate the value of g, the gravitational field strength: (i) at the surface of the Earth, and (ii) at the surface of the Moon.
(b) How do the values of g in (a) (i) and (ii) above compare? 15 (a) At what distance along a straight line between the Moon and the Earth is the net gravitational field strength zero? (b) A space capsule orbits the Earth in the Earth's gravitational field. How is it, then, that astronauts and objects inside the capsule orbiting the Earth experience 'weightlessness' even though the orbiting is taking place in a non-zero gravitational field of the Earth? Gravitation and uniform circular motion 16 (a) Estimate the near-surface orbital speed of a satellite near the surface of: (i) the Moon (ii) Mars. (b) Discuss qualitatively how close to the surface of each body in (a) (i) and (ii) the orbital speed) can be applied.
equation (v = GM R2 (c) Discuss one use of near-Earth satellites and one limitation.
17 Apply Newton's law of universal gravitation.
(a) Derive the equation for the orbital speed of a near-Earth satellite. (b) Determine the time taken for a near-surface satellite to orbit: (i) the Earth (ii) the Moon. Chapter 6 Circular motion and gravitation 18 (a) (i) What is meant by a 'geostationary' satellite? (ii) What is the period in seconds of such a satellite around the Earth. (b) Why are geostationary satellites possible only above the equator?
(c) Determine for a geostationary satellite orbiting the Earth: (i) the height of the satellite above the Earth (ii) the orbital speed of such a satellite. 19 Satellites are reported to be falling back to the Earth. (a) Explain the reason for their falling back to Earth.
(b) Which type of satellite is more likely to fall back to the Earth: a near-Earth satellite or a geostationary satellite? Explain your answer fully. (c) Answer part (b), but replace the word 'Earth' with 'Moon'. (d) Answer part (b), but replace the word 'Earth' with 'Mars'. Mixed exercises Practical activities and challenges 1 Using circular motion to determine coefficient of static friction 2 Investigating centripetal force Caution: wear protective eye goggles and perform these experiments well away from fragile objects or people. The centripetal force, F, on an object being whirled in a horizontal circle is given by mv2 r 2πr m × 4π2r (since v = F=) T 2 T F= where T is the period of the motion. 20 A 20 g mass, A, is placed at a varying distances, R, from the centre of a horizontal disc rotating at 40 rpm. The coefficient of static friction between the mass and the disc is 0.45. (a) At what maximum distance from the centre must the mass be placed so that it does not slip on the rotating disc? (b) If a 20 g mass were to be struck on top of the mass A, at what distance from the centre must this combination of masses be placed so that there would be no slippage on the rotating disc? Explain your answer. 21 A conical pendulum is 2.0 m long and the bob has a mass of 800 g. The string is inclined at 30° to the vertical and the bob moves at constant speed. (a) Draw a free-body diagram showing the forces acting on the pendulum bob. (b) Calculate: (i) the tension in the string (ii) the period of the motion (iii) the percentage difference between the period calculated in part (b) (ii) and that of a simple pendulum of same mass and length that oscillates through a small angle. Devise and carry out an experiment to determine the coefficient of static friction between two surfaces by using a mass on a horizontally rotating turntable. Using apparatus such as shown in Figure 6.25 (or otherwise), investigate how the centripetal force, F, 1 varies with T2 for the motion.
By whirling the ball in an approximately horizontal circle, and allowing it to move in a horizontal circle, measure the period and the corresponding tensions in the string (the latter would equal the weights of the paper clips, if the top edge of the hollow glass tube is very smooth) keeping the radius of the motion constant. The period, T, can be determined by using a digital video camera to photograph the motion. light polystyrene ball thread F hollow glass tube – smooth at the top edge to reduce friction paper clips Figure 6.25 A light ball being moving in a horizontal circle.
By plotting a suitable graph, comment on the extent to which your experimental results agree with the theoretical predictions and discuss possible sources of error. 103 104 Unit 1 Mechanics Answers to Review questions 1 The axis of rotation is perpendicular to the plane of the page, i.e. is perpendicular to the x- and y-axes and passes through the intersection of these axes. 3 (a) 20.9 rad s−1 (b) 3.1 m s−1 2 (a) Since the force of tension at A is given by mv2 FTA = mg + r , and at B is FT = mg, it follows that from A to B the force of tension is decreasing. 2 − mg, hence (b) At C the force of tension FTC = mv r 2 mv2 − mg, as the the tension decreases from mv to r r ball moves from B to C. 3 Law 1: the object is at rest and remains at rest, therefore the net force on the object is zero. This means that the normal contact force on the object FN is equal but opposite to its weight (since FN + mg = 0). 7 (b) (i) 9.1 m s−1 (ii) The mass cancels out in the equation. Thus, the speed does not depend on the mass. (c) 1200 N Law 2: the net acceleration of the object is zero. Hence, FN + mg = ma = 0, Hence, FN is equal but opposite to its weight. NOT Law 3: Law 3 refers to two and only two objects.
FN is the force of the table on the object, whereas mg is the force of the Earth on the object. Three objects, not two, as Law 3 requires, are involved here: object, table and Earth. 4 Yes, since this safe speed depends only on r, μs and (g). If r and μs are the same for all the vehicles, then the maximum safe speed would be the same. However, please see Figures 6.9a and b which address another safety issue, toppling. 5 Newton's third law of motion.
1 1 6 ( ) × F (since F is proportional to 2). 9 d 7 24 hours.
5 (b) 8.6 m s−1 11 (a) (i) 640 N (ii) (i) 250 N (anticlockwise) (ii) 360 N towards the person's shoulder 13 (c) 2640 km 15 (a) 346 × 103 km (b) Both are experiencing the same (centripetal) acceleration as the capsule. Hence they are not attracted to each other, nor to the capsule. Therefore the astronauts appear 'weightless' in the capsule. 105 Module 2 Oscillations and waves Chapter 7 Harmonic oscillations Learning objectives ■ Define simple harmonic motion (SHM). ■ Derive equations for acceleration (a), displacement (x), velocity (v) and period (T) for a particle undergoing SHM. ■ Apply SHM equations involving acceleration, displacement, velocity and period to solve problems. ■ Outline the conditions necessary for SHM. ■ Sketch and interpret the following types of graphs for SHM: displacement versus time velocity versus time ●● velocity versus displacement ●● acceleration versus displacement. ●● ●● (g), and period for m) . of a mass-and-spring system (T = 2π k Derive and use equations for energy in SHM. Describe, with the aid of graphs, the potential energy and kinetic energy changes that occur during SHM. Describe the effect of forced oscillations on resonance.
Discuss cases in which resonance is desirable and not desirable. Describe, with the aid of graphs, over-damping, underdamping and critical damping. Describe, with explanations, how damping is achieved in various applications. ■ Derive and apply formulae for the period of a simple pendulum (T = 2π ■ ■ ■ ■ ■ ■ ■ Simple harmonic motion (SHM) Repetitive motions that occur in a straight line, or nearly straight line, are called oscillations. A small boat, tied near to the shore, bobs up-and-down. A child on a swing moves to-and-fro in an arc. The surface of the Earth shakes side-to-side during an earthquake (Figure 7.1). All of these motions are oscillations. This chapter focuses on a special type of oscillation called simple harmonic motion (SHM). Figure 7.1 Earthquakes involve harmonic oscillations of the Earth's surface and can lead to catastrophic destruction. But why was the upper level of the roadway destroyed, but not the lamp posts, the tree or the fence under the roadway? A study of harmonic oscillations might give us a clue as to the answer. 106 Unit 1 Module 2 Oscillations and waves Consider a point, Q, which moves at constant angular speed, ω, anticlockwise from the x-axis. The angular velocity, ω, of the particle is defined as the angle swept out per unit time. In terms of an equation, heavy mass (resists motion) ω= paper (moves in direction of arrow) θ t ....(7.3) pen y Q v M Earth motion due to earthquake x1 Figure 7.2 A seismograph records the side-to-side motion of the Earth's surface. This motion approximates to simple harmonic motion. Note that the heavy mass shown is not a pendulum. Rather, the heavy mass remains practically stationary as the Earth and the rest of the apparatus sitting on it shakes from side to side. Simple harmonic motion Simple harmonic motion is motion along a straight line in which acceleration is directly proportional to displacement and is always directed opposite to the displacement. Since, in SHM, acceleration, a, is directly proportional to displacement, x, then a∝ x, or a = constant × x Since acceleration is always directed opposite to displacement, then, a = − constant × x...............(7.1) The '−' sign indicates that if x is positive, a is negative, and vice-versa, i.e. a is always directed opposite to x. It is customary to use ω2 (angular velocity squared) as the constant when defining SHM in terms of a and x. Equation 7.1 then becomes a = −ω2x...............(7.2) Angular velocity, ω, is defined in terms of circular motion, as the angle swept out per unit time. SHM, however, does not involve angles since this motion takes place in a straight line. We therefore need to establish the connection between SHM, which is linear, and uniform circular motion, which is angular.
SHM and uniform circular motion Consider a particle, M, moving at constant speed v in a circle of radius r in the x-y plane (Figure 7.3).
Suppose the clock is started when the particle is crossing the x-axis at T at time t0 = 0 s, and that at time t the radius has swept θ1 θ1 O θ x B P x Figure 7.3 A particle moving in a circle at constant speed v in the x-y plane. Getting it right! In circular motion, the angle θ is in radians, not degrees (see Chapter 6 for the definition of the radian).
Angular velocity is in radians per second (rad s−1).
Some text books use the term 'angular frequency' or 'angular speed' instead of the term 'angular velocity'. In this chapter, they all mean the same thing. In Chapter 6, we obtained the following equations for uniform circular motion of period T or frequency f, ω= 2π T ....(7.4) or, ω = 2πf................(7.5) A close look at the projection, B, of the particle, M, on the x-axis shows that both the particle and its projection have the same period (i.e. the time taken for M to complete one circle is equal to the time taken for B to make one complete to-and-fro motion from P). Hence, by equation 7.4, they have the same 'angular velocity', ω. We can imagine the projection of M on the x-axis as the 'shadow', B, of M if a light were coming from above and then from below the circle. The to-and-fro motion of B along the diameter of the circle is another way of defining SHM (but in terms of uniform circular motion).
ITQ 1 In Figure 7.3, a clock was started at t0 = 0 s when the particle M was crossing the y-axis at Q. Show that we would get, for an angle θ1 and time t1, x1 = A sin (ωt1)...............(7.7) where A is the radius, r, of the circle. Chapter 7 Harmonic oscillations Simple harmonic motion in terms of circular motion Simple harmonic motion is the projected motion, along a diameter, for a particle moving in a circle at constant speed.
Therefore, the period of SHM, defined this way in terms of projection motion in a straight line, could be given by equation 7.4, by T = 2π ω where ω represents the angular velocity of the SHM in rad s−1, even though the SHM is linear and not circular.
The equations of SHM The minus sign indicates that the direction of vx is opposite to that of x. If we omit subscript x, with the understanding that the motion of B is along the x-axis, we get v = −v0 sin θ or, v = −rω sin θ...............(7.9) θ or, since ω = , t v = −rω sin (ωt)...............(7.10) or, v = −v0 sin (ωt)...............(7.11) Velocity, v, as a function of displacement in triangle OMB (Figure 7.4) MB OM Displacement, x sin θ = Referring to Figure 7.3, we see that the displacement, x, of B (the projection of M) is given by Therefore, MB = OM sin θ = r sin θ = r cos θ θ But since ω = (equation 7.3), we get t x = r cos (ωt) Therefore, r sin θ = √r2 − x2 But, by equation 7.9, v = −ωr sin θ And since r = A, we get The maximum value of x, called the amplitude, A, of the SHM, is equal to r. Therefore, x = A cos (ωt)...............(7.6) Velocity, v, as a function of time v = −ω √r2 − x2 Therefore, v = ±ω√A2 − x2...............(7.13) When x = 0, i.e. when B is passing through the centre of the SHM along the x-axis, we therefore get v2 = ω2A2 or v = ωA...............(7.13) vy M B v = −ω √A2 − x2 Equation 7.12 tells us that the velocity, v, is a maximum as the projection, B, passes through the centre of the SHM, i.e. when x = 0. x Figure 7.4 Components of velocity for a particle undergoing uniform circular motion. In Chapter 6 we obtained the equation for the magnitude of the velocity, v0, of a particle undergoing uniform circular motion as v0 = rω...............(7.8) Applying equation 7.8 to Figure 7.4, we get vx = −v0 sin θ ITQ 2 What will be the speed of a particle at the amplitude position of its SHM? Explain your answer in terms of equation 7.12. Getting it right! The angle, 4.197, in the calculation in Worked example 7.1 is in radians, not degrees. This is because angles in SHM are defined in terms of radians. Hence your calculator must be switched from deg to rad when finding sin 4.197.
107 108 Unit 1 Module 2 Oscillations and waves Worked example 7.1: Velocity of a particle undergoing SHM Q A A particle moves in SHM with a period of 2.41 s and amplitude 4.53 cm. The particle is at the positive amplitude position when timing begins at t0 = 0 s. Calculate: (a) the maximum speed of the particle (b) the velocity of the particle at time t = 1.61 s.
= 2.41 s ■ amplitude, A = 4.53 cm = 0.0453 m ■ time, t = 1.61 s (a) According to equation 7.12, the maximum speed occurs when x = 0. Therefore, maximum speed = ωA (by equation 7.13) 2π = × A T 2π = × 0.0453 2.41 = 0.118 m s−1 = 11.8 cm s−1 (b) velocity, v = −v0 sin (ωt) 2π × 1.61 s = −0.118 m s−1 × sin 2.41 s = −0.118 m s−1 × sin 4.197 = −0.118 m s−1 × −0.870 = 0.103 m s−1 = 10.3 cm s−1 Worked example 7.2: Particle coming off a piston Q M A ■ period, T ( ) Acceleration, a, as a function of time In Chapter 6 we obtained the equation for acceleration, a, of a particle moving at constant speed in a circle of radius r as a = −ω2r...............(7.14) for a particle moving at constant speed in a circle of radius r.
The acceleration, a, always points towards the centre of the circle and is called a centripetal (i.e. centre-seeking) acceleration (Figure 7.5). O 16.0 cm = 2 Figure 7.6 A mass resting on an oscillating piston. A The mass, M, would just lose contact with the piston when the downward acceleration is equal to the acceleration due to gravity. ■ amplitude, A = 16.0 cm ÷ 2 = 8.0 cm = 0.080 m ■ for SHM, a = −ω2x ■ at the positive amplitude position, a = −ω2A since a = g at this position, so −g = −ω2A The minus sign in front of g indicates the downward direction of g, with the downward direction being taken as negative.
we get, −9.80 m s−2 = −ω2 × 0.080 m 9.8 m s−2 ω2 = 0.080 m 4π2 T2 = 122.5 s−2 since ω = 2π T by equation 7.5 f = 3.1 Hz (since f Hz = 1 s−1) From Figure 7.5, we get ax = a cos θ ay θ Therefore, ax = −ω2r cos θ, and, by equation 7.6 (equation 7.14) Therefore, ax = −ω2A cos (ωt), since r = A and ω = θ t Therefore, ax = −ω2x (since x = r cos θ) M ay a O A, rests on a piston that is undergoing SHM through a displacement of 16.0 cm as shown in Figure 7.6. At what frequency, f, will the mass just lose contact with the piston at the amplitude position? θ B x Omitting the subscript x as before, bearing in mind that the motion of B is along the x-axis, we get Figure 7.5 Acceleration of a particle undergoing uniform circular motion.
a = −ω2A cos (ωt)...............(7.15) and, a = −ω2x...............(7.16) Chapter 7 Harmonic oscillations +A x O Graphs of SHM x = A cos (ω t) π 4 π 2 3π 4 π 5π 4 3π 2 7π 4 2π 0 –A +v0 v v = – v0 sin (ω t) O π 4 π 2 3π 4 π 5π 4 3π 2 7π 4 2π 3 ωt –v0 a = – ω 2 A cos (ω t) +ω 2A The key to sketching graphs of SHM in terms of time is to use times that are fractions of the period, like 0, T/4, T/2, 3T/4 and T during one cycle (these values of time correspond to values of θ (or ωt) of 0, π/2, π , 3π/2 and 2π radians). Figure 7.7 shows variations of displacement, velocity and acceleration as functions of time during SHM. Figure 7.8 shows the variation of acceleration and velocity with displacement. Equations are given for each graph. Mass-spring oscillators Horizontal mass–spring oscillator O a π 4 π 2 3π 4 π 5π 4 3π 2 7π 4 2π −ω 2A Figure 7.7 Graphs of displacement, x, velocity, v, and acceleration, a, as functions of time, t, during SHM. F O a a Figure 7.9 Horizontal mass–spring oscillator. a = −ω 2x 2 +ω A +A –A x –ω 2A b v +v0 –A +A x –v0 Figure 7.8 a A graph of acceleration, a, versus displacement, x. b A graph of velocity, v, versus displacement, x. Figure 7.9 shows a mass, m, connected to a horizontal spring. ... By Newton's second law F = ma Therefore, ma = −kx v = ± ω √A2 − x2 ... = −A Figure 7.9 shows a mass, m, connected to a horizontal spring.
The other end of the spring is fixed and the mass rests on a frictionless surface. The mass is pulled a distance x from its equilibrium position, O. At the distance x the only horizontal force acting on the mass at the time of release is due to the spring and is given by Hooke's Law as F = −kx (the minus sign indicates that the direction of F is opposite to that of x). k x k m = ...(7.17) Equation 7.17, with the constant, k/m, is of the same form as equation 7.16 (a = −ω2x). Hence the motion of the mass is SHM. This means that for the mass attached to the spring, k...............(7.18) m 2π Since ω = (equation 7.4), equation 7.18 becomes T 2 4π = k T 2 m k T 2 m Hence, by equation 7.4, we can obtain the period, T, of the motion of the mass as ω2 = −aA Figure 7.8 Variation of acceleration, a, and velocity, v, with displacement, x, during SHM. T = 2π Equation 7.16 is identical to equation 7.2, which is the equation that defines SHM in terms of acceleration and displacement.
This result also demonstrates that the identification of the constant in the original definition of SHM (equation 7.1) with ω2 is justified. m k ....(7.19) We can now arrive at yet another definition of SHM, since it is a force, F, that obeys Hooke's law that results in this type of motion. SHM is the motion of an object that results from a force that obeys Hooke's law acting on it. 109 110 Unit 1 Module 2 Oscillations and waves Worked example 7.3: Mass–spring oscillator L O F1 Q A 2.0 kg mass is hung from the free end of a spiral spring. The spring stretches 3.0 cm and the mass settles in this equilibrium position.
At time t = 0, the mass is pulled an additional 5.0 cm and allowed to oscillate. Determine: (a) the spring constant, k, of the spring (b) the equation of motion of the displacement, x, as a function of time, t (c) the total energy of the system, assuming zero energy at the equilibrium position. A ■ mass, m e O F2 x mg mg Figure 7.10 Vertical mass–spring oscillator Vertical mass–spring oscillator Equation 7.19 for the period of a horizontal mass–spring oscillator is the same for a vertical mass–spring oscillator, assuming the mass of the spring is negligible. We show this in the derivations below. We take the upward direction as positive and downward as negative. Figure 7.10a shows an unstretched spring, of negligible mass, suspended from a firm support. In Figure 7.10b, a mass, m, hung on the spring, causes the latter to extend by an amount e. The mass settles at an equilibrium position, O, due to the weight of the mass and the force, F1, of the spring on the mass. Applying Newton's second law to Figure 7.10b, we get F1 − mg = 0 i.e. −ke − mg = 0...............(7.20) In Figure 7.10c, the mass is pulled down and an additional displacement x from O. At the point of release, and applying Newton's second law, we get F2 − mg = ma i.e. −k(e + x) − mg = ma or −ke − kx − mg = ma i.e. 2.0 kg ■ extension, ΔL = 3.0 cm = 0.030 m ■ amplitude, A = 5.0 cm = 0.050 m (a) By Hooke's law, mg = kΔL mg Therefore, k = ΔL 2.0 kg × 9.8 m s−2 = 0.030 m = 650 N m−1 (b) The mass oscillates in a vertical direction when released after being pulled down. We take the downward direction as positive.
Since a mass–spring system displays SHM, the equation of motion would be of the form y = A cos (ωt). The cosine is chosen rather than the sine, since at time t = 0, the mass, when pulled and released, was at the amplitude position, A = 5.0 cm (0.050 m). At t = 0, cos (ωt) = cos 0 = 1, and therefore y = A. The sine function would have given us y = 0 rather than y = A at time t = 0. To obtain the equation of the motion specific to this situation, we need to calculate ω. Based on equation 7.18), we get Using ω2 = (k/m) 650 N m−1 = 326.5 (rad s−1)2 ω2 = 2.0 kg Therefore, ω = √326.5 rad s−1 = 18.1 rad s−1 = 18 rad s−1 (to 2 sig. fig.) Therefore, y = 0.050 cos (18t) is the required equation. (c) total energy = 12 kA2 = 12 × 653 N m−1 × 0.0502 m2 = 0.82 J or −kx = ma (since, by equation 7.20, −ke − mg = 0) We therefore get, a= − k x m which is the same as equation 7.17, obtained in the case of the horizontal mass–spring. Hence, the period of the vertical mass–spring oscillator is also given by T = 2π m k ITQ 3 A car frame rests on four, vertical, identical springs each of spring constant 8 000 N m−1. What is the net effective spring constant of the springs supporting the car? Chapter 7 Harmonic oscillations Energy of a mass–spring oscillator The energy of a mass–spring oscillator consists of potential energy, PE, stored in the stretched (or compressed) spring and kinetic energy, KE, of the moving mass. At any instant of time, if the spring is stretched by an amount x and the mass is moving at a speed v the total energy, E, of the system is given by E = PE + KE 1 h k1 F2 F2 k2 W According to equation 4.15 (Chapter 4), the energy stored in an elastic spring of spring constant k and stretched by 1 an amount ΔL is given by PE = 2 kΔL2. If the spring was originally stretched a distance A from O, just before release, the total energy would 1 be 2 kA2, which is the energy put into and stored in the stretched spring. By conservation of energy, 1 kA2 = 12 kx2 + 12 mv2...............(7.21) 1 2 2 kA F W Figure 7.12 Springs hung: (a) in parallel and (b) in series. Series arrangement 1 = 2kx2 + 2 mv2...............(7.22) Figure 7.11 shows the variation of potential energy and kinetic energy with displacement, x, of a mass during SHM. For springs in series (Figure 7.12b), the force, F, throughout the system is the same.
However, the springs extend by different amounts, x1, and x2, respectively. We get, F = −k1x1 and F = −k2x2 Springs in series and parallel Therefore, x1 = − Parallel arrangement Consider a force W which causes two springs arranged in parallel (Figure 7.12a) to stretch the same amount, x. By Hooke's law, and using the symbols displayed in the diagram, we get F1 = −k1x and F2 = −k2x Therefore the net restoring force, F, is given by But the total displacement, x, of the mass is given by x = x1 + x2.
Therefore, x = − = −k1x − k2x Therefore, F = −(k1 + k2)x Since F = −kx, when k is the effective spring constant, we get, for a parallel arrangement, k = k1 + k2...............(7.23) − F k F F F =− − k k1 k2 F 1 1 = −F k k1 k2 (Therefore,) 1 1 1 = + k k1 k2 ( ) ....(7.24) The simple pendulum E KE PE −A F F − k1 k2 But, as before, F = −kx, where k is the effective spring constant.
The mass attached to the string in a simple pendulum (Figure 7.13) exhibits SHM for small angles of swing, θ (i.e. θ < 10°). For a bob of mass m, the restoring force F = −mg sin θ (the minus sign indicating that the direction of the force is opposite to the direction of increasing θ). For small θ, sin θ ≈ θ (see Appendix 3) 111 112 Unit 1 Module 2 Oscillations and waves Worked example 7.4: Simple pendulum Q A chandelier suspended from a hotel ceiling makes 10 complete oscillations in 32.4 s. (a) Estimate the oscillating length, L, of the chandelier. (b) Justify why equation 7.26 can be used in arriving at the estimate. A (a) period, T = 32.4 s ÷ 10 = 3.24 s l Since T = 2π g (equation 7.26) l B A x F C θ mg cos θ mg 4π2l g T 2g therefore l = 4π2 3.242 s2 × 9.80 m s−2 = 4π2 = 2.6 m (b) Equation 7.26 is valid for small angles of oscillation, θ, and the angle the chandelier makes is likely to be small. Equation 7.26 is valid for small angles of swing (θ < 10°). For the chandelier, since θ is justified in the case of the chandelier, then, T 2 = Figure 7.13 A simple pendulum. Note: you can verify this yourself for an angle of 10°. 10 π° = rad = 0.1745 rad 180 You will get sin (0.1745 rad) = 0.1736 = 0.1745 (Remember to switch your calculator to the radians mode when finding sin 0.1745.) Therefore, F = −mg sin θ = −mgθ AC l mg × AC Therefore F = − l Forced oscillations and resonance But θ = arc length Here is something to try, as an introduction to the subject of resonance. But for small θ, AC = x ■ Cut a piece of thin cardboard in the shape shown in mg Therefore F = − l By Newton's second law, we get mgx ma = − l g Therefore, a = − x l ( ) Figure 7.14. Note that the base is folded over so that the base can be firm. 4 cm Q ....(7.25) thin cardboard As, a = −ω2x for SHM (equation 7.2) Equation 7.25 shows that the pendulum bob undergoes SHM, such that ω2 =l 4π2 i.e. 2 = T l or T = 2π l g 4 cm P ~25 cm ....(7.26) base folded over for firmness Figure 7.14 Thin cardboard shape for demonstrating resonance. ITQ 4 On Earth, a mass hung on a particular vertical spring and a particular simple pendulum have the same period. Will their periods be the same on the Moon?
Explain your answer. Chapter 7 Harmonic oscillations ■ Hold the base firmly with one hand and pluck the ■ Mechanical resonance is used in the ultrasound shattering of tiny bone outgrowths called spurs. When the frequency of the ultrasound matches that of one of the tiny bone outgrowths, the outgrowth resonates with large amplitude and breaks off the bone. end P and then pluck the end Q. Note that the natural frequency of vibration of P is greater than that of Q.
■ Now shake the system by its base, starting from a low frequency and increasing to a higher frequency. You will see, quite dramatically, that P vibrates with large amplitude at a particular frequency of vibration of the hand close to the natural frequency of P. Q vibrates with large amplitude when energy is being supplied at a particular lower frequency which is close to its natural frequency of vibration. This simple exercise shows that when energy is supplied to an object at its natural frequency of oscillation, the object vibrates at large amplitude. This phenomenon is called resonance. ■ Electrical resonance is used in the tuning of radios and TVs. When the natural frequency of a radio is 'tuned' to a particular frequency, only broadcast signals coming from a station at that frequency will be received loudly. Stations broadcasting at any other frequency will not cause resonance in the tuned radio and therefore will not by heard at all, or heard only faintly.
■ Magnetic resonance imaging (MRI) is used in diagnostic imaging of various parts of the body. Basically, radio frequencies are aimed to the desired part and the amount of resonance occurring in hydrogen nuclei is detected, which gives rise to a special type of image of that part of the body. A f0 f Figure 7.15 A system with natural frequency f0 at resonance. When resonance occurs, the amplitude, A, of vibration is at a maximum. Figure 7.15 shows the variation of amplitude, A, with the frequency, f, of the energy supplied. At a particular frequency, a natural frequency of the object, vibration takes place at large amplitude – i.e. resonance occurs. Importance of resonance Earthquakes occur at very low frequencies.
Tall buildings and large bridges also sway at low frequencies naturally. One of the reasons why bridges and tall, heavy buildings collapse during earthquakes, whereas shorter buildings do not, is resonance. When the frequency of the energy supplied, whether by earthquake waves, water waves or by wind gusts, matches the natural frequency of a structure, resonance occurs and the structure vibrates at large amplitudes, often with resulting damage. Forms of resonance can also find good use. Damped oscillations Observations of a mass–spring oscillator and a simple pendulum show that their amplitudes of motion decrease with time. This decrease of amplitude with the passage of time is called damping. Graphs of damping +A 0 A –A no damping +A x O –A critical damping +A ITQ 5 Explain why it is not uncommon for glass windows to rattle only when certain bass notes are coming from the loudspeakers at a party. O –A over-damping At the instant of time of impact, ... to the frequency of the lamp that damped oscillations x, of a ... In Figure 7.1 collapsed, but the lamp post near the roadway didn't. d T 4 T 2 3T 4 T 5T 4 3T 2 7T 4 t Figure 7.16 Graphs of displacement, y, versus time, t, to illustrate various types of damping. 113 114 Unit 1 Module 2 Oscillations and waves Figure 7.16a shows the variation of displacement, y, with time, t, for an object undergoing SHM, assuming no damping. In Figure 7.16b, the amplitude of the oscillating object decreases with the passage of time as the object makes several oscillations.
This is called under-damping. Figure 7.16c shows the graph of an oscillating object reaching the equilibrium position in the shortest possible time and remaining in that position. This kind of damping is called critical damping.
The damping shown in Figure 7.16d is so huge that the object does not oscillate at all, but very slowly reaches the equilibrium position. This kind of damping is called over-damping. Damping in practical situations Un-damped oscillations are not common in mechanical systems. This is because, with movement, a system loses energy doing work to overcome friction. Un-damped oscillations occur at the atomic and molecular levels when these particles are very close to each other and exhibit vibrations. Under-damping occurs in the mass-spring and pendulum systems.
As the masses move through the air they do work against friction and energy is lost as heat, resulting in the decreasing amplitude. Strings in stringed instruments (e.g. guitars and pianos) also show underdamping. The under-damping makes the sounds pleasant to the ears. In some pianos, however, a special pedal is connected to a damper which causes the damping time to shorten and produce a quicker decay of the sound, to produce a special musical effect. Critical damping is employed in situations where it is desirable to reach equilibrium position in the shortest possible time. Spring-operated doors in air-conditioned rooms are often critically damped for this reason. Shock absorbers on cars also make use of critical damping. To achieve critical damping in cars a piston with holes provides the frictional damping force as it moves through a special fluid contained in a cylinder (Figure 7.17). The same method of critical damping is applied in tall buildings in earthquake-prone areas to reduce swaying during an earthquake. Over-damping is often used on exit doors to buildings in which large numbers of people are gathered (e.g. auditoriums), or to accommodate people who are physically challenged. This is because the door must close slowly for safety reasons, especially in an exit emergency. A quickly closing door can injure persons trying to exit. The fluid used in the spring damper mechanism is very viscous, providing a large amount of friction as the door tries to close. This causes the door to close very slowly and without oscillation. spring holes in the piston viscous oil Figure 7.17 A motor car shock absorber. ITQ 7 Should an exit door to an air-conditioned cinema be overdamped or critically damped? Explain your answer.
Chapter 7 Harmonic oscillations Summary ✓ Simple harmonic motion (SHM) is motion along a straight line in which acceleration, a, is directly proportional to displacement, x, and is always directed opposite to the displacement. In equation form, a = −ω2x...............(7.2) where ω is the angular velocity of the motion. (Angular velocity ω = 2π for motion with period T) T ✓ In terms of a force obeying Hooke's law, SHM is the motion of an object that is being acted on by a force which obeys Hooke's law. ✓ In terms uniform circular motion, SHM is the projected motion, along a diameter, for a particle moving in a circle at constant speed.
✓ Equations for SHM, assuming that at time t = 0 the object is passing the positive amplitude position, A, are as follows: ■ ■ displacement Note: if at time t = 0, the object is passing through the equilibrium position, the equation for the displacement is x = A sin (ωt)...............(7.7) ■ acceleration a = −ω2A cos (ωt)...............(7.15) Equations in terms of displacement, x ■ acceleration a = −ω2x...............(7.16) ■ ■ period m k ■ ■ potential energy T = 2π ....(7.19) 1 PE = 2 kx2 ■ kinetic energy 1 KE = 2 mv2 ■ total energy 1 1 1 E = 2 kA2 = 2 kx2 + 2 mv2...............(7.22) ✓ The effective (net) spring constant k of springs with spring constants k1 and k2 respectively are: ■ arrangement k = k1 + k2...............(7.23) ■ series arrangement (1 1 1 = + k k1 k2 ....(7.24) through a small angle, the motion is very nearly SHM and the period is given by T = 2π l g ....(7.26) ✓ Resonance occurs when energy is supplied to an object at its natural ✓ velocity v = −v0 sin (ωt)...............(7.11) ■ ■ ■ ✓ For a simple pendulum of length l swinging x = A cos (ωt)...............(7.6) ✓ For a mass–spring oscillator, the equations are: a system at one of its natural frequencies of oscillation. When resonance occurs, the system vibrates with very large amplitude. ✓ Damping is the decreasing of the amplitude of motion with the passage of time. In mechanical systems, damping is caused by friction.
■ ■ In under-damping, the system makes several oscillations before settling in the equilibrium position. ■ ■ In critical damping, the system does not oscillate, but settles in the equilibrium position in the shortest possible time, usually within a quarter of a period. ■ ■ In over-damping, the system does not oscillate, but takes a very long time to reach the equilibrium position. velocity v2 = ω2(A2 − x2)...............(7.12) or, v = ω√A − x 2 2 115 116 Unit 1 Module 2 Oscillations and waves Review questions Simple harmonic motion 1 (a) (i) Define simple harmonic motion (SHM) in terms of acceleration and displacement. (ii) Express part (a) (i) above in equation form, explaining all the symbols used.
(b) What are the conditions necessary for SHM? (c) A mass undergoing SHM has the acceleration of the particle when it is 3.6 cm from the equilibrium position. 2 A particle undergoing SHM passes through the equilibrium position at time t0 = 0 s. The particle oscillates with a frequency of 21 Hz and amplitude of 3.7 cm. (a) Determine the angular frequency of the SHM. (b) Obtain an equation, making use of the data given, for the displacement, x, as a function of time. (d) Calculate the velocity, v, at time t = 5.9 s. (e) If the particle was, instead, at the positive amplitude position at time t0 = 0 s, what would have been the equation for the displacement in terms of time, t? 3 (a) Using equations of SHM, derive expressions for: (i) acceleration, a, in terms of displacement, x (ii) velocity, v, in terms of displacement, x (b) Sketch graphs for each of parts (a) (i) and (a) (ii). (c) A mass undergoes SHM with amplitude 6.0 cm and period 3.1 s. At time t0 = 0 s, the mass was at the positive amplitude position. Determine at time 7.3 s: (i) the displacement of the mass (ii) the acceleration of the mass (iii) the velocity of the mass. 4 A point P on a guitar string vibrates with a frequency of 200 Hz and amplitude 0.30 cm. At time t0 = 0 s, the point P was passing through the equilibrium position. (a) Sketch a graph of displacement, y, versus time, t, for a time of 0 to 10 ms. (b) Write an equation for the displacement, y, of P, as a function of time, t, making use of data given in the question. (c) Find the maximum acceleration of the point P. 5 (a) Write one SHM equation each for the following: (i) displacement, x, versus time, t (ii) velocity, v, versus time, t (iii) acceleration, a, versus time, t (b) Sketch a graph in each of parts (a) (i), (a) (ii) and (a) (iii) over two complete cycles of the SHM, in terms of period, T. (c) A mass undergoes SHM according to the 3n equation x = 0.042 cos (2)t, where x is in metres. Determine: (i) the angular frequency of the SHM (ii) the velocity of the mass at time t = 1.5 s (iii) the acceleration of the mass at time t = 1.5 s.
Mass-spring oscillations 6 (a) A horizontal spring of spring constant k is attached at one end to a fixed wall and to a mass, m, at the other end. The mass rests on a frictionless surface. The mass is pulled a distance A from its equilibrium and then released. (i) Show that the mass executes SHM upon release. (ii) Derive an expression for the period of the SHM in terms of m and k. (b) For mass m = 150 g, spring constant k = 28 N m−1 and distance A = 4.5 cm, determine: (i) the period of the SHM (ii) the total energy of the mass–spring system. 7 A mass of 250 g hung on a spring of length 11.8 cm causes the spring to stretch to 14.2 cm. Determine: (a) the force constant, k, of the spring (b) the frequency with which the spring oscillates when stretched and then released. 8 (a) Derive the equations: (i) k = k1 + k2, and (ii) 1 = 1 + 1 k k1 k2 for the effective (net) spring constant of springs of force constants k1 and k2, arranged (i) in parallel and (ii) in series. (b) Two identical springs of force constant 30 N m−1 are connected first in series and then in parallel.
A 110 g mass, hung from each arrangement in turn, undergoes SHM. Determine: (i) the effective force constant in each arrangement (ii) the period of the SHM in each case. Chapter 7 Harmonic oscillations 9 (a) Show that the energy, PE, stored in a spring compressed (or extended) a distance x is given by 1 PE = 2 kx2. (b) A toy dart gun is set by pushing a safety plastic, rubber-tipped dart (mass 40 g), compressing a spring in the barrel of the gun. If a force of 2.0 N is required to compress the spring 5.5 cm, determine (i) the energy stored in the compressed spring (ii) the speed with which the dart leaves the gun when the trigger mechanism is released. 12 A vertical spring is compressed a distance x = 6.0 cm, and a 50 g mass is then placed on a light platform attached at the top of the spring (Figure 7.18). If the spring force constant is 160 N m−1, estimate the speed with which the mass leaves the platform when the platform is released. (Note: in this case there are two forms of potential energy to consider – elastic PE in the spring and gravitational PE of the mass.) m = 50 g x 10 In an experiment with masses m hung from a spring of force constant k, a student obtained the following results. Mass, m No. /g 1 90 2 3 Time for 20 oscillations t1 t2 / s kg 0.090 t2 / s 7.50 7.53 100 8.04 7.99 110 8.36 8.41 4 120 8.76 8.80 5 130 9.13 9.06 6 140 9.49 9.45 tav / s 7.52 Period Period2 T/s T2 / s2 0.376 spring 0.141 (a) Complete the table. The first row has been done for you. (b) Plot a graph of T2 (s2) versus m (kg).
(Remember: T = 2π m) k (c) From the graph in part (b), determine the spring constant, k. Show clearly how you arrived at your answer. 11 (a) The total energy of a mass–spring oscillator is given by 1 E = 2 kA2 1 light platform 1 = 2 kx2 + 2 mv2 (i) Explain the meaning of each symbol (letter) used. (ii) State what each of the terms represents. (b) Using the same pair of axes, sketch graphs of potential energy Ep, kinetic energy Ek, and total energy E, of the system as a function of displacement x, for x = −A to x = +A. Figure 7.18 13 (a) With the aid of a diagram, derive the equation T = 2π gl for a simple pendulum. State assumptions made at appropriate steps of the derivation. (b) Determine the period of a 1.4 m long simple pendulum on Earth.
(c) The equation in part (a) was obtained from the horizontal component of the tension in the string. Determine the period of the pendulum in part (b) if it was suspended in a lift that was accelerating downward with an acceleration of 0.40 m s−2. 14 A 200 g mass attached to a spring oscillates with a period of 1.5 s and amplitude 5.8 cm. Determine: (a) the angular frequency of the mass (b) the kinetic energy of the mass at the amplitude positions (c) the kinetic energy of the mass as it passes the equilibrium position. Forced oscillations, Damped oscillations 15 (a) What is meant by the term 'resonance'?
(b) What is the condition for resonance to occur? (c) Sketch a graph of amplitude, A, versus frequency, f, for a system that has two pronounced resonant frequencies of 60 Hz and 120 Hz. 117 118 Unit 1 Module 2 Oscillations and waves 16 Describe and explain one situation in which resonance is: (a) desirable or helpful (b) undesirable or destructive. 17 (a) What is meant by the term 'damping'? (b) With the aid of suitable graphs, describe what is meant by 'under-damping', 'over-damping' and 'critical damping'. (c) Give one argument for and one argument against using critical damping of a spring-operated supermarket door. (d) Describe, with the aid of a diagram, how damping is achieved in a car shock absorber. 3. If you love model-making, here is one you can try. Our theory about hearing is based on the fact that the cochlea inside the ear has hairs of various lengths, each connected to a specialised cell. Each length resonates at a different frequency.
The resonance causes the basal cell to trigger electrical impulses. Construct a model cochlea using various lengths of thin cardboard that resonate at different frequencies (see Figure 7.19). An audio oscillator connected to an amplified speaker, or a musical keyboard in the organ mode, can serve as the source of various audio frequencies. thin cardboard Practical exercises and challenges 1 You are given a spiral spring, a metre rule, a set of hanging masses and a clamp and stand. (a) Determine the spring constant of the spring by a Hooke's law method. (b) Using the value obtained in part (a), determine the mass of an unknown mass using a mass–spring oscillation method. Note: in each of parts (a) and (b), plot suitable graphs to analyse the data obtained. 2 Liquid dampers are used as damping fluids in shock absorbers or damping mechanisms in spring-operated doors. Design and carry out an investigation to compare the damping effectiveness of motor oils of the same grade (e.g. SAE 30 or 40) but from different manufacturers. Suggested equipment could include: a mass hanger, a slotted mass, a spring and a stopwatch. An estimate should be made of the percentage error in the results so that a meaningful comparison can be made. variable frequency sound source Figure 7.19 Model of a cochlea unravelled. Answers to ITQs 1 Referring to Figure 7.3, sin θ1 = x1 r Therefore, x1 = r sin θ1 = A sin θ1 (Since r = A) But, ωt1 = θ1 Therefore, x1 = A sin (ωt1) 2 When x = ±A, v = ±ω√A2 − A2 = 0 Therefore, the speed v = 0 at the amplitude position. 3 The springs are in parallel. Therefore net spring constant, k = k1 + k2 + k3 + k4 k = 4 × k1 = 4 × 8000 N m−1 = 32 000 N m−1 4 You have to look at the equations for the periods in order to decide. The period of the mass-spring system would be the same on the Earth and Moon since the equation, T = 2π m does not involve g, even though g k on the Moon is less than g on Earth. 5 Rattling is the glass vibrating with large amplitude. This occurs when the frequency of the sound hitting the glass is equal to one of the natural frequencies of vibration of the glass. Chapter 7 Harmonic oscillations 6 The natural frequency of vibration of the lamp is very low, since it has a small mass. The natural frequencies of vibration of the lamp post is much higher, since its mass is small. Since earthquakes occur at low frequencies, the roadway would more likely resonate at large amplitude and therefore suffer damage. 7 Critical damping would help save on air-conditioning energy costs since the door shuts quickly. The quick closing might turn out to be a hazard, though, especially if a crowd is trying to exit in an emergency. Over-damping might be the better option from a safety point of view. Answers to Review questions 1 (c) (i) 0.65 s (ii) 3.8 m s−1 (toward the equilibrium position) 3 (c) (i) −3.9 cm (ii) 16 cm s−2 (iii) −10.3 cm s−1 5 (c) (i) 4.7 rad s−1 (ii) −3.3 m s−1 (iii) −0.66 m s−2 7 (a) 102 N m−1 (b) 1.86 Hz 9 (i) 1.7 m s−1 13 (b) 2.38 s = 2.4 s (c) 2.42 s = 2.4 s 119 120 Chapter 8 Waves and their behaviours Learning objectives ■ Define and use the following terms to describe wave motion: period, frequency, displacement, ■ ■ ■ ■ ■ ■ ■ ■ amplitude, wave speed, particle speed, phase difference, intensity, polarization. Use the relationship: intensity ∝ (amplitude)2 ■ Differentiate between transverse and longitudinal waves.
Differentiate between stationary and progressive waves. Represent waves graphically: transverse, longitudinal, progressive. Derive and apply the wave equation, v = fλ. Perform calculations based on the properties of stationary waves. Describe experiments to demonstrate the following behaviours: rectilinear propagation, reflection, refraction, diffraction, interference, polarization. Describe applications of wave behaviours. Apply the laws of reflection and refraction to wave behaviours. Waves and energy transfer In 2010 a tsunami originated near Concepcion, Chile. How were scientists able to predict with great accuracy the time of arrival of the waves 6785 miles (10 920 km) away in Hawaii? And why did only very mild waves, less than 1 m high, reach Hawaii?
We will have an idea of the answers as we study waves and their behaviours.
the travelling waves (Figure 8.2a). (The drift speed of the barrel, however, is much lower than the wave speed.) Most waves, however, need a material medium that must oscillate. The medium (e.g. water) itself does not carry energy along with itself; rather it is the wave that travels in the medium that transfers energy. Waves are characterised by energy transfer. A barrel bobs up-and-down in sync with ocean waves and moves forward due to the energy supplied by The medium in which sound waves travel is air. Again, the medium (air) itself does not travel and transfer energy; it is the wave itself.
Waves also travel along solid media such as bars or straight strings or ropes. When we shake one end of a rope, the material itself does not travel – the wave does. a b Hawaiian islands Energy can also be transferred by a particle method (Figure 8.2b). In this method, Hawaiian islands Concepcion Concepcion Figure 8.1 (a) The location of Concepcion, Chile, and (b) a diagrammatic representation showing the position of the tsunami wave every 3 hours as it travelled to Hawaii. Chapter 8 Waves and their behaviours a b Figure 8.2 Wave and particle methods of energy transfer. (a) In the wave method, water (not the flowing of the water medium) is the material, in this case) nudges the barrel forward.
Much energy is transferred by waves. Waves move through the water, i.e. along matter. (b) In the particle method, the particle (the ball) carries kinetic energy along with it. This energy moves along with matter and is used for doing work in dislodging the cricket stumps. energy is transferred along with matter as kinetic energy. However, in the wave method, energy is transferred through, rather than along with, matter. As we shall see in Chapter 9, not all waves require a medium in which to travel. A special class of waves, called electromagnetic waves, can travel through empty space. In fact, light and other electromagnetic waves travel through space in a vacuum (empty space). The speed of all such waves is 3.0 × 108 m s−1, which is much greater than the speed of any other kind of wave or particle. Terms used in describing waves Figure 8.3 shows 'snapshots' of a wave travelling from left to the right along a taut string. The wave is generated by a source that causes a point on the string to undergo simple harmonic motion (SHM – see Chapter 7). The solid, curved line shows the wave at a particular instant of time. The broken line shows the wave a short time later. Displacement (y) and amplitude (A) The displacement of a point on a wave is the distance from the centre line (the x-axis in Figure 8.3) to that point. Thus, the displacements of points P, Q, R and S are all the same.
The amplitude is the maximum displacement. The highest points of the wave (e.g. Q1) are called crests; the lowest points (e.g. U) are called troughs. 2 y / cm 1 A P Q1 P1 1 2 The period of the wave is the time taken for one complete oscillation of any point on the wave. If a point on the wave, e.g. the point A, a point x = 1.0 cm, t makes n oscillations in t seconds, then the period T = n seconds. The frequency is the number of oscillations made per second by any point on the wave. Thus, f = 1 The S.I. unit for frequency is the hertz (Hz). n 1 1 Because T = n and f = t , it means that T = and f = ...(8.4) Phase (Φ) and wavelength (λ) Two points on a wave are said to be in phase if they have the same state of motion at the same time and their motions at that time are the same. The points P and Q on the wave in Figure 8.3 have the same displacement. However, as the wave moves to the right, P moves downward to P1 but Q moves upward to Q1. Hence, P and Q are not in phase. However, P and R are in phase. B R Q O Period (T) and frequency (f) 3 4 S 5 6 7 W1 8 W2 −1 U1 D1 U2 ITQ 1 For the wave shown in Figure 8.3: (a) what is the displacement of a point such as P? (b) what is the amplitude of the wave? −2 Figure 8.3 A wave travelling in the x-direction to the right. ITQ 2 Identify two other points that are in phase on the wave in Figure 8.3. 121 122 Unit 1 Module 2 Oscillations and waves The wavelength is the length between two consecutive points that are in phase on a wave (i.e. that have a phase difference of 2π).
Thus, for Figure 8.3, PR = QS = the wavelength of the wave, since these two successive points on the wave are in phase. Similarly, the distance between two consecutive crests or two consecutive troughs is one wavelength, since each pair of these consecutive points are in phase. The wavelength of the wave shown is 4.0 cm.
We measure phase in radians. The phase of a point is the angle corresponding to its displacement. Since, in one cycle, the phase angle swept out in SHM is 2π rad and the wavelength is one cycle of the wave, a point, A, at a distance x from the start of the wave, or has a phase angle (Φ) and with a displacement (see Chapter 7 on SHM and circular motion), the phase Φ is given by 2πx Φ= λ ....(8.5) Wave speed (v) and particle speed (vp) The distance travelled per unit time is called the wave speed. Thus, if a crest travels a distance x in a time t, x the speed of the wave is given by v = t . As a particle at A completes one down-up cycle (during one period, T), the crest of a wave travels forward a distance of one wavelength. That is, as x = λ, t = T. Then we get the equation for wave speed, x v= t and, ...(8.6) λ T and since v = fλ...............(8.7) 1 = f T ....(8.8) Note that since the time, T, taken for particle A to complete an oscillation is the same as the time taken for the crest at A to reach the point B, the period is also the time taken for a point on a wave (e.g. a crest) to travel one wavelength; the frequency is also the number of crests passing a given point per second. Wave speed is different from particle speed. Wave speed, v, is constant for a wave travelling along a uniform medium that is not subject to constant physical conditions. The particle speed is not constant as the particle oscillates with SHM. Particle speed, vp, for a wave of period T is given by equation of the form, vp = A sin ( ) 2πt T where 2π = ω is called the angular frequency of the SHM (see Chapter 7 for equations of SHM). Wave intensity (I) The wave intensity is defined as the energy carried by a wave through unit area per unit time. v x Figure 8.4 A wave travelling through a small volume V = As × x. Consider a wave (Figure 8.4) travelling with speed v. Assume that a portion of the wave passes through a volume V of cross-section area As and a small thickness x, as shown, in a time t. The total energy, E, of particles of total mass m undergoing SHM with amplitude A as the wave passes is given by 1 Worked example 8.1: Speed of water waves E = 2 kA2 Q E = 2 4π2f2mA2 A An observer notices that waves pass a 4.0 m long stationary boat, one crest touches the front and the next touches the back at the same time.
If 10 crests pass the front of the boat in 20 s, estimate the speed of the waves. wavelength (distance between successive crests), λ = 4.0 m ■ frequency, f = 10 ÷ 20 s = 0.50 Hz ■ speed, v = fλ = 0.50 Hz × 4.0 m = 2.0 m s−1 ■ ITQ 3 The Chilean tsunami (of 2010 (see Figure 8.1) took 15 hours to reach Hawaii. (a) Estimate the speed of the tsunami in km/h. (b) How does the speed compare with that of a jet plane? As 1 (see Chapter 7, equation 7.22) where f is the frequency of the wave (refer to Chapter 7 on SHM and equations 7.5 and 7.18). Since E = 2π2f2mA2 and, k = mω2 = m(2πf)2 by ω2 = k m , i.e. k = m(2πf)2 then, I = E At ■ ITQ 4 The intensity of a 10 m high tsunami wave near Concepcion is 3.0 × 102 W m−2. Estimate the intensity of the wave when it arrives at Hawaii if the height of the wave there is 1.0 m. Suggest two reasons why the intensity at Hawaii is much less than that near Concepcion. Chapter 8 Waves and their behaviours Transverse and longitudinal waves Transverse waves Transverse waves Transverse waves are waves in which the oscillations of the particles are at right angles to the direction of travel of the wave (Figure 8.5). A transverse wave on a rope held horizontally as in Figure 8.5a oscillates vertically (at right angles to the direction of travel) and moves along the rope from left to right, for example. 2π2f2ρAsxA2 Therefore, I = At since m = ρV = ρAsx x I = 2π2f2ρvA2 since = v t I ∝ A2 2 A Transverse waves in rope a wave which oscillations of a rope in a vertical plane.
When we shake a rope up and down, the oscillations of the particles of the rope take place in a vertical plane. We say that such a wave is polarized in the vertical plane. Such a polarized wave will pass through a vertical slit, but not through a horizontal slit (Figure 8.5). Therefore intensity (energy per unit area per unit time) is given by 2π2f 2AsvρA2 I= At Therefore, I = 2π2f 2ρvA2...............(8.10) Ordinary light waves (e.g. from the Sun) are polarized in all planes. A Polaroid sheet behaves like a set of slits – it allows light to pass through only in one plane. Thus all components of light in that plane (e.g. all vertical components) will pass through it. A second Polaroid sheet is added such that the plane of polarization is at right angles to the first, then no light gets through (Figure 8.6). Polaroid sheets are used in photography. The light reflected off glass or water is polarized in one plane. Thus, for the intensity of a wave is directly proportional to the square of the amplitude of the wave, which is reflection is seen, rotating a Polaroid sheet in front of the lens 90° to the plane of polarization would block the reflection from entering the camera. Worked example 8.2: Intensity of a tsunami wave lens.
This enables a clear picture to be taken of what's behind the glass. Estimate the intensity of a tsunami wave with a wavelength of 250 km and travelling at 750 km/h along seawater of density 1.03 × 103 kg m−3. Polarized sunshades make use of 5 polarization in that they block out one ■ wavelength, λ = 250 km = 2.50 × 10 m 3 750 × 10 m λ = 1 ■ of the two net components of light, ■ wave speed, v = 750 km/h = 208 m s 3600 s reducing the light intensity to about ■ amplitude, A = 10.0 m 50%. In polarized sunshades the page is 3 −3 ■ density, ρ = 1.03 × 10 kg m −1 of polarization is usually vertical since frequency, f = v = 208 m s5 = 8.32 × 10−4 Hz 2.50 × 10 m light reflected off water, e.g. at beaches, is intensity, I = 2π2f 2ρvA2 polarized in the horizontal plane. = 2π2 × (8.32 × 10−4Hz)2 × 208 m s−1 × 1.03 × 103 kg m−3 × (10.0 m)2 Warning: sunshades should never be used = 2.93 × 102 W m−2 to look at the Sun or at welding lights. The amount of UV light that only vertical components can still get through these let through shades can inflict permanent damage to the delicate cells of our light comes the retina of the eye. Polarized sheet Polarized sheet ...Figure 8.6 Action of Polaroid sheets. 3D cinema uses a different Polaroid sheet rotated at 90° on the glasses. ...are so polarized light. During filming, two lenses are used, separated a distance roughly equal to the separation of the eyes. A 123 124 Unit 1 Module 2 Oscillations and waves a observer b degrees scale Polaroid (analyser) Polaroid angle measurer eye positioned above Polaroid when viewing test tube plastic tube Polaroid sugar solution Polaroid (polarizer) light source Figure 8.7 (a) The basic arrangement in a polarimeter. (b) A simple polarimeter: yellow LED Polaroid sheet is placed in front of each lens so that their planes of polarization are perpendicular to each other. Two images are recorded from slightly different angles, just like the eyes would see them. The two images are projected separately through Polaroids onto the screen and made to overlap. During viewing, persons wear Polaroid spectacles whose polarizing axes are aligned just like those of the camera and projector. The left eye sees only what the left camera lens 'sees' and the right eye vice-versa. Thus a 3D image is reconstructed and seen by the viewer. Polarimetry is used industrially in measuring sugar concentration, e.g. in the manufacture of sugar or in determining sugar content in fruits and vegetables. The more concentrated the sugar solution is, the more it rotates a polarized beam of light. A beam of polarized light is passed through an empty test-tube. The light is viewed through a second Polaroid sheet, which can be rotated, at the other end (the analyser). The analyser is rotated until the light is seen. The sugar solution is then placed in the test-tube and the light is not seen because the solution has rotated the polarized light. The analyser is now rotated until the light is again seen. The angle through which the analyser is rotated gives a measure of the concentration of the sugar solution (Figure 8.7).

Polarized light is also used in pixels in LCDs (liquid crystal displays). A thin film of a Polaroid material is placed in front and at the back of each pixel. Each pixel contains a liquid whose molecular orientations, and hence optical properties, are controlled electrically. ITQ 5 What would a person viewing a 3D movie see if the right lens of the 3D spectacles were placed over the left eye and viceversa? Longitudinal waves A longitudinal wave is one in which the oscillations take place along the same line as the line of travel of the wave. Figure 8.8a shows a slinky model of a longitudinal wave made up of compressions and rarefactions of air. Since the oscillations take place along the line of wave travel, slits placed horizontally and vertically, as in Figure 8.5, will not block the oscillations. This means that sound waves are not polarized. The wavelength is the distance between two consecutive compressions or two consecutive rarefactions. In general, transverse waves (e.g. waves on a rope) are polarized; longitudinal waves (e.g. sound waves) are not. Getting it right! Surface water waves The waves on the surface of water are transverse. The barrel mentioned in the opening paragraph of this chapter and shown in Figure 8.2a bobs up and down as the waves move forward. However, a closer look at the barrel will show it oscillating also, forward and backward with each wave, i.e. the surface wave also displays some longitudinal wave motion. These combined motions produce a somewhat circular motion of the barrel.

However, as was mentioned before, the energy of the transverse wave nudges the barrel forward. ITQ 6 Using equation 8.11, determine the value of y when x = λ/4, at a time t = 0 s. Chapter 8 Waves and their behaviours a a v C R C R C R C λ A y O λ 4 λ 2 3λ 4 y(x) = A sin (kx) λ 5λ 4 x –A b b wavelength air molecules vibrating Figure 8.8 Longitudinal waves (a) On a slinky and (b) in the air. C = compression, R = rarefaction. Progressive and stationary waves Progressive waves Progressive waves transfer energy from one location to another. Surface waves in the ocean are progressive waves. Waves along a rope (or piece of string) are also progressive waves (though 'stationary waves' can also be generated on a rope, as will be discussed in the next section). The wave equation is generally given by $y = A\sin(kx - \omega t)$..............(8.11) for a wave travelling to the right. Then at time t = 0: $y = A\sin(kx)$..............(8.12) ω 2π is called the propagation constant v for the where k = λ wave – see equation 8.5 and at point x = 0 T + A y O T 4 T 2 3T 4 y(t) = A sin (ωt) T 5T 4 t –A Figure 8.9 Graphs of displacement versus (a) position, x, and (b) time, t, for a progressive wave travelling to the right. (Note that the graph of y = A sin (ωt) is a reflection of the graph of y = –A sin (ωt) about the t axis.) Figure 8.9a shows a graph of displacement (y) versus position (x) at a time t = 0 for a progressive wave. Figure 8.9b shows a graph of displacement (y) versus time (t) at a point x = 0. Getting it right! One cycle of a displacement versus time graph does not give wavelength, but period. The trace of y versus t (Figure 8.9b) is not an actual wave, but is a graph of y versus t at a particular value of x, e.g. the y-axis. The trace of y versus x (Figure 8.9a) is a picture of an actual wave at a particular instant of time. Stationary waves With a stationary wave (sometimes called a standing wave), energy is not transferred but rather is contained.

A stationary wave occurs when a progressive wave is being reflected back-and-forth in a small region of space. Figure 8.10 shows the first four stationary waves 2L, 2L v v v n = 1n =λ11 = 2L f1 = f1 = = λ 1 = 2L = formed on a taut string of length λ 1 L21 1 1 2L 1 L that is fixed at both ends. When such a string is plucked, a variety of waves of different wavelengths 2L 2L v v v v n = 2n =λ22 = L f2 = f2 = = 2 λ 2 2L = L = λ 2 λ 2 2L 2L travel back-and-forth, cancelling 2 f each other and quickly dying y = 4in (+=ωt......................(8.13) 2π – from equation 8.7. where ω = T antinode antinode antinode which means (8.13) 2λ1 2 λ2 λ2 2L 2L n = 2n =λ33 = 2233Ln== 23 L n = 3 v v v v f3 = f3 = 3 λ 3 λ 3 2L 2L 2L 2L n = 4n =λ44 = λ14L= L = 4 4 v v v v f4 = f4 == 4 = 4 λ 4 λ 4 2L 2L λ3 λ3 λ4 λ4 L L Figure 8.10 Stationary waves on a string. Note: in each case, the wavelength is the length that would be occupied by one complete wave cycle, e.g. λ3 as shown. A wavelength is not the distance from one antinode to the next antinode, or from one node to the next node; such a 1 distance is 2 × wavelength. 125 126 Unit 1 Module 2 Oscillations and waves Worked example 8.3: Speed of wave along a string Worked example 8.4: Speed of sound in air Q The vibrating length of a guitar string is 90 cm. If the string is sounding a frequency of 256 Hz in the first harmonic mode (one loop), determine the speed of the waves on the string. Q A ■ When a sound source of frequency 256 Hz is placed at the open end of a uniform tube containing water, the first two loud sounds are heard when the lengths of the air column are 0.30 m and 0.98 m, respectively. Determine the speed of the sound in the air column. A Note: although the open end is an antinode and the 0.30 m mark is at the first node, we do not use 4λ = 0.30 m for our calculation, since there is usually an 'end error' at the open end. We use the distance, instead, between the first two nodes as a half-wavelength (Figure 8.13): i.e. λ = 0.98 m − 0.30 m = 0.68 m 2 vibrating length, L = 90 cm = 0.90 m ■ harmonic mode, n = 1 frequency, f1 = 256 Hz λ1 = 2L (from equation 8.14, below) 1 2 ∴ = 0.90 m 1 = 1.80 m v = f1 λ1 = 256 Hz × 1.80 m = 460 m s−1 (to 2 s.f.) 0.30 m down. However, waves such as the ones shown persist much longer. These waves have points which do not oscillate (nodes) and points which oscillate with maximum amplitude halfway between the nodes (antinodes). A loop appears to be formed between successive nodes as waves travel back-and-forth very rapidly. Wave energy is trapped by reflection between the two fixed end-points. 0.98 m The first possible mode of vibration is called the fundamental or the first harmonic mode and is designated as the n = 1 mode. Three other harmonic modes of vibration are shown (n = 2, 3, 4). In terms of L, the wavelengths, λn, in the given modes follow a pattern (see Figure 8.10). The frequencies also follow a pattern. The two patterns are λn = 2L n water Figure 8.13 ■ distance between two consecutive nodes, λ = 0.98 m − 0.30 m = 0.68 m 2 ■ wavelength, λ = 2 × 0.68 m = 1.36 m ■ frequency, f = 256 Hz Sound wave speed, v = f λ = 256 Hz × 1.36 m = 348 m s−1 = 350 m s−1 ....(8.14) fn = nf1 where f1 = v 2L ....(8.15) Stationary waves on a strip Figure 8.11 shows two modes of vibration of a thin metal or cardboard strip, fixed at one end. Here the free end is an antinode, and the fixed end is a node. a A B L n=1 C n=2 Figure 8.11 The first two modes of vibration of a thin strip fixed at one end. Figure 8.12 Stationary waves in air columns within pipes. The curved lines show the relative variation of particle displacement. (a) One end closed; (b) both ends open. n=1 n=2 b n=2 n=1 λ2 4 λ1 λ1 4 λ2 2 λ2 2 λ2 Chapter 8 Waves and their behaviours Stationary waves in air columns At the open end of an air column, air molecules can vibrate freely; at the closed end, the air molecules can not vibrate (Figure 8.12).

Hence at particular frequencies, antinodes are formed at open ends and nodes at closed ends. These frequencies are called resonant frequencies because maximum vibrations of air molecules are occurring due to energy supplied at those particular frequencies. At resonance, sounds are heard rather loudly due to the large amplitudes of vibration. The phenomenon of resonance is discussed in Chapter 7. Worked example 8.4 contains sample calculations. When waves reflect off a concave surface, they are brought to a focus. This effect is employed in a TV 'dish', which collects a large amount of a weak signal and focuses it on to an antenna element placed at the focus of the dish (Figure 8.15b). Rectilinear propagation vibrating strip wave fronts some time later Figure 8.14 Rectilinear propagation of water waves. In a homogeneous medium, a wave travels at a constant speed in a straight line, i.e. shows rectilinear propagation. Figure 8.14 shows plane water wave crests travelling in a straight line in water of uniform density b plane barrier i Reflection is the bouncing of waves off a boundary.

Figure 8.15a shows that when water waves reflect from a plane barrier, the angle of incidence (i) is equal to the angle of reflection (r). All waves behave in this way.

Note that both i and r are measured from the normal (the perpendicular) to the plane barrier. When waves reflect off a place surface, the angle of incidence is equal to the angle of reflection. This is a statement of one of the laws of reflection. A ripple tank can be used to demonstrate several behaviours shown by waves. Heavy and light strings can also be used to demonstrate certain wave behaviours. A reflected Law of reflection Behaviours shown by waves wave fronts generated and constant depth. Note that the direction of travel of each crest (or wave front, as the collection of points on each crest, or that are in phase, is called) is at right angles to the wave fronts. Figure 8.16 shows reflection of waves along linear media. In Figure 8.16a, a wave is reflected out of phase when it bounces off a fixed barrier. In Figure 8.16d, the wave is reflected in phase when there is no barrier (e.g. when you crack a whip). Both Figure 8.16b and Figure 8.16c show that waves can undergo partial reflection and partial transmission. (Indeed, this is one of the behaviours that distinguishes waves from particles as models. Particles are either wholly reflected or wholly transmitted; they do not undergo partial reflection or partial transmission.) Note that transmission is always in phase with the incident wave.

However, reflection from a heavier medium is out of phase with the incident wave, and reflection from a lighter medium is in phase with the incident wave. concave barrier reflected wave fronts focus incident wave fronts reflected wave fronts ITQ 7 How many loops are expected in the 6th harmonic mode of vibration? (Look back at Figure 8.10.) incident wave fronts Figure 8.15 Reflection of water waves: (a) off a plane surface; (b) off a curved surface. ITQ 8 In terms of the length, L, of the strip in Figure 8.11, what is the wavelength in the 2nd harmonic mode? 127 128 Unit 1 Module 2 Oscillations and waves fixed barrier a total reflection (out of phase) I R T partial transmission (in phase) partial reflection (out of phase) b I R T partial transmission (in phase) partial reflection (in phase) c I R d no barrier total reflection (in phase) I R light string heavy string Figure 8.16 Reflection and transmission of waves along linear media. Refraction Refraction is the change of direction of a wave as it crosses a boundary and undergoes a change in speed. Figure 8.17 shows a wave front, DB, approaching a boundary between deep and shallow water. The speed of waves in deep water, v1, is greater than the speed, v2, in shallow water.

In a given time, t, the point D on the incident wave travels a distance DA, while the end B travels a shorter distance BC on account of the slower wave speed in the shallow water. From the geometry shown, we get sin θ1 = sin θ2 = ⇒ D deep region v1 X O A θ1 a boundary B θ2 v2 C shallow region DA AB v1t y AB Figure 8.17 Refraction of water waves. BC AB v2t incidence and refraction measured from the normal to the boundary. Since wave travel. All sin θ1 v1 therefore, = sin θ2 v2t sin θ1 v1 or, sin θ2 v2 ....(8.16) Referring again to the geometry in Figure 8.17 we get, θ1 + a = 90° hence, a + θ2 = 90° therefore, x = θ1 and, by a similar argument, y = θ2 Thus, we see that the angles of inclination that the incident and refracted wave fronts make with the boundary are equal, respectively, to the angles of Equation 8.16 can be called the wave speed equation for refraction, where θ1 and θ2 are the angles of incidence and refraction, respectively, when a wave travels with speeds v1 and v2, respectively, from a medium 1 to a medium 2. Diffraction When a wave passes through a gap that is of the order of its wavelength, or smaller, the wave spreads out in various directions. This kind of spreading out is called diffraction.

In Figure 8.18a, water waves keep travelling straight through a gap (e.g. hardly undergo diffraction) since the wavelength of the wave is much less than the width of the gap. In Figure 8.18b, the waves spread out, since the wavelength is greater than the width of the gap.

Chapter 8 Waves and their behaviours a b incident wave Figure 8.18 Diffraction through a gap (a) when the wavelength is short, and (b) when the wavelength is of the order of, or large, as compared to the width of the gap. diffracted waves As with water waves, high-pitched sounds from a police car siren have short wavelengths. As such, the sounds are heard very far down a street, since most of the sound energy travels straight and hardly diffracts into the side streets. However, as the vehicle approaches an intersection, and a low pitch (long wavelength) sound is made, the sound diffracts around the corners of the intersection, thus warning on-coming intersection traffic. Worked example 8.5: Refraction of water waves Q Plane water waves, travelling from a shallow region, make an angle of 35° with the boundary between a shallow region and a deep region. If the speed of the waves in the shallow and deep regions are 2.2 m s−1 and 2.8 m s−1, respectively, determine the angle the refracted waves make with the boundary. A ■ Some examples of diffraction involving electromagnetic waves are discussed in the next chapter. Interference When two (or more) waves are at the same place at the same time, their displacements add algebraically to produce a resultant wave. This is called the principle of superposition of waves. Figure 8.19a shows that two waves which are in phase, their displacements add to produce a wave of larger amplitude. We say that the waves are showing constructive interference. In Figure 8.19b, the waves are half a wavelength out of phase with each other and the resultant wave is zero. This is destructive interference. a b y y two waves in phase resultant wave x waves out of phase with each other angle of incident waves with boundary, θ1 = 35° ■ angle of refracted waves with boundary, θ2 = ? ■ speed of incident waves, v1 = 2.2 m s−1 ■ speed of refracted waves, v2 = 2.8 m s−1 sin θ1 v1 By equation 8.16, = sin θ2 v2 v × sin θ1 Therefore, sin θ2 = 2 v1 2.8 m s−1 × sin 35° 2.2 m s−1 θ2 = 47° Note: this angle is larger than 35° and shows that the wave would be refracting away from the normal for a change of wave speed from slow to fast.

= resultant wave x Figure 8.19 Superposition of two waves: (a) of equal amplitudes and in phase; (b) of equal amplitudes but π radians (i.e. half a wavelength) out of phase, with each other. 129 130 Unit 1 Module 2 Oscillations and waves b A suspended vibrating beam metal barriers gap D gap C D D C D Figure 8.20 Interference pattern generated by two sets of circular waves that are in phase with each other: (a) a drawing; (b) a photograph. C = line of constructive interference; D = line of destructive interference. Figure 8.20a shows a drawing of circular waves generated in a ripple tank by diffraction through two narrow slits. The two sets of waves are in phase with each other. Verify for yourself that the line labelled C is a line of constructive interference, and the lines labelled D are lines of destructive interference. Figure 8.20b shows the pattern produced by the two sets of waves interfering with each other. The humps are areas of constructive interference; the straight dark lines are areas of destructive interference. Where crests meet (or troughs meet) constructive interference takes place since these waves are in phase. Where crests meet with troughs, however, destructive interference occurs. Lines of constructive interference (C) and destructive interference (D) are shown spreading out from the sources. Along C, waves of large amplitude travel outward. Along D, a flat 'wave' occurs. An interference pattern is therefore produced, since the waves are coherent. If the phase between the two waves keeps changing, a fixed pattern, such as shown in Figure 8.20, is not attained. Figure 8.21 shows the interference pattern produced by the same set of waves when the sources are further apart (Figure 8.21b). a b sources C D C D C D C D C D Figure 8.21 Interference pattern shown by water waves: (a) when sources of the waves are close; (b) when sources of the waves are further apart. C = line of constructive interference; D = line of destructive interference. (one zone moving outward while the other is moving inward) there will be destructive interference along the middle line and the result will be soft sounds there. For this reason, loudspeaker terminals are usually colourcoded black and red so that they can be connected to amplifier in phase with each other. Also, it should be noted that if these speakers are placed in an auditorium, or even outdoors, there will be positions where the sound will be heard loudly and positions where the sound is heard softly. Acoustic applications involving interference Figure 8.20 shows that the intensity of a wave is strongest along the centre line perpendicular to the line separating the two sources. This is true, however, only if the waves are in phase. For this reason, pairs of loudspeakers connected to a single source in phase (i.e. both cones pushing outward at the same time), will give a strong sound along the middle line. If connected out of phase ITQ 9 Two loudspeakers are placed on a stage in an auditorium.

Where would the sounds from the speakers most likely be heard: (a) loudly? (b) softly? (Assume the loudspeakers are connected in phase.) Chapter 8 Waves and their behaviours Summary ✓ Waves transfer energy by means of oscillations. ✓ Stationary waves (also called standing waves) are formed by multiple reflections of progressive waves, of particular wavelengths, within a pair of boundaries.

In mechanical waves, energy is transferred through a medium (which can be solid, liquid or gas) as particles of the medium oscillate. ✓ Transverse waves are waves in which the direction of energy transfer (the direction of the travel of the waves) is at right angles to the oscillations of the points that make up the waves (e.g. waves on a rope). Electromagnetic waves are also transverse waves, even though these waves can transfer energy through empty space. ✓ Waves travel at a speed v in a homogeneous medium according to the equation where f and λ are the frequency and wavelength of the wave, respectively. ✓ In standing waves, the distance between successive nodes is half a wavelength (a node is a point on the wave that does not oscillate).

✓ In longitudinal waves, the direction of energy transfer is along the line in which oscillations take place (e.g. sound waves in air). ✓ The main difference between a transverse wave and a longitudinal wave is that a transverse wave exhibits polarization, e.g. oscillations of all points on a transverse wave take place in a single plane at the same time, whereas this is not the case in a longitudinal wave. ✓ Progressive waves (also called travelling waves) transfer energy from one location to another. ✓ f= ....(8.8) ✓ For mechanical waves, the intensity, I, is given by I = constant × A2..................(8.10) ✓ When waves reflect off a plane barrier, the angle of incidence is equal to the angle of reflection. This statement is one of the laws of reflection. ✓ When a wave travels from medium 1 to medium 2, the angles of incidence and refraction (θ1 and θ2, respectively) are related to the wave speeds (v1 and v2) by the equation sin θ1 v1 = ...(8.16) sin θ2 v2 131 132 Unit 1 Module 2 Oscillations and waves Review questions Waves and their behaviours 1 Lightning and thunder occur at the same time in a cloud.

Estimate the time difference between a lightning strike occurring 5.0 km away being seen and being heard (speed of light = 3.0 × 108 m s−1, speed of sound in air = 330 m s−1). 2 (a) Define intensity of a wave. (b) The intensity of a water wave of amplitude 5.0 m is 2.0 × 102 W m−2. Estimate the intensity of a water wave of amplitude 1.2 m if the frequency and wave speed are the same. (c) Estimate the power of a wave of intensity 2.0 × 102 W m−2 as it passes through an opening 20 m long and 5 m high in a breached wall. 3 Prior to the arrival of a tsunami wave, persons on the shore can see the bottom of the sea for considerable distances (Figure 8.22). Tsunami waves typically have wavelengths of 250 km and travel at jet plane speeds of about 750 km/h. tsunami wave v P normal sea level land Figure 8.22 A tsunami wave approaching a shore. Estimate how much time a person, P, standing on the shore, as in Figure 8.22, has to get out of the way from destruction by the approaching tsunami wave. 4 Ultrasound wave pulses are used for measuring the depth of the ocean. Figure 8.23 shows oscilloscope traces of a pulse (T) transmitted from the bottom of a boat and the pulse (R) received on reflection from the sea bed. If the time base (horizontal axis) on the oscilloscope is 0.5 s cm−1, estimate the depth of the ocean. (Speed of sound in sea water = 1500 m s−1).

T R 5.0 cm Figure 8.23 5 (a) Calculate the wavelength of radio waves of a local AM and a local FM radio station using their broadcast frequencies. In AM the frequencies are given in kHz; in FM they are given in MHz. (b) An AM station and an FM station, near to each other, are broadcasting from a similar type of location at approximately sea level, several km from a hill. Explain in terms of wavelengths and diffraction why the AM station is more likely to be heard louder than the FM station in homes at approximately sea level on the opposite side of the hill. Transverse and longitudinal waves 6 (a) Distinguish between a transverse wave and a longitudinal wave in terms of their oscillations and direction of travel. (b) When a sheet of Polaroid material is placed in the path of a beam of light, about half of the intensity of the light is blocked. When the sheet is rotated by 90° on its axis, the amount blocked stays the same. Explain these two phenomena. (c) How can the light be completely blocked using a second Polaroid sheet? 7 (a) Classify the following as longitudinal or transverse waves: ultrasound, ultraviolet light, radio waves, surface waves on the ocean, waves in the depth of the ocean. (b) A small radio with its FM antenna extended vertically, is tuned to and faces a station that is heard weakly. When the radio antenna is rotated 90° about a horizontal axis with the radio still facing the station, the station is heard louder. Explain this occurrence in terms of polarization. Stationary waves 8 A child, playing in a bath tub, taps the water at a constant rate 8 times in 10 s. A standing wave is produced with a node in the middle and antinodes at the sides of the tub. (a) Draw a wave diagram to show the standing wave produced in the tub. (b) If the width of the tub is 75 cm, estimate the speed of the water waves.

9 (a) Draw labelled apparatus for determining the speed of waves on a light string from which a 15 g mass hangs. (b) In an experiment, such as mentioned in part (a), the vibrating length, L, of the string was 92.3 cm. Chapter 8 Waves and their behaviours The table shows data collected concerning the standing waves obtained. The first row of the table is complete. Complete the other three rows.

No. of loops, n Frequency of stationary wave, fn / Hz Wavelength, λn / m 1 13 1.846 2 26 3 42 4 55 (c) For waves of speed v, frequency f, and wavelength λ, v = fλ. By plotting a suitable graph using frequency and wavelength data from the table above, determine the speed of the waves on the string. (d) Discuss three sources of error in this experiment. 10 (a) Draw a diagram showing the first and second modes of vibration of standing waves in a pipe: (i) open at both ends (ii) open at one end only. (b) Estimate the lengths of pipe that would give rise to the standing waves in part (a) for sound of frequency 440 Hz: (i) open at both ends (ii) open at one end only. 11 (a) Explain how stationary waves are formed on a string. (b) Two adjacent nodes on string vibrating at 440 Hz are 75 cm apart. Find the speed of waves on the string. (c) If the vibration of the string in part (b) consisted of only 4 loops, at what frequency would the string vibrate in 6 loops, all other physical factors remaining the same. 12 In an experiment with microwaves (see Figure 8.26), 7 nodes were detected over a length of 15 cm. Determine the frequency of the microwaves. (Speed of light = 3.0 × 108 m s−1). Refraction of waves 13 The speed of sound in air, v, varies with temperature according to the equation v = (330 + 0.600) m s−1 where θ is the air temperature in °C. A sound wave travels from air at 20 °C to a layer of air at 34 °C. (a) Calculate the speed of the wave in the two types of air. (b) Draw a diagram, labelling each layer of air, the direction of wave travel in the two media and the angles of incidence and refraction. (c) The angle of incidence of the sound wave is 44°.

Calculate the angle of refraction. 14 Earthquake waves, travelling at 8.0 km s−1 in a certain type of rock, make an angle of incidence of 50° at the boundary with another type of rock. If the speed of the wave in the latter rock is 7.2 km s−1, calculate the angle of refraction of the wave. Include a labelled diagram displaying the above information. 15 (a) What is meant by the superposition principle? (b) What is the condition for an interference pattern to be formed from two circular waves in water. (c) Draw a diagram showing two circular waves producing an interference pattern. Label areas of constructive interference and destructive interference. Practical exercises and challenges Safety first! In all of these exercises, you should discuss with your teacher before attempting them and where necessary carry out them under the direct supervision of your teacher. 1 You can determine the speed of light quite easily using stationary waves. Go online and look up 'speed of light using microwaves and marshmallows'. ■ Put a shallow, non-metallic, microwave-safe dish containing one layer of marshmallows in the microwave. ■ After a short while, take the dish out and you will notice 'hot-spots' where the marshmallow melts. The distance between two consecutive hot spots (antinodes) is half the wavelength of the microwaves. Find the average half wavelength. ■ Find the frequency of the microwaves; this is usually stamped at the back of the microwave oven. ■ From the frequency and the measured half wavelength, you can determine the speed (v = fλ) of the microwaves, which is the same as the speed of light. Note: take care since the microwave oven is operated from mains electricity. 2 You can go online and learn how to make a simple polarimeter. You will need a light source (e.g. a small bulb connected to a battery), a transparent tube to contain sugar solutions, two small Polaroid sheets (if you cannot get the sheets you might be lucky 133 134 Unit 1 Module 2 Oscillations and waves from discarded Polaroid sunshades) and a degree scale.

The idea is to send polarized light through the solutions and measure the degree of rotation of the polarized light that has passed through the solution. A sketch of the basic idea is shown in Figure 8.7. Calibrate your polarimeter in degrees corresponding to known made-up concentrations of sugar solution. Test your calibrated instrument on a sugar solution of 'unknown' concentration. (See also the section in this chapter on polarization.) 3 Use the apparatus shown in Figure 8.24 to find the velocity of waves on a string. Adjust the signal generator to produce standing waves of 1, 2, 3, 4 or 5 loops for a given load. For a given load, the speed of waves on a string is constant. Since v = fλ, what variables will you plot on your graph, and how will you arrive at v from your graph? Make an estimate of the percentage error expected in the value you obtained for v. signal generator vibrator pulley string load Figure 8.24 4 Use the apparatus in Figure 8.25 to determine the speed of sound in air. Move the microphone to locate consecutive nodes (where the sound is the faintest). Using various frequencies and the measured wavelengths (the distance between consecutive nodes is λ/2), plot a suitable graph to obtain the value of v. Why would you expect the value of v to be constant? If your experiment did not yield a constant, make suggestions why this might have happened. signal generator amplifier plane reflector loudspeaker microphone to CRO or datalogger 5 Determine the frequency of microwaves using the arrangement shown in Figure 8.26. Determine the average position of consecutive nodes by moving the probe and counting the number of consecutive nodes. Explain how this technique gives a better result for λ/2 than measuring the distance between two consecutive nodes directly. Use c = fλ (c = 3.00 × 108 m s−1) to find f. Estimate the percentage error in your experimental value of f. aluminium sheet (reflector) microwave transmitter probe to receiver Figure 8.26 6 For those of you who are more adventurous, make your own 'Polaroid'.

You can coat a glass slide with various chemicals and then investigate for polarization of light passing through. Figure 8.25 Chapter 8 Waves and their behaviours Answers to ITQs Answers to Review questions 1 (a) 0.6 cm (b) 1.2 cm 1 15 s 2 (a) and 5. 5 (a) Use λ = for each station, where c = 3.0 × 108 m s−1. f 9 (b) 0.923 m, 0.615 m, 0.462 m (c) plot f against 1 . Slope = v = 24.8 m s−1 3 11 (b) 660 m s−1 (c) 660 Hz 3 13 minutes t 10920 km = 13 h (a) v = 730 km/h 4 (b) let aircraft travel at these kinds of speeds across oceans. In: kAc2 = IH kAH2 Therefore I × kAH2 IH = kAc2 × 3.0 × 102 × 1.02 102 = 3.0 W m−2 The intensity at Hawaii is less than that near Conception since the waves would lose energy (e.g. by friction with air) as it travels over the Pacific; also, the waves are spreading over a larger area, resulting in intensity = (energy per unit time / area) becoming smaller the further the wave gets from Conception. 5 The person would see the image intended for the left eye with the right eye, and vice versa – the 3D effect would be confused. 6 ( ) ( ) 2π λ × 4 λ π = A sin y = A sin = A 7 6 loops. On a string fixed at both ends, the number of loops is equal to the harmonic mode number. 3 4 8 Since L = 4 λ then λ = 3 L.

9 (a) Along lines of constructive interference. (b) Along lines of destructive interference. (See Figures 8.20 and 8.21.) 13 (a) 342 m s−1; 350 m s−1 (c) 45.3° 135 136 Chapter 9 Light: rays and waves Learning objectives ■ Discuss evidence supporting a ray model and a wave model of light. ■ Apply the laws of reflection of light at plane surfaces. ■ Define refractive index.

■ Derive and use Snell's law. ■ State the laws of reflection of light at plane surfaces. ■ Discuss the formation of a critical angle and the conditions necessary for total internal reflection. ■ Discuss applications of total internal reflection. ■ Draw ray diagrams to show image formation by lenses. ■ Apply lens formulae: 1 + 1 = 1 u v f 1 ●● P = f hi ●● m = = v ho u Discuss image formation in the simple camera and magnifying glass. Discuss the formation of focused images by the eye. Discuss optical defects of the eye and how they can be corrected. Discuss the formation of fringes in two-slit experiments. Derive, the expression for fringe spacing in Young's experiment, where slit-to-screen distance D ≫ ai ●● y = ●● λD ●● n λ = a sin θ (for the nth bright fringe) Determine the wavelength and frequency of light waves using a diffraction grating. Outline the position of light in the electromagnetic spectrum. ■ d sin θ = nλ Ray optics Ray optics is the study of light travelling as 'rays'.

A ray can be defined as the line indicating the direction of travel of light. By studying light from a ray point of view, we are able to explain how mirrors and lenses form various types of images. A wave model does not enable us to do this. Wave optics enables us to account for characteristically 'wave' behaviours of light such as diffraction, interference, polarization and refraction. A ray model of light is not able to explain these phenomena. Figure 9.1 A microscope can make very tiny things appear much larger. This is explained using the idea of light travelling as rays. However, there is a limit to the extent that things can be enlarged to show detail. This is explained using the idea of light travelling as waves. Rays or waves? How does light really travel? Chapter 9 Light: rays and waves Propagation of light The rays in Figure 9.2 shows clearly that light travels in straight lines. (Actually, light travels in a straight line only a homogeneous medium, such as air or water of uniform density.) The streaks of light in Figure 9.2 are not, however, rays, but narrow beams of light. Each beam is made up of an infinite set of rays. A ray of light, treated this way, is like the mathematical concept of a line that has no width. Figure 9.2 also shows light displaying a behaviour typical of waves – waves passing through each other. Each ray that passes through each other emerges unaffected by each other is a typical wave phenomenon. This kind of behaviour can easily be demonstrated using water waves. Figure 9.2 Light showing ray and wave behaviours. The optical lever If you were to study Figure 9.4 and the following proof, you will find that, on account of law 2, when a mirror is rotated P degrees, the reflected beam rotates an angle, Q = 2P degrees. This doubling of the angle rotated by the mirror in the angle rotated by the beam is called the optical lever principle. Reflection from plane surfaces Laws of reflection of light Experiments show that light undergoing reflection from a plane surface obeys two laws. These two laws are illustrated in Figure 9.3. Proof of optical lever principle In Figure 9.4a: 1 The incident ray, the reflected ray and the normal (perpendicular) to the surface at the point of incidence all lie in the same plane. ■ M1 and M2 are positions of the mirror before and after rotation ■ P is the angle of rotation of M1 2 The angle of incidence (θi) is equal to the angle of reflection (θr). ■ I1 is the incident beam of light on M1 and M2 ■ R1 is the beam of light reflected off M1 In equation form, law 2 is expressed as θi = θr ■ R2 is the beam of light reflected off M2 ....(9.1) Q is the angle of rotation of the reflected beam ■ N1 and N2 are the normals, respectively, to M1 and M2. a incident ray θi θi M1 N1 θ1 P Figure 9.3 Reflection of light off a plane surface: (a) ray diagram; (b) wave diagram. plane smooth reflecting surface N2 fibre b R1 mirror scale P N2 reflected wavefronts reflected ray for plane smooth reflecting surface θr reflected wavefronts b normal R2 ε o M2 Q θ2 θ1 light source (narrow beam) Figure 9.4 (a) Effect on a beam of light when a mirror is rotated. (b) Schematic diagram of apparatus used by Henry Cavendish to measure tiny angles of rotation of a mirror in his experiment to determine the universal gravitation constant.

137 138 Unit 1 Module 2 Oscillations and waves Figure 9.4 At the first reflection:If the mirror is rotated an angle P the normal is rotated through the same angle, P. At the first reflection (i.e. off M1) θ1 = P + a At the second reflection (i.e. off M2) θ1 + P = a + Q Combining these two equations, we get P+a+P=a+Q Therefore, Q = 2P...............(9.2) Image formation by a plane mirror Applications of plane mirror reflection Henry Cavendish used the optical lever (Figure 9.4b) to magnify the angles of rotation of a mirror by a factor of 2, since the angles resulting from the rather weak gravitational forces in his experiment were very tiny.

Charles Coulomb used the same device to measure angles while investigating forces between point charges.

Stealth aircraft are constructed with flat panels angled so as to reflect incident radar signals away from the radar station (Figure 9.6). (Later in the chapter we shall discuss a wave technique that is also employed in stealth aircraft to further reduce reflected radar from reaching the radar station.) Ray optics enables us to account for image formation of a point object by a plane mirror. Two rays are drawn from the point object. The rays obey reflection law 2 when they reflect from the mirror.

The eye perceives an image of the object at the point where the rays from the mirror appear to meet when projected backwards behind the mirror. Figure 9.5a shows that a virtual image is formed the same distance behind the mirror (IM) as the object is in front of the mirror (OM). The image is virtual since it is not formed by real rays of light converging to a point, but by rays appearing to diverge from that point. Such an image cannot be seen on a screen. Applying the result, IM = OM, point by point, to an extended object (Figure 9.5b), we obtain an image with the following characteristics: ■ nature – virtual ■ size – same as object ■ position – the same distance behind the mirror as the object is in front of the mirror ■ orientation – laterally (sideways) inverted. a O M I b Figure 9.6 A Stealth aircraft with angled, flat, reflecting panels. Plane mirrors are quite commonly used as dressing mirrors since the image seen is the same size as the person standing in front of it.

The internal mirrors in motor vehicles are also plane mirrors. This means that the image is the actual size of the object, hence enabling the driver to judge fairly accurately the distance an object is behind the vehicle. Refraction at plane surfaces extended object virtual image of extended object Figure 9.5 (a) Virtual image, I, of a point object, O, formed by a plane mirror. (b) Lateral inversion produced by a plane mirror.

ITQ 1 A beam of light makes an angle of incidence of 30° with the normal of a plane mirror. If the mirror is then rotated an angle of 10°, by how much will the reflected beam rotate? We learned in Chapter 8 that when waves approach at boundary between two media in which the wave speeds differ, there is a change in direction of the wave as it crosses the boundary (see Figure 8.17).

We call this change in direction, refraction. Figure 9.7 shows a beam of light undergoing a change in direction as it travels from air to glass. This change in direction suggests that light is exhibiting a wave nature. However, in our calculations, we will focus on the 'ray' aspects of these waves, i.e. the directions in which the waves travel. Wave fronts are superimposed on the incident and refracted beams. Chapter 9 Light: rays and waves and Snell's law normal air θ1 θ2 glass When light undergoes refraction between two media, the ratio of the sine of the angle of incidence to the sine of the angle of refraction is a constant. This is a statement of Snell's law. Snell's law is one of the laws of refraction of light.

Figure 9.7 A beam of light refracting as it travels from air to glass. Plane waves are superimposed on this beam.

Refractive index Different materials refract light by different amounts. The amount of refraction is represented by the refractive index of the material. The smaller the angle of refraction produced within a material with respect to the angle of incidence in free space (i.e. the greater the deviation of the ray travelling from free space into a material), the greater the refractive index of the material. 1 The incident ray, the refracted ray and the normal (perpendicular) to the surface at the point of incidence all lie in the same plane. 2 The ratio of the sine of the angle of incidence to the sine of the angle of refraction is a constant. Worked example 9.1: Refraction from light underwater Q An underwater spotlight L at the bottom of a pool 1.8 m deep, throws a spot at a point P on a wall 3.0 m from O, i.e. OQ = 3.0 m. The distance LN = 1.2 m. How high above Q is the point P? Figure 9.8 shows the situation. See Table 9.1 (page 140) for data on refractive index. P The refractive index, n, of a material is defined as the ratio of the speed of light in free space to the speed of light in the material.

In equation form, this is given by n2 c = ...(9.3) v where c represents the speed of light in free space and v is the speed of light in the material. By equation 9.3, c n1 = or v1 = v1 c n1 c v2 = or v2 = c n1 c n2 Substituting these into equation 8.16 we get n1 sin θ1 = n2 sin θ2...............(9.4) Since n1 and n2 are constants for the separate media, equation 9.4 tells us that sin θ1/sin θ2 is a constant. This relationship is known as Snell's law.

Equation 9.4 is called the Snell's law equation. Note also the fact that the angle of refraction in glass (Figure 9.7) is less than the angle of incidence in air tell us about the speed of light in glass?

Q water θ1 Consider the wave equation for refraction (see equation 8.16) sin θ1 v1 = sin θ2 v2 In the case of light, let us designate the refractive indices of the two media as n1 and n2, respectively, where the incident and refracted rays occur. x O L 1.8 m N Figure 9.8 Light refracting from underwater on to a wall. A ■ depth of pool, ON = 1.8 m ■ LN = 1.2 m LN tan θ1 = ON 1.8 Therefore θ1 = 33.7° By Snell's law, n1 sin θ1 = n2 sin θ2 Therefore 1.33 × sin 33.7° = 1.00 × sin θ2 This gives θ2 = 47.5° and so x = 42.5° PQ But tan x = OQ Therefore PQ = OQ × tan 42.5° = 3.0 × tan 42.5° = 2.7 m ITQ 3 Using the data in Table 9.1 (page 140), estimate the speed of light in water. 139 140 Unit 1 Module 2 Oscillations and waves X1 a B N2 A θ1 R3 θ2 R3 I1 I3 R3 R2 I2 I1 N R1 Table 9.1 Refractive indices of various materials Material Refractive index, n free space 1.000 air 1.0003 water 1.33 glass 1.5 (average) diamond 2.42 lucite 1.51 N Worked example 9.2: When light does not exit water Q For what angles of incidence does light from an underwater spotlight in a pool not exit the water (Figure 9.10)?

normal air refracted ray water θ2 Table 9.1 shows refractive indices of various materials. The fact that, to 4 significant figures, the refractive index of free space and air are the same means that we can, for most calculations, assume that the speed of light in air and in free space are practically the same. Critical angle and total internal reflection Figure 9.9a shows incident rays of light, I1, I2 and I3, sent to pass through the centre, O, of a semicircular block of a transparent material. The rays travel straight to O without undergoing refraction. The ray I1O undergoes partial refraction at the plane boundary AOB, as well as partial reflection within the block. The refracted ray is OX1 and the reflected ray is OR1. As the angle of incidence, θ1, is increased to a certain value, θ2, there is partial reflection and partial refraction again. The angle of refraction of the refracted ray is 90°. The angle of incidence, θ2, made by the ray I2O is called the critical angle. Note that there is also a partially reflected ray, OR2.

The critical angle is the angle of incidence when the angle of refraction is 90°. If the angle of incidence is made larger than the critical angle, there is no longer any refracted ray, only a reflected ray, OR3. We call this phenomenon total internal reflection, since all the incident light energy is reflected. ITQ 4 In Figure 9.9a, why do the incident rays that enter the block and pass through O not undergo refraction (a change in direction) when entering the semicircular block? Figure 9.9 (a) Formation of critical angle (θ2) and occurrence of total internal reflection. (b) Wavefronts superimposed on direction of travel of light in total internal reflection. incident ray A Figure 9.10 Critical angle for light from an under water spotlight. A ■ Light will not exit the air as medium 2. Light will not exit the water for angles of incidence above the critical angle. By Snell's law equation, n1 sin θ1 = n2 sin θ2 At critical angle, n1 sin θc = n2 sin 90° 1.33 × sin θc = 1.00 × 1 ⇒ sin θc = 1.00 ⇒ θc = 1.33 = 48.8° Therefore light does not exit the water at any point on the water surface where the angle of incidence is greater than 48.8°. ( ) within the diagram (Figure 9.9b). In making calculations involving the rays and angles shown in Figure 9.9, we apply the laws of refraction whenever refraction occurs, and the laws of reflection whenever reflection occurs. Conditions necessary for total internal reflection The preceding discussion and Worked example 2 illustrate that there are two conditions necessary for total internal reflection. ITQ 5 In Figure 9.9a, which of the reflected rays OR1, OR2 or OR3) is expected to be brightest? Explain your answer. Chapter 9 Light: rays and waves ■ First, the light must be travelling from a medium to one of lower refractive index. In such a case, the angle of refraction will always be greater than the angle of incidence and an angle of refraction of 90° will be possible. ■ Second, the critical angle between the two media must be exceeded. Applications of total internal reflection Retroreflectors are used on vehicles and roads to direct light back to the source. One type consists of a truncated corner of a transparent cube made of plastic or optical glass. Light enters one face perpendicularly and makes an angle of incidence, θ, greater than the critical angle, θc, between the plastic or glass on the second face and the surrounding air (Figure 9.11). Total internal reflection occurs twice and practically all the light, incident from the vehicle's headlamps, is reflected back towards the driver, causing the reflected light to appear very bright. Front and rear reflectors on bicycles, as well as some types of road retroreflectors, make use of total internal reflection in this manner. modulated laser beam carrying many signals can be sent along one fibre, and the information passed along is less distorted along the way than if carried (electrically) by wires. Fibres, being so thin, can be bundled, thus carrying a vast amount of information along a narrow cable. Also, the fibre optic cables are lighter than metal wires and are less affected by weathering. Bundled optical fibres are also used in endoscopes in medicine. A physician can 'look' into the stomach, for example, and examine the lining of the stomach. Light is sent along some of the fibres of the endoscope, reflects from the stomach wall and travels along the optical fibres to a viewing screen (Figure 9.12c). optical fibre a light solid plastic b c transparent cladding illuminating light optical fibres θ>θc to screen monitor incident light reflected light image Figure 9.11 Total internal reflection in one type of retroreflector. The sparkle of specially cut diamonds is due, in part, to total internal reflection. The critical angle for diamond is only about 25° since the refractive index is high (2.42). (You can calculate the critical angle using the Snell's law equation, 9.4.) For a given incident beam of light, the critical angle is easily exceeded by slight movements of the hand, which reorients the diamond. Total internal reflection occurs, resulting in the emerging light having a bright, sparkling appearance. Optical fibres also make use of total internal reflection. Each fibre consists of a solid, transparent plastic core surrounded by a solid transparent cladding of refractive index less than that of the plastic core (Figure 9.12a). Light entering a fibre at a small angle of incidence exceeding the critical angle along the length of the fibre. Hence total internal reflection occurs (Figure 9.12b). Multiple reflections occur along the fibre with little loss of light.

Optical fibres are preferred over wires for carrying signals in the internet.

A object Figure 9.12 (a) Cut-away showing total internal reflection within an optical fibre. (b) Total internal reflection occurring along a bent optical fibre. (c) Schematic diagram showing the operation of an endoscope. Refraction at curved surfaces A lens is a transparent material that has at least one curved surface. Two of the simplest lens shapes are the spherical biconvex and spherical biconcave lenses. The following discussions refer to thin lenses, i.e. lenses whose thicknesses are small as compared with their widths. Treating light as travelling by rays, rather than by waves, enables us to account for the various types of images formed by lenses.

141 142 Unit 1 Module 2 Oscillations and waves a b F O F A b O F a Figure 9.13 Effect on rays of light parallel and close to the principal axis by (a) a biconvex (converging) lens and (b) a biconcave (diverging) lens. b F f Terms used in describing lenses The following descriptions are based on Figures 9.13a and 9.13b. ■ Optical centre (O): the physical centre of the lens. ■ Principal axis (AB): the line that passes through the optical centre and is perpendicular to both surfaces of the lens. ■ Principal focus (F): the point on the principal axis where rays parallel and close to the principal axis converge (Figure 9.13a) or appear to converge (Figure 9.13b). There are two such points, one on either side of the lens, at equal distances from the optical centre, since light can be made to enter from either side. ■ Focal length (f): the distance OF between the optical centre and the principal focus. ■ Focal plane: the plane passing through F; all rays parallel to each other, whether parallel to the principal axis or not, converge on this plane.

■ radius of curvature (R): the radius of the sphere of which a surface of a lens is part. Since a lens has two surfaces, it will have two radii of curvature, R1 and R2. The relationship between focal length and radius of curvature is given by (1 1 1 = (n − 1) + f R1 R2) The S.I. unit for power is m−1, called a dioptre (unit symbol, D). Equation 9.5 shows us that for the focal length of a lens to be small, the radii of curvature of its surfaces must also be small (i.e. the lens must be curved sharply). Image formation by converging lenses Figure 9.15 displays ray diagrams showing image formation of an upright object, O, placed at various distances from a converging lens. For accurate drawing of ray diagrams, the lens is drawn very thin, as a line. Shapes are drawn on the two ends of the line to indicate the shape of the lens. The principal foci are indicated on both sides of the lens. We draw two rays coming from the tip of the object and follow these two rays as they pass through the lens. The image of the tip would be formed where the rays meet, or appear to meet. For a converging lens, for example: ■ a ray of light parallel to the principal axis passes through the lens, is refracted and then passes through the principal focus on the other side of the lens ■ a ray of light passing through the principal focus on one side of the lens ends up parallel to the principal axis on the other side of the lens ■ a ray of light passing through the optical centre of a ....(9.5) ■ Power (P): the reciprocal of the focal length (in m) of a lens. The power of a lens refers to the converging (or diverging) ability of the lens and has nothing to do with mechanical power. In equation form, P= Figure 9.14 (a) A thin lens has a large focal length. (b) A thick lens has a short focal length. 1 ....(9.6) f ITQ 6 lenses going straight through without deviation Some rules about ray tracing. Some rules about image formed by lenses are: (a) Which of the lenses in Figure 9.14 has surfaces of greater radii of curvature than the other? ■ the images are formed on the other side of the lens as (b) Which of the lenses has the greater power? ■ the images are always inverted. the object Chapter 9 Light: rays and waves a c b I F O I F O Figure 9.15 Image formation by a biconvex (converging) lens. (a) object further than 2f from the lens; (b) object between a distance f and 2f from the lens; (c) object nearer than f from the lens.

O I O ho v O u hi I Figure 9.16 Image formation by a biconcave (diverging) lens.

Figure 9.17 Similar triangles formed when a real image occurs. Lens equations and conventions and image are real). Since the triangles are similar, then by geometry, hi v = ho u Using u as the distance between the object and the lens, v as the distance between the image and the lens, it can be shown, using rays and geometry of triangles, that 1 1 1 + = ...(9.7) u v f If ho represents the height of the object and hi represents the height of the image, the magnification, m, of the lens is defined as hi m = ...(9.8) ho Lenses are capable of producing both real and virtual images. Real images are formed when rays converge at a point on the other side of the lens, i.e. convex lenses. However, virtual images can be formed when the image is virtual, and images could be upright or inverted, sign conventions are used when applying equations 9.7 and 9.8. Convention applied to distances (equation 9.7) ■ All distances (of objects or images) are measured from the optical centre of the lens. For the equation above to be consistent with the sign conventions, however, we must have hi = v ho u Hence, m = hi ho = −v u ....(9.9) Lens usage in optical instruments The human eye as an optical instrument Figure 9.18 shows the structure of the human eye.

The cornea, the transparent, curved part of the tough, protective white sclera, does most of the focusing of the light entering the eye. ■ Distances to real images are designated positive; distances to virtual things are designated negative. Convention applied to orientation (equation 9.8) sclera choroid aqueous humour ■ Upright heights are designated positive. ■ Inverted heights are designated negative. Using the lens sign conventions For the similar triangles shown in Figure 9.17: hi is positive; hi is negative (orientation convention); u is positive; v is positive (distance convention – both object ITQ 7 What difference is there between virtual images, formed by converging and diverging lenses? (Refer to Figures 9.15c and 9.16.) retina cornea pupil lens vitreous humour iris ciliary muscle Figure 9.18 The structure of the human eye. optic nerve 143 144 Unit 1 Module 2 Oscillations and waves Worked example 9.3: Image formed by a converging lens Q A An object is placed 3.0 cm from a converging lens of focal length 5.0 cm.

Determine, analytically (i.e. by calculations only): (a) the distance of the image from the lens (b) the magnification produced (c) the power of the lens. (e) Give a practical use of this arrangement of object and lens, giving reasons.

■ object distance, u = +3.0 cm (applying the distance convention: object distance is positive) ■ image distance, v = ?

■ focal length f = +5.0 cm (applying the distance convention: f is positive since the focus of a converging lens is real, i.e. formed from real rays converging to a point) 1 1 1 (a) = + u v f 1 1 1 3-5 = − = − cm-1 1 v 5.0 cm 3.0 cm 15 Therefore v = –7.5 cm The negative sign indicates that the image is virtual. The image is therefore formed 7.5 cm from the lens and on the same side as the object. You have to say this in your answer and not just write −7.5 cm. Getting it right! Position of images Avoid using terms like 'in front of' or 'behind' the lens in your answer, unless you specify that the object was in front of or behind the lens to start with. What some persons call 'in front' others call 'behind'. The safe answer is to say that the image is on the same side of the lens as the object or on the opposite side of the lens as the object. (b) The image orientation is the same as that of the object, since the image is virtual. v (c) magnification, m = −(−7.5 cm) = +3.0 cm = +2.5 1 (d) power, P = (but f must be in metres) f 1 P = +0.050 = +20 D (e) Use could be as a magnifying glass, since the image is larger that the object, and upright as the object. ■ A lens, elastic during much of one's lifetime, does the fine focusing so that really sharp images are formed on the retina. ■ The retina is a layer of cells that transform light they receive into electrical impulses which are carried to the brain via the optic nerve to be interpreted as 'sight'. In normal vision, the ciliary ring of muscles around each lens is relaxed and the elastic lens is in a thin state. Parallel rays of light (e.g. from a point at infinity) are focused on the retina as a sharp point. We say that the far point of this eye (the furthest distance in which objects can be seen sharply) is infinity. As objects are brought closer to the eye, the point image would have been formed behind the retina, causing a blurred image on the retinal itself. However, the ciliary muscles contract, squeezing the lens so that it bulges. This bulging causes the light to be focused on the retina. The ability of the lens to change shape so that focusing on the retina can take place is called accommodation. The closest distance the normal eye can see without undue strain on the ciliary muscles is about 25 cm. This distance is called the near point of the normal eye. Common eye defects and their correction Short-sightedness Figure 9.19a shows a condition called short-sightedness (myopia or near sight). The condition results from either (i) the eyeball being too long, or (ii) the lens or the cornea being too thick (i.e. too 'converging').

A person with short-sightedness can see objects closer than 25 cm from the eye because of the relatively short focal length of the relaxed lens system. However, such a person cannot see distant objects clearly (e.g. writing on street signs) clearly, since images are focused in front of the retina.

Making the lens thicker by contraction of the ciliary muscles causes the lens to bulge more and only worsens the focus on the retina. To correct this condition, a diverging lens is used, as shown in Figure 9.19b. Long-sightedness In long-sightedness (hypermetropia or far sight), a person can see distant objects clearly (e.g. writing on billboards) but not nearby objects (e.g. printed matter in a book). Figure 9.20 shows how this happens. Either the eyeball is too short or the lens system is not sufficiently curved, even when the ciliary muscles contract. Rays from a point object near to the eye converge after passing Chapter 9 Light: rays and waves through the cornea and lens system before coming to a focus, resulting a blurred image. A converging lens corrects this defect by increasing the power of the lens system of the eye. Note that the unaided long-sighted eye will not see distant objects clearly in the relaxed state. However, the ciliary muscles can contract automatically and increase the converging power of the eye, causing parallel rays entering the eye to be focused on the retina, resulting in seeing distant objects clearly. Getting it right! Distances from the eye In Worked example 9.4, it is assumed that the lens is practically in contact with the eyeball (like a 'contact' lens). If the person is wearing glasses a distance of 1.0 cm in front of the eyeball, then, in part (a), v would be −19 cm, and in part (b), v would be −7.0 cm, since u and v are measured from the corrective lens in applying the lens equation. A short-sighted eye. Correcting a short-sighted eye using a diverging lens. a diverging rays from near object retina b corrective converging lens Figure 9.20 (a) Long-sightedness (hypermetropia). (b) Correction of long-sightedness using a converging lens. Worked example 9.4: Power of lens needed to correct short-sightedness Q A The near point and far points of a short-sighted eye are 8 cm and 20 cm, respectively, from the eyeball. (a) What power of corrective lens is needed for this eye to see distant objects clearly? (b) What will the resulting near point then be? (Assume that the lens is placed very close to the eye.) a b ■ u = ∞ ■ v = −20 cm (negative because image seen is virtual) ■ f = ? ■ P = ? power, P, if measurements are in metres) u v f 1 1 + = 1 P =0.20 m = P ⇒ P = −5.0 D (the negative value is in agreement with a diverging lens needed to correct short-sightedness) u v = −8.0 cm ■ P = −5.0 D ■ u = ? The corrective lens should have a power of −5.0 D. (b) The object is to be placed at the near point of the lens-aided eye (object distance, u). The virtual image is seen at the near point of the unaided eye, where v = −8.0 cm = −0.080 m 1 1 1 + = (= power, P, if measurements are in metres) u v f 1 1 1 1 + = −5.0 ⇒ u = −8.0 cm Figure 9.21 Correcting a short-sighted eye. Finding (a) the power, P, of lens needed; (b) the position of the resulting near point. (b) The object is to be placed at the near point of the lens-aided eye (object distance), u = +? ⇒ u = 0.080 m 1 1 ⇒ = −5.0 m−1 ⇒ −1 = 12.5 m−1 − 5.0 m−1 = 7.5 m−1 u Therefore u = 7.5 m−1 = 0.13 m This is 13 cm from the corrective lens, the new near point, and lens and infinity. 145 Unit 1 Module 2 Oscillations and waves a b 1 10 12 2 line image 11 4 8 point 3 9 14 7 6 Astigmatism If the surface of a person's cornea or lens is shaped somewhat cylindrically (rather than being part of a sphere), the person sees blurred images. This condition is called astigmatism. Figure 9.22a shows how a vertical cylindrical lens gives rise to a line image. The lines will overlap and the image appear blurred. Figure 9.22b shows a test for astigmatism. Lines appearing blurred or faint indicate astigmatism along a specified direction. Corrective lenses are shaped with a compensating cylindrical surface. Presbyopia As a person gets older the lens loses its elasticity. If the lens becomes set (e.g. somewhere between its thinnest and thickest states), accommodation cannot occur. This condition is called presbyopia, derived from the root 'presbyter' which means 'elder'. To correct presbyopia, bifocal lenses are usually prescribed. The upper part is a diverging lens to enable clear distance vision, and the lower part is converging to enable close vision as in reading a book. Varifocus or progressive lenses can also be used instead. The power of the latter lenses varies gradually from diverging to converging, enabling focus on the retina from infinity to about 25 cm through different parts of the lens. Cataract Another eye defect is cataract, in which the eye lens becomes opaque (usually with aging).

Vision becomes blurred as the milky lens scatters light. Ultimately, blindness may result if no light can reach the retina. One common treatment for cataract is to remove the lens by laser surgery and replace the cloudy lens with an artificial one. Some patients, however, have experienced damage to the retina following this process. An older treatment ITQ 8 Instead of using corrective lenses, can anything be done to the cornea to correct short-sightedness, long-sightedness or astigmatism? 5 Figure 9.22 (a) A point object at the lens image formed by a cylindrical lens. (b) A test for astigmatism – lines appear blurred or faint in one direction. After having the lens removed, special, thick lenses along with sunshades that block harmful UV light. There has been some research into eye drops that can slow down or even dissolve cataracts. However, it is claimed that cataract formation is also slowed down by not smoking, restricting one's intake of toxic chemical substances and reducing one's exposure to UV light.

The simple single-lens camera Figure 9.23a shows the structure of a simple single-lens camera. Focusing is done by rotating the lens barrel which causes the lens to be positioned at varying distances from the film or other light-sensitive device such as a CCD (charge-coupled device). When the lens is positioned at a distance equal to the focal length, f, from the film, parallel rays from a point object will give rise to a sharp image on the film. For objects closer to the camera, the lens is moved a distance greater than f from the film. In Figure 9.23b, the camera lens is focused on a nearby point object. An ideal lens would produce a point image, but a real lens produces a spot due to aberrations and diffraction effects. The size of this spot when the lens focus is optimum is the 'least circle of confusion'. In Figure 9.23b, the circle of confusion due to the arrangement of nearby point object, lens and film (screen) is small since the film is at a position for the arrangement shown, the circle of confusion is at a minimum. The least circle of confusion at the image position of a large size, additionally, at film less than the lens. An extended nearby object on which the rays of the camera focuses would therefore look sharp. However, a distant object (further far closer) would look blurred (Figure 9.23c). If the lens is moved such that the film is at an intermediate position, C (Figure 9.23b), there would be the least circle of confusion overall. The nearby object as well as the distant object would cast circles of confusion of approximately the same size at C and the picture will Chapter 9 Light: rays and waves a b viewfinder lens c C rays from distant object D = lens opening 'circle of confusion' for distant object (greatly exaggerated) film or CCD shutter rays from nearby object (in focus) iris diaphragm or 'stop' Figure 9.23 (a) Structure of simple single-lens camera. (b) Focussing in single-lens camera. (c) Appearance of picture when camera is focused on a nearby object. ■ from Figure 9.24c look moderately sharp. For a lens set to photograph a nearby object to give a sharp image, the range of distances over which the image is seen acceptably sharp is called the depth of field for that setting. Depth of field can be increased by using a lens opening of small diameter, D, since only rays near to the principal axis will enter, resulting in smaller circles of confusion. MN = but h N θN h = ÷ = ...(9.13) u N u θ 1 1 1 + = u v f i.e. The magnifying glass and the microscope = Figure 9.24c shows how a magnifying glass forms a virtual, magnified, upright image, I, of an extended object, O. The angular magnification, M (or magnifying power, not to be confused with power, P, which is equal 1 to) produced by the lens is defined by the equation, f θ M = N ....(9.10) θ where θ is the angle subtended at the naked eye by the object placed at the near point (Figure 9.24a), and θN is the angle subtended at the magnifying lens being used. therefore, a h θ∞ h = ÷ f N θ h ....(9.12) b object c to ∞ N f N(N + f) (N + f) = fN f MN = N +1 f ....(9.15) N θN θ∞ θ MN = The lens equation (equation 9.8), and consequently equations 9.12 and 9.15, are based on a ray model of light.

The latter two equations indicate that the shorter the focal length of a magnifying glass, the greater the magnification. However, lenses of very short focal length tend to produce distortions (called aberrations) ■ from Figure 9.24b N f ....(9.14) Equations 9.12 and 9.15 show that a greater angular magnification is obtained if the object is positioned such that the image is formed at the near point rather than if the object is positioned at the principal focus of the lens. h ....(9.11) N therefore, M∞ = N+f fN Combining equation 9.14 with equation 9.13, we get ■ from Figure 9.24a M∞ = 1 1 − (−N because the image is virtual) f −N 1 1 1 = + u f N = With reference to Figure 9.24, we arrive at the following for small angles: θ= 1 1 1 = − u f v f Figure 9.24 (a) Object placed at near point, showing angle, θ, subtended by the naked eye. (b) Object placed at the principal focus so that the image is seen at infinity. (c) Object placed closer than f to the lens such that the image is seen at the near point. 147 148 Unit 1 Module 2 Oscillations and waves that are due to the spherical shape of their surfaces. This means that it is not practical to achieve very high, useful magnifications using a single lens. As a result, two lenses are used to achieve large magnifications, as in microscopes. The first lens produces a magnified, real image, and the second lens produces a virtual magnified image of the real image produced by the first lens. There is a second problem, however, with lenses being used for magnification – there is a limit to the magnification that can be produced by a lens. This is explained on the basis of light propagation by waves rather than by rays. This limitation will be taken up in the next section, which looks at diffraction. Diffraction of light Diffraction is a wave phenomenon (see Chapter 8) and is the spreading out of a wave as it passes through an opening that is of the same order as, or smaller than, the wavelength of the wave. Diffraction through a single slit Using laser light and a very narrow slit, a definite pattern of bright and dark bands (called fringes) is easily seen (Figure 9.25). Also, the spreading of light from the slit can a bright central fringe θ points on an object Figure 9.26 Diffraction pattern of two points placed in front of a lens showing the limit of resolution of a lens.

be seen on a screen far away. Laser light is light in which the waves are coherent, i.e. in phase with each other, and are also of a single wavelength. Bright fringes are located where, on average, the waves which have passed through different parts of the slit arrive in phase and reinforce each other (constructive interference). The dark fringes are located where the interference is, on average, destructive. Fringe formation will be discussed in more detail in the section on diffraction through a double slit. Limit of resolution of a lens A lens lets light through it and therefore acts like a slit or a hole. For a point object, the image produced by a lens is not a point, but a circular diffraction pattern (Figure 9.25b). For two points, with a tiny angular separation, θ, the diffraction patterns might overlap (Figure 9.26). Increasing the magnification simply enlarges the diffraction patterns and does not make seeing the details of the tiny points any better. It can be shown that for a circular lens of diameter D, the smallest angular separation between points that can just be resolved as separate images depends on the wavelength, λ, of light as follows: θ≈ b bright central spot lens diffraction pattern at limit of resolution 1.22 × λ D ....(9.16) Thus, the 'wave' nature of light limits the useful magnification that the 'ray' nature of light predicts. Diffraction and interference through a double slit Figure 9.27 shows two slits, A and B, separated by a tiny distance AB = a. Monochromatic light (light of a single colour, i.e. of a single wavelength) is incident normally on the slits. Light from a single source is used to ensure that light passes through A and B simultaneously in phase. Light whose waves are in phase diffracts through each slit Figure 9.25 The diffraction pattern formed using monochromatic light passing through (a) a vertical slit and (b) a circular hole. ITQ 9 What sort of wavelength of light, long or short, should be used in order to be able to distinguish very tiny details when using a microscope. Chapter 9 Light: rays and waves Worked example 9.5: Smallest distance barely distinguishable by a lens Q A What is the smallest size of object that can just be resolved (i.e. seen as a separate object) through a 1.5 cm diameter lens when placed 1.2 cm from the lens if light of wavelength 500 nm is used? (Assume no lens aberrations due to shape.) By equation 9.16, the smallest angle, θ ≈ θ≈ 1.22 × λ D 1.22 × 500 × 10−9 m 1.5 × 10−2 m ≈ 4.1 × 10−5 rad The smallest size, s = rθ ≈ 1.2 × 10−2 m × 4.1 × 10−5 rad ≈ 4.9 × 10−7 m Compare this with the size of a red blood cell (~5 × 10−6 m), Staphylococcus bacterium (~5 × 10−7 m) and Rhinovirus (~2 × 10−8 m). waves from different regions of the bulb filament are not coherent.

The single slit only allows light from a small region of the filament to reach the double slit. Alternatively, a bulb with a narrow vertical filament can be used if the bulb is placed so that the filament is parallel to the slits. These issues do not arise when a laser is used as the light source.) Waves also arrive in phase at Q if the path difference, AC, from the two slits to the screen is an integral multiple, n, of the wavelength, λ, of the light. Thus, for constructive interference when a wavefront, e.g. BC, makes an angle θ with AB, sin θ = AC (see Figure 9.27b) AB sin θ = nλ a ....(9.17) nλ = a sin θ ....(9.18) The detail in Figure 9.27b also shows that as semicircular wavelets. A screen is placed far away, a distance D from the two slits. Along the mid-line from the slits, constructive interference occurs at P. Because AP = BP, and the waves left A and B in phase, they arrive at P in phase. As a result, a bright fringe is formed on the screen at P along the centre line OP. (Note: if two incandescent bulbs are used, one in front of each slit, a pattern displaying constructive interference is not obtained on the screen. This is because light from each of these bulbs are in random phase and will cancel each other on the screen.

However, if a single incandescent bulb is used as the light source, it is usually necessary to place a single slit between the bulb and the double slit to produce a light source with sufficient coherence to produce a constructive interference pattern at regions like P on the screen (Figure 9.27a). This is because light θ + x = 90° and α + x = 90° so α = θ According to Figure 9.27a, for small angles θ, the distance, y, of the first bright fringe from the central fringe is given by y for the n = 1 bright fringe D ....(9.19) sin θ ≈ tan θ = Getting it right! For the formation of bright fringes, the path difference, AC, must be an integral number of wavelengths. Thus nλ = a sin θ. For dark fringe formation, AC must be 1 (n + 2)λ so that the two waves are out of phase. Hence, 1 dark fringes occur when (n + 2)λ = a sin θ. a b monochromatic light A a C central fringe O θ to P α a O P (n = 0) y θ B A n=1 D C Q (n = 1) R (n = 2) x θ to Q B Figure 9.27 (a) Interference pattern produced by a double slit. The graph on the screen represents relative brightness of the fringes. The central fringe is the brightest. Note: because AB is tiny as compared to OP, this means that AQ, OQ and BQ are practically parallel to each other. (b) Detail of direction of light from a double slit when constructive interference occurs on a screen far away. 149 150 Unit 1 Module 2 Oscillations and waves For D >> a, θ is very small. Stealth aircraft and interference Therefore sin θ ≈ θ ≈ tan θ, where θ is in radians. Two-source interference is also applied in making stealth aircraft difficult to detect (Figure 9.6). The aircraft is coated with a partially reflecting thin layer of thickness t (Figure 9.29). Combining equations 9.17 and 9.19, we get, for small θ, y nλ ≈ a D nλD Therefore y ≈ a ....(9.20) Worked example 9.6: Distance of second-order fringe from central fringe Q Light of wavelength 600 nm, from a single source, passes through two slits 0.100 mm apart. An interference pattern is seen on a screen 1.10 m from the slits. How far is the secondorder bright fringe on the screen from the central fringe? A First we draw a diagram (Figure 9.28). Note: for simplicity in calculations, we omit the lines from the slits and draw only the line from O. λ D O θ2 y n=2 body of aircraft B t layer of paint A When a radar wave is C incident at a point A on the incident radar wave from special paint layer, partial station reflection occurs (AD) and partial transmission (AB). The wave A→B is reflected D E at the boundary between reflected waves the special paint layer and the surface of the aircraft Figure 9.29 Anti-radar coating on stealth aircraft. and emerges as wave CE. For there to be practically no detection of the reflected waves, AD and CE must be 180° out of phase with each other. This can be achieved if the path difference between 1 1 the two waves, AB + BC, is 2 λ or (m + 2)λ where m is a whole number. Therefore, for destructive interference of the reflected waves, 1 Figure 9.28 Double-slit interference. wavelength of light, λ = 600 nm = 600 × 10−9 m ■ distance between slits, a = 0.100 mm = 1.00 × 10−4 m ■ slits-to-screen distance, D = 1.10 m ■ order of bright fringe, n = 2 ■ distance between second-order bright fringe and central fringe, y =? For second-order bright fringe (constructive interference) nλ = a sin θ2 nλ therefore, sin θ2 = a 2 × 600 × 10−9 m = 1.00 × 10−4 m = 1.20 × 10−2 θ2 is a very small angle (in radians) and so sin θ2 ≈ θ2 ≈ tan θ2 (see Appendix 3, Trigonometry). y But, tan θ2 = (equation 9.19) D therefore, y ≈ Dθ2 ≈ D sin θ2 ≈ 1.10 m × 1.20 × 10−2 ≈ 1.32 × 10−2 m ≈ 13.2 × 10−3 m or 13.2 mm ■ (m + 2)λ = AB + BC = 2t 1 (m + 2)λ or t = 2 The minimum thickness of the anti-radar coating must then be λ 1 t = (2 λ ÷ 2) = 4 Note: the above discussion assumes that both at A and at B, reflections occurred with the same change in phase, i.e. either in-phase or out-of-phase in both cases (see Chapter 8, Figure 8.16).

Diffraction and interference through multiple slits – the diffraction grating A diffraction grating (Figure 9.30) consists of many slits very closely spaced.

If the spacing, a, between slits is the same in the grating as for the double slit, the bright and dark fringes are formed at the same spots on the screen, since a wave from one slit undergoes the same path difference in the same direction as a wave from the adjacent slits. The difference on the screen, however, is that the bright fringes are less tapered out and so the light is concentrated, resulting in fringes much brighter than Chapter 9 Light: rays and waves a b fringe from grating path difference λ λ fringe from double-slit light λ λ to screen far away λ those produced by the double slit. A very large dark space results between adjacent bright fringes. Since path differences are the same in both the double and multiple slits, the equations are the same, i.e. nλ = a sin θ...............(9.18) tan θ = y ....(9.19) D and for D ≫ a y≈ nλD ....(9.20) a Worked example 9.7: Highest order of fringe observable Q What is the highest order of bright fringe that can be seen if light of wavelength 600 nm is incident normally on a grating that has 8000 lines per cm? A 1 cm = 1.25 × 10−6 m 8000 ■ wavelength of light, λ = 600 nm = 600 × 10−9 m = 6.00 × 10−7 m For an order to be seen on a screen, the angle θ must be less than 90°.

nλ < a sin 90° a n< λ 1.25 × 10−6 m n< 6.00 × 10−7 m n < 2.08 Therefore the highest order visible would be n = 2 since orders are in whole numbers. ■ slit separation, a = ITQ 10 A diffraction grating has 300 lines per mm (i.e. 300 slits per mm).

What is the separation, a, between adjacent slits (give your answer in m). ITQ 11 What is the order of magnitude for visible light (a) in frequency? (b) in wavelength? Figure 9.30 (a) Waves in a particular direction from a diffraction grating. (b) Interference pattern produced by the grating, with double-slit interference pattern (faint lines) superimposed for the same slit spacing, a. Polarization of light The polarization of light is discussed in Chapter 8 in the section 'Transverse waves.' Polarization is vibration of particles in a single plane as the wave moves along. Light travelling as rays cannot account for polarization, only light travelling as waves, and even then only as transverse waves. Position of light in the electromagnetic spectrum Light belongs to a class of waves that have electric and magnetic properties. These waves form the electromagnetic spectrum. The waves all travel in free space at the same speed, the speed of light (3.0 × 108 m s−1, to 2 significant figures). Figure 9.31 shows the position of visible light in the electromagnetic spectrum. The wavelengths of visible light extend from about 400 nm (ultraviolet) to 700 nm (red). Light: rays or waves? Our discussions in this chapter seem to point overwhelmingly to a 'wave' rather than a 'ray' nature of light. A wave model accounts for many of the behaviours shown by light, such as rectilinear propagation, reflection, refraction, diffraction, interference and polarization. However, the wave model does not account for types of images formed – something which the 'ray' model does very well. Also, whereas the 'ray' model suggests that all sizes of magnification are possible with lenses, the 'wave' model predicts a limit, imposed by wavelengths of light, to useful magnifications. We will revisit the ideas of 'rays' and 'waves' in a later chapter of this book, since we will come across evidence that light consists of a stream of particles (called 'photons'), and we know that, according to Newton's first law, particles travel in straight lines ('rays'). However, these same photon 'particles', we will find, have 'wave' properties also. 151 152 Unit 1 Module 2 Oscillations and waves Frequency, f / Hz gamma rays Wavelength, λ / m 23 10 –14 10 1022 10–13 gamma rays 21 10 X-rays 10–12 20 10 10 19 10 ultraviolet 10–10 1018 X-rays 10 17 10 10 15 10 10–7 visible 10–6 14 10 10 infrared 1013 1 THz 10 10 –3 10 –2 11 10 1 GHz 109 108 microwaves short radio waves 10–1 1 101 television and FM radio 7 10 102 AM radio 6 1 MHz 10 10 105 104 –5 10–4 12 microwaves –9 10–8 ultraviolet 16 1010 –11 3 104 105 long radio waves 3 1 kHz 10 106 radio waves Figure 9.31 Electromagnetic spectrum. Summary ✓ Light is part of a spectrum of electromagnetic radiation; visible light extends from 400 nm to 700 nm in wavelength. Light displays behaviours characteristic of waves (such as refraction, diffraction, interference and polarization). In certain situations (e.g. the photoelectric effect), light behaves as if it were a stream of particles (photons) but these 'particles' themselves display wave characteristics such as frequency and wavelength. This will be covered in Chapter 28. ✓ The refractive index, n, of a medium is defined by the equation c ....(9.3) v where c is the speed of light in free space and v is the speed of light in the medium. n= ✓ When light undergoes refraction between two media, the following laws of refraction apply: the incident ray, the refracted ray and the normal at the point of incidence all lie in the same plane sin θ1 ■■ the ratio sin θ is a constant, where θ1 and θ2 2 are the angles of incidence and refraction, respectively. This latter statement is known as Snell's law. ■■ ✓ A 'ray' treatment of light is necessary to account for the various types of images formed by reflection and refraction. ✓ The laws of reflection state that: ■■ the incident ray, the reflected ray and the normal at the point of incidence all lie in the same plane ■■ the angle of incidence (θi) is equal to the angle of reflection (θr). ✓ According to the Snell's law equation, n1 sin θ1 = n2 sin θ2...............(9.4) Chapter 9 Light: rays and waves ✓ The critical angle is the angle of incidence when the angle of refraction is 90º. For total internal reflection to occur light must be travelling from one medium to another whose refractive index is less and the critical angle for the two media must be exceeded. ✓ Optical defects of the eye can be corrected by using converging lenses (for long-sightedness), by using diverging lenses (for short-sightedness) or by removing and replacing the eye lens (for cataract). ✓ When light passes through two slits, diffraction ✓ Four lens equations are: 1 1 1 + = u v f hi −v m= = ho u P= occurs through each slit and an interference pattern is formed on a screen if: ....(9.7) ....(9.9) 1 f ....(9.6) (1 1 1 = (n − 1) + f R1 R2) ....(9.5) ✓ For the lens equations to work properly, lens conventions must be used. For example, distances to real things and upright heights are designated positive and vice-versa.

✓ For a magnifying glass, angular magnification, M, is given by: θ M= N θ ....(9.10) M∞ = N f ....(9.12) MN = N +1 f ....(9.15) Review questions Rays and wave optics; Rectilinear propagation of light; Reflection of light from plane surfaces 1 A 'geosynchronous' satellite orbits the Earth at an approximate height of 36 000 km. A TV signal sent from one part of the Earth reflects off the satellite and is received at another location on the Earth. Estimate the time taken between the sending and the reception of the TV signal. (Have you ever observed this 'time-delay' in international TV broadcasts?) 2 Both rays and waves travel in straight directions in a homogeneous medium. Explain why we cannot 'see' around corners but we can 'hear' around corners. 3 Our nearest star (other than the Sun) is approximately 4.0 × 1016 m away. How long does light from this star take to reach the Earth: (a) in seconds? (b) in years? (c) When we look at it, are we seeing it as it is now? ■■ the slits are very close together (a is small) ■■ the light is coherent as it enters both slits ■■ the screen-to-slit distance (D) is large (interference patterns, known as Fresnel patterns, are also formed closer to the slits, but the simple equations produced when D is large do not apply to these). ✓ The equations for double-slit (and multiple slit) interference are: nλ = a sin θ...............(9.18) nλD for D ≫ a a y tan θ = D y≈ ....(9.20) ....(9.19) 4 Two plane mirrors, M1 and M2, are at an angle of 120° to each other (Figure 9.32). A ray of light makes an angle of 40° with M1. What angle does the reflected ray make with M2? 40˚ M2 M1 120˚ Figure 9.32 5 In a Cavendish-type lamp/mirror/scale experiment (see Figure 9.4b), the distance from the mirror to the scale is 3.00 m. If the spot of light moves 1.4 cm when the mirror rotates, determine the angle of rotation of: (a) the beam of light (b) the mirror. (Give your answers in radians and in degrees.) 153 154 Unit 1 Module 2 Oscillations and waves Refraction at plane surfaces a 6 (a) Define refractive index of a material in terms of wave speeds. (b) A fibre optic cable consists of a central, transparent core surrounded by a transparent cladding (Figure 9.33). air 35˚ water 2.5 m incident light P b W3 W2 W1 wavefront air core cladding Figure 9.33 Fibre optic cable. (i) Explain how light is able to travel within the cable with very little loss of light energy. (ii) What two conditions must be met for the light to be able to travel along the cable with little loss of energy. (iii) The materials making up the 'optical' fibre have refractive indices of 1.42 and 1.50. Determine the minimum angle of incidence necessary between the core and the cladding for the situation in part (i) to take place. 7 (a) With the aid of suitable diagrams explain: (i) how light is able to travel along an optical fibre even if it is bent (ii) why very little light will travel along the fibre if it is bent very sharply. (b) Give an optical explanation why optical fibres are made with a cladding and not just with the core left surrounded by air.

(c) Briefly describe one practical application of fibre optics. 8 A beam of light, travelling from air, makes an angle of 35° with the surface of water in a pool 2.5 m deep (Figure 9.34). (a) (i) On Figure 9.34a, label the angle of incidence, θ1, and the angle of refraction, θ2. (ii) State the Snell's law equation in terms of θ1 and θ2, and the refractive indices, n1 and n2 of air and water, respectively. (ii) Calculate the angle of refraction. (iii) Determine the distance from P to the spot when the beam of light reaches the bottom of the pool. water Figure 9.34 A beam of light entering water in a pool. (b) A wavefront, W1, is superimposed on the incident beam of light (Figure 9.34b). At equal times later, wavefronts W2 and W3 are shown partially. (i) Complete Figure 9.34b to show the complete wavefronts W2 and W3. (ii) Draw the 4th wavefront W4 using the same time interval between wavefronts.

9 Figure 9.35 shows a beam of light, A, travelling from Perspex to air. The beam in the air is labelled B.

a Perspex B θ2 air A θ1 b θ1 / ˚ 40 30 20 10 0 0 20 40 60 80 100 θ2 / ˚ Figure 9.35 (a) Light travelling from a semicircular Perspex block to air. (b) Sketch of graph of angle of incidence, θ1, versus angle of refraction, θ2. Chapter 9 Light: rays and waves (a) In an actual set-up like Figure 9.35, another beam of light, C, appears along with B. (i) Accurately draw and label beam C on the diagram. (ii) Explain how the simultaneous appearance of beams B and C suggest a wave nature of light.

(b) (i) Define critical angle. (ii) In an experiment, values of θ1 from 10° to 38° and the corresponding angles of θ2 were obtained. A graph of θ1 versus θ2 was plotted (see Figure 9.35b). How can the graph of θ1 versus θ2 be used to determine the critical angle θc for the Perspex– air boundary? (c) Explain how you can use the value of the critical angle obtained in (b) to determine the refractive index of Perspex. Refraction at curved surfaces; Lenses in optical instruments 10 The focal length of a camera lens is 100 mm. The maximum distance between the lens and the plane of the film is 130 mm. (a) What is the closest distance from the lens that an object can be placed and photographed in sharp focus? (b) How far from the film should the lens be when photographing an object 15.0 m from the lens? 11 A film projector has a focal length of 100 mm. The distance from the projector lens to the screen is 20 m. (a) How far from the lens should a 35 mm wide film be placed in order to obtain a sharp image on the screen. (b) What will the width of the image on the screen be? 12 An aged person has an eye defect such that she cannot focus on objects less than 100 cm from her eye (Figure 9.36). 100 cm 25 cm Figure 9.36 (a) (i) Give the most likely name for the eye defect. (ii) Suggest three possible causes for the eye defect. (b) (i) Draw a suitable lens and corresponding rays on the diagram to show how the eye defect may be corrected so that she can read normal text placed closer to the eye. (ii) Calculate the focal length and power of the lens needed such that the person can see objects clearly at 25 cm from the eye (the typical near point for the human eye).

(c) Calculate the farthest distance that the lens-aided eye in part (b) (ii) would be able to see clearly. 13 Figure 9.37a shows how an object O, when placed in air, forms a sharp image, I, in the eye. (a) On the diagram, draw the path of the same two rays from O when the object is viewed with the naked eye under water instead of in air (refractive indices of air, water and cornea, respectively, are 1.00, 1.33 and 1.38).

Explain why the image seen would be blurred. (b) Explain why, in Figure 9.37b, if goggles are worn with air trapped between the goggles and the eye, clear vision underwater could be restored. a air I O goggles b water O air Figure 9.37 Vision: (a) in air and (b) under water, using goggles. 14 The objective and eyepiece lenses of a compound microscope each produce magnified images, but produce them differently. (a) Explain how each of the two lenses produces its magnified image. (b) Figure 9.38 shows the arrangement of objective and eyepiece lenses of focal lengths 5.0 cm and 10.0 cm, respectively, in a compound microscope model. The lenses are 20.0 cm apart.

A tiny object is placed 7.0 cm in front of the objective lens (i.e. on the left of the objective lens). 155 156 Unit 1 Module 2 Oscillations and waves objective lens 20 cm f = 5.0 cm eyepiece lens f = 10.0 cm Figure 9.38 Compound microscope (simplified) showing arrangement of lenses. (i) By means of a diagram, drawn accurately to scale, draw rays showing the formation of the image by the objective lens and the final image seen by the eye. (ii) By taking appropriate measurements from the diagram, determine the magnifications produced by 1 the objective lens only 2 the eyepiece lens (the eyepiece further magnifies the image produced by the objective lens) 3 the lens combination. Diffraction and interference of light 15 (a) According to Huygens' principle, a wavefront can be considered as being made up of tiny point sources that give off spherical wavelets travelling in the forward direction at the same speed as the wave itself. A subsequent wavefront formed is tangential to the spherical wavelets (Figure 9.39a). (b) Figure 9.39b shows a plane wave approaching a barrier. Applying Huygens' principle to the wave as it passes the barrier, show that it is possible to have a strong wave behind the barrier along the barrier's perpendicular bisector. a wavefront b subsequent wavefront wavefront approaching barrier barrier perpendicular bisector to barrier wavefront at barrier Figure 9.39 16 (a) With the aid of a diagram, explain how bright and dark fringes are formed on a screen far away from a diffraction grating (slits-to-screen distance, D ≫ a, spacing of slits). (b) A narrow beam of white light is incident normally on a 6000 lines per cm grating. A screen is placed 4.00 m from the grating. (White light is made up of wavelengths from 400 nm to 700 nm.) (a) (i) Draw a labelled diagram showing the beam of light, the grating, the screen, the zeroth-order fringe and the first-order fringe. (ii) Describe and explain the appearance of the zeroth- and first-order fringes. (iii) Determine angular deviations from the normal for the extremes of first-order fringe. (iv) Determine the width (in cm) of the first-order fringe. 17 Two slits are separated a distance a. A screen is placed a distance D from the double slit. When a narrow beam of monochromatic light of wavelength λ is incident normally on the slits, a zeroth-order and several orders of bright fringes appear on the screen. (a) With the aid of a suitably labelled diagram, derive the following equations: (i) λ = a sin θ (for the first-order fringe, where θ is the angle of deviation of the fringe from the normal) y (ii) tan θ = D (where y is the distance of the first bright fringe from the zeroth-order fringe) (iii) y ≈ λD a (where y is only an approximate value) (iv) nλ = a sin θ (for the nth-order bright fringe formed on the screen). (b) Monochromatic light of wavelength 680 nm passing through a double slit forms a first-order bright fringe 4.6 cm from the zeroth fringe on a screen 3.75 m away. Determine the approximate separation between the two slits. 18 When monochromatic light is incident normally on a pair of closely spaced slits, a bright central fringe is formed on a screen far away along the perpendicular bisector between the slits. When a thin sheet of transparent material is placed to cover one of the slits, the bright central fringe shifts and a dark fringe takes its place on the screen. Explain how this change happens.

Chapter 9 Light: rays and waves Polarization of light; Position of light in the electromagnetic spectrum; Light: rays or waves? a light box with red filter 19 (a) What is meant by polarization of light? (b) Describe, with the aid of diagrams, practical activities that show that the following are polarized. You may need to refer to Chapter 8: (i) visible light (ii) microwaves. b screen grating n=1 y n=1 fringes red light beam a b grating collimator source S L θ telescope eye Practical exercises and challenges 1 You are given a lens with a stated value of its focal length, f. Use the arrangement in Figure 9.40 to determine f by plotting values of 1/v against 1/u. Explain why the intercepts on both axes will give the value of 1/f.

Estimate the percentage error expected in your experiment and compare it with the percentage error you obtained between your value and the given value of f. 4 Use a semicircular block of Perspex in air to determine the critical angle, and hence the refractive index, of Perspex, by plotting the angle of incidence, θ1, against the angle of refraction, θ2 (see Review question 9). converging lens screen Figure 9.41 5 Determine the thickness of a tiny object (e.g. a hair or a sheet of paper) using a plane mirror, a narrow beam of light and a screen. The formula Q = 2P might be helpful, and an arrangement is suggested in Figure 9.42.

light source image u scale v Figure 9.40 2 Convince yourself by this simple activity that light demonstrates wave properties. Shine a narrow beam of light (e.g. from a flashlight) on to a flat piece of transparent plastic or glass at an acute angle to the beam in a fairly dark room. Notice partial reflection and partial transmission of the light – a characteristic of waves. 3 Using either of the arrangements shown in Figure 9.41 to determine the wavelength, and hence the frequency, of a given colour of light (e.g. red). To find an average, fringes on both sides of the zeroth-order fringe can be used. collim source θ light order box grating 20 (a) By what factor (in of magnitude, i.e. 10x) screen with red filter are wavelengths of cellular phone electromagneticn = 1 waves greater than those of light waves (see θ Figure 9.31)?

y n=1 (b) Estimate, by calculation, the wavelength of: fringes red light (i) a 900 MHz cellular phone wave beam (ii) red laser light of frequency 4.7 × 1014 Hz. illuminated object 157 incident beam reflected beam Q hair mirror t P m Figure 9.42 S 158 Unit 1 Module 2 Oscillations and waves Answers to ITQs Answers to Review questions 1 20º. Q = 2P. The angle of incidence of the light does not matter, only the angle of rotation of the mirror. 1 0.24 s 2 The speed of light in glass must be less than the speed of light in air (see Figure 8.17). 3 c n= v c n 3.00 × 108 m s−1 = 1.33 Therefore v = = 2.26 × 108 m s−1 4 Any straight line from the circumference of a circle to its centre is a radius. A radius is perpendicular to the circumference at the point of intersection of the two. Hence the angle of incidence of rays entering the block is zero. Therefore the incident rays in air go straight through the block to the point, O. 5 OR3, since none of the incident light energy exits by refraction, and energy is conserved. 6 (a) Lens (a) (b) Lens (b) 1 since P = f 7 Virtual images are always formed on the same side of the lens as the object and are always upright.

However, virtual images formed by converging lenses are larger than the object, whereas virtual images formed by diverging lenses are smaller than the object. 8 Since the cornea is a lens in itself, its outer surface could be 'shaved' using a laser to make it more curved or less curved as needed for the correction to the vision defect. 9 Very short wavelength light (from equation 9.16, for θ to be small, λ must also be small). Note also: increasing D, the width of the lens, can also result in higher resolution. 10 a = 1 mm = 3.33 × 10−6 m 300 11 (a) (Hz) 1014 (b) (m) 10−7 3 (a) 1.3 × 108 s (b) 4.2 years 5 (a) 4.7 × 10−3 rad, 0.27° (b) 2.3 × 10−3 rad, 0.13° 9 (b) (ii) Extrapolate the graph to θ2 = 90°. Then θ1 = critical angle, c ≈ 42°. (c) n = 1 ≈ 1.5 sin c 11 (a) 100.5 mm (b) 7.0 m 17 (b) 0.055 mm 159 Chapter 10 Sound Learning objectives ■ Identify physical quantities closely associated with pitch, loudness and quality of sound. ■ List the major sections of the sound spectrum. ■ Discuss wave properties of sound. ■ Describe activities and solve problems based on reflection and interference of sound waves. ■ Define the following and solve problems involving them: P sound intensity (I = A) ●● threshold intensity of hearing (I0) ●● threshold intensity of discomfort/pain (Ip) I ●● sound level β (dB) = 10 log I0 ■ Discuss the response of the ear in terms of: ●● frequency ●● sound intensity ●● sound level ●● loudness level. ■ Discuss application of sound in: ●● musical instruments ●● industry ●● medicine. ●● () vibrating loudspeaker cone rarefaction sound waves compression Figure 10.1 Sound travels through the air by waves of compressions and rarefactions. The air itself does not travel – only the waves do. Describing sound Production and transmission of sound Sound frequency We perceive sound when our eardrums vibrate and pass on these vibrations to specialized sensory cells within our ears. The vibrations of our ear drums are due to mechanical pressure waves that travel through the air. Sound requires a material medium (e.g. air) and a mechanically vibrating source (e.g. a vibrating loudspeaker cone). The source produces compressions and rarefactions within the medium. These compressions and rarefactions travel through the medium as longitudinal waves. We call these longitudinal waves sound. We often describe sound by its pitch. The sound of a chirping bird, or from a small whistle, is said to be of high pitch; that from a bellowing cow, of low pitch.

The pitch of a sound is closely related to its frequency. The frequency, f, is the number of vibrations per second made by the sound waves. Sound can travel through solids, liquids or gases (but not through an empty space since nothing would be available to compress or rarefy in an empty space). Figure 10.1 shows regions of compressed and rarefied air (called compressions and rarefactions) travelling as sound waves through air. The areas of compressed and rarefied air are at high and low pressures, respectively. Figure 10.2 shows how an oscilloscope connected to a microphone can be used to measure the frequency of sound. The microphone converts pressure variations received on its diaphragm into electrical voltages which are displayed on the oscilloscope screen as voltage (y-axis) versus time (x-axis) variations. One and a quarter voltage cycles are shown in the diagram. With the time base set at 0.5 ms cm−1, one cycle is seen occupying 4.0 cm. On the time axis, this cycle represents a time period of 4.0 × 0.5 ms or 2.0 ms. Recall that the period, T, is the time taken to complete one cycle and the frequency, f, 160 Unit 1 Module 2 Oscillations and waves Sound intensity 1 1 Y-input To a large measure, sound intensity corresponds with the 'loudness' of a sound. However, as physical quantities, sound intensity and loudness of a sound are not the same thing. We shall be making this distinction as we go through this chapter and especially in the section dealing with hearing and the ear. 1 cycle ON OFF 5 0.5 5 0.5 10 0.1 10 0.1 Y-gain Sound intensity Time base (ms / cm) Sound intensity, I, is defined as the sound energy crossing perpendicularly across unit area per unit time. loudspeaker In equation form, if sound of energy E crosses an area A perpendicularly in a time t I= microphone to signal generator Figure 10.2 Determining the frequency of sound using a microphone connected to an oscilloscope. is the number of cycles per second. T and f are reciprocals of each other – see Chapter 8 and equations 8.3 and 8.4. From the period, therefore, we can determine the frequency of the sound. Sound frequencies form a spectrum that can be divided into three major sections. The human ear can detect sound frequencies from about 20 Hz to 20 kHz. Sound within this frequency range is called audible sound or just 'sound'. Sounds of frequencies less than 20 Hz are called infrasound (or subsound); sounds of frequencies higher than 20 kHz are called ultrasound. Table 10.1 shows typical sound frequencies that are detected by various animals.

Earthquake frequencies are typically less than 20 Hz, so humans do not hear them but only feel them. E/A ....(10.1) t Since energy per unit time is power, P, then I= P ....(10.2) A From equation 10.2, it follows that the S.I. unit for sound intensity is watts per square metre (W m−2). Worked example 10.1: Musician near to a loudspeaker Q A musician stands 4.0 m from a 1000 W loudspeaker which is at full blast (Figure 10.3). (a) What intensity of sound does his ear experience?

(b) Is he is likely to experience damage to the ear?

(Assume the sound is spread equally in all directions from the loudspeaker.) a b r S Table 10.1 Approximate range of frequencies heard by various animals Animal Approximate hearing range bat 20 Hz to 110 000 Hz cat 30 Hz to 64 000 Hz dog 40 Hz to 50 000 Hz dolphin 100 Hz to 150 000 Hz elephant 15 Hz to 12 000 Hz mouse 800 Hz to 75 000 Hz ITQ 1 What is the frequency of the sound being displayed on the oscilloscope in Figure 10.2? ITQ 2 Can mice communicate with each other using vocal waves without a cat hearing them? Explain your answer. 4m Figure 10.3 (a) Musician performing near to a 1000 W loudspeaker. (b) Sphere at a radius of r = 4.0 m from the loudspeaker, S, treated as a point source of sound. A ■ power, P = 1000 W radius of sphere through which sound passes, r = 4.0 m ■ curved area of sphere, A = 4πr2 = 4π × 4.02 = 64π m2 P 1000 W (a) Intensity, I = = A 64π m2 ≈ 5.0 W m−2 (b) Most likely! Ear damage is a likely result since above an intensity of 1.0 W m−2 the ear begins to experience pain. See the section 'Threshold intensity of discomfort (pain)' on page 161. ■ Chapter 10 Sound a b c Figure 10.4 Different wave forms: (a) a tuning fork sounding a pure note, (b) a violin and (c) noise.

For a point source of power P that radiates sound equally in all directions, the sound passes through a sphere of area A = 4πr2 at a distance r from the source. The intensity at a distance r from the source is therefore given by I= P 4πr2 Threshold intensity of hearing The threshold intensity of hearing, I0, is the smallest intensity of a 1 kHz sound that the human ear can typically detect. This intensity, I0, has a value of approximately 1.0 × 10−12 W m−2.

When the ear is subject to high intensities, discomfort is experienced, usually in the form of pain. As the intensity increases further, permanent damage to the ear may result. Threshold intensity of discomfort (pain) The threshold intensity of discomfort, Ip, is the largest intensity of a 1 kHz sound that the human ear can typically detect without the discomfort of pain. The value of this intensity is approximately 1.0 W m−2. Sound quality Two sounds may have the same frequency and intensity, yet they do not sound the same.

The difference is in their quality. Another word for sound quality is the French word timbre (pronounced 'tahmbr'). Displays of sound waveforms using a microphone and oscilloscope (Figure 10.4) show waveform shapes that are characteristic of various sounds. Pleasant sounds usually show shapes that are a regular repetition of a basic shape (Figures 10.4a and 10.4b). In Figure 10.4c, the wave shape is random and we hear this as a 'shhhhhh' kind of noise. Figures 10.4a and 10.4b show waveforms from different sources that are emitting a sound of the same frequency. The simplest waveform is a sine wave shape and is ITQ 3 Estimate the speed of sound in air at 25 °C (use equation 10.3). ITQ 4 Suggest why, although at a party where very loud music of high and low frequencies is playing, it is the bass sounds, rather than the high pitched sounds, that tend to be heard more loudly down the street from where the party is going on. produced by a tuning fork. Analysis of the waveforms of the other sound sources have shown that these waveforms result from a fundamental sine wave combined with sine wave harmonic frequencies. (Recall: Chapter 8 discusses that strings and air columns show standing waves at a fundamental frequency and also at higher harmonic frequencies called overtones.) The dominant frequency perceived by the ear is usually the fundamental frequency. Sound quality refers to the presence of harmonic frequencies that gives a sound its distinctive identity. Speed of sound The speed of sound in air varies with the temperature of the air. The equation for the speed, v, at an air temperature of θ °C is given by v ≈ (330 + 0.60 θ) m s−1 ....(10.3) At 25 °C, sound travels approximately 4 times as fast in water as in air, and 15 times as fast in iron as in air. Sound waves In Chapter 8, we discussed the following behaviours shown by waves: rectilinear propagation, reflection, refraction, diffraction, interference and polarization. Sound displays all these behaviours, except polarization. Sound does not show polarization because sound waves are longitudinal and the oscillations are not confined to single planes, as, for example, are waves on a rope (see the section on 'Transverse and longitudinal waves' in Chapter 8). Rectilinear propagation Sound of high frequency travels in straight lines through wide openings, with very little spreading occurring, i.e. with hardly any diffraction taking place. You can check for yourself, using the equation, v = fλ, that a whistle sound of frequency 15 000 Hz, travelling at 340 m s−1 through the air, will have a wavelength of about 2 cm.

Because this wavelength is much smaller than the width of a door (which is about 80 cm), the sound waves will 161 162 Unit 1 Module 2 Oscillations and waves reflected pulse bat insect SONAR (ultrasound waves) transmitted from ship incident pulse Figure 10.5 Navigation of bats by ultrasound waves. SONAR (ultrasound waves) reflected back to ship from school of fish pass straight through the door with very little diffraction. A bass drum sound of frequency 80 Hz will have a wavelength of about 4 m which is much larger than the width of the door. This sound will diffract (spread out) as it passes through the door opening. Reflection Sounds reflect very well off hard surfaces. The reflections are called echoes. Bats make use of echoes in navigating (Figure 10.5). They emit ultrasonic beams, which travel very straight with hardly any diffraction, and therefore the wave energy can travel a long distance. When the reflected beams reach the ears of the bat, the bat is able to locate the object – how far away it is and in what direction. It is believed that the bat is able to pinpoint direction by using the slight phase difference of arrival of the reflected sound at its two ears. SONAR (Sound Navigation And Ranging) pulses of ultrasound frequencies are used in determining ocean depths and locating shoals of fish and submarines. Figure 10.6 shows SONAR being used to locate a shoal of fish. Frequent use of SONAR, as in naval exercises involving detection of submarines, is believed to cause certain sea creatures (such as whales and porpoises) to become disoriented – even leading to their death. The use of SONAR in locating shoals of fish has become so successful that overfishing sometimes occurs, leading to a disturbance in the marine food chain. Some types of medical imaging also make use of reflection of ultrasound.

Figure 10.7 shows how the ultrasound pulse-echo imaging technique works using a single transducer. (A transducer emits ultrasound pulses in the 1 MHz range and detects the reflected pulses as well.) Figure 10.8 shows the reflection dot pattern resulting from moving the transducer along the abdomen or by using an array of transducers. Figure 10.6 Using SONAR to locate shoals of fish. Worked example 10.2: Using SONAR to measure ocean depth Q A sonar pulse leaves a ship and returns 4.8 s later, having been reflected off the seabed directly below. If the average speed of sound in the sea water is 1560 m s−1, estimate the depth of the ocean. A ■ depth of ocean, d = ?

■ distance travelled by SONAR pulse = 2d (since reflection occurs) ■ speed of sound, v = 1560 m s−1 ■ time taken, t = 4.8 s 2d v = t vt Therefore d = 2 1560 m s−1 × 4.8 s = 2 = 3744 m ≈ 3700 m Figure 10.7a shows an ultrasound transducer placed against the abdominal wall. The transducer emits pulses of ultrasound and detects the pulses reflecting off tissues along the path. Reflections from soft tissues, such as the stomach wall, are weak; reflections from hard tissues, such as the vertebra, are strong. Figure 10.7b shows intensity of reflected pulses detected by the transducer as a function of time. In Figure 10.7c, Chapter 10 Sound stomach wall a transducer abdominal wall vertebra pulse Q echoes echoes Worked example 10.3: Refraction of sound waves in air b Figure 10.9 shows sound from a source, S, making an angle of incidence θ1 = 60°, at a point, P, between two layers of air at 20 °C and 30 °C as shown. (a) Determine the speed of sound in the lower and upper layers of air. (b) Determine the angle of refraction of the sound in the upper layer. Strength of reflected pulse (I) normal θ2 P air at 30 ˚C Time θ1 S (II) air at 20 ˚C c Figure 10.7 How ultrasound pulse-echo imaging works. a transducer abdominal wall stomach wall vertebra Figure 10.9 Sound travelling from one layer of air to another at a higher temperature. A b Figure 10.8 Ultrasound dot image (a one-line scan) resulting from moving a single transducer along the abdomen. A picture scan is obtained by moving an array of transducers across the abdomen. the reflected pulses are plotted electronically as dots to produce an image on a monitor.

Reflections from soft tissues show up as faint dots; reflection from hard tissues shows up as bright dots. Figure 10.8 shows the dot image formed by moving the transducer along the abdomen or by placement of an array of transducers. With a larger number of transducers angle of incidence, θ1 = 60° ■ angle of refraction, θ2 = ? ■ temperature of lower layer of air, T1 = 20 °C Note that we use T rather than θ in equation 10.3 to distinguish between angle and temperature, as this example involves both. ■ temperature of upper layer of air, T2 = 30 °C (a) Speed of sound in lower layer v1 ≈ (330 + 0.60T1) m s−1 (by equation 10.3) ≈ (330 + 0.60 × 20) m s−1 ≈ 342 m s−1 Speed of sound in upper layer v2 ≈ (330 + 0.60T2) m s−1 ≈ (330 + 0.60 × 30) m s−1 ≈ 348 m s−1 (b) By the refraction equation (equation 8.16) sin θ1 v1 = sin θ2 v2 sin θ1 × v2 sin θ2 = v1 ■ = sin 60° × 348 m s−1 342 m s−1 = 0.8812 Therefore θ2 ≈ 62° ITQ 5 Around what time of day is the shouting of someone least likely to be heard far away due to refraction? Explain your answer.

163 164 Unit 1 Module 2 Oscillations and waves more closely spaced in the array, a better resolution image is obtained.

To obtain a picture scan, the array is moved across the abdomen. Worked example 10.4: Interference of sound waves Q For the situation depicted in Figure 10.10 (see page 165), the loudspeakers were separated by a distance of 1.2 m. The distance between maximum intensities occurring at O and P, respectively, was determined experimentally to be 2.2 m. Determine: (a) the distance AP (b) the distance BP (c) the wavelength of the sound (d) the distance of y (= OP), using the equation y = λD a (equation 9.20) and explain why this calculated value of y is different from the value of y = 2.2 m as shown in the experimental diagram.

A (a) In the right-angled triangle ANP: AN = 5.0 m NP = NO + y = 0.6 m + 2.2 m = 2.8 m (NO = 12 AB = 0.6 m) AP = √AN2 + NP2 m = √(5.02 + 2.82) m ≈ 5.7 m (b) In right-angled triangle BMP: BM = 5.0 m MP = 2.2 m – 0.6 m = 1.6 m BP = √AN2 + NP2 m = √(5.02 + 1.62) m ≈ 5.2 m (c) wavelength = path difference, AP – BP ≈ 5.7 m − 5.2 m = 0.5 m λD (d) using y = we get a 0.5 m × 5.0 m y= 1.2 m Using ultrasound for medical imaging is advantageous over using X-ray imaging in certain ways. X-rays can damage tissue by ionization; ultrasound is less damaging. X-rays also are very penetrating and therefore do not distinguish between various types of soft tissues as well as ultrasound does. Note: X-ray images are really shadows of tissues formed when tissues are in the path of X-rays; ultrasound images are reflections of ultrasound waves. Focused high-energy ultrasound is used as shock waves in shattering kidney stones and gallstones so that the fragments can be voided by the patient. This use of ultrasound is often employed instead of invasive surgery. Ultrasound is also used in certain massage therapies involving elastic tissues. The tissues resonate when their natural frequencies match the ultrasound frequencies used. (See the section on 'Resonance' in Chapter 7.) The echo technique in ultrasound imaging is also used to detect cracks within metal structures.

Refraction The speed of sound in warm air is greater than that in cold air (see equation 10.3). Thus, when sound waves travel from one layer of air into another layer where the temperature, and hence the speed, is different, refraction, a change in direction of travel, takes place. Now look at Worked example 10.3. Note that the angle of refraction in the warmer air (Figure 10.9, sound beam I) is greater than the angle of incidence. Total internal reflection will occur within the lower layer if θ1 exceeds the critical angle between the two layers (see sound beam II). Sound can then be heard a long distance from the source. This effect is likely to occur early in the morning when the ground or the surface of the ocean is cold and upper layers of air are much warmer. (Refer to Chapter 8 for a discussion of the critical angle and total internal reflection.) Diffraction and interference Figure 10.10 shows two loudspeakers, A and B, connected in parallel and in phase to a signal generator. A directional microphone, connected to an oscilloscope, detects a maximum intensity at O and the next maximum at P when moved along a line NP. Sound waves from A and B diffract in concentric circles from each source. A few of the circular diffracted waves are shown. Since wave speed is the same for sounds = 2.1 m The difference between the actual and the calculated values for y is due to the fact that the equation y = λD a (equation 9.20, where n = 1) is derived on the assumption that D ≫ a. This is clearly not the case here, since D = 5.0 m and a = 1.2 m. Hence, equation 9.20 is not applicable in this situation. A derivation of the small angle approximation equation, y = nλD , is found in Chapter 9 in the section on diffraction and a interference through a double slit. For the first maximum from the central maximum, n = 1, which gives y = λD a. travelling in the air from each speaker, the waves travel equal distances in the same time. Along the perpendicular bisector of AB, waves travel equal distances from the loudspeakers to O.

Since the waves leave A and B in Chapter 10 Sound D = 5.0 m A a = 1.2 m N microphone O C B M y = 2.2 m P phase and travel equal distances to O, they arrive in phase at O and constructive interference occurs. Hence a maximum sound intensity occurs at O. Constructive interference also occurs at P if the path difference AP minus BP is one wavelength (or an integral number of wavelengths), since waves would again be arriving in phase. At a point, M, mid-way between O and P, waves from A and B cancel each other. This is because the path difference, AM minus BM is half a wavelength. Destructive interference occurs at M and hardly any sound is detected there by the microphone. Response of the human ear to sound levels We use three different terms to describe the 'loudness' of a sound. They are: ■ intensity (measured in watts per square metre) ■ sound level (measured in decibels) and ■ loudness level (measured in phons). Intensity The intensity of a sound was defined earlier as the sound energy crossing perpendicularly across unit area per unit time. Sound intensity is measured in W m−2.

The human ear can comfortably detect sounds of intensities of the order of 10−12 W m−2 to 1 W m−2. Although instruments have been developed that can measure sound intensities over the normal human hearing range, for practical purposes, a ratio of intensities is used for measuring sound levels. Some form of ratio is necessary when measuring human hearing, since hearing does not involve a true zero of sound intensity. The average 'zero' intensity of human hearing is actually 1.0 × 10−12 W m−2. Other intensities are therefore expressed in terms of this threshold intensity of hearing as sound levels. Figure 10.10 Diffraction and interference of sound waves. Sound level The ratio of comfortable hearing of sound of greatest intensity to the threshold intensity of hearing is approximately 1 W m−2 to 10−12 W m−2 , or 1012 (1 trillion). Sound level, β, measured in bels (B), corresponding to a sound of intensity I is defined by the logarithmic ratio β (bel)= log10 () I I0 ....(10.4) Worked example 10.5: Range of sound levels, in bels, detected by the ear Q What are typically the largest and smallest sound levels, in bels, detected comfortably by the human ear? A ■ largest intensity, Ip = 1.0 W m−2 ■ threshold intensity, I0 = 1.0 × 10−12 W m−2 Ip Largest sound level βp (bel) = log10 I (equation 10.4) 0 ( ) ( ) 1.0 10−12 = log10 1012 = 12 B I0 Smallest sound level β0 (bel) = log10 I 0 = log10 1 = log10 100 =0B = log10 ( ) Worked example 10.5 shows that using sound level as a logarithmic ratio of sound intensities results in a small range of numbers of bels (0 to 12) corresponding to the entire human range of hearing. This contrasts with the rather huge range of 1 to 1012 if a simple ratio of intensities is used to define sound level. Users can therefore comprehend sound level measurements much more easily on the logarithmic scale. To measure sound 165 166 Unit 1 Module 2 Oscillations and waves levels with greater precision, a decibel (dB) scale is used, where 10 dB = 1 B. On this scale, sound levels detected comfortably by the human ear range from 0 dB to a 120 dB. Thus, sound level in decibels is defined as follows: Worked example 10.6: Sound levels near two jet engines Q A dB meter reads 130 dB at a certain position from a jet plane when one engine is turned on. (a) What is the sound intensity at that position? (b) What would be the sound level at that same position if two engines are turned on, each having the same intensity as the first? A ■ sound level, β, measured in decibels (dB), corresponding to a sound intensity I is defined as 10 times the logarithmic ratio of the intensity, I, of the sound to the threshold intensity of hearing, I0. β (dB)= 10 log10 () I I0 ....(10.5) Equation 10.5 is used as the practical defining equation for sound level. Decibel (dB) meters have been developed to measure sound levels over this range. Note about logs (logarithms) The log (to base 10) of a number is the power to which 10 must be raised to obtain that number. See also Appendix 3, Logarithms. Here are some examples: ■ 1000 = 103. Therefore, log10 of 1000 = 3 ■ 100 = 102. Therefore, log10 of 100 = 2 sound level produced by one engine, β1 = 130 dB ■ sound level produced by two engines, β2 = ? ■ sound intensity produced by one engine, I1 = ? ■ sound intensity produced by two engines, I2 = ? I1 (a) β (dB) = 10 log I , according to equation 10.5 0 I1 130 = 10 log 10−12 13 = log I1 – log 10−12 13 = log I1 − (−12) = log I1 + 12 1 = log I1 I1 = 10 W m−2 ( ) ( ) Getting it right! Combining sound levels For two engines we do not simply add sound levels since sound levels are logarithmic ratios and, therefore, do not add. Rather, we add intensities since the latter represents energy (per unit area per unit time) and energy can be added. Working with logs 1000 × 100 = 103 × 102 = 105 log10(1000 × 100) = log (103 × 102) = log10 of 105 = 5 (but note that 5 is 3 + 2) This example shows that log10 (103 × 102) = log10 103 + log10 102 = 3 + 2 If we let A = 103 and B = 102, we see that log10 (AB) = log10 A + log10 B It can be shown that log (A) = log A − log B B To obtain the log of most numbers you must use the log function on your calculator. Conversely, if you know the log of a number, you use the inverse log function on your calculator to find the number. For example, if log10 x = 1.2, then x = 15.8 Logarithms to different bases can be used, so it is necessary to state the base.

However, in this chapter all the logarithms are to base 10, and so the '10' will be assumed from here on. Loudness The human ear can detect sound frequencies from 20 Hz to 20 kHz and sound levels from 0 dB to 120 dB (the latter corresponding to sound intensities from 1.0 × 10−12 W m−2 (b) I2 = 2 × I1 = 2 × 10 W m−2 = 20 W m−2 ( ) I2 β2 (dB) = 10 log I 0 = 10 × (log 20 − log 10−12) = 10 × [1.3 − (−12)] = 133 dB Useful tip! This example shows that a doubling of sound intensity results in an increase of only 3 dB in sound level.

This is a useful point to remember. ITQ 6 Using your calculator, determine: (a) log 10 000 (b) log 0.01 (c) log 500 ITQ 7 The sound level produced at a point P by a singer at point D is 90 dB. What would the sound level be at P if three more singers joined in at D, singing just as loudly? Chapter 10 Sound 100 Sound level / dB 100 10 –2 80 80 10 –4 60 60 10 –6 40 40 threshold of hearing 20 10 –8 20 10 –10 0 0 0 50 100 500 1000 Frequency / Hz –2 120 120 phons Intensity / W m threshold of pain 10 –12 5000 10,000 Figure 10.11 Curves of equal loudness for the human ear. to 1.0 W m−2). However, the ear does not perceive different frequencies of the same sound level as being just as loud. This is because the ear is more sensitive to certain sound frequencies than to others. Figure 10.11 shows curves of equal loudness as perceived by the human ear, based on frequencies and sound levels. Note that in order to be able to display the full audio range on a graph that is of reasonable size, both the intensity and frequency scales used are logarithmic and not linear. Hence the ear's response to sound intensities and frequencies is best displayed on a logarithmic scale. Loudness level Loudness level in phons is numerically equal to sound level in dB of a 1 kHz sound with the same apparent loudness. Musical instruments example, the large portion of the hollow body resonates at lower frequencies predominantly; the upper, smaller portion resonates at predominantly higher frequencies. Thus, in the sound of a guitar, only certain harmonic frequencies given off by the string would be amplified by the hollow body through resonance, giving the guitar its distinctive sound. The harmonic wavelengths, λn on a string are given by λn = 2L (see Chapter 8, equation 8.14) n where n = 1, 2, 3, … represents the harmonic mode of vibration and L is the vibrating length of the string. The harmonic frequencies, fn, are given by fn = nf1 where f1 = v = v (see Chapter 8, equation 8.15) λ1 2L where f1 is the first harmonic frequency (corresponding to one loop on the string) and v is the speed of sound along the string. It can be shown that the speed of sound along a stretched string is given by F v = μT ....(10.6) where FT is the force of tension in the string and μ is its mass per unit length (called its linear density). It follows from equation 8.15 that, in order to obtain a wide range of frequencies using the same lengths of strings (e.g. on a guitar), the speed of sound on the string must be varied. This could be done, for example, by changing the force of tension in a string (see equation 10.6). Screws at the head of the guitar are turned to adjust the tension when the guitar is being 'tuned'. Equation 10.6 tells us that a large μ Stringed instruments L A review of the section on 'Standing waves' in Chapter 8 will prove quite useful at this point. Some of the concepts and equations that follow were introduced there. When strings are plucked, or bowed, stationary waves, formed by waves travelling in opposite directions, persist. Figure 10.12 shows three of these stationary wave patterns. The quality of sound heard depends on where, and how, the string is plucked or bowed, as this determines which harmonic frequencies predominate on the string. For example, beginners often produce squeaky sounds when learning how to bow on a violin string due to the predominance of the higher harmonic frequencies. The final quality of the sound heard from an instrument also depends on the shape and size of the hollow body of the instrument (called the resonator). In a guitar, for 2L 1 n=1 λ1 = n=2 λ2 = L = n=3 λ3 = 1 2λ1 2L 2 λ2 2L 3 λ3 Figure 10.12 The first three standing waves on a string. ITQ 8 What is the loudness level of a 100 Hz sound at a sound level of 60 dB? Refer to Figure 10.11. (Hint: go to 100 Hz, 60 dB and follow the curve of equal loudness till it intersects 1 kHz.) 167 168 Unit 1 Module 2 Oscillations and waves n=1 a n=2 n=3 b n=1 A n=2 n=3 A N 1 λ 4 1 3 λ 4 2 5 λ 4 3 A 1 λ 2 1 N 3 λ 2 3 λ2 λ3 λ3 N Figure 10.13 Stationary wave patterns in air columns in pipes of same length, L. Some antinodes, A (regions of maximum vibrational displacement of air), and nodes, N (regions where the air is always stationary), are shown. (a large linear density) will also result in a small speed and hence low harmonic frequencies. The different strings on a guitar have different linear densities and therefore the guitar can produce sounds of a wide range of frequencies.

Hence the resonant frequencies for a pipe open at one end are f1, 3f1, 5f1 etc. (called odd harmonics) where f1 is the first harmonic frequency (called the fundamental frequency). Wind instruments A similar analysis applied to Figure 10.13b shows that for a pipe open at both ends: The topic of stationary waves in air columns was discussed in Chapter 8 (see Figure 8.12). Figure 10.13 shows the first three stationary wave patterns in air columns within pipes of same length, L. The stationary waves occur at particular frequencies called resonant frequencies. It is these frequencies that are heard predominantly when wind instruments are played. Pipe open at one end Antinodes, A, are not formed exactly at the open ends of pipes, but rather at an extra distance, e, from the end. This extra distance, e, is called the end error. According to Figure 10.13a, the fundamental wavelength, λ1 is given by: 1 4 λ1 =L+e The second harmonic wavelength, λ2, is given by: 3 4 λ2 =L+e The third harmonic wavelength, λ3, is given by: 5 4 λ3 =L+e Since the speed v of sound of frequency f is given by v = fλ, then the first, second and third harmonic frequencies are given by f1 = v v = λ1 4(L + e) f2 = v v = = 3f1 λ2 4 (L + e) 3 f3 = v v = = 5f1 λ3 4 (L + e) 5 Pipe open at both ends f1 = v v = λ1 2(L + 2e) f2 = v v = = 2f1 λ2 2(L + 2e) f3 = v v = 2 = 3f1 λ3 3 (L + 2e) Hence the resonant frequencies for a pipe open at both ends are f1, 2f1, 3f1 and so on, the same pattern as the resonant frequencies on a tight string fixed at both ends. Since the quality of a sound depends on the harmonic frequencies present, it follows that sounds from pipes of the same length, open at one end, would sound different from those from pipes open at both ends. Some pipe-based musical instruments ■ Pipe organs make use of pipes open at one end as well as pipes open at both ends. An organ pipe resonates with sound when air is blown over a sharp edge which results in vibrations of the air within the pipe. ■ The vibrations of air in a trumpet are created by vibrating the lips at the mouthpiece. ■ In a clarinet, air is blown at the mouthpiece and this causes a reed to vibrate, resulting in vibrations of the air within the body of the instrument. ■ In a flute, air blown over a sharp edge produces the vibrations. ITQ 9 According to equation 8.15, how can the frequency of a note from a given string at a given tension be increased during playing? Chapter 10 Sound Worked example 10.7: Calculation of end error, e, of a sounding pipe of length, L Q For a pipe closed at one end, the first harmonic wavelength, λ1, is given by: 14 λ1 = L + e v (a) Show that the first harmonic frequency is given by: f1 = 4(L + e) 1 (b) Using the equation in part (a), derive a linear equation making f 1 the subject of the equation. 1 (c) Figure 10.14 shows a graph of f versus L for a pipe closed at one 1 end. From the graph, determine: (i) the speed of sound in air (ii) the end error, e. 0.0050 0.0040 0.0030 1 / Hz –1 f 0.0020 0.0010 1 Figure 10.14 Graph of f versus L for a pipe closed at one end. 1 A (a) Since 14 λ1 = L + e, then λ1 = 4(L + e) But v = fλ, i.e. f = v λ v f1 = or λ1 v f1 = 4(L + e) v , we get 4(L + e) 1 4(L + e) = v f1 1 4 4 Therefore, = × L + × e f1 v v This is a linear equation of the form y = mx + c 1 4 4 where y = , m = and c = × e f1 v v (b) From f1 = 0.20 0.30 L/m 0.40 0.50 (c) (i) For the equation, 1 4 4 = ×L+ ×e f1 v v 4 represents the slope. v 4 (0.0050 − 0.0020) Hz−1 Therefore, = v (0.40 − 0.15) m from which we get v = 330 m s−1 (to 2 s.f.) (ii) According to the graph, when 1 = 0.0020 Hz−1, L = 0.15 m f1 −1 Substituting v = 330 m s into the equation 1 4 4 = × L + × e we get f1 v v 4 4 0.0020 Hz−1 = × 0.15 m + ×e 330 m s−1 330 m s−1 which gives e ≈ 0.015 m ≈ 1.5 cm 1 Note that an alternative method for obtaining v and e would be to use two values of f and corresponding values of L from the graph 1 4(L + e) to form two equations which could be solved simultaneously. and put these into 1 = v f1 ITQ 10 Use Figure 10.15.

Are the following musical instruments open at one end or open at both ends? (a) clarinet (b) flute 0.10 ITQ 11 A moaning or groaning sound is sometimes heard distinctly in the quiet of night coming from old trees. Some people say that these are sounds from jumbies or bacoos since no animals are seen near to the tree. Close observation shows that the tree branches are broken at one end and hollow. Suggest an alternative explanation for the moans and groans. 169 170 Unit 1 Module 2 Oscillations and waves Worked example 10.8: Fundamental frequency of a flute Q The vibrating length of air in a flute is 0.640 m.

What is the fundamental frequency of the flute: (a) at 30 °C (b) at 20 °C? A ■ length of vibrating air column, L = 0.640 m ■ speed of sound in air, v = 330 + 0.60θ ■ a flute is open at both ends ■ λ1 = 2L = 2 × 0.640 m = 1.28 m (a) At 30 °C, v = (330 + 0.60 × 30) m s−1 = 348 m s−1 v f1 = λ1 348 m s−1 1.28 m ≈ 272 Hz (b) At 20 °C, v = (330 +

0.60 × 20) = 342 m s−1 v f1 = λ1 = 342 m s−1 1.28 m = 267 Hz From this example we see that frequencies produced by wind instruments, and by stringed instruments too, will vary slightly in different temperature situations. Therefore musicians often retune their instruments in a concert hall prior to a performance to compensate for the temperature within the hall. = Figure 10.15 Organ pipes, clarinet and flute.
All wind instruments achieve tuning by having different lengths of vibrating air columns. In the organ, separate pipes are used for each fundamental resonant frequency. In the clarinet, flute and the recorder, the effective length of the vibrating air column is achieved by covering or uncovering holes on the instrument so that an antinode can occur at a hole. Percussion instruments Sound from percussion instruments, such as drums and steel pans, can be accounted for in a similar manner to sounds from other musical instruments. The smaller the vibrating surfaces and the smaller the resonant cavities, the higher the resonant frequencies attained. The way in which the instruments are struck also determines the quality Figure 10.16 The Trinidad and Tobago steel pan has gone international and can now be found all over the world! of sound. This is just like the way, or point at which, a guitar string is plucked, or how sharply one blows on a flute, determines the quality of the note being played.
Figure 10.16 shows steel of drums of the same diameter but different heights. The playing surfaces of the taller pans are divided up into a few large areas. The combination of large vibrating surfaces and large resonant cavities produce deep bass sounds. The steel pan is hailed as the only musical instrument to have been developed in the 20th century. Trinidad and Tobago is often regarded as the home of the steel pan. Chapter 10 Sound Summary ✓ Sound is produced by pressure waves travelling ✓ Sound level, β measured in decibels (dB), corresponding to a sound intensity I is defined as 10 times the logarithmic ratio of the intensity, I, of a sound to the threshold intensity of hearing, I0. in materials that are solid, liquid or gas. ✓ Humans can hear sounds ranging in frequencies from approximately 20 Hz to 20 kHz. Sounds below 20 Hz are called infrasound or subsound; above 20 kHz sounds are called ultrasound. ✓ The speed of sound in air at Celsius temperature, θ, is given by v = (330 + 0.60 θ) m s−1. ....(10.3) ✓ The speed of sound in a taut string is given by v= FT μ ....(10.6) where FT is the tension in the string and μ is its linear density. ✓ Sound intensity is defined as sound energy per unit time (sound power) per unit area: P I= A ....(10.2) ✓ The threshold intensity of hearing, I0, is the smallest intensity of sound detectable by the human ear. This intensity is defined as 1.0 × 10−12 W m−2 for a 1 kHz sound. ✓ The threshold intensity of pain or discomfort, Ip, is the largest intensity of sound that the human ear can withstand without discomfort. This intensity is defined as 1.0 W m−2 for a 1 kHz sound. Severe pain in the ears usually occurs when the threshold intensity of discomfort is exceeded. Review questions Describing sound 1 (a) What is sound? (b) Give the sound frequency ranges of: (i) audible sound (ii) infrasound (or subsound) (iii) ultrasound. (c) A device, silent to the human ear when turned on, has been advertised that stops dog from barking. (i) Suggest how the device might work. (Hint: see Table 10.1.) (ii) What might be an undesirable side effect of this device in operation? β (dB)= 10 log10 (I I0 ....)(10.5) Equation 10.5, in conjunction with the values of the threshold intensities, tells us that the average human ear can comfortably detect sound levels from 0 dB to 120 dB. ✓ The ear is not equally sensitive to sounds of various frequencies. Very low frequency and very high frequency sounds have to be at large sound levels in order to be heard equally loud. ✓ Loudness level is measured in phons. The number of phons is numerically equal to the sound level in decibels of a 1 kHz sound on a curve of equal loudness.
✓ The sound quality from musical instruments results from harmonics present in vibrating strings, air columns and surfaces, and also depends on resonant frequencies of hollow parts of the instrument. ✓ Ultrasound is used in medical scans and in depth-sounding in oceans. Both applications make use of the echo effect of sound waves. 2 Sound waves from a tuning fork are displayed on an oscilloscope connected to a microphone (Figure 10.17). A B Figure 10.17 Sound waves displayed on a 1 cm × 1 cm grid oscilloscope screen.
171 172 Unit 1 Module 2 Oscillations and waves If the time base (x-axis) is set at 1.0 ms per cm, and the Y-gain is set at 25 °C, determine: (a) the speed of the sounds in the air (b) the frequency of sound A (c) the wavelength of sound A. (Note: the wavelength is not the distance between two 'crests' here. Why not?) (d) the intensity of sound B, if the sound intensity of A is 2.0 × 10−8 W m−2. (Hint: see equation 8.10.) 3 Two tenors are singing a note of the same frequency and loudness. Explain why their voices sound different. Sound waves 4 (a) Describe the conditions under which waves show: (i) partial reflection and partial transmission (ii) a change of phase upon reflection. (b) In making ultrasound scans of the abdomen, why is a gel placed on the abdominal wall and the ultrasound transducer placed in the gel?
(c) It is desired to be able to resolve defects of a scan for objects at least 2 mm in size. Assuming that for this to happen, the wavelength of the ultrasound must be 2 mm or less, calculate the smallest ultrasound frequency that must be used. (Assume the speed of sound in body fluids and tissues is approximately 1500 m s−1.) (d) For best results, should the front end of the transducer be of the order of 2 mm, much greater, or much less? Explain your answer? 5 (a) What is meant by 'coherent' waves? (Hint: you may need to review Chapter 8, Waves.) (b) Two loudspeakers are connected in parallel from a single source. Explain whether the sound waves coming from the loudspeakers would be necessarily: (i) in phase (ii) coherent. (c) Figure 10.18 shows two loudspeakers, P and Q, connected in parallel, in phase, to a signal generator. P and Q are a distance a = 1.2 m apart. A unidirectional microphone, moved along a line XY, a distance D = 8.0 m from P and Q, detects maximum sound intensities at A, C and E, and minimum sound intensities at F and G. X D = 8.0 m A P F a = 1.2 m C G Q E Y Figure 10.18 Two loudspeakers connected in parallel from a single source. (i) Derive the equation y ≈ λD/a (equation 9.20) for a first-order maximum intensity to occur at A using waves of wavelength λ. γ Chapter 10 Sound A B 9 (a) Define: (i) threshold intensity of hearing (ii) threshold intensity of discomfort (pain).
(b) The sound level at a certain distance from an exploding firecracker is 100 dB.
What would the sound level be at the same point if three identical firecrackers exploded simultaneously from the same spot as the first firecracker? (Remember: sound levels do not add; sound intensities do.) 10 Figure 10.20 shows two curves of equal loudness for a particular person whose ears are being tested. - 26 To find the depth of the ocean, an ultrasound transducer is connected to an oscilloscope where the time base (x-axis) is set at 5.0 ms per cm. A pulse of ultrasound is sent from the surface of the ocean and the reflected pulse is detected on the oscilloscope by the transducer (Figure 10.19). Response of the human ear to sound levels Sound intensity / W m (ii) What assumptions are made in deriving the above equation? (iii) If the frequency of the sound signal used is 500 Hz, calculate the distances CE and CG, using y ≈ λD/a and assuming that the setup is arranged in a room at 25 °C. (Remember: the speed of sound in air is given by v = (330 + 0.60 θ) m s−1.) 10 –4 10 –6 10 –8 10 –10 10 A –12 C 1 cm 10 Figure 10.19 Ultrasound pulses detected on an oscilloscope. (a) Identify, giving a reason, the transmitted and reflected pulses shown on the oscilloscope screen (Figure 10.19). (b) (i) Estimate the time between the two pulses. (ii) Determine the depth of the ocean (assume that the speed of sound in sea water is 1540 m s−1). (c) Why are ultrasound frequencies, and not audio sound frequencies, used in determining ocean depths?
7 (a) Define sound intensity. (b) A 100 mW pair of headphones is covered over two ears. If all of the power delivered to the headphones is converted into sound, and sound from each headphone at full blast reaches the eardrum (diameter = 10 mm), estimate the sound intensity at each eardrum. (c) Is the user in a situation like that in part (b) above likely to experience ear damage?
Explain your answer. 8 (a) Determine the sound intensity at a point 6.0 m from a 1000 W loudspeaker operating at full blast. (Assume the sound spreads equally around the loudspeaker.) (b) How much further from the loudspeaker operating at 1000 W must a listener move to experience a more comfortable sound intensity of 10−9 W m−2? 100 1000 Frequency / Hz 10,000 100,000 20,000 Figure 10.20 Two curves of equal loudness for a particular person's hearing. (a) What intensity of sound is defined as the threshold intensity of hearing and at what frequency? (b) Give two reasons why the response of the human ear to sounds is considered logarithmic. (c) Estimate the sound level (in dB) at the point labelled A. (d) State the loudness level (in phons): (i) at A (ii) along the lower curve. (e) Both curves end at a frequency C. (i) Estimate the frequency C. (ii) What does this tell you about the hearing of the person whose ears were tested. (f) What is the name given to the lower curve? (g) What sound frequencies at 20 dB are heard equally loud? Musical instruments 11 (a) Draw a diagram to show the third harmonic mode in a pipe open at one end. Label node, antinode, wavelength and end error. (b) A pipe of length L, open at one end, resonates in the third harmonic mode at a sound frequency of 807 Hz. 173 174 Unit 1 Module 2 Oscillations and waves (i) Calculate the wavelength, λ3, of the sound in the third harmonic mode. (Assume speed of sound in air = 340 m s−1) (ii) Write a formula for λ3, the third harmonic wavelength, in terms of the speed of the sound, v, the length, L, and the end error, e. Hence, determine the end error, e, if L = 50.0 cm. (iii) Determine the fundamental resonant frequency of the pipe. 12 Repeat question 11 for a pipe of same length but open at both ends. 13 A stretched guitar string resonates in the fundamental (first harmonic) mode at a frequency of 262 Hz. (a) What would be the second and third harmonic frequencies of the string? (b) Upon analysis of the sound from the string using a microphone and oscilloscope, only the first and second harmonic frequencies showed up prominently. Give two possible explanations for this. 14 What two design features in a steel pan contribute to sounds of low frequency being heard.
Explain your answer. Practical exercises and challenges 1 Using a dB meter (or a sound intensity meter), measure sound levels (or sound intensities) at various locations, e.g. in classrooms, at a school party or on the sidewalk. 2 Measure the range of frequencies detected by the human ear by gender or by age.
■ Make sure that the loudspeakers or headphones used do indeed produce a wide range of sound frequencies. ■ Take care to use only comfortable levels of sound so as not to cause damage to the ears. ■ Identify variables that should be controlled in each kind of investigation. 3 Use the internet to find out about curves of equal loudness. There are sites which allow you to test your ears with actual sounds at various frequencies and sound levels and automatically plot curves of equal loudness for your ears. 4 Using a set of calibrated tuning forks of frequency f and a tube of variable length L, open at one end (e.g. by adjusting either a piston or water level at one end), determine the lengths, L, of hollow tube when the first harmonic is produced, and corresponding fundamental frequencies, f. Rearrange equation 10.7 to make l/f the subject in terms of v, L and e. Plot a graph of l/f versus L. From the graph, determine v and e. 5 Using an arrangement such as shown in Figure 10.10, determine the wavelength, and hence the frequency of sound, coming from two loudspeakers connected in parallel. 6 Calculate the lengths of pipe open at both ends needed to make a musical scale (doh, ray, me, fah, soh, lah, tee, doh – C, D, E, F, G, A, B, C) for a wind instrument. You will need first to find out what frequency corresponds to each individual note. Then use lengths of pipe, open at both ends, along with a set of tuning forks, to make a percussion instrument. Do the experimental lengths of the pipes agree with the lengths calculated above for a pipe open at both ends?
Comment on the results you obtain. 7 Make a one-stringed 'guitar' with various shapes of hollow bodies. Investigate the prominent harmonic sound frequencies obtained, using a microphone connected to an oscilloscope. Two possible shapes of boxes are shown in Figures 10.21a and 10.21b. A possible arrangement is shown for varying the tension in the string (Figure 10.21c). a b c string resonating box weights Figure 10.21 Investigating harmonics from a single-stringed instrument using different shapes of resonating boxes. Chapter 10 Sound Answers to ITQs Answers to Review questions 1 Period (time for 1 cycle), T = 4 cm × 0.5 ms per cm = 2.0 ms 5 (c) (iii) CE = 4.6 m; CG 2.3 m Frequency, f = 1 1 = = 500 Hz T 0.0020 s 3 (a) Dogs can hear sounds of frequencies beyond 65 000 Hz, cats won't hear them since cats typically hear only up to 64 000 Hz mice can hear up to 75 000 Hz. 3 v = (330 + 0.60 × 25) m s−1 = 330 + 15 = 345 m s−1 4 Bass sounds are of low frequency, i.e. large wavelength, and hence diffract (spread out) much more than high frequency sounds. 5 In the afternoon when the air near the ground is hotter than the layer of air above it. Since the speed of sound in the upper layer is less than that in the lower layer sound rays from the source are refracted towards the normal, i.e. upwards. 6 (a) 4 (b) −2 (c) 2.70 (to 3 significant figures) 7 96 dB two singers increase the sound level by 3 dB to 93 dB; four singers increase the sound level of the two singers by another 3 dB, resulting in 96 dB) 8 40 phons (a sound frequency of 100 Hz at sound level of 60 dB intersects the 40 phon curve of equal loudness). 9 By reducing the vibrating length, L. This can be done by pressing the string against one of the frets on the neck of the guitar. 10 (a) Clarinet – open at one end only (closed at the mouthpiece end).
(b) Flute – open at both ends. 11 The hollow branch is like a pipe open at one end. When wind blows across the open end, vibrations are set up which can produce standing waves in the branch which are heard as groans or moans. 7 (a) 640 W m−2 (b) Yes. (If all the assumptions in the question are correct.) 9 (b) 105 dB 11 (b) (i) 0.421 m (ii) 2.6 cm (iii) 161 Hz 13 (a) 524 Hz; 786 Hz 175 176 Module 3 Thermal and mechanical properties of matter Chapter 11 Thermometry Learning objectives ■ Recall the thermodynamic interpretation of temperature. ■ Determine temperatures in the Kelvin, Celsius and empirical centigrade scales. ■ Discuss the variation of physical properties of thermometric substances with temperature. ■ Describe design features of selected thermometers. ■ Evaluate the use of selected thermometers. Introduction A thermometer measures temperature, that is, the degree of hotness or coldness of a substance. This chapter focuses on the design and use of thermometers. An example of a modern thermometer is shown in Figure 11.1. Thermodynamic interpretation of temperature The kinetic theory of matter (see Chapter 14) offers a thermodynamic interpretation of temperature.
The particles that make up a gas are in a state of continuous random motion. For an ideal gas, the Kelvin temperature, T, is related to the mean square speed, $\langle c^2 \rangle$ of the particles of such a gas according to the equation: 1 2 2 m 3 = 2 kT..................(14.15) A temperature scale, according to this equation, depends on the mean square speed of the particles of an ideal gas, where m is the mass of each particle of the gas and k is the Boltzmann's constant. Figure 11.1 A laser-guided semiconductor thermometer. This thermometer can read the person's temperature at a distance, which is an advantage when screening for a disease such as Ebola. [Advanced theory, however, shows that the temperature defined by the kinetic theory of an ideal gas, as just described, is identical to the temperature as defined by the absolute thermodynamic scale (described below) which is completely independent of the physical properties of any particular substance.] The gas in a constant volume gas thermometer, being at a relatively low pressure, closely approximates an ideal gas. Such a thermometer can therefore not only Chapter 11 Thermometry measure temperature as defined by the kinetic theory but can also be used to check on the accuracy of other thermometers.
Temperature scales Any form of measurement assumes a scale. A practical temperature scale consists of two fixed points (temperatures that are relatively easy to reproduce) and a temperature interval between the fixed points. The interval is divided up into smaller equal intervals (called degrees in the Celsius and Fahrenheit scales). The Celsius scale The lower fixed point on the Celsius scale is defined as the temperature of pure melting ice at standard pressure and is assigned a value of 0 °C. The upper fixed point, by definition, is the temperature of pure boiling water at standard pressure and is assigned a value of 100 °C. There are therefore 100 Celsius degrees (C°) between the fixed points. Hence, this scale is termed a centigrade scale because there are 100 divisions between the two fixed points. Note: the Celsius scale applies only to a gas thermometer that is calibrated using the ice point and the steam point. The temperature unit in any other centigrade scale calibrated at these same points (e.g. on a resistance thermometer) is not degrees Celsius but degrees centigrade. In common usage this distinction has become blurred, and we refer to all such temperatures simply as degrees Celsius. In the S.I. 1 C° (Celsius degree) is defined as exactly equal to 1 K and so only a gas thermometer gives a perfectly accurate Celsius scale. Centigrade scales that use alternative thermometric properties will differ slightly from the Celsius scale, as discussed in the next section. The Fahrenheit scale The lower fixed point on the thermodynamic scale is the theoretical temperature at which old molecular motion would cease.
According to quantum theory, there is a small amount of molecular motion at 0 K.) This temperature is assigned a value of zero kelvin (0 K) and is referred to as the absolute zero of temperature. At this temperature, not only the theoretical temperature, but also the theoretical pressure, of an ideal gas is zero, since each depends on molecular motion. Temperatures on the absolute scale have the unit K, not °K, whereas temperatures on the Celsius scale have the unit °C. Kelvin (thermodynamic) temperature intervals since, by definition, a temperature interval of 1 C° = 1 K. The kelvin is the S.I. base unit of temperature. It can be shown that 0 K = −273.15 °C. The relationship between absolute (thermodynamic) temperature T in kelvin and Celsius temperature θ in °C therefore is T = θ + 273.15 We can apply a simple empirical ratio to Figure 11.2. If θ and F, respectively, represent the same temperature on Celsius and Fahrenheit scales, then ....(11.1) ....(11.2) The thermodynamic scale (unlike the Celsius scale) uses as its second fixed point the triple point of water (rather than the ice point). The triple point of water is the temperature at which ice, water and water vapour coexist in thermal equilibrium. This temperature can be reproduced reliably as it does not depend on pressure, but the ice point does slightly. By definition, the triple point of water is 273.16 K, as this defines the thermodynamic (Kelvin) scale (and hence the degree Celsius). The triple point temperature is 0.01 C° higher than ice point. Celsius scale 100 °C The same fixed points used in the Celsius scale also characterize the Fahrenheit scale, still being used in some countries.
However, the lower fixed point is given the value of 32 °F and the upper fixed point a value of 212 °F. There are 180 Fahrenheit degrees between the fixed points. F − 32 F − 32 θ−0 = = 100 − 0 212 − 32 180 The absolute (thermodynamic) scale (Kelvin scale) steam point 212 °F F °F θ °C 0 °C Fahrenheit scale ice point 32 °F Figure 11.2 Celsius and Fahrenheit temperature scales. ITQ 1 According to the ratios shown in equation 11.1, what would be a simple formula for Celsius temperature, θ, in terms of Fahrenheit temperature, F? ITQ 2 What is the ice point temperature in kelvin? 177 178 Unit 1 Module 3 Thermal and mechanical properties of matter Variation of physical properties with temperature Physical properties of certain substances, or arrangements of substances, vary with temperature. The property used is called the thermometric property. For instance, with increasing temperature: ■ the volume of liquids generally (though not always) increases ■ the pressure of a gas at constant volume increases Worked example 11.1 Q Measured from the bulb of a liquid-in-glass thermometer, the lengths of the thermometric liquid column are as follows: l0 = 5.4 cm at ice point, l100 = 18.6 cm at steam point, and lθ = 8.6 cm at room temperature. What is the temperature of the room, according to this thermometer? A We use equation 11.3 as follows: l−l θ−0 = θ 0 100 − 0 l100 − l0 ■ the electrical resistance of a conductor increases 8.6 cm − 5.4 cm θ = 100 °C 18.6 cm − 5.4 cm ■ the electrical resistance of a semiconductor decreases ■ the e.m.f. (voltage) of some thermocouples increases, but that of others decreases. Variations such as these form the basis for constructing practical thermometers. All thermometers properly calibrated at ice point and steam point would agree at these two temperatures. However, because the physical properties utilised may not vary uniformly with temperature, different types of thermometers might give slightly different readings corresponding to the same temperature. The only thermometers that come closest to giving temperatures predicted by the kinetic theory. Physical properties and the Celsius scale Figure 11.3 shows a Celsius temperature scale beside a scale based on a physical property, p. By a simple empirical ratio, if pθ is the measured value of a physical property (e.g. pressure) at θ °C, then p −p θ−0 = p θ − p0 100 p θ −p θ = p0 × 100 ....(11.3) 100 °C 0 °C Physical properties and the Kelvin scale Figure 11.4 shows a Kelvin scale next to a scale based on a physical property, e.g. pressure p.
The pressures p0K, ptr and pT are pressures at absolute zero of temperature, the triple point temperature of water and the Kelvin temperature T, respectively. Applying equal ratios to the Kelvin scale and the scale based on the physical property, we obtain an empirical formula for Kelvin temperature, T, in terms of triple point temperature, Ttr, as follows: pT − p0K T−0 = p tr − p0K 273.16 − 0 tr p0K pT or, T = p × 273.16 ....(11.4) tr The p0K term disappears since the pressure of a gas at 0 K is zero. Kelvin gas pressure, p 373.15 K T pT Ttr = 273.16 K ptr p100 p0 pθ 0K p0K Figure 11.4 Kelvin temperature scale compared with a scale based on a physical property, p. 0.01 °C 0 °C −273.15 °C p0 Figure 11.3 A Celsius temperature scale compared with a scale based on a physical property, p. ITQ 3 What is the triple point temperature of water: (a) in °C? (b) in kelvin? Chapter 11 Thermometry Worked example 11.2 Q A metre rule The pressure of a fixed mass of gas in Kelvin temperature T is 1.4135 Pa. The pressure of that gas at the triple point of water is 1.0266 Pa. Determine: (a) the Kelvin temperature of the gas (b) the corresponding temperature on the Celsius scale. Since we are given the pressure at the triple point, and not at the ice point, we cannot use equation 11.1 to find either Kelvin or Celsius temperature. We must use equation 11.4 instead. (a) Substituting the values given into equation 11.4, we get, 1.4135 × 273.16 T= 1.0266 = 376.11 K (b) Since T = θ + 273.15 (equation 11.2), then θ = 376.11 − 273.15 = 102.96 °C mercury reservoir capillary bore in liquid stem mercury X dry air flask flexible tube electric hot plate Figure 11.5 A simple form of constant volume gas thermometer. Practical thermometers: design and usage We will keep in mind the following – accuracy, sensitivity, readability, response time, robustness, storage of readings, safety, convenience of use and range – as we describe the construction features of five types of thermometers. The types chosen are gas, liquid-in-glass, resistance, solid-state and thermocouple. Constant volume gas thermometer The gas in a constant volume gas thermometer, being at a relatively low pressure, closely approximates an ideal gas. As such, the gas pressure is directly proportional to the Kelvin temperature. As pointed out earlier, the gas thermometer has been used for checking the accuracy of other types of thermometer. Figure 11.5 shows a simple form of constant volume gas thermometer. Dry air is sealed inside a flask. The flask is made of a material with a low coefficient of expansion (i.e. that expands very little with increasing temperature). The volume of air is kept constant by raising or lowering an open reservoir containing mercury such that the end of the air column in the capillary tube remains at a mark, X. The pressure of the gas in the bulb is equal to atmospheric pressure + hρg, where h is the height of the mercury column above X and ρ is the density of mercury. (A digital pressure transducer, connected to a specially constructed constant volume chamber, can easily measure gas pressures at different temperatures. The transducer is connected to a data-logger which reads out the pressure sensed by the transducer.) Since change in gas pressure is proportional to change in height of the column above X, we can use equation 11.1 to determine the Celsius temperature, θ, of the gas (and hence of the liquid bath) as follows: h − h0 × 100 °C θ= h100 − h0 The Kelvin temperature can be found from equation 11.2: T = θ + 273.15 Note the following design features of this thermometer and precautions taken to ensure accuracy: ■ in calibrating the thermometer, a bath of pure, melting ice is used for ice point determination of h; a bath of pure boiling water is used for steam point determination (temperature of 100 °C must be corrected if the atmospheric pressure is not standard) ■ dry air is used because we do not wish water vapour to condense in the bulb and thus change the mass of gas that is in the bulb ■ the liquid bath goes up the neck of the bulb so that nearly all of the air, excepting the tiny amount in the connecting capillary tube, receives the same temperature, i.e. that of the bath ■ just before a measurement of h is taken, the bath is stirred gently to bring all of the surrounding liquid at the same temperature; the open reservoir is also adjusted to ensure that the end of the air column in the capillary tube is at X.
179 180 Unit 1 Module 3 Thermal and mechanical properties of matter Liquid-in-glass thermometer °C 100 90 bore 80 70 stem 60 50 scale 40 30 20 mercury thread 10 0 bulb Figure 11.6 A liquid-in-glass laboratory thermometer. Figure 11.6 shows a typical liquid-in-glass thermometer, the type you are likely to use in a laboratory. The thermometric property used in this kind of thermometer is volume since the volume of a liquid changes with changing temperature.
The thermometric liquid is usually mercury or ethyl alcohol (ethanol).
Table 11.1 shows a comparison between mercury and alcohol as a thermometric liquid. The bulb of a liquid-in-glass thermometer is made large to contain a large volume of liquid so that a small change in temperature can cause the liquid to expand appreciably. In using such a thermometer, one must ensure that the temperature being measured surrounds the entire bulb. The capillary bore is very narrow so that even a small change in volume of the liquid shows up as an easily readable change in length of the thread in the bore. To increase visibility of the liquid in a capillary bore, especially in clinical thermometers, the surface of the stem above the capillary bore is shaped to serve as a magnifying lens. a b °C 43 41 39 37 normal body temperature 37 °C 35 construction bulb Figure 11.7 (a) A liquid-in-glass clinical thermometer. (b) A digital clinical thermometer. The glass of the bulb is made thin so that heat exchanges across the bulb to and from the liquid can be quick.
However, one must allow time for the maximum reading to be reached. It will take a while, not only for heat to be exchanged across the walls of the bulb, but also for the mass of mercury or alcohol to absorb the heat and expand. Glass is also fragile and therefore these thermometers have to be handled with great care. The length of these thermometers can make them cumbersome to carry around or use.
In liquid-in-glass clinical thermometers (Figure 11.7a) a constriction on the bore prevents the liquid from returning to the bulb after the temperature of a patient is taken. Hence the temperature reading is stored on the stem and can be read later. Nowadays, liquid-in-glass clinical thermometers are increasingly being replaced by solid-state thermometers Table 11.1 Mercury and alcohol thermometers compared Features Properties compared Comments Sensitivity Alcohol expands about five An alcohol thermometer is more sensitive times as much as mercury than a mercury thermometer of the same for the same temperature dimensions. change. Operating range Alcohol freezes at −115 °C An alcohol thermometer is better suited than and boils at 85 °C. a mercury thermometer for use involving low Mercury freezes at −40 °C temperatures (such as in the Arctic regions). and boils at 360 °C. Visibility Safety A mercury thermometer is more advantageous for use in higher temperature measurements up to about 350 °C. Alcohol is colourless. Alcohol, though colourless, can be made Mercury is silvery and can easily visible by adding a dye to it. be seen relatively easily in to further enhance the visibility of the the capillary tube. mercury, a yellow background is sometimes used. Mercury vapour is toxic. Alcohol thermometers are therefore Alcohol vapour is not toxic. sometimes preferred over mercury thermometers if the possibility of breakage is being considered. Figure 11.8 A liquid-crystal thermometer in use. ITQ 4 The sensitivity of a thermometer refers to its ability to display tiny changes in temperature. Chapter 11 Thermometry (Figure 11.7b). as these can be made very sensitive, are digital and hence easy to read, and can store readings in their memories. (See page 182 for a discussion on solid-state thermometers.) resistances. Can you identify at least three design features that contribute towards a very sensitive liquid-in-glass thermometer?
Resistance thermometers give quite accurate measurements when used at proper temperature ranges. Figure 11.9 shows a resistance thermometer circuits. The material used for the resistance coil is usually platinum.
This is mainly because platinum does not corrode. Platinum has a high melting point and therefore this type of thermometer can operate over a very wide range of temperatures.
Platinum also has a moderate temperature coefficient of resistivity (see Table 11.2), resulting in appreciable changes in resistance for small changes in temperature. A new technology of liquid-crystal thermometers has also been developed (Figure 11.8). The liquid crystal in each circle changes colour at a specific temperature. Strips of these liquid crystal thermometers can be placed on a patient's forehead and temperature easily monitored. Thermometers like these are particularly suited to use in the home. Resistance thermometer The thermometric property used in a resistance thermometer is electrical resistance, since the resistance of a conductor varies with temperature. Over limited temperature ranges, the resistance, R, of a conductor varies linearly with temperature according to the equation: Rθ = R0 + αR0(T − θ0) or Rθ = R0 + αR0(θ − θ0).....................(11.5) The coil size can be made quite small and therefore temperatures can be measured in somewhat small or hard to reach areas. Because the thermometer is electrical, the readings can be taken quite conveniently from a location different from the coil by means of a wire or wireless connection. This feature is useful when monitoring temperatures during manufacture involving high temperatures.
Readings can be easily read on digital or analogue meters, or graphed and stored on computer. Digital ammeters and voltmeters are preferred since such instruments can be very sensitive. Digital voltmeters have very high resistance and can therefore measure voltages accurately. Alternatively, the resistance values can be measured by connecting the coil in one arm of a Wheatstone bridge (Figure 11.9b). See Chapter 19, page 305, for the method of calculating resistance. where R0 is the resistance at a reference temperature, T0, corresponding usually to the temperature 0 °C, and α is the temperature coefficient of resistivity of the material of the conductor. Temperature coefficient of resistivity is defined as the change in resistance per unit resistance at 20 °C per degree change in temperature.
In equation form, R0 − R0 α = R0 (θ − θ0)..............(11.6) 0 Table 11.2 Melting points, MP, and temperature coefficients of resistivity, α, of selected materials The resistance of a material with a large temperature coefficient of resistivity will therefore show large changes in resistance per degree change in temperature. Material Table 11.2 shows the melting points of various materials, as well as their temperature coefficients of resistivity at 20 °C (293 K). Equation 11.1 can be used to determine the temperature corresponding to various a Melting point, MP / K Temperature coefficient of resistivity, α / K−1 silicon (pure) 1687 −7.5 × 10−2 gold 1336 3.4 × 10−3 platinum 2041 3.9 × 10−3 aluminium 933 4.3 × 10−3 silver 1234 6.1 × 10−3 copper 1356 6.8 × 10−3 metre-bridge wire digital voltmeter V I IT standard resistance A digital ammeter RT = V I Rs RT resistance coil at balance Rs R = T IS IT RT resistance coil Figure 11.9 A resistance thermometer.
181 182 Unit 1 Module 3 Thermal and mechanical properties of matter Worked example 11.3 Q A platinum resistance thermometer has a resistance of 3.00 Ω at 20.0 °C. What would be the temperature corresponding to a resistance of 2.67 Ω? A ■ resistance, R0 = R20.0 = 3.00 Ω ■ resistance, Rθ = 2.67 Ω ■ temperature coefficient of resistivity of platinum, α = 0.0039 (C°)−1 Using Rθ = R0 + αR0(θ − θ0), we get 2.67 Ω = 3.00 Ω + 0.0039 (C°)−1 × 3.00 Ω (θ − 20.0) (C°) 2.67 = 1 + 0.0039(θ − 20.0) 3.00 0.890 − 1 = 0.078 − 0.0039 θ = θ = −8.2 °C Solid state (thermistor) thermometer Some types of thermistor thermometers make use of semiconductors (see Chapter 22). When the temperature of a semiconductor rises, the number of charge carriers per unit volume increases, resulting in lower resistance and hence increased current for a given voltage. Linear behaviour, assumed by equations 11.5 and 11.6, holds for semiconductors over only a relatively restricted range of temperatures. Unlike the case in conductors, the temperature coefficient of resistivity for silicon is negative (see Table 11.2). This means that the resistance decreases with increasing temperature. 0 1 cm 2 Figure 11.10 shows a thermistor and Figure 11.10 A thermistor can 11.11 shows a simple be very small. semiconductor (thermistor) thermometer circuit.
The semiconductor often used is silicon. digital milliammeter thermistor mA Figure 11.11 Thermistor thermometer circuit diagram. Thermistor thermometers can be made to be very small: the sensitive element can be smaller than the head of a matchstick and the connecting leads a few cm long. These types of thermometers can ITQ 5 Suggest three advantages that the electrical nature of a thermistor thermometer gives to such a thermometer? therefore measure temperatures in tiny areas. Their small size also contributes to their very fast response time. Depending on the semiconductor material used, small temperature changes can greatly affect the number of charge carriers per unit volume and hence change the resistance (or current) substantially. (This effect can be greatly enhanced by introducing small quantities of impurities called dopants – see Chapter 22.) Further, when fed to an amplifier circuit, the changes are easily detectable. Hence, these thermometers can be made extremely sensitive. Semiconductors also respond very quickly to temperature changes. Semiconductors cannot, however, be used in direct contact with high temperatures.
The increasing number of charge carriers produced at such temperatures result in increasing heating effects of currents which can themselves result in generating more charge carriers. The heating effect due to current from an avalanche of these carriers can destroy the semiconductor of a thermistor thermometer. A typical operating temperature range is −90 °C to 120 °C. Thermistor thermometers, because of their extreme sensitivity, can measure temperatures of objects from a distance by the thermal radiation emitted from the objects. In one model, a laser beam guides the user to aim at the tiny spot where temperature is being measured. In Figure 11.1, a laser-guided solid state thermometer is used to read the temperature of a person's forehead. The solid state element in this type of thermometer detects the intensity of thermal radiation emitted from a spot on the person's forehead, and the associated electronic circuitry gives a digital readout of the corresponding temperature at the spot. Since the thermometer does not touch the person, it could advantageously be used in situations where body temperature is being measured but the health worker must not come into contact with the person. An example is if the individual is being screened for some dangerous virus, e.g. Ebola. Temperatures of furnaces, kitchen utensils, food being cooked and at various points in buildings can also be read this way. Like resistance thermometers, thermistor thermometers are electrical and therefore digital ammeters provide the most sensitive results. Thermocouple thermometer Two different metals that are joined at one end constitute a thermocouple (Figure 11.12a). An e.m.f. (i.e. a voltage), whose size depends on the temperature Figure 11.12 (a) A thermocouple thermometer; (b) a thermopile.
V / mV V output is electrical, the voltage can be amplified, making it additionally easier to monitor tiny changes in temperature. 0.05 0.04 One disadvantage of thermocouple thermometers, however, is that the e.m.f.s do not vary linearly with temperature, as shown in Figure 11.13. Generally, e.m.f. increases with temperature non-uniformly. Above a certain temperature, called the inversion temperature, the e.m.f. then decreases with temperature. 0.03 0.02 θi 0.01 Figure 11.13 Inversion temperature, θi, of a thermocouple thermometer. The tip of the temperature probe can be made very tiny and therefore temperatures can be measured at pin point areas. Further, since the tip is small, and is made of conductors, the response time for temperature changes is extremely short. difference between the junction and the free ends, appears across the free ends. In practice, an arrangement involving two or more junctions is used (Figure 11.12b). One junction, the reference junction, is kept at a reference temperature, such as the temperature of melting ice or of the ambient temperature, and the other forms the temperature probe. An arrangement consisting of several thermocouples connected together is called a thermopile. Several thermocouples can be hooked up in series to form a thermopile (see Figure 11.12b). Thermopile output voltages vary in proportion to the number of thermocouples connected together. Temperatures of furnaces can be measured by directing the thermal radiation they emit on to the 'hot junction' of a thermopile. Thermopiles are used in this way in measuring temperatures from flames, furnaces – and even inside the ear.
Since modern (solid-state, digital) meters can register very tiny voltages, the small e.m.f.s typically produced by a thermocouple are easily detected. Further, since The choice of metals determines the maximum p.d. each thermocouple produces.
Metals with very high melting points also enable a thermocouple thermometer to 100 200 300 400 500 θ / °C Gas p 0 Liquid-in glass l 0 / °C Resistance R θ / °C 0 Resistance R θ / °C 0 ITQ 6 (a) What is the inversion temperature of the thermocouple shown in Figure 11.13? (b) Why is the inversion temperature behaviour a disadvantage in the thermocouple thermometer? Thermocouple E θ / °C θ / °C Figure 11.14 Variation of thermometric property with temperature, θ, for five types of thermometer. ITQ 7 Using the graphs of thermometric property variation with temperature (Figure 11.14), which of the five thermometers is the most accurate? Give the reasons for your choice. 183 184 Unit 1 Module 3 Thermal and mechanical properties of matter operate over a fairly large range of temperatures when the thermopile itself is in direct contact with the hot object. An iron-copper thermocouple has a range of 0 °C to 1700 °C, that of a platinum-rhodium thermometer has a range of 0 °C to 700 °C. Comparison of thermometric properties The graphs in Figure 11.14 give a qualitative comparison in the variation of thermometric property with temperature, θ, for the five types of thermometer discussed in this chapter. As you can see, their responses vary from linearity in different ways.
Summary ✓ Temperature can be defined qualitatively as a measure of hotness or coldness. According to the kinetic theory, temperature is directly proportional to the mean square speed of the particles that make up an ideal gas. ✓ A temperature scale consists of two fixed points (temperatures that are easy to reproduce) and sub-divisions between these two points. If there are 100 equal sub-divisions between the fixed points, the scale is called a centigrade scale. The Celsius scale, whose fixed points are ice point and steam point, is an example of a centigrade scale. ✓ Ice point is the temperature of pure, melting ice at standard pressure. ✓ The Kelvin scale is based on the thermodynamic theory of the heat engine as developed by Carnot; it can be shown to be identical to the ideal gas scale.
The kinetic theory of an ideal gas shows us that temperature, T, in kelvin, is proportional to the mean square speed of the gas molecules. ✓ The Kelvin scale defines its lower fixed point as 0 K, the lowest theoretically possible temperature. The second fixed point on this scale is the triple point of water, assigned a value of 273.16 K. On the Kelvin scale, ice point is 273.15 K. ✓ Temperature scales based on physical properties of matter are not quite linear with respect to the Kelvin scale but agree with values of the theoretically derived thermodynamic scale only at the fixed points. ✓ The relationship between Kelvin temperature, T, in kelvin, and the Celsius temperature, θ, is: T = θ + 273.15. In the S.I. a temperature interval of 1 C° is defined as being equal to 1 K. ✓ The equation for calculating Celsius temperature, θ, based on its pressure, p, is given by: p −p θ = p θ − p0 × 100..................(11.3) 100 0 ✓ The equation for calculating temperature on the thermodynamic scale is: T = pT ptr × 273.16..................(11.4) ✓ Assuming a linear relationship, the temperature, T, in a resistance thermometer is obtained from the equation: RT = R0 + αR0 (T − T0) where α is the temperature coefficient of resistivity of the material at 20 °C and is defined by the equation: Rθ − R0 α= R0(θ − θ0) ✓ A range of physical properties of matter vary with temperature. Some of these properties and the thermometers based on them are summarised here: ■ in a constant volume gas thermometer, the pressure of the gas varies with temperature ■ in a liquid-in-glass thermometer, the volume of the liquid varies with temperature ■ the resistance of a conductor varies with temperature in a resistance thermometer ■ the resistance of a semiconductor varies with temperature in a thermistor thermometer ■ the e.m.f. varies with temperature in a thermocouple thermometer. ✓ Factors such as the following may need to be considered in the construction and selection of thermometers: range, sensitivity, accuracy, convenience of usage and safety. Chapter 11 Thermometry Review questions Thermodynamic interpretation of temperature 1 How does the kinetic theory interpret temperature? 2 Why, according to the kinetic theory, is there a finite temperature below which no lower temperatures are believed to exist? Temperature scales 3 (a) What are meant by fixed points on a thermometric scale? (b) Define the fixed points used on the Celsius scale. (c) Why is the Celsius scale referred to as a 'centigrade' scale?
4 (a) Distinguish between ice point and the triple point of water. (b) Why is the triple point preferred to the ice point when calibrating a thermometer on the thermodynamic scale? 5 (a) Human body temperature is 37.6 °C. Express this temperature in kelvin. (b) The boiling point of liquid nitrogen is 177.4 K. How much is this in °C? Variation of physical properties with temperature 6 Define temperature.
7 (a) Discuss two advantages of a platinum resistance thermometer over a thermistor (solid-state) thermometer. (b) Discuss two disadvantages of a platinum resistance thermometer when compared with a thermistor (solid-state) thermometer. 8 (a) Draw a diagram to show a thermocouple thermometer. (b) What physical property of this thermometer varies with temperature? (c) Is the value of this property the same at the same temperature for different choices of materials? Practical thermometers 9 Liquid-in-glass laboratory thermometers usually have a range of −10 to 110 °C. Give three reasons why water is not a suitable liquid for a laboratory liquid-in-glass thermometer. 10 A platinum resistance thermometer has a resistance of 30.0 Ω at 20 °C. When placed in a clear, colourless, liquid boiling at standard pressure, the resistance is 35.6 Ω.
(a) Determine the boiling point of the liquid. (b) Is the liquid water? Explain your answer. 11 (a) Draw a circuit diagram of a resistance thermometer that uses a voltage- and a currentmeasuring device. (b) Why must the voltage-measuring device have a high resistance in order to obtain accurate values of the resistance? (c) The voltage and current for a particular temperature reading using a resistance thermometer are 6.84 V and 17.3 mA, respectively. Calculate the resistance of the coil at that temperature to the correct number of significant figures. 12 It is desired to monitor the temperature of a critically ill patient from the nurse's desk. Which from among resistance, liquid-in-glass, thermistor and thermocouple thermometers would be most suitable?
Give three reasons for your choice.
13 The ideal gas scale is based on the theoretical behaviour of gases at constant volume. (a) What is the thermometric property of this scale? (b) In what way, if any, is the scale affected by the nature of the gas or the volume of the gas? 14 The output voltage (or e.m.f.), V, of a certain thermocouple thermometer varies with Celsius temperature, θ, according to the equation: V = 45 × 10−6 × θ − 0.40 × 10−6 × θ 2 V (a) Complete the table below and plot a graph of V versus θ.
(This is a good opportunity to use Excel® or some similar program to check on your calculated values and on the manual plot of your graph.) V / V (× 10−6) θ / °C 10 20 30 40 50 60 70 80 90 100 (b) What temperatures on the graph correspond to a voltage reading of 900 × 10−6 V? Confirm these values algebraically (i.e. by using the quadratic formula – see the MATH HELP section in Appendix 3). (c) From the graph, determine the inversion temperature (the temperature corresponding to the maximum voltage) of this thermometer. 15 The table below shows how the output voltage (or e.m.f.), V, of a certain thermocouple thermometer varies with Celsius temperature, θ. V / V (×10−6) 840 1360 1560 1440 1000 θ / °C 20 40 60 80 100 185 186 Unit 1 Module 3 Thermal and mechanical properties of matter (a) Plot a calibration graph of V versus θ. (b) Using the paired data in a graphing calculator or a computer, obtain a regression second degree polynomial equation that best fits the data. (See the MATH HELP section in Appendix 3.) (c) Describe one precaution that can help in obtaining accurate results when using a thermocouple thermometer. Practical activities and challenges 1 Resistance thermometer Construct a resistance thermometer and calibrate it at ice and steam points. Using a centigrade scale, determine room temperature on this thermometer (see equation 11.1).
How does this measurement compare with that on a liquid-in-glass laboratory thermometer in terms of percentage difference? 2 Thermocouple thermometer Construct a thermocouple thermometer (a thermocouple in circuit with a sensitive voltmeter, e.g. a digital millivoltmeter). Using a liquid-in-glass laboratory thermometer as your 'standard', calibrate your thermometer by plotting a voltage versus temperature graph. (a) Does the voltage vary linearly with temperature read from the laboratory thermometer? (b) Using the paired data obtained, obtain a best fit regression equation that describes the relationship. Some calculators and computers have a feature that enables you to do this; see the MATH HELP section in Appendix 3. (c) Another activity you can consider is an investigation to compare the voltages produced by different metal combinations.
3 Thermistor thermometer Construct a thermistor thermometer (e.g. a thermistor in series with a suitable d.c. voltage source and a digital milliammeter). Compare the sensitivity of this thermometer with that of a liquid-in-glass laboratory thermometer. 4 Weather watch Log on to a website that gives weather data for a city or region nearest you.
Select a given month of the year and obtain a graph of the average daily temperatures (maximum and minimum) for that month from meteorological data that has been published. Using a stated type of thermometer (e.g. liquid-in-glass, bimetallic home, solid-state home) take daily measurements for that month and plot graphs of daily maximum and minimum temperatures. Are your graphs in keeping with the average trend shown on the website? If so, or if not, comment on your results. 5 What is the limit?
Do internet searches for 'thermometers'. You may wish to include ancient as well as modern thermometers. You may also find out about thermometers that can measure thermonuclear temperatures (millions of degrees) or near absolute zero temperatures. Answers to ITQs 1 5 θ = (F − 32) 9 2 273.15 K (since ice point temperature θ = 0 °C, and T = θ + 273.15) 3 (a) 0.01 °C (since triple point is 0.01 C° higher than ice point) (b) 273.16 K (since T = θ + 273.15) 4 Large bulb, fine capillary bore and a thermometric liquid with a large coefficient of expansion. 5 Since the readings are electrical: ■ they can be easily amplified and hence very tiny temperature changes can be read ■ temperatures can be read from a remote location – only a wire or wireless connection is needed ■ readings can be read on analogue or digital meters, graphed or stored electronically. 6 (a) 250 °C (b) There will be two temperatures for the same p.d., and this would create an ambiguity. 7 All are accurate at the fixed points. However, the gas thermometer has the best overall accuracy since it is the thermometer that comes closest to producing a linear relationship between the thermometric property and the Celsius/Kelvin temperature scales. Answers to Review questions 5 (a) 310.75 K (b) −95.75 °C 11 (c) 395 Ω 15 (b) V = 50 θ − 0.4 θ2 187 Chapter 12 Thermal properties of matter Learning objectives ■ Explain heat as a transfer of internal energy. ■ Express internal energy of a system in terms of molecular potential and kinetic energies of the system. ■ Relate temperature rise to an increase in internal energy. ■ Differentiate between 'heat capacity' and 'specific heat capacity'. ■ Describe experimental methods of determining specific heat capacity. ■ Differentiate between 'latent heat' and 'specific latent heat'. ■ Describe experimental methods of determining specific latent heats. ■ Explain 'melting', 'boiling' and 'evaporation' in terms of energy and temperature. ■ Determine freezing, melting and boiling points from temperature against time graphs. ■ Use the equations EH = mcΔθ and EH = mL to solve numerical problems. Introduction Matter is made up of particles that are in continuous motion. The particles also exert forces on each other. These two factors contribute towards important thermal properties of matter. Matter has a capacity to absorb heat, with a resulting rise in temperature. In some circumstances, matter can also absorb heat without a change in temperature. This latter occurs during a change of state, such as melting or boiling, and the heat is said to be 'latent'. In addition, matter exhibits properties associated with heat transfer by conduction, convection and radiation. The focus of this chapter is on properties of matter relating to heat capacity, latent heat and change of state.
y z x Figure 12.2 Rotation (about one of three mutually perpendicular axes), vibration and translation movements of diatomic molecules. Heat and internal energy According to the kinetic theory (see Chapter 14), the particles that make up matter are in continuous motion. In monatomic gases the free motion of particles (molecules) is along a straight line, i.e. is translational. In diatomic (and polyatomic) gases, the motion may be translational, rotational and vibrational (Figure 12.2). In solids, since particles are not free to move around, the motion is only vibrational. Internal energy and thermal energy Figure 12.1 What are two thermal properties of water that makes it so effective at fighting fires? Since the particles of a gas are in motion, they all contribute to the total kinetic energy of a gas. There are, however, forces between the particles that make up matter. In gases at very low pressures, the forces between molecules are practically negligible; the forces between vibrating particles that make up polyatomic molecules, however, are not. In solids and liquids, too, the forces of attraction and repulsion between neighbouring molecules are appreciable. These forces result in potential energies 188 Unit 1 Module 3 Thermal and mechanical properties of matter among the system of molecules, in much the same way as potential energy is stored in a spring when it is stretched or compressed. Worked example 12.1: Mother and baby in a car Q A mother of mass 50 kg sits waiting in a car in the midday sunshine with the car windows closed and air-conditioning not functioning. (Perhaps she closes the windows because the area is unsafe.) The inside of the car heats up. Neglecting biological temperature regulation and assuming a specific heat capacity of the body as 3500 J kg−1 (C°)−1: (a) what quantity of heat, gained by the mother in the car, could cause her temperature to rise by 0.20 C°? (b) what could the baby's temperature rise be if that same quantity of heat was transferred to her 5.0 kg baby instead? A ■ The sum total of the potential energies and kinetic energies within a system of molecules is called the internal energy of the substance. The term thermal energy is sometimes used to refer to kinetic energy associated with the motion of molecules. Temperature and internal energy Temperature, on the basis of the kinetic theory, is directly proportional to the average kinetic energy per molecule of a substance. A temperature rise means an increase in average kinetic energy per molecule. Thus, the internal energy, which is the sum of kinetic and potential energies, increases with increasing temperature.
Heat as internal energy transfer Heat is a net, spontaneous transfer of energy from a hotter to a colder object until both objects attain the same temperature (Figure 12.3). The average kinetic energy per molecule of the hotter object decreases while the average kinetic energy per molecule of the colder object increases until their average kinetic energies per molecule are the same.
When the average kinetic energies per molecule become the same, i.e. their temperatures are the same, we say that the objects are in thermal equilibrium with each other. Getting it right! Heat is not a form of energy as such, so it is not like chemical potential energy or nuclear energy. Heat is a transfer of energy, much like mechanical work is a transfer of energy using forces. colder object hotter object quantity of heat, Q = ? mass of mother, m1 = 50 kg ■ mass of baby, m2 = 5.0 kg ■ specific heat capacity of body, c = 3500 J kg−1 (C°)−1 ■ temperature rise of mother, Δθ1 = 0.20 C° ■ temperature rise of baby, Δθ2 = ? Using the general equation c = Q , we get mΔθ (a) For the mother Q = m1cΔθ1 = 50 kg × 3500 J kg−1 (C°)−1 × 0.20 C° = 35 000 J (b) For the baby, assuming the same amount of heat transfer: Q = m2cΔθ2 Q Δθ2 = m2c 35 000 J = 5.0 kg × 3500 J kg−1 (C°)−1 = 2.0 C° Because the baby's mass is one-tenth of the mother's mass, the baby's temperature rise could be 10 times that of the mother – enough to cause serious harm or even death to the baby. ■ ITQ 1 A 1 kg block of copper and a 2 kg block of copper are at the same temperature. (a) Which block has more internal energy? (b) Will there be a transfer of heat if the two blocks are brought into thermal contact?
Explain your answers. heat energy ITQ 2 What would be the SI units for: (a) heat capacity? (b) specific heat capacity? Figure 12.3 Heat is internal energy that is transferred when there is a temperature difference between two substances. ITQ 3 Based on Table 12.1, what do you notice about the specific heat capacities per mol of the metals? (1 mol is 6.02 × 1023 particles of atoms or molecules of a substance. We look at the mole in Chapter 14.) Chapter 12 Thermal properties of matter Specific heat capacity Heat capacity and specific heat capacity In many cases, heat transfer to or from a substance results in a temperature change in the substance.
Some objects require much heat to result in a given temperature rise; others require little heat. Experience tells us, for example, that a large mass of water on a stove requires a large quantity of heat to produce a significant temperature change. Table 12.1 Specific heat capacities, c, of various substances (Note: °C refers to an actual temperature; C° refers to a change in temperature) Substance Specific heat capacity at constant pressure and room temperature (approx.
25 °C) J kg−1 K−1 J mol−1 K−1 J g−1 (C°)−1 gold 129 25.4 0.13 lead 130 26.4 0.13 copper 390 24.5 0.39 iron 450 25.1 0.45 aluminium 900 24.2 0.90 mercury 1170 28.0 1.17 gasolene 2200 228 2.20 ethanol 2440 112 2.44 water 4200 75.3 4.20 air 1010 29.2 1.01 oxygen 918 29.4 0.918 nitrogen 1040 29.1 1.04 hydrogen 1430 28.8 1.43 carbon dioxide 835 36.9 0.835 thermometer We define the heat capacity, C, of an object as the quantity of heat, Q, required to change the temperature of the object by 1 K (or by 1 C°). Thus, for a temperature change ΔT Q C= ....(12.1) ΔT The heat capacity of 1 kg of a substance is called the specific heat capacity of the substance. Thus, considering m kg of a substance, Q c= ....(12.2) mΔT Worked example 12.1 contains calculations that use specific heat capacity. Determination of specific heat capacity by method of mixtures Specific heat capacity of a solid by the method of mixtures The method of mixtures can be illustrated by this experiment. Figure 12.4 shows a hot metal solid just having been placed in a liquid at a lower temperature. The liquid is contained in an aluminium can (called a calorimeter).
A stirrer is used to agitate the mixture gently just before a temperature reading is taken on a thermometer. Assuming no energy enters or leaves the system, the net change in internal energy of the solid, liquid, stirrer, calorimeter and thermometer is zero. (In calculations, however, we shall be ignoring the small thermal contribution of the thermometer.) The solid loses heat; the rest of the system gains heat. The system attains a final temperature (called the equilibrium temperature). During this process, the total change in internal energy is zero. Hence, energy change in solid + energy change in rest of the system = 0 polystyrene lid i.e. ΣΔQ = 0 aluminium container aluminium stirrer metal cylinder with thread attached Figure 12.4 Specific heat capacity by method of mixtures. ITQ 4 According to which law is the net change in internal energy of the system zero? ....(12.3) Worked example 12.2: Method of mixtures Q What is the equilibrium temperature when 4.0 kg of copper, originally at 100 °C, is mixed with 8.0 kg of water, originally at 20 °C? (Refer to Table 12.1 for values of specific heat capacities.) A equilibrium temperature = θ2 heat change in copper + heat change in water = 0 4.0 kg × 390 J kg−1 (C°)−1 × (θ2 – 100) C° + 8.0 kg × 4200 J kg−1 (C°)−1 × (θ2 – 20) C° = 0 1560 × θ2 − 156 000 + 33 600 × θ2 − 672 000 = 0 35160 × θ2 = 828 000 θ2 ≈ 24 °C (to 2 s.f.) ■ 189 190 Unit 1 Module 3 Thermal and mechanical properties of matter This can be put another way: ■ Specific heat capacity, assumed precise to 10 J kg−1 K−1 heat lost by solid = heat gained by the rest of the system However, students often experience difficulties in using this latter approach. Equation 12.3 is simpler to apply, as long as two simple conventions are followed. Getting it right! To apply equation 12.3, correctly, we must follow two simple conventions: ■ heat lost is negative; heat gained is positive ■ change in temperature = final temperature – initial temperature Applying the temperature change convention will automatically give the correct sign for the resulting heat loss/gain. This is shown in Worked example 12.2. Specific heat capacity of a liquid by the method of mixtures We revisit the scenario described at the beginning of this section, involving a hot solid placed in a liquid of lower temperature. We use the following, hypothetical, experimental data to illustrate the determination of the specific heat capacity of the liquid and to make an uncertainty estimate in our result. We treat uncertainty in 1 a measurement as ± 2 of the precision of the instrument (see Appendix 3 'Percentage errors and uncertainties' and Chapter 1). We consider a copper solid, an aluminium stirrer and an aluminium calorimeter. Calculation of specific heat capacity ■ Digital balance, precise to 1 g mass of solid, m1 = 136 g ± 0.5 g (±0.7%) mass of calorimeter + stirrer, m2 = 200 g ± 0.5 g (±0.3%) ●● mass of calorimeter + stirrer + liquid, m3 = 314 g ± 0.5 g (±0.2%) ●● mass of liquid, m4 = m3 – m2 = 114 g ± 1 g (±0.9%) ●● ●● ■ Laboratory thermometer, precise to 1 C° initial temperature of solid, θ1 = 100 °C ± 0.5 C° (±0.5%) ●● initial temperature of liquid, θ2 = 25 °C ± 0.5 C° (±2.0%) ●● initial temperature of calorimeter + stirrer, θ3 = 25 °C ± 0.5 C° (±2.0%) ●● final equilibrium temperature, θ = 35 °C ± 0.5 C° (±1.5%) ●● copper solid, c1 = 390 ± 5 J kg−1 K−1 = 0.390 ± 0.005 J g−1 (C°)−1 (±1.3%) ●● aluminium calorimeter + stirrer, c2 = 900 ± 5 J kg−1 K−1 = 0.900 ± 0.005 J g−1 (C°)−1 (±0.6%) ●● liquid X, c4 = ? ± ? % ●● ■ Temperature differences – note that the uncertainty in a difference, or sum, is the sum of the individual uncertainties θ1 − θ = 100 − 35 = 65 °C ± 1 C° (±1.6%) θ − θ3 = 35 − 25 = 10 °C ± 1 C° (±10%) ●● θ − θ2 = 35 − 25 = 10 °C ± 1 C° (±10%) ●● ●●
Do not confuse uncertainty in a measurement with percentage uncertainty. For example, in the last line of the data above, the uncertainty in the temperature difference is ±1 C°. However, the percentage uncertainty is ±10%. Applying equation 12.1, we get, heat change of copper solid + heat change of liquid X + heat change of calorimeter = 0

m1c1(θ1 − θ) + m4c4(θ − θ2) + m2c2(θ − θ3) = 0 Therefore, m c (θ − θ) − m2c2(θ − θ3) c4 = 1 1 1 m4(θ − θ2) ....(12.4) Substituting the values given above into equation 12.4, we get c4 ≈ 1.4 J g−1 (C°)−1 (to 2 s.f.) ≈ 1400 J kg−1 K−1 (to 2 s.f.) Since mass is given in grams, it might sometimes, as in this case, be easier to do calculations using grams rather than kilograms and then do a final conversion to S.I. units. Estimation of overall uncertainty due to instrument precision (random error) Table 12.2, based on equation 12.4, shows how an overall estimate of uncertainty in the numerical results is obtained (refer to data above for percentage uncertainties).

There is an overall random error uncertainty of 30% in the value obtained for c4: 19.4% in the numerator of equation 12.3 and 10.9% in the denominator. Table 12.2 shows that the major percentage uncertainty comes from the temperature differences, θ − θ2 and θ − θ3, each contributing approximately 10%. Chapter 12 Thermal properties of matter Table 12.2 Estimating overall uncertainty in specific heat capacity In the case being discussed, systematic error could result from heat exchange with the environment. This kind of error could be minimised by: Calculated numerical value Percentage uncertainty / ±% Uncertainty m1c1(θ1 − θ) 3448 J 0.4 + 1.3 + 1.6 = 3.6 ±124 J ■ using a well-insulated calorimeter and a shiny outer m2c2(θ − θ3) 1800 J 0.3 + 0.6 + 10 = 10.9 ±196 J calorimeter (shiny objects are poor radiators of heat and good reflectors of heat) m1c1(θ1 − θ) 1648 J − m2c2(θ − θ3) 124 + 196 1648 × 100 ■ ensuring a quick transfer of the hot solid to the liquid ±320 J ■ using an initial temperature below that of the room = 19.4 m4(θ − θ2) 1140 J 0.9 + 10 = 10.9 c4 1400 J kg−1 K−1 19.4 + 10.9 = 30 ±424 J ±420 J kg−1 K−1 Note that the percentage uncertainty in a product or quotient involving two numbers is the sum of the individual percentage uncertainties (see 'Percentage errors and uncertainties', Appendix 3, as well as Chapter 1). To reduce this percentage uncertainty: ■ the temperature differences must be made larger, or ■ a thermometer of greater precision (e.g. 0.1 C°) should be used. A lot of solid of large mass could be used as this would result in a larger final temperature, θ. This, along with lower initial temperatures, θ2 and θ3, would also make the temperature differences larger.

A large percentage uncertainty also arose because of the different term in the numerator of equation 12.3. The dominant percentage uncertainty results from θ − θ3. Ways of reducing this percentage uncertainty have just been discussed above. Uncertainty due to systematic errors A systematic error is an error arising from the parts of the measuring system. For example, an instrument may be wrongly calibrated. A systematic error can be regarded as an error in the zero of the measuring system. Unlike a random error a systematic error – whatever its magnitude – is always in one direction, and a final temperature the same number of degrees above that of the room; in this 'Rumford correction', the apparatus gains heat from the environment in the first part of the experiment and loses approximately the same amount of heat to the environment in the second part, resulting in a net heat exchange of zero within the system. Systematic errors could also arise from a user positioning the eye the same way, incorrectly, when reading the thermometer. Determination of specific heat capacity by electrical methods In electrical methods, heat supplied electrically brings about the change in temperature of a substance. Specific heat capacity of a liquid The arrangement shown in Figure 12.5 can be used for determining the specific heat capacity of a liquid by an electrical method.

An insulated electrical heater coil supplies energy to the liquid and calorimeter. For a steady current I and voltage V the energy, EH, supplied to the liquid and calorimeter in a time t is given by EH = IVt...............(12.5) (See also Chapter 19, Worked example 19.2.) thermometer A polystyrene lid variable low-voltage D.C. power supply polystyrene container V electric heater polystyrene stirrer liquid Figure 12.5 Arrangement for determining the specific heat capacity of a liquid by an electrical method. 191 192 Unit 1 Module 3 Thermal and mechanical properties of matter If the energy, EH, results in a temperature rise, Δθ, of the liquid and calorimeter, then heater leads thermometer in slot in block IVt = mcΔθ + mc1Δθ...............(12.6) where m is the mass of the liquid and c its specific heat capacity, and m1 and c1 are the mass of calorimeter and stirrer and specific heat capacity of calorimeter and stirrer, respectively (assuming the stirrer is made of the same material as the calorimeter).

Rearranging equation 12.6 we get IVt = mc1Δθ c= mΔθ V A metal block surrounded by insulation material ....(12.7) In practice, if polystyrene is used as the insulator, both m1 and c1 are small and their contributions can be neglected. As in the method mixtures just described, a well-insulated calorimeter, a shiny outer calorimeter and a 'Rumford correction' (see page 191) would also assist in reducing errors due to heat exchanges between the calorimeter and the room environment. heater in slot in block Figure 12.6 An ammeter/voltmeter method of determining the specific heat capacity of a metal solid.

ON OFF 0 1 2 3 5 6 4 (N) joulemeter thermometer in slot in block OUT Specific heat capacity of a conducting solid by an electrical method Figure 12.6 shows an arrangement that can be used to determine the specific heat capacity of a solid. Two holes are drilled in the solid: one for the thermometer and the other to accommodate the heating element. The placement of the ammeter and voltmeter are shown.

In Figure 12.7, a joulemeter replaces the ammeter and voltmeter and measures the energy supplied directly. In both cases, the highest temperature reached by the block is recorded after the current is switched off.

This is to allow for the heat generated by the coil to reach the thermometer. Petroleum jelly surrounding the thermometer bulb helps improve the heat flow. A graph of temperature against time would be helpful here in the determination of the highest temperature reached. For a steady current flowing during a time t, since there is no calorimeter, ignoring the thermal capacity of the thermometer, equation 12.6 reduces to IVt = mcΔθ...............(12.8) Rearranging equation 12.8, we get c= IVt mΔθ ....(12.9) Estimation of overall percentage uncertainty Since only products and a division are involved in equation 12.9, the overall percentage uncertainty is the sum of the percentage uncertainties in each measurement (see Appendix 3 'Percentage errors and uncertainties'). heater in slot in block OFF low-voltage power supply Figure 12.7 A joulemeter used instead of an ammeter and a voltmeter. Reducing in percentage uncertainties in this case could be achieved by: ■ using an ammeter with high precision ■ using a voltmeter with a high precision ■ using a large time ■ using a large Δθ ■ using a large m (or using a small m with a high precision balance). Using a large time (which would also help in obtaining a large Δθ) would mean a prolonged experiment. This aggravates the problem of heat exchange between the block and the room environment, as more time means more energy exchange with the environment. ITQ 5 With reference to equation 12.7, what are three ways of reducing uncertainty (random error) due to instrument precision? Reducing systematic error due to heat exchange with the environment: ■ the metal block can be enclosed by a good insulator Temperature / °C Chapter 12 Thermal properties of matter D water + steam 100 (e.g. polyurethane), minimizing heat exchanges by conduction ■ the insulator could also be surrounded by a shiny outer calorimeter, minimizing heat exchanges by radiation ■ in addition, a 'Rumford correction' technique (see page 191) can be employed. The value of c could also be obtained graphically. (See Appendix 3 for tips on using and interpreting graphs.) Worked example 12.3: Obtaining the value of c graphically Q How can the value of c be obtained graphically using the electrical method described above? A Rearranging equation 12.8, we get IVt IVt Δθ = or θ − θ0 = mc mc IVt Therefore, θ = + θ0 ....(12.10) mc where θ is the temperature at time t and θ0 is the initial temperature. A linear graph should be obtained if I and V are kept constant during the experiment and heat exchanges with the surroundings are negligible (see Figure 12.8). From this graph we get IV = slope mc IV and therefore, c = m × slope θ slope = θ0 IV mc t Figure 12.8 Graph of temperature θ versus time t for heating of a metal block. ITQ 6 The electrical method works well with blocks of conductors but not blocks of insulators. Why is this? E steam C water A ice B ice + water 0 Time Figure 12.9 Graph of temperature, θ, versus time, t, for heating of crushed ice. Specific latent heat We can see the effect of latent heat by looking at Figure 12.9. This shows a graph of temperature versus time for heating some crushed ice, initially at a temperature below 0 °C. We will discuss the five regions shown. Region A The temperature of the crushed ice rises as heat is being supplied. This is because the average vibrational energy of the molecules of ice increases. The total energy (internal energy) of the ice increases. Region B The temperature of the crushed ice remains constant. Hence there is no increase in average vibrational energy of the ice molecules. The energy supplied increases the internal energy of the ice, on account of the work being done in breaking intermolecular bonds of attraction and rearranging the ice molecules. The average potential energy per molecule increases. During this period, solid ice is changed into water – a process called melting. Since the heat supplied does not show as a change in temperature, the heat is referred to as latent heat (or 'hidden' heat).

By conservation of energy, the heat required to change a given mass from solid to liquid (melting) at a given temperature is equal to the heat that must be removed from the same mass of liquid to convert it to solid (freezing) at the same temperature. The term 'fusion' refers to the change from solid to liquid, and the temperature at which it occurs is called the melting point. (The same quantity of heat must be removed from a liquid to change it into a solid at the liquid's freezing point.) 193 194 Unit 1 Module 3 Thermal and mechanical properties of matter Table 12.3 Specific latent heats of fusion, LF, of various substances Specific latent heat of fusion, LF Substance J kg Ice 334 000 Lead 25 000 25 ammonia 33 000 ethanol 104 000 water 330 000 −1 Substance Q Specific latent heat of vaporization, LV J kg−1 J −1 ammonia 140 000 140 33 ethanol 850 000 850 104 lead 870 000 870 330 water 2260 000 2260 −1 In equation form, we get LF = Table 12.4 Specific latent heats of vaporization, LV, of various substances ....(12.11) where m represents the mass undergoing melting or freezing, Q represents the heat required for the solid/ liquid change and LF represents the specific latent heat of fusion of the solid. Region C At the start of region C all the ice has been changed to water. As heat energy is supplied, the water temperature keeps rising due to an increase in average vibrational energy of the molecules of water. However, at the border between regions C and D, the temperature stops rising. Region D The temperature is constant throughout this region. This indicates that the average kinetic energy per molecule remains constant. The energy being supplied results in an increase in the internal energy of the water, on account of the work done to break intermolecular bonds and the work done against the external atmosphere. The water molecules can therefore move about freely. During this process, called boiling, water is converted to water vapour (steam). As in region B, latent heat is involved since there is no change in temperature. The specific latent heat of vaporization of a liquid is the quantity of heat required to change unit mass of liquid to vapour (gas) at the liquid's boiling point at standard pressure (1 atm). Since more work has to be done in this latter case to overcome intermolecular forces, the specific latent heat of vaporization of a liquid is far greater than the specific latent heat of fusion. Region E The increasing temperature shows that the average kinetic energy per molecule is increasing. The total (internal) energy of the vapour keeps increasing as heat is supplied.

Worked example 12.4: Steam added to water Q When 20 g of dried steam at 100 °C is added to 400 g of water at 30 °C in an insulated container of negligible heat capacity, what is the final equilibrium temperature of the water? A Let the final equilibrium temperature = θ Assuming no net heat exchange with the environment, we get heat change in steam at 100 °C to water at 100 °C + heat change of water at 100 °C to final temperature + heat change of water originally in insulated container = 0 Therefore, −20 g × 2260 J g−1 + 20 g × 4.2 J g−1 (C°)−1 × (θ − 100) C° + 400 g × 4.2 J g−1 (C°)−1 × (θ − 30) C° = 0 or, −45200 + 84 θ − 8400 + 1680 θ − 50400 = 0 from which we get θ ≈ 59 °C (to 2 s.f.) In equation form, we get LV = Q m ....(12.12) where Q represents the quantity of heat required to convert a mass m from liquid to vapour and LV represents the specific latent heat of vaporization. In converting a solid to a liquid, the average distance between the molecules increases slightly. However, in converting a liquid to vapour, the average separation distance between molecules is increased much more. Getting it right! Note that the quantity of latent heat transferred is not negative because the steam loses heat (see conventions stated in 'Getting it right', page 190). For the specific heat terms, we simply use final temperature minus initial temperature in our calculation of temperature change as this takes care of heat loss and heat gain. Chapter 12 Thermal properties of matter Determination of specific latent heat by method of mixtures Specific latent heat of fusion of ice by the method of mixtures Figure 12.10 shows an arrangement that can be used to determine the specific latent heat of fusion of ice. Dried, melting ice is added quickly to water in a polystyrene container. The melting ice has to be dried (e.g. using an absorbent tissue) and transferred quickly so that ice only, and not water and ice, are added to the water in the calorimeter. Ice has a large specific latent heat of fusion as compared with the specific heat capacity of water, and the introduction of water with the ice could therefore cause a large error in determining the specific latent heat of fusion. ■ initial temperature of ice, θ3 = 0 °C ± 0.5 C° ■ final equilibrium temperature, θ = 18 °C ± 0.5 C° ■ temperature difference (water), θ1 − θ = 16 °C ± 1 C° (±6.3%) ■ temperature difference (ice), θ − θ3 = 18 °C ± 1 C° (±5.6%) ■ specific heat capacity of polystyrene, c1 = 1.3 J g−1 K−1 ± 0.05 J g−1 K−1 (±3.8%) ■ specific heat capacity of water, c2 = 4.2 J g−1 K−1 ± 0.05 J g−1 K−1 (±1.2%) ■ specific latent heat of fusion of ice, LF = ? J g−1 Assuming no net heat exchange with the environment, again ignoring the thermal contribution of the thermometer, and assuming a final equilibrium temperature θ we get heat change of ice, during melting polystyrene + heat change of ice water from melted ice thermometer + heat change in cup, stirrer and water lid m2c2(θ − θ3) = m1c1(θ1 − θ) + m2c2(θ1 − θ) = 0 polystyrene container on polystyrene stirrer Figure 12.10 Determining the specific latent heat of fusion of ice by the method of mixtures. (Note the '+' sign used in the first term since the ice is gaining heat.) Therefore, m c (θ − θ) + m2c2(θ1 − θ) − m3c2(θ − θ3) LF = 1 1 1 m3 ....(12.13) Substituting the values given above into equation 12.13, we get LF ≈ 360 J g−1 (to 2 s.f.) (or 360 000 J kg−1) The following sample data and calculation show how the specific latent heat of fusion of ice is obtained experimentally. A discussion follows concerning estimation of overall uncertainty in the specific latent heat of fusion of ice ■ mass of polystyrene cup + polystyrene stirrer. This result compares favourably with the accepted value of the specific latent heat of fusion of ice, which is 330 J g−1 (Table 12.3). m1 = 6 g ± 0.5 g (±8.3%) ■ mass of polystyrene cup, stirrer and water, m4 = 337 g ± 0.5 g ■ mass of water, m2 = m4 − m1 = 331 g ± 1 g (±0.3%) ■ mass of polystyrene cup, stirrer, water and ice, m5 = 388 g ± 0.5 g ■ mass of ice, m3 = m5 − m4 = 51 g ± 1 g (±2.0%) ■ initial temperature of cup + water + stirrer, θ1 = 34 °C ± 0.5 C° Calculated numerical value Percentage uncertainty / ±% Uncertainty (a) m1c1(θ1 − θ) m3 436 J g−1 (0.3 + 1.2 + 6.3 + 2.0) = 9.8 ±42.73 J g−1 (c) m3c2(θ − θ3) m3 75.6 J g−1 (2.0 + 1.2 + 5.6 + 2.0) = 10.8 ±8.16 J g−1 (±51.39 J g−1 LF = (a) − (b) − (c) 363 J g−1 ±51.39 J g−1 ±51 J g−1 (b) m2c2(θ1 − θ) m3 26 J g−1 (8.3 + 3.8 + 6.3 + 2.0) = 20.4 ±0.50 J g−1 (b) m2c2(θ1 − θ) m3 = 1.00 J g−1 ±5.00 J g−1 14 ± 363 195 196 Unit 1 Module 3 Thermal and mechanical properties of matter Percentage uncertainty due to instrument precision (random error)

Although there is a percentage uncertainty of ±8.8% in the mass, m1, of the polystyrene, and ±3.8% in its specific heat capacity, a calculation omitting the polystyrene yields the same final result of 360 J g−1 to 2 significant figures. This is because the other masses are large as compared with the mass of the polystyrene. The major percentage uncertainties result from the temperature differences θ1 − θ and θ − θ3. These could be reduced by using a thermometer of greater precision (e.g. 0.1 C°). Specific latent heat of vaporization of water by a method of mixtures A method of mixtures similar to that described for specific latent heat of fusion of ice can be used for determining the specific latent heat of vaporization of water. Steam from a steam generator can be piped via tubing into a calorimeter containing water at room temperature (Figure 12.11). As with the ice experiment, steam has to be piped out of the generator until only steam (and not steam and water) is flowing through the tubing. Then the tubing can be placed in the water in the calorimeter. Caution: since heat is involved in this experiment, the teacher should take the necessary safety precautions.

wet steam tubing dry steam digital thermometer heat source (the negative sign in front of LV indicates that the steam loses heat) Therefore, m c (θ − θ3) + m1c1(θ − θ1) + m2c2(θ − θ1) LV = 3 2 m3 ....(12.14) Determination of specific latent heat by electrical methods Specific latent heat of fusion of ice by an electrical method connecting leads heater ice cubes plastic or glass funnel beaker water Figure 12.12 An electrical method of determining the specific latent heat of fusion of ice. See Review question 10 for a sample calculation of LF involving this method. Specific latent heat of vaporization of a liquid by electrical methods water calorimeter −m3LV + m3c2(θ − θ3) + m1c1(θ − θ1) + m2c2(θ − θ1) = 0 Heat is supplied electrically to some melting ice in a large funnel (Figure 12.12). A mass, m1, of water collected in the beaker in the same time interval, but with the heater not turned on, is measured. The mass, m2, of water collected in the beaker in the same time interval, with the heater turned on, is measured. The value of LF for the ice is determined using equation 12.11. steam trap stirrer Using the same symbols as before, the equation for calculating LV becomes 10 °C lagging Figure 12.11 Steam from a steam generator being added to water. ITQ 7 Would your recommend the method shown in Figure 12.14 for a liquid like ethyl alcohol? Give a reason. Figures 12.13 and 12.14 show arrangements that can be used experimentally to determine the specific latent heat of vaporization of a liquid by electrical methods. The liquid chosen in this case is water. In Figure 12.13, the liquid is placed in an inner flask that has holes in its neck. An electric heater element brings the liquid to a boil.

The vapour escapes through the holes in the neck into an outer flask and then into a condenser. Chapter 12 Thermal properties of matter For the arrangement in Figure 12.14, if P represents the power of the heater and m the mass of water vaporized, equation 12.15 becomes connecting leads heater element holes in neck Pt = mLV...............(12.16) liquid in inner flask vapour Boiling and evaporation According to the kinetic theory (Chapter 14), the molecules of a liquid have a wide range of average speeds. Near the surface, some molecules with very high speeds possess enough kinetic energy to leave the liquid and form vapour. When they escape, the internal energy of the bulk of the liquid reduces, the mean square speed of the remaining molecules becomes lessened, so the temperature of the liquid falls. water outflow condenser cooling water in Figure 12.13 An electrical method of determining the specific latent heat of vaporization of water. condensed vapour The condensed vapour (now liquid) is collected in a beaker below the condenser. After the liquid has started boiling and condensed vapour emerges at a steady rate, the condensed vapour is collected for a given length of time. The latent heat equation becomes: heat supplied by heater in time t = mass of condensed vapour × specific latent heat of liquid IVt = mLV ....(12.15) for a steady current I, steady voltage V, time t, mass of condensed vapour m, and specific latent heat of vaporization LV. The escaping of molecules from the surface of a liquid takes places at any temperature. This surface phenomenon is called evaporation. Some of the molecules of vapour may return to the liquid surface and condense back to liquid. Figure 12.15 shows molecules leaving and re-entering the surface of a liquid. As a comparison, Figure 12.16 shows boiling, a phenomenon in which bubbles of vapour, containing molecules, are formed throughout the liquid. As a liquid is heated from below, molecules near the bottom gain enough energy to form vapour. The vapour is enclosed by the surrounding liquid in bubbles. A dynamic equilibrium is set up where molecules of liquid are becoming vapour and molecules of vapour are returning to the liquid.

We describe such a vapour in the bubble as a saturated vapour, and the pressure it creates in the bubbles as the saturated vapour pressure (SVP) at that temperature. The bubbles rise to the surface and burst open if the external pressure is equal to (or less than) the SVP, releasing the vapour. molecules of vapour to a.c. supply electric immersion heater (100 W) liquid surface bubble containing molecules of vapour beaker containing liquid can water molecules of liquid liquid balance heat Figure 12.14 Another electrical method of determining the specific latent heat of vaporization of water. Figure 12.15 Evaporation. Figure 12.16 Boiling. 197 198 Unit 1 Module 3 Thermal and mechanical properties of matter The conversion of liquid to vapour at a particular temperature when the SVP is equal to the external pressure is called boiling. Water boils at 100 °C at a pressure of 1 atm (103 × 105 Pa). Consider what happens inside a pressure cooker (Figure 12.17). During use, the lid of a pressure cooker is firmly fastened to the pot. As the liquid is heated, the fastest molecules leaving the surface create a high pressure inside the cooker. The liquid will therefore not boil unless the temperature is high enough to produce a large enough SVP within the bubbles being formed in the liquid. Pressure cookers are useful since they can produce high boiling temperatures and foods, such as meats and vegetables, cook much faster at higher temperatures. Many pressure cookers work at about 100 kPa over atmospheric pressure. This gives a temperature of about 120 °C. weight for obtaining desired pressure safety valve steam water food Figure 12.17 A pressure cooker. The safety valve releases steam when the pressure inside the cooker gets dangerously high. Practical importance of thermal properties Fighting fires Let us revisit the opening photograph (Figure 12.1). Water is very effective in fighting fires, not only because it is not combustible, but also because of its large specific heat capacity and its very large specific latent heat capacity. Food processing and preservation The pressure cooker (Figure 12.17) is an example of applying the thermal properties associated with boiling at high pressures to produce temperatures that can accelerate cooking and also help to sterilize food. The storage and trans-shipment of food makes use of refrigeration. Techniques such as freezing and freezedrying are used extensively today in mass distribution of food. Refrigerants make use of latent heat effects in the removal of heat from foods. To protect the environment, however, there is a ban on the use of CFC (chlorofluorocarbon) refrigerants as these are known to contribute to the destruction of the protective ozone layer around the Earth once they get into the atmosphere. An ideal refrigerant would be a volatile liquid (a liquid that is easily converted to vapour), is not flammable, has a large specific latent heat of vaporization and is environmentally friendly. Thermal waste Large bodies of water, e.g. rivers, are often used in cooling power plants. Water has a high specific heat capacity and can therefore remove the large quantities of heat generated in a power plant. However, in spite of the high specific heat capacity of water and the large masses of water in rivers, temperature rises do take place in rivers – with consequences such as reduced oxygen content and increased metabolic activity of living organisms, resulting in ecological disruptions such as death of fish and plankton.

Further, in very hot weather, when demand for air-conditioning, and hence for electrical power, increases dramatically, the warm water drawn from rivers is too hot to cool down large power plants. Some power plants therefore resort to air-cooling using large, tall, towers (Figure 12.18).

■ On account of its large specific heat capacity, water absorbs huge amounts of heat for a small rise in temperature. This helps to lower temperatures and so inhibit combustion. ■ The very large specific latent heat of vaporization results in much heat being absorbed for converting water into steam (which itself forms a fire blanket). Because the heat is absorbed, it is not available to destroy other materials at the scene. Figure 12.18 Cooling towers at a power station. Chapter 12 Thermal properties of matter However, hot air rising helps to spread air-borne pollutants generated from the power plants. Some 'evaporative cooling' towers have been employed, taking advantage of the large specific latent heat of water. But these towers have also provided an ideal environment for pathogens, causing the spread of the fatal Legionnaire's disease. Hence, we see that solving energy supply problems is not a straightforward application of thermal physics but must also take into consideration biotic, environmental factors as well. Thermal waste from power plants has also been put to good use since the large specific heat capacity of water makes it a large store of heat. In Iceland, waste waters from a geothermal plant have been used, not only to heat water within the power plant and nearby houses, but also diverted to form scenic swimming pools.

Summary ✓ The sum total of the potential energies and kinetic energies within a system of molecules is called the internal energy of a substance. ✓ Absolute temperature is proportional to the average kinetic energy (translational, rotational and vibrational) per molecule of a substance. ✓ Heat is transferred spontaneously from a hotter to a colder object until both objects attain the same temperature. The average kinetic energy per molecule of the hotter object decreases while that of the colder object increases until their average kinetic energies per molecule are the same.

When the average kinetic energies per molecule are the same, we say that the objects are in thermal equilibrium with each other. ✓ The heat capacity, C, of an object is the quantity of heat, Q, required to change the temperature of the object by 1 K (or by 1 C°). Thus, for a temperature change, ΔT, Q C= ΔT ....(12.1) ✓ The heat capacity per kilogram of a substance ✓ The specific latent heat of fusion of a solid is the quantity of heat required to change unit mass of the solid to liquid (or liquid to solid) at the melting or freezing point. In equation form, we get, LF = Q m ....(12.11) ✓ The specific latent heat of vaporization of a liquid is the quantity of heat required to change unit mass of liquid to vapour at the liquid's boiling point at standard pressure. In equation form, we get, LV = Q m ....(12.12) ✓ Boiling takes place at a particular temperature and external pressure and takes place throughout a liquid, whereas evaporation takes place at any temperature and is a liquid surface phenomenon. ✓ Applications of thermal physics include fighting fires, food processing and preservation, and dealing with thermal waste. is called the specific heat capacity, c, of the substance. Thus, considering m kg of a substance, c = Q m ΔT ....(12.2) Review questions Heat and internal energy Specific heat capacity 1 Explain, in terms of particles that make up substances, what is meant by: (a) internal energy (b) temperature (c) heat. 2 Distinguish between heat capacity and specific heat capacity. 199 200 Unit 1 Module 3 Thermal and mechanical properties of matter 3 A mass balance is marked 0–300 g × 1 g.

(a) What is the precision of such a balance? (b) What is the uncertainty (random error) in a reading on this balance?

(c) How can a systematic error arise in using this balance? 4 A student measured the mass of a given quantity of water, using a balance of precision 1 g, and reported it as 56.0 g ± 0.5 g. Has the student correctly written the measurement? Explain your answer. 5 A student used an electric heater to heat a liquid in a polystyrene cup of negligible heat capacity. The specific heat capacity, c, of the liquid was determined using a graphical representation of the equation IVt = mc(θ − θ0). The data obtained in the experiment is shown in the table. ■ power of heater = 100 W ■ mass, m, of liquid = 800 g Time, t /s 0 100 200 300 400 500 Temperature, θ / °C 25 33 40 50 56 65 6 A solid metal block was heated electrically. The energy supplied, EH, was measured by a joulemeter. Consider the equation: EH = mc(θ − θ0). (a) Complete the table below to show estimated uncertainties (due to random error) and percentage uncertainties in the measured values. Quantity Measured value Instrument Uncertainty ± precision % energy, EH 5500 J 10 J mass, m 500 g 1g initial temperature, θ0 30 °C 1 C° final temperature, θ 40 °C 1 C° θ − θ0 20 C° 7 Distinguish between latent heat of fusion and specific latent heat of fusion. 8 (a) Distinguish between boiling and evaporation. (b) Why does a pressure cooker cook food faster? 10 Why does evaporation result in cooling of a liquid? 9 What changes in internal energy take place when water at 25 °C are placed in a freezer which removes heat at an average rate of 90 J s−1 from the water. The graph, ABCD, shows the temperature, θ, against time, t, of the water as it is cooled (Figure 12.19). θ /°C 25 A 0 B C E Us (a) In the equation, IVt = mc(θ − θ0), what is the value of IV used in this experiment? (b) Using the value of IV in (a), rearrange the equation IVt = mc(θ − θ0) to make θ the subject. (c) What would the slope of a θ versus t graph of the rearranged equation in (b) represent? (d) Plot a graph of temperature, θ, versus time, t, for the liquid, assuming a starting temperature of θ0. (e) Using the slope of the graph, determine the specific heat capacity of the liquid.

(f) Suggest one precaution that should be taken in order to get good results.

5) Specific latent heat 0.1 (b) Calculate the specific heat capacity of the block. (c) Determine the uncertainty in the result from (b). –10 D Figure 12.19 (a) Which portion of the graph represents: (i) water only? (ii) ice only? (iii) the time at which ice first appears? (b) Determine the time represented by BC. (c) Determine the time represented by CE. 10 Refer back to Figure 12.11. With the heater off, water from melting ice was collected for 200 s and found to have a mass of 45 g.

With the heater on, when the ice was melting at a steady rate, water from the melt was collected for 200 s and found to have a mass of 106 g. If the heater operating power was 100 W, determine the specific latent heat of fusion of ice. Practical importance of thermal capacities 11 Suggest one reason in each case, based on thermal properties, why (a) water is used as a coolant in motor car engines (b) alcohol is used in skin coolants. 12 Suggest one way in which each case how: (a) the tremendous 'waste' heat generated in fossil fuel car engines can be put to practical use in the Caribbean (b) the heat dumped into a river from an electricity generating plant can have adverse ecological effects on the river. Chapter 13 Thermal properties of matter Practical activities and challenges Answers to ITQs 1 Boiling versus steaming (a) Compare cooking times (or textures) of a selected vegetable by boiling and steaming. Vegetables could be selected on the basis of relatively low water content (e.g. cassava, eddoe or plantain) or high water content (e.g. squash or egg-plant). (b) Investigate the effect on vitamin C content by boiling or steaming a leafy vegetable. 1 (a) The 2 kg block, since it has more particles and the average kinetic energy per particle is the same in both cases. (b) No. A transfer of heat occurs only if the temperature (average kinetic energy per molecule) in one block is different from that in another. 2 Investigating coolants (a) Compare the specific latent heats of evaporation at body temperature for different brands of skin coolants. Note that this is different from the specific latent heat of vaporization, which is measured at the boiling temperature. (It would require some ingenuity to figure out how to determine, experimentally, the specific latent heat of evaporation.) (b) Engine coolants must have large specific heat capacities. How does one brand of coolant compare with water (or even coconut water) in specific heat capacity? 2 (a) J K−1 (J/K) (b) J kg−1 K−1 (J/kg K) 3 The specific heat capacities per unit mass are about the same (≈ 25 J mol−1 K−1) for metals. 4 The law of conservation of energy. According to this law, energy is neither created nor destroyed but merely undergoes transfer or may be transferred. Hence the total internal energy of the system does not change. 5 Make Δθ large so that the percentage uncertainty (random error) with m1cIΔθ so that the difference between the two is large, and hence the percentage error small; use instruments of higher precision. 3 Heat capacity and specific heat capacity (a) Does heat capacity of a 'tawa' affect the cooking time (or texture) of roti? (b) Determine the specific heat capacity of a substance by the method of mixtures. (c) Does heat capacity play a part in 'the sauce'? (c) Determine the specific heat capacity of a particular building material or grade of material (e.g. concrete, wood, claybrick). 6 With conductors, heat supplied by the heating element transfers speedily throughout the block. With insulators, the heat spreads so slowly that one is never sure that the temperature taken at a spot at a given time is representative of the temperature of the block. 4 Product development A 'pressure cooker' boils food at temperatures higher than 100 °C. Can you make a 'low pressure cooker' which boils food at temperatures lower than 100 °C? Suggest a use for such a cooker. Answers to Review questions 5 Site visits Visits could be made to places where thermal properties of materials are employed.

Examples are: ■ a sugar factory (to observe 'pan-boiling') ■ a coconut oil or cassreep manufacturing establishment ■ a foundry.

7 No! Alcohol vapour is flammable and any spark from the electrical apparatus can ignite it. 3 (a) 1 g (b) ±0.5 g 5 (a) 100 W IV (see Worked example 12.3) (c) mc (e) 1600 J kg−1 (C°)−1 9 (a) (i) AB (ii) CD (iii) B (b) 370 s (c) 23 s 201 202 Chapter 13 Heat transfer Learning objectives ■ Describe the mechanism of thermal conduction. Q = −kA Δθ . t Δx Use the concept of equivalent conductor to solve numerical problems involving composite conductors. Outline principles involved in the experimental determination of thermal conductivity of good and bad conductors. Relate convection to density changes. Explain the origin and nature of thermal radiation. Solve problems on blackbody radiation using the Stefan's equation P = AσT4. Explain the greenhouse effect. Discuss applications of heat transfer especially with regards to vacuum flasks and solar water heaters. Discuss development of heating and cooling systems that may help reduce Caribbean dependency on fossil fuels. ■ Solve problems in one-dimensional heat flow using the equation ■ ■ ■ ■ ■ ■ ■ ■ Introduction a b Figure 13.1 A saucepan and a frying pan. What do you think about the thermal design of the handles? You may be in for a surprise ... aluminium foil Heat transfer takes place from one object to another when there is a temperature difference between them.

In Chapter 12 we referred to heat as a transfer of internal energy by vibrations of molecules when two objects are brought into thermal contact. Heat transfer, however, can take place by being carried along with matter, and also independent of a material medium. This chapter focuses on heat transfer through matter (by conduction), along with matter (convection) and independent of matter (radiation). Thermal conduction cardboard metal bar wax Figure 13.2 shows an arrangement for demonstrating heat transfer by conduction through a metal bar. The aluminium foil and cardboard prevent heat transfer around the bar from the flame to the wax and thumb tack. But after a while, the thumb tack falls off. The wax has melted due to the right-hand end of the bar becoming hot by heat transfer along the bar. flame Figure 13.2 Demonstrating heat conduction along a metal bar. Mechanism of thermal conduction ITQ 1 Why are liquids and gases generally poorer conductors of heat than solids? Experience teaches us that metals are good conductors of heat whereas non-metals generally are not. Why is this so? The atomic structures of these materials give an explanation. Chapter 13 Heat transfer a hot end cold end (a) In solid conductors: free electrons impart energy to vibrating ions. (b) In solid non-conductors: only vibrations of molecules take place. net flow of heat energy b free electron hot end cold end Figure 13.4 Thermal conduction mechanisms. (a) In solid conductors: free electrons impart energy to vibrating ions. net flow of heat energy b free electron Free electrons (Band theory) Key: vibrating ion Figure 13.4 Band theory and energy levels. vibrating molecule Heat conduction Band valence band energy conduction band (many unoccupied states) (a) energy electrons The atomic structure of metals is such that there are 'free' electrons that move about quite easily through a vibrating lattice of ions (Figure 13.3a). Band theory (see below) accounts for the high mobility of 'free' electrons, when a conductor is heated at one spot, electrons in that vicinity gain translational kinetic energy. They travel very fast and collide with nearby ions in the lattice. These ions increase in vibrational kinetic energy and pass the vibrations on to neighbouring ions, resulting in a flow of thermal energy. The vibrating ions might also pass on some energy by colliding with other 'free' electrons, which, in turn, collide quickly with other ions. The overall result is a much quicker rate of energy transfer than if molecular or ionic vibrations alone were responsible. This is why metals are such good conductors of heat. In insulators, there are no 'free' electrons to move about the material. Hence the only means of passing on thermal energy is by vibrations from particle to particle, a much slower process than in metals (Figure 13.3b).

Band theory explanation of conduction The single outermost 'valence' electron in an atom of a metal such as copper occupies a single energy level (see Chapter 22 and Figure 13.4a). However, for a large number of atoms bound together in the solid state, the valence electron energy levels split to form an outermost band of energy levels called the conduction band, and one just below it called the valence band. Energy band theory. Insulating and semiconducting materials have band gaps. In metals there is no gap between the valence band and the conduction band. In metals there is no gap between the valence and conduction band (Figure 13.4b). The energy required to jump from the valence band into the conduction band is greater than an electron can acquire from normal heating, and so the conduction band is empty. Electrons cannot therefore readily move from one energy level to another; in other words, conduction does not take place in the material.

Thermal conductivity Experiment shows that the rate of flow of heat (Q/t) is directly proportional to the cross-section area A of the material through which the heat flows and the temperature gradient, (θ1 − θ2)/x, across the material (assuming there is no heat loss or gain from the sides of the material).

In equation form: Q kA(θ1 − θ2) = ——— ....(13.1) t x where θ1 is the temperature at the hotter end, θ2 is the temperature at the colder end and x is the length of the material between θ1 and θ2. 203 204 Unit 1 Module 3 Thermal and mechanical properties of matter Table 13.1 Thermal conductivities, k, of selected substances Substance Thermal conductivity, k / W m−1 K−1 air 0.023 water 0.56 wood 0.10 bakelite 0.23 concrete 0.84 glass 0.84 steel 40 aluminium 200 copper 385 The constant, k, is called the thermal conductivity of the material and is large for conductors and small for non-conductors (see Table 13.1). Getting it right! We are treating Qt as positive if heat travels from a hot region to a colder region. Hence, in equation 13.1, Q/t is positive since θ1 > θ2. Equation 13.1 could also be written: Q −kA(θ2 − θ1) = t x since (θ2 − θ1) is negative and the sign in front of the k is negative. Worked example 13.1: Heat transfer through a door Q Calculate the rate of heat flow through a solid wooden door (2.5 m high, 1.0 m wide, 4.0 cm thick) if the outer surface temperature is 35 °C and the inner surface temperature is 25 °C. A ■ area, A = 2.5 × 1.0 = 2.5 m2 ■ thickness, x = 4.0 cm = 0.040 m ■ temperature difference, θ1 − θ2 = (35 − 25) °C = 10 °C = 10 K ■ thermal conductivity of wood (see Table 13.1), k = 0.10 W m−1 (C°)−1 Using equation 13.1, we get Q 0.10 W m−1 (C°)−1 × 2.5 m × 10 C° = 0.040 m = 63 W Getting it right! Make sure you understand the difference between °C and C°. °C refers to an actual temperature reading; C° refers to a temperature change. Conceptual example 13.2: The saucepan and the frying pan Q Comment on the thermal designs for the handles of the saucepan and the frying pan (Figure 13.1). A At first glance, it looks as if a metal handle for a saucepan is a poor design, thermally. Won't heat be easily conducted along the handle, making it hot, since both pot and handle are made of metal? In contrast, the insulating Bakelite handle for the frying pan restricts the flow of heat and therefore its handle would not get as hot. Let us analyse the thermal design involving each of these handles. Let us treat both saucepan handle and pot handle as being approximately at a room temperature. According to equation 13.1, for a given room temperature θ2 and pot temperature θ1 Q kA(θ1 − θ2) = t x In the case of the frying pan (a Bakelite handle), k is extremely small. Hence, Q is small, and therefore it takes quite a while t before a significant amount of heat reaches the far end of the handle. During this time, much of the heat passing within the handle would have been lost to the surrounding air by radiation or convection (radiation and convection are discussed later in the chapter). Thus the temperature of the handle further from the heated pan is much smaller than the temperature, θ1, nearer to the hot end. A closer look at the saucepan handle reveals that it is hollow. So the metal handle is also made of steel which has a relatively low thermal conductivity when compared with metals such as aluminium or copper. Hence, because of a somewhat smaller k (though not as small as that for the Bakelite handle on the frying pan), as well as the small A, Q for the saucepan handle is small. ITQ 2 Why is a large area A undesirable? Again, Using t the previous argument for the Bakelite handle, the temperature of the saucepan, θ2, at the far end of the handle being comfortable to the hand. (It is also quite possible that the manufacturer may have placed a thin insulator between the saucepan handle and the pan itself, thus further reducing Q.) t This metal Figure 13.5 Saucepan handle, showing the cross-section. ITQ 2 Referring to Figure 13.6b, what do you notice about the magnitude of the temperature gradient (the slope of θ versus x) along the bar of high thermal conductivity as compared with that along the metal of lower thermal conductivity? ITQ 3 Why are cooling coils of copper (rather than, say, steel) used in the Searle's bar method (see Figure 13.7)? Chapter 13 Heat transfer Worked example 13.3: Compound bar Q 01 An aluminium bar of length 30.0 cm and a steel bar of length 20.0 cm and same cross-section area are joined thermally end-to-end (see Figure 13.6a).

The aluminium end is kept at 250 °C and the steel end at 100 °C.

(a) Find the temperature where the two metals are joined. (b) Draw a graph of temperature versus length, x, for the compound bar. A Let the temperature of the junction be θ °C. We assume that there is no heat loss along the sides of the bar. Applying equation 13.1 we get Q kAl(250 − θ) ksteelA(θ − 100) = = t 0.300 0.200 Multiplying across gives us 0.500 × kAl(250 − θ) C° = 0.300 × ksteel(θ − 100) C° Inserting values for k from Table 13.1, we get 0.500 W m−1 (C°)−1 × 200 W m−1 (C°)−1 × (θ − 100) C° 25 000 − 100θ = 12θ − 1200 1120 = 26 200 θ = 234 °C = 234 °C Figure 13.6b shows the temperature variation with length along the compound bar aluminium steel θ = 250 °C 100 °C b θ2 θ3 θ4 water inflow to copper cooling coils wrapped in lagging to reduce heat losses by radiation and convection Figure 13.7 Searle's bar method for determining thermal conductivity of a metal. aluminium steel x Figure 13.6 Determining temperature along a compound bar. The bar's surface/volume ratio is kept small so that relatively less surface is available for heat loss.

This is achieved by having a large diameter (e.g. about 3 cm). 3 Because of the high conductivity of metals, a length, x, of at least 20 cm is used so as to obtain a temperature difference of reasonable size. Too small a temperature difference would result in a large percentage error in (θ1 − θ2). 4 The contacts between the cooling coil and the bar must be clean. Dirt, for example, will reduce the rate of heat flow from the bar into the coil. 250 θ / °C When steady temperatures are achieved in all four thermometers, the mass, m, of water flowing through the cooling coils in a time t is measured. 100 0 10 20 30 40 50 60 70 80 x / cm Figure 13.6 (a) The compound bar of aluminium and steel. (b) Graph of temperature, θ, versus position along the compound bar, x. Experimental determination of the thermal conductivity of a conductor: Searle's bar Figure 13.7 shows a Searle's bar method of experimentally determining the thermal conductivity of metal. The metal bar is thermally insulated (for example by polyurethane) and heated at one end by an electrical heating coil. Water flowing through copper cooling coils wrapped tightly around the other end of the bar removes heat conducted along the bar. We make use of the specific heat capacity equation: Q = mcΔθ...............(13.2) Interpreting the results Applying equation 13.1 to the section of the bar between θ1 and θ2 and assuming no heat losses from the sides of the bar (since it is lagged), we get The rate of removal of heat from the bar is then given by mc(θ3 − θ4) = t t The thermal conductivity of the bar can then be calculated using equation 13.1. See Review questions, question 4, for an exercise in this regard. Figure 13.5 (a) What is the value of the specific heat capacity of the bar? 14 Which of these has a cylinder of a 1 length l and radius r the surface to volume ratio is 2/r. (b) In what way does r affect the surface/volume ratio? 205 a steam in b steam chest steam out θ2 brass base insulation brass disc θ insulator under test θ1 Figure 13.8 Lee's discaapparatus for determining thermal conductivity of an insulator. Lee's disc method Experimental determination of thermal conductivity of an insulator: Lee's disc method The insulator is shaped into a thin, smooth clean disc. It Q = mc Δθ is sandwiched between a steam chest on the top and a brass base at the bottom (Figure 13.8). The whole apparatus is suspended by three small chains. Steam is passed through the steam chest. When the thermometer readings are steady, θ1 and θ2 are read. These are assumed to be the temperatures on the top and bottom of the insulating disc. Using equation 13.1, along with the thermometer readings and the dimensions of the disc, we can calculate the thermal conductivity of the disc if the rate of flow of heat is known (see Review questions, question 5). Experimental issues 1 On account of its very small k, the rate of heat conducted through an insulator is extremely small, giving much more time for heat to be lost from its side compared with a conductor. The disc is therefore made very thin, so as to reduce the surface area of its side and corresponding heat loss.

2 Since a thin disc is used, its thickness would need to be measured by a Vernier callipers or micrometer screw gauge so as to reduce the percentage error in determining thickness x. However, typically a disc of diameter about 10 cm is used, and this can be measured with a mm ruler. 3 For a temperature difference of (θ1 − θ2) ≈ 10 °C, using a thermometer of 0.1 °C would result in only about 1% error in the measurement of the temperature difference. θ1 Δθ Δt Δθ = mc Δθ t Δt x brass disc Temperature / θ 206 θ1 Δt Δθ At Time / s Figure 13.9 Determining rate of cooling of brass disc.

a Insulated disc
Insulated pad on top of lower disc.

(b) Graph of θ versus t.

The rate of flow of heat, Q, through the insulating disc is equal to the rate at heat loss through the lower brass disc. To determine this rate of flow of heat, the steam chest and insulating disc are removed (using gloves to prevent burns!). The chest is then placed directly on the brass disc until the temperature of the brass disc is about 5 C° above θ2. Next, the chest is removed from the top of the brass disc and replaced by a thick insulator (e.g. felt or polyethane) (Figure 13.9a). A graph of temperature versus time is plotted for the disc as it cools and the slope, Δθ, at a temperature equivalent to θ2 is determined. Figure 13.9b Δt The rate of heat flow Q at that temperature is given by Q mcΔθ = t Δt where m is the mass of the brass disc and c its specific heat capacity. ITQ 6 The rate of flow of heat through the disc is given by Q mcΔθ = t Δt = kA(θ1 − θ2) This means that k = mc ITQ 5 Why is it important that the sides of the insulating disc and the brass disc be smooth? xΔθ At πr2(θ1 − θ2) Assuming negligible error in m and c, write an equation for Δk/k that can be used for estimating the overall experimental error. Chapter 13 Heat transfer Conceptual example 13.4: Equivalent conductor Worked example 13.5: Double-glazed window Q Q A At Time / s What length, l2, of conductor of thermal conductivity k2 is equivalent thermally to a length, l1, of conductor of same cross-section area, A, and thermal conductivity k1? glass air θ l k1 , l1 A k1 , l1 , A k2 , l2 , A k2 , l2 , A 30 °C Figure 13.10 Equivalent thermal conductors Worked example Based on Figure 13.10, and applying equation 13.1, we get Q kA(θ − θ2) kA(θ − θ2) = 1 1 = 2 t 11 l2 Simplifying the equation we get l l2 = ...(13.3) l1 k2 k1 = k2 Thermal convection Earlier, we pointed out that liquids and gases are generally poor conductors of heat. We see, however, that heat travels much more rapidly than solids by a method called convection. This is because, in convection, large numbers of molecules move rapidly carrying thermal energy with them. In solids, molecules can only vibrate over small distances, passing on their vibrations to neighbours – a much slower way of transferring heat as compared to convection. When a liquid is heated (Figure 13.12) it expands because molecules are vibrating more rapidly and pushing against each other, causing an increase in their mean separation distance. The density of the liquid near the heat source becomes less than the density of the surrounding liquid. The hot liquid, having lower density than the cooler surrounding liquid, therefore rises and cooler liquid rushes in to take its place, as shown by the arrows. Convection also takes place in gases. The effect is similar to that in liquids, hot gases also rise, causing convection currents. However, expansion of heated gases is due mainly to an increase in average translational speeds of molecules resulting in an increase in mean separation distance. Figure 13.11 Double-glazed window 25 °C A double-glazed 1.0 m × 1.2 m window is made of glass panes of thickness 3.0 mm separated by a 2.0 mm thick layer of still air (Figure 13.11).

What is the rate of heat flow through this window when the temperatures inside and outside of the window are 25 °C and 30 °C, respectively? A We first convert the layer of air between the panes to a thermally equivalent slab of glass. By equation 13.2, we get kglass lglass = × lair kair 0.84 W m−1 (C°)−1 × 2.0 × 10−3 m 0.023 W m−1 (C°)−1 = 7.3 × 10−3 m (or 7.3 mm thickness of glass) The total equivalent thickness of glass now becomes xglass = 3.0 mm + 7.3 mm + 3.0 mm = 13.3 mm Applying equation 13.1 to this equivalent window, we get Q kglassA(θ1 − θ2) = t xglass = 0.84 W m−1 (C°)−1 × 1.0 × 1.2 m2 × (30 − 25) C° 13.3 × 10−3 m = 3.8 × 102 W (i.e. 380 W) Instead of a solid glass window, 13.3 mm thick, a thinner glass/air/glass window, 8.0 mm thick, serves the same purpose thermally. ■ beaker Land and sea breezes can be explained by convection. ■ During the day, the land heats up faster than the sea.

Hot air rises above the land causing a current of cooler air to rush in from the sea to take its place. This current of air is called a sea breeze (Figure 13.13a). potassium permanganate crystal heat Figure 13.12 Convection currents in water. 207 208 Unit 1 Module 3 Thermal and mechanical properties of matter a afternoon night b 1 km sea breeze land breeze warm land cool water Figure 13.13 Formation of land and sea breezes. (a) Sea breeze during the day. (b) Land breeze during the night. cool land warm water 50 N 30 N polar high h North Pole polar easterlies northern hemisphere L westerlies L = low pressure H = high pressure subtropical high H north-east trades northeast trades Equator L south-east trades intertropical convergence zone (ITCZ) equatorial low southeast trades H L South Pole subtropical high westerlies southern hemisphere polar easterlies polar high Figure 13.14 (a) Vertical section of part of the Earth, showing convection air currents in the northern and southern hemispheres. (b) The north-east and south-east trade winds. The arrows directed toward the left show the effect of the Coriolis effect. ■ During the night, the land cools down faster than the sea. A land breeze forms as the hot air over the sea rises and cooler air from the land rushes to take its place. Convection also plays a part in the formation of major wind and ocean currents. Heated air in the tropics rises and cooler air from northern latitudes moves in to take its place (Figure 13.14a). The results are our year-round winds on the surface of the Caribbean. The surface winds would have been from north to south, but they are deflected westward by a Coriolis force effect, which is due to the rotation of the Earth towards the east. The winds end up blowing toward a south-westerly direction. These winds are called north-east trade winds (since they blow from the north-east) (Figure 13.14b). A convectional North Equatorial Current of water, stretching from north-west Africa to the northern coastal ITQ 7 Why are islands like Barbados and Trinidad and Tobago such ideal locations for international surfing competitions?

waters of South America, is believed to be formed partly by the shearing forces of the north-east trade winds. The current merges with the Gulf Stream (Figure 13.15a). The Gulf Stream flows northward towards the colder latitudes from the very warm waters of the Gulf of Mexico, along the shorelines of the southern United States and continues towards Europe. From Europe, the current returns to the equatorial regions via the west of Africa (Figure 13.15b). The Gulf Stream is huge, approximately 80 km wide and 1 km deep, and is responsible for the mild weather along the shores of the eastern United States and western Europe during winter. The rather large speed (approx. 1–2 m s−1) of the massive Gulf Stream has led to serious speculation about harnessing its kinetic energy as a major alternative source of energy in the future. Chapter 13 Heat transfer a b Greenland Newfoundland Europe North America 2 Atlantic Ocean deep water 1 Gulf Stream 3 warm, fresh, less dense, shallow water cold, salty, dense, deep water a c b Figure 13.15 Global convection water currents. The North Equatorial Current (1), the Gulf Stream (2) and the South Equatorial Current (3) are part of a global system of convection currents at the surface of seas and under water. dry air cools hot moist air low pressure, warm ocean north-east trade winds low pressure, warm ocean rotation of clouds low pressure, warm ocean Figure 13.16 Convection plays a significant part in hurricane formation. (a) Hot air carrying moisture rises. Clouds are formed. An area of low pressure is also created below, due to the rising air. (b) Clouds swirl anticlockwise in the northern hemisphere due to the interaction with the north-east trade winds coupled with the Coriolis effect. (c) As more warm air rises and clouds are formed, latent heat of condensation is given to the clouds making them swirl faster.

Dry air from above rushes towards the area of low pressure. As warm waters reach cooler regions, some of the water may also sink due to an increase in density and return towards the equatorial regions via an underwater current. Massive convection loop current currents have been identified worldwide involving surface and underwater currents. Locally occurring underwater currents can pose dangers to swimmers and divers, since they involve flow of large masses of water. The Gulf Stream also heats the air above it, aiding formation of tropical cyclones (hurricanes). Most hurricanes in the Caribbean, however, originate from near the equator where north-east trade winds and ITQ 9 Why do most hurricanes in the Caribbean form between June and October? south-west trade winds converge, creating low pressure regions, and where the waters are very warm. Hurricane formation is complex. In a simple description, moist, rising air forms clouds (Figure 13.16a). The clouds, influenced by the north-east trade winds and the Coriolis effect, swirl anticlockwise in the northern hemisphere.

As moisture carried by the rising air condenses, latent heat is given up to the clouds, making them swirl faster about a central 'eye' of low pressure (Figure 13.16b). Dry air from above the clouds rushes towards the area of low pressure (Figure 13.16c). The air picks up moisture from beneath and becomes heated by the warm waters, and the cycle continues. 209 210 Unit 1 Module 3 Thermal and mechanical properties of matter Thermal radiation Worked example 13.6: Hot metal ball Heat can be transferred from one body to another by electromagnetic waves. This type of transfer is called thermal radiation.

Electromagnetic radiation is emitted by molecular vibrations since molecules consist of electric charges. Oscillating electric charges are accelerating and therefore emit electromagnetic waves of varying wavelengths (see Chapter 25). A useful concept in the study of radiation is the blackbody. A blackbody is a body that absorbs all radiation incident on it. Matt black objects come very near to this ideal. An enclosure with a small hole is also a blackbody since any radiation entering is reflected multiple times within the enclosure and eventually absorbed (Figure 13.17). Experiment shows that the surface of a perfect blackbody is also a perfect radiator of heat. Other surfaces do not absorb or radiate heat as well. The ratio of the intensity of radiation given off by a given surface of a body to that given off by the blackbody at the same temperature is called the emissivity, e, of the surface. Q The surface temperature of a metal ball of diameter 2.0 cm and emissivity 0.24 is 500 °C. (a) Calculate the power output due to radiation. (b) Calculate the power absorbed by the ball if it was placed in an enclosed, insulated room of temperature 25 °C. A (a) Using P = eσAT4 we get P (0.24 × 5.67 × 10−8 W m−2 K−4 × 4π × (1.0 × 10−2)2 m2 × 7734 K4 = 6.1 W (b) The ball will also absorb radiation from the room. Assuming that the ball is at TI and room temperature is T2, we get Pnet = 0.24 × 5.67 × 10−8 W m−2 K−4 × 4π × (1.0 × 10−2)2 m2 × (7734 − 2984) K4 ≈ 5.97 W ≈ 6.0 W We define the term 'thermal equilibrium' temperature. The net power that radiated by a body at temperature T in an enclosure of temperature T0 is given by Pnet = eσA(T4 − T04)...............(13.5) entrance hole Quality of thermal radiation Figure 13.18 shows that not all wavelengths of radiation from a black body are emitted at the same intensity. The peak intensity wavelength decreases with increasing temperature of the blackbody. hollow object Figure 13.17 The concept of an ideal blackbody. The power, P, of radiation emitted by a blackbody was studied by Josef Stefan (1835–1893), who arrived at the equation P = σAT4 ....(13.4) where σ is a constant called Stefan's constant. ITQ 10 For a body whose surface is not perfectly black, equation 18.3 becomes P = eσAT4...............(13.4) When tungsten wire is heated to about 1400 K, it looks red. This is because the peak intensity is in the infrared region (2000 nm) and only a small fraction of the radiation lies in the visible band, mainly in the red-orange region (550–700 nm). At about 3000 K, the peak intensity is still in the infrared region (1500 nm), but the radiation spread across the entire visible spectrum (400–700 nm), making the tungsten glow a reddish white. where e is the emissivity of the surface, A is the surface area, T the absolute temperature and σ is a constant called the Stefan constant. The value of σ is 5.67 × 10−8 W m−2 K−4. The Sun appears white because its surface temperature is about 6000 K. At this temperature, the peak intensity of wavelengths radiated lies in the middle of the visible spectrum, and there are large intensities of all the wavelengths of the visible spectrum in the light radiated. If the body is surrounded by an enclosure at temperature T0, the enclosure will also radiate heat to the body. The heat radiation exchanges will result in a net of heat temperature of one body and a net drop in the other until ITQ 10 Why does absorption of heat radiation by a body result in a temperature rise of the body? Chapter 13 Heat transfer Intensity (relative) 3000 K 6000 K 1400 K 500 1000 1500 visible spectrum Some stars, however, appear blue.

This is because their peak intensities (radiated at about 30 000 K) occur in the ultraviolet region and wavelengths of the visible spectrum radiated are predominantly in the blue region, and to a far lesser extent in the rest of the visible spectrum. Greenhouse effect Certain materials, for example glass, are quite transparent to short-wavelength electromagnetic radiation but less so to longer wavelength radiation. 2 Wavelength / nm Figure 13.18 Intensity distribution of radiated blackbody radiation. The glass allows short-wavelength waves from the Sun to enter the greenhouse. Substances within the greenhouse absorb the waves, with an accompanying rise in temperature. These substances then emit mainly longer wavelength radiation, which is trapped within the greenhouse since they do not pass through glass easily. The net result is that the temperature within the greenhouse becomes higher than the temperature outside. This allows enough warmth for growth of plants even during winters – hence the name 'greenhouse effect'. On a global scale, carbon dioxide and other 'greenhouse' gases, such as methane, are believed to contribute to a greenhouse effect. Like glass, these greenhouse gases are more transparent to short-wavelength radiation than to long-wavelength radiation. Without the greenhouse effect produced by naturally occurring greenhouse gases, the temperature of the Earth would be −18 °C. Global warming, the rising of the temperature of the planet, is believed to be the result of an enhanced greenhouse effect due to human activities. In part, this enhanced effect is thought to be linked to an increase in the amount of greenhouse gases, especially carbon dioxide, in the atmosphere. This increase is due to human activities, such as the burning of fossil fuels which add carbon dioxide to the air, and also to massive cutting 2000 Wavelength nm Figure 13.18 Peak intensities of blackbody radiation and corresponding wavelengths for radiation emitted at different blackbody temperatures. down of forest trees, which, through photosynthesis, help to remove carbon dioxide from the air. There is much debate as to the causes of global warming, and the United Nations has set up a special committee, The Intergovernmental Panel on Climate Change, since it is believed that the rising temperature on the planet seriously affects weather patterns, due to the fact that convection currents in both wind and seawater are brought about by temperature effects. Applications involving energy transfer Vacuum flask Figure 13.19 shows the construction features of a vacuum flask. These features help to minimize heat transfer between the contents of the flask and the environment. Let us consider a hot liquid in the flask to see how this transfer is minimized.

system case double glass walls silvering on both sides of both walls vacuum felt pads Figure 13.19 The structure of a vacuum flask. 211 212 Unit 1 Module 3 Thermal and mechanical properties of matter glass cover a sunlight to domestic hot water system hot water insulating material polyurethane backing heat exchanger tube glazing b black collector plate (heatabsorbing surface) glazing frame inlet connection outlet connection flow tubes cold panel exchanger pump cold water enclosure absorber plate insulation Figure 13.20 Solar water heaters: (a) solar thermal water heating system; (b) flat-plate collector detail. ■ The double-glass walls minimize heat loss by conduction since glass is a poor conductor. ■ The vacuum between the glass walls also minimizes heat loss by conduction, since conduction cannot take place in a vacuum. The vacuum also minimizes heat loss by convection, since heat transfer by convection requires a liquid or gas. ■ The interior silvered surface of the glass walls reflects heat back into the liquid; the exterior silvered surface of the glass walls reflects heat back into the liquid, as silvering is a poor radiator of heat; these two minimize radiation heat loss. ■ The stopper is made from air-filled plastic or cork.

Both materials reduce heat loss by conduction since they are insulators. They also prevent hot air from escaping from above the liquid, thus minimizing heat loss by convection. Further, the stopper also prevents heat loss by evaporation from the surface of the hot liquid. ■ The felt pads protect the glass from mechanical shock and are themselves thermal insulators, very transparent to longer wavelength radiation and therefore heat is trapped in the collector, resulting in a rise in temperature. ■ The collector plate is blackened to absorb heat radiation readily. It is made of a good conductor (e.g. copper) so that heat is easily transferred to metal tubes bolted on to the plate. The metal flow tubes are also made of a good conductor. Heat is conducted from the metal collector plate to the metal tubes and into the water that flows within the tubes. ■ Polyurethane insulation minimizes heat escaping by conduction from the base of the collector. ■ The outlet is placed above the inlet during use. Heated water in the tubes will rise and carry heat by convection to this outlet. Hot water from the collector outlet passes through an insulated tube to a metal heat exchange tube (Figure 13.20a) ■ The case is made of tough plastic to help protect the glass container. Plastic is also a poor conductor and this further helps to minimize heat loss. ■ The layer of air between the glass container and the case helps to reduce heat loss by conduction since air is a poor conductor of heat. Solar water heater Figure 13.20a shows a flat-plate collector solar water heater connected to the storage tank of a domestic water heating system. Figure 13.20b shows construction details of the flat-plate collector. ■ The glass allows short wavelength electromagnetic waves from the Sun to pass easily into the collector. The waves heat up the collector plate, which in turn radiates electromagnetic waves, but of longer wavelengths (refer to Figure 13.18). Glass is not Figure 13.21 A solar water heater with roof-top storage tank. Note that the collector is black but the storage tank is shiny. Can you say why? Chapter 13 Heat transfer within an insulated water-storage tank. Heat from the hot water within the heat exchange tube is conducted across the metal tube to the water in the storage tank. By convection, the heated water rises in the storage tank. An outlet at the top of the tank allows this heated water to be used to supply hot water for domestic purposes. As in the collector, the cold water inlet to the storage tank is placed near the bottom of the tank.

The water exiting from the heat exchanger tubing is piped back into the inlet of the collector plate.

The heating cycle results in very hot water being produced in the storage tank. In some solar heating systems, the storage tank is placed above the collector plate (Figure 13.21). In others, a pump assists in the circulation if the storage tank is too far below the collector. The pump, itself, can be operated by solar (electric) energy. Other applications involving energy transfer In Guyana, it had been the practice to dry foods such as fish and padi commercially using the direct heat (electromagnetic radiation) of the Sun. Moisture from the heated foodstuff escapes by convection into the open air. The practice of drying padi was gradually being superseded by using more 'modern' silos, powered by electricity generated from fossil fuel power plants – with the latter being implicated in adding the greenhouse gas, carbon dioxide, to the air due to the combustion of the fossil fuel. However, with the worldwide trend towards using 'cleaner' (i.e. non-carbon dioxide producing), renewable sources of energy, solar drying might well be encouraged and developed in the Caribbean, where solar energy is in the greatest abundance in the world. In a similar way, the Sun is used to produce sun-dried tomatoes (Figure 13.22). Figure 13.22 Sun-dried tomatoes. One medical application of infrared radiation transfer is in thermograms.

A thermogram is a picture of an object formed from the infrared radiation it emits.

Thermograms have been useful in diagnosing areas of breast cancer as these areas characteristically emit more intense infrared radiation. A thermogram taken inside a house can identify areas where most heat enters. Infrared security cameras can now take thermograms of the human face. Each face exhibits a unique pattern, which can be compared with a pattern stored in a computer memory. It is claimed that these cameras can distinguish between identical twins. Heating and keeping cool in the Caribbean It takes energy to provide heating or cooling. In the sunny Caribbean, there is heavy reliance on fossil fuel for this energy. Trinidad and Tobago is a source of fossil fuel. Its natural gas supplies could help reduce the amount of carbon dioxide produced by combustion of fossil fuels such as oil and coal. However, the world's supply of fossil fuel is finite and is running low. Further, fossil fuel usage has been linked to environmental pollution (through chemicals and particles released into the air) and to enhanced global warming (through release of carbon dioxide into the air). At a summit held in Italy in July 2009, the G-8 nations agreed to reduce CO2 emissions by 50% by the year 2050 through less usage of fossil fuel. There is increasing interest in the Caribbean in using energy alternatives such as solar and wind, especially in attempting to meet its heating and cooling needs. Some hotels and homes in Barbados are already using solar energy to heat pools and to provide hot water for domestic use through solar panels on roof tops of their buildings (Figure 13.21). Aero generators convert wind energy to electricity, some of which is used for refrigeration and air-conditioning. The Island School, Bahamas, has committed itself to being 'carbon neutral' through measures like utilizing solar and wind power to offset the carbon dioxide emissions that would have resulted from their use of fossil fuel energy. A hospital in the Guyana hinterland, at Orealla, has been powered by solar energy using solar photo-voltaic cells and storage batteries. Refrigeration is achieved using electricity generated from these solar cells. However, the use of each of these alternatives has a downside on the environment. The reader is urged to do internet searches to be aware of these and their long term effects. 213 214 Unit 1 Module 3 Thermal and mechanical properties of matter Summary ✓ Through careful application of heat transfer principles, energy wastage associated with heating and cooling can also be reduced.

■ Using shiny pots for cooking can help reduce domestic heat wastage that can occur by radiation. ■ Insulating building materials (e.g. hollow tiles that trap air) can help reduce heat getting into homes by conduction. ■ Proper positioning of windows and vents on a building can ensure that cool air enters the building (from below) and hot air exits (from above). ■ Although wearing of light-coloured clothing helps to reduce heat absorption from the Sun, Caribbean folk can still be comfortable if not-so-light-colour clothing is worn loosely as this would allow heat to escape by convection between the clothing and the body. ■ In the painting of buildings, a light colour can help reduce absorption of heat radiation and hence reduce air-conditioning demands. Summary ✓ Conduction is a process in which thermal energy ... ✓ Radiation is the transfer of heat by electromagnetic waves. The power, P, of radiation emitted by a body is given by Stefan's equation: is transferred from hotter areas to colder areas by means of molecular vibrations. ✓ In conductors, free electrons, being very mobile, transfer thermal kinetic energy readily as they move through the lattice structure of solid conductors and collide with the vibrating ions.

The vibrating ions might also pass on some energy by colliding with other 'free' electrons, which, in turn, collide quickly with other ions. The overall result is a much quicker rate of energy transfer than if molecular or ionic vibrations alone were responsible. ✓ The order of decreasing conductivity is generally solid, liquid, gas. This is due to the fact that the mean separation distances between molecules increase in that order and therefore the time taken for energy to pass on also increases. ✓ The rate of heat transfer by conduction is given by the equation: Q kA(θ1 − θ2) = t l ....(13.1) t where θ1 is the temperature at the hotter end, θ2 is the temperature at the colder end, l is the length of the material between θ1 and θ2. The constant, k, is called the thermal conductivity of the material. ✓ The transfer of heat bodily by a fluid is called convection. The density of the fluid near the heat source becomes less than the density of the surrounding fluid. The hot fluid therefore rises and cooler fluid rushes in to take its place. Convection currents play a significant role in wind and ocean current formation. P = σAT4...............(13.3) For bodies that are not perfect blackbodies, Stefan's equation becomes: P = εσAT4...............(13.4) where ε is the emissivity of the surface, A is the surface area emitting the radiation, T is the absolute temperature of the surface, and σ is a constant called the Stefan constant. The net power Pnet radiated by a body at temperature T in an enclosure of temperature T0 would be given by Pnet = εσA(T4 − T04)...............(13.5) ✓ Applications involving energy transfer by conduction, convection and/or radiation include the vacuum flask, the solar water heater, solar drying and thermograms. ✓ Heating and cooling systems, based on clean, renewable, alternative sources of energy such as solar and wind, which are in abundance, can help reduce the Caribbean dependency on fossil fuels. However, the diffuse nature of these energy sources, as well as the requirement for land space, both for scenic and housing purposes, are only two of many problems that must be taken into consideration when developing these sources for use in society. Principles involved in conduction, convection and radiation can also be applied in the design of buildings and water- and space-heating systems in collecting energy from these sources as well as for minimizing wastage of energy. Chapter 13 Heat transfer Review questions Thermal conduction, Thermal conductivity 1 (a) What is meant by 'thermal conduction'? (b) Explain why thermal conductivity in solid metals is usually much greater than thermal conductivity in solid insulators. 2 (a) How can one explain the fact that, generally, decreasing thermal conductivity is shown in the order: solids, liquids and then gases? (b) Suggest an explanation for the thermal conductivity of wood (k = 0.10 W m−1 K−1) being less rather than greater than that of water, a liquid (k = 0.56 W m−1 K−1) (see Table 13.1). 3 (a) Estimate the rate of heat conduction from within the human body to the outer portion of the skin. Assume the following data: temperature 1.0 cm below the skin = 37 °C temperature at surface of skin = 35 °C surface area of skin = 1.5 m2 average thermal conductivity of body tissue = 0.20 W m−1 K−1 (b) What effect would a layer of fat just under the surface of the skin have on the rate of outward heat conduction. 4 In a Searle's bar experiment (see Figure 13.6) to determine the thermal conductivity of a metal, the following data was obtained: θ1 = 70.0 °C θ2 = 65.0 °C θ3 = 58.0 °C θ4 = 55.0 °C length, x = 20.0 cm diameter of bar, d = 4.0 cm mass of water collected, m = 200 g time of water collection, t = 100 s (a) Determine the thermal conductivity of the metal. (b) What is a major source of experimental error that could arise in a Searle's bar experiment? Discuss, in terms of the length of the bar, how such error can be minimized. 5 In a Searle's bar experiment (see Figure 13.7) to determine the thermal conductivity of a metal (a) why is the heating coil insulated but not the copper cooling coil? (b) why is the rate of flow of heat along the bar calculated from the rate of cooling by the cooling coil rather than simply power × voltage of the heating coil? 6 An iron bar of length 20.0 cm is joined to a copper bar of length 30.0 cm.

The free end of the iron bar is kept at a temperature of 100 °C while the free end of the copper is kept at 0 °C. What is the temperature at the junction of the metals? 7 0.20 kg of water in an iron pot of diameter 24.0 cm and thickness 3.0 mm boils off in 5.0 minutes. (a) Calculate the rate at which energy is supplied to the water. (b) Estimate the temperature on the underside of the pot. 8 A thin slice of insulator is sandwiched between two identical metal bars.

One end of the compound bar is placed in a flame; the other end is kept in ice. Sketch a graph of temperature versus length along the bar, starting from the hot end. (Assume that the bar is insulated to prevent heat loss from the sides.) 9 In a typical Lee's disc experiment to find the thermal conductivity, k, of a kind of wood (see Figure 13.8), the following data was obtained: diameter of wooden disc, d = 10.0 cm thickness of wooden disc, x = 2.0 mm temperature θ1 = 85 °C temperature θ2 = 95 °C slope of temperature versus time graph at temperature θ1 for the brass disc: = 8.5 × 10−3 C° s−1 mass of brass disc, m = 1.50 kg specific heat capacity of brass, c = 370 J kg−1 K−1 (a) Use the above data to calculate k. (b) Why are the brass discs highly polished? 10 A double-glazed window of area A is made of plastic of thickness l1 and thermal conductivity k1, glass of thickness l2 and thermal conductivity k2, and a layer of air of thickness l2 and thermal conductivity k3. (a) What thicknesses of plastic, lp2, would have the same thermal effect as air of thickness l2? (b) What thicknesses of plastic, lp3, would have the same thermal effect as glass of thickness l3? (c) Show that the magnitude of the rate of flow of heat through the window is given by Q A(θ1 − θ2) = t l1/k1 + l2/k2 + l3/k3 where θ1 and θ2, respectively, the inside and outside temperatures of the window. (Hint: treat the window as being made of plastic, using plastic equivalents of air and glass.) 215 216 Unit 1 Module 3 Thermal and mechanical properties of matter Thermal convection 11 Describe how heat loss by convection can be minimized when boiling vegetables in a pot. 12 An electric oven has a heating element, E, at the bottom, as shown in Figure 13.23. Two racks are shown above the heating element. Figure 13.23 Sketch the convection current when the heat of the oven is turned on. Blackbody radiation 13 Consider the solar water heater panel (Figure 13.7) to determine the thermal conductivity... Thermal radiation 13 Explain why a glass of hot water cools down faster in a vacuum flask than a lead through a double-glazed window. Explain your answer.

Explain this observation. 14 (a) As the Sun 'moves' from east to west, what direction should the incline of a fixed solar water heater panel face. Explain your answer.

Heating and keeping cool in the Caribbean 18 (a) What is meant by 'fossil fuel'? (b) Give two reasons why the Caribbean should develop alternative sources of energy to lessen dependence on fossil fuels in order to address its heating and cooling needs.

(c) One energy alternative within a country is to construct a solar voltaic power station to provide electricity for various purposes, including for heating and cooling. What area of collector is needed to construct such a 20 MW power station? (Assume an average normal incidence solar power of 800 W m−2 and solar panel efficiency of 30%.) (d) Is it better for one large solar voltaic power station to be constructed in your country, or to encourage householders to have individual roof-top solar voltaic units? Give one reason for and one reason against the position you take. Thermal radiation 14 (a) Estimate the power radiated from the skin (assume a skin temperature of 35 °C, area 0.30 m2 and emissivity 0.70).

(b) Estimate the power of air-conditioning needed to remove power radiated from the skin in a room of 100 persons if the efficiency of the air-conditioner is 40%. 15 Compare the rate of heat conducted to the surface of the skin (see question 3a) with the rate of heat radiated from the skin (see question 14a). How can you account for the large difference? Applications involving energy transfer 16 (a) Draw a labelled diagram of a vacuum flask.

(b) Describe, with reference to conduction, convection and radiation, how the design features of a vacuum flask help to keep ice from melting too quickly. 17 Refer to the solar water heater panel (Figure 13.20b). (a) Describe, with reference to conduction, convection and radiation, how the design features help to maximize the rate of heating up and the temperature of the water. (b) Solar water heater panels are usually mounted at an incline rather than horizontal. Give one reasons for this. Practical activities and challenges Some suggested challenges based on concepts covered in this chapter are: investigations, model-building, internet/ library activities and stargazing. Which of these capture(s) your imagination?

As usual, using the internet or library will enrich your approaches in all of the challenges.

1 Investigations (a) Thermal conductivity of local building materials (e.g. clay bricks, hollow tiles, wood). (b) Thermal conductivity of lean meat versus fatty meat. (c) Thermal conductivity of leaves, such as banana leaves used for wrapping food (e.g. 'kankey' in banana leaf) or layering of pots with leaves, for cooking certain local delicacies. 2 Ideas for model building (a) Solar cooker. (b) Solar water heater. (c) A flat model in a tank to illustrate ocean currents (e.g. 'continents' made out of sand, ice placed at the 'poles', a heater at the equator and a fan to simulate the 'north-east trade winds'. (d) A simple computer model to illustrate convection. (e) Model 'houses' and checking the effects of paints and other variables on internal temperature. Chapter 13 Heat transfer 3 Internet/library activities to widen your knowledge (a) The Gulf Stream.

(b) Ocean gyres. (c) Cyclogenesis in the oceans (above and below). (d) R-values and building materials. (e) Causes of global warming. (f) Stabilization mechanisms in the Earth's temperature. (g) Formation of hurricanes. 4 Stargazing Identify high, low and medium temperature stars by sight and estimate their temperatures by the colour of light they emit. Obtain a star map and identify the stars.

Then find out the current estimated temperatures of these stars. Answers to ITQs 1 The mean separation of molecules is least in solids, and a little greater in liquids. Hence, the time taken for vibrations to be passed on, and hence the conductivity, is greater in solids than in liquids. In gases, the molecules move about fairly freely and the mean separation of molecules is very great. The rate at which energy is passed on by collisions, and hence the conductivity, is small compared to solids and liquids. 2 The magnitude of the temperature gradient along the metal of higher thermal conductivity is smaller than that along the metal of lower thermal conductivity. The θ versus x graph is less steep in the former case than in the latter. 3 Copper has a much greater thermal conductivity than steel and would therefore remove the heat being conducted along the bar far more efficiently than the steel as water flows through the coils. 4 (a) surface area, A divided by volume, V = 2πrl 2 = r πr2l (b) The smaller the value of r, the greater will be the surface area relative to the volume and hence, the greater the heat loss from surface of the bar. 5 If the disc surfaces in thermal contact were not smooth, air, itself an insulator, could be trapped in pockets between them, resulting in a spurious value of the conductivity of the insulating disc. 6 ħk ΔAθ(Δt) An 2πr ΔθI = Q = + + k t+ (ΔθI) n 02 7 These islands lie in the path of the powerful and year-round north-east trade winds which produce great coastal surfs when interacting with the Atlantic Ocean. 8 The westward direction of the South Equatorial Current is believed to be due largely to the sweeping force of the south-east trade winds on the ocean surface. 9 This is the hottest season of the year and therefore convection is at its greatest. Convection plays a significant role in hurricane formation. 10 When molecules absorb heat radiation, they gain energy. The molecules vibrate faster, resulting in a rise in temperature, since temperature is directly related to average kinetic energy of molecules. Answers to Review questions 3 (a) 60 W (b) 101.3 °C (d) 1567 W 9 (a) 0.03 W m−1 K−1 15 (a) power radiated power conducted = 535 W 60 W =9 (Note: the skin is also receiving power from the surroundings. Hence, net power radiated from skin is much less than 535 W.) 217 218 Chapter 14 Kinetic theory of gases Learning objectives ■ Discuss the basic assumptions of the kinetic theory of gases. ■ Explain gas pressure in terms of molecular motion. 1 Nm. 3 ■ Use the equations of state for an ideal gas, pV = nRT and pV = NkT. ■ Derive the equation, Ek = 3 kT, for the average translational kinetic energy of a 2 monatomic gas.

■ Derive the equation, Ek = 3 nRT for the total kinetic energy of a monatomic gas. 2 ■ Derive and use the equation, pV = Basic assumptions of the kinetic theory of gases The kinetic theory of ideal gases makes a range of assumptions. Comments on these assumptions are given in the brackets. 1 Gases are made up of numerous tiny particles, called molecules. (The large numbers enable the behaviours of gases to be analysed and described statistically.) gas are moving, they must be doing so with a random distribution of speeds and in random directions to produce a net gas velocity of zero.) 4 Molecules have negligible volume.

(The volume of molecules in a real gas, however, becomes significant at high pressures, when the molecules of the gas are close together, and the total volume in which they can move about freely becomes small.) 2 Molecules of a given gas are identical. (Real gases, however, may have molecules containing isotopes and therefore not all molecules in such gases may have the same mass.) 5 Molecules are far apart from each other and therefore do not attract each other. (The attraction, however, becomes significant for real gases at high pressures when the total volume of a gas becomes small and molecules are close to each other.) 3 Molecules of a gas are continually moving independently of each other with random speeds and in random directions.

(The net velocity of the gas in a stationary container must be zero. Since molecules of 6 Collisions between molecules of a gas, or involving molecules of a gas and those of the walls of the gas container, are perfectly elastic. (This means that the collisions do not involve loss of kinetic energy.) 7 The duration of a molecular collision is very small as compared with the time between successive collisions. (We can therefore ignore potential energies associated with molecular collisions as molecules of a real gas at ordinary temperatures and pressures would display kinetic energies most of the time.) A molecule is the smallest particle of a substance that can exist by itself. It is also the smallest particle of a compound that retains the properties of the compound. Figure 14.1 According to the kinetic theory, gases are made up of identical particles moving freely at random speeds in various directions. This chapter explores the connection between the kinetic theory of gases and experimental data on pressure, volume and temperature of gases. Gas pressure and the kinetic theory The kinetic theory can explain why a gas exerts a pressure on a surface in contact with the gas. Chapter 14 Kinetic theory of gases A molecule rebounds upon colliding with a surface. Hence, there is a change in momentum of the molecule brought about by the collision. According to Newton's second law, the change in momentum per unit time of the molecule represents the force on the molecule (due to the surface). Hence, there is an average force exerted by the surface on a set of rebounding molecules. This average force is constant since numerous molecules would be bombarding the surface at any given small interval of time. The rebounding molecules exert an equal and opposite force on the surface. The surface experiences a pressure, since pressure is force per unit area. Let us first focus on the component of movement in the x direction.

Assuming a perfectly elastic collision between the molecule and the wall, kinetic energy is conserved. This means that the magnitude of the initial velocity, v1, is unchanged. This, in turn, means that since the y and z components of its velocity are unchanged by the collision, then the x component must be reversed. The magnitudes of v1x and v'1x are therefore equal. Taking directions into consideration, v1x = −v'1x since the molecule rebounds from the wall. The time, Δt, between successive collisions of this molecule with the wall is the time taken to traverse a distance 2l, and is given by Δt = surface molecule v –v Figure 14.2 A molecule rebounding from a surface undergoes a change in momentum (normal incidence of velocity component is shown).

change in momentum = mv – (m(–v) = 2mv 2l v1x (The y and the z components of v are at right angles to l and therefore do not affect Δt.) By Newton's second law, the average force, F1, per collision on the molecule is given by Δ(mv1) F1 = Δt = (mv'1x − mv1x) ÷ 1 Derivation of the equation pV = 3 Nm = −2(mv1x) ÷ We now derive a formula for the pressure, p, exerted by a gas on the walls of a container of volume V.

= Consider a molecule of mass m inside a rectangular container of length l (Figure 14.3). The diagram shows a molecule moving in an arbitrary direction with a velocity v1. The components of the molecule's velocity before collision with one of the walls of the container, situated along the y-z plane shown, are v1x, v1y and v1z. After the collision, the molecule's velocity has components of v'1x , v'1y and v'1z. y 2l v1x 2l (since v1x = −v'1x) v1x −mv1x2 l By Newton's third law, the average force exerted by the one molecule (on the wall) is therefore given by mv1x2 F1 = l The net force, F, on the surface due to all the molecules in the container is given by F = F1 + F2 + F3 + ... + FN where each molecule is assigned a number from 1 to N. Since each molecule has identical mass, then l F= v1 ( ) A m(v1x2 + v2x2 + v3x2 + ... + vNx2) l i.e. x F= z Figure 14.3 A molecule of gas with velocity v1 inside a rectangular container of length l and cross-section area A. ITQ 1 According to which law would the rebounding molecules exert an equal and opposite force on the surface? mN l ....(14.1) where = (v1x2 + v2x2 + v3x2 + ... + vNx2) N ....(14.2) The symbol is the mean value of the squares of the magnitudes of the x components of the speeds of the molecules. Since the motion of molecules of the gas 219 220 Unit 1 Module 3 Thermal and mechanical properties of matter is assumed to be random, then = = . We can use Pythagoras' theorem to find the effective mean-square speed per molecule. We denote this mean square speed as . We get It can be shown from equation 14.5 that, for a gas of density, ρ, p= 1 ρ 3 = 1 ρv 2 3 rms = + + = 3 (since = =) Substituting = 3 into equation 14.1, we get mN F= 3l ....(14.3) Since the volume, V = lA, equation 14.3 becomes F= AmN 3V i.e. Worked example 14.1: rms speed of oxygen molecules Q The density of oxygen at a pressure of 102 kPa is 1.30 kg m−3. Calculate the rms speed, vrms, of the oxygen molecules. A Substitute in the equation 1 p = ρ 3 1 102 × 103 N m−2 = × 1.30 kg m−3 × 3 F mN = A 3V Therefore, p= 1 Nm 3 V 3 × 102 × 103 N m−2 1.30 kg m−3 vrms = √235 384.6 m s−1 = 485 m s−1 = ...(14.4) since p= ....(14.6) F A Distribution of speeds of molecules in a gas Or 1 pV = Nm 3 ....(14.5) The square root of (the mean-square speed of the molecules of a gas) is called the root-mean-square speed, vrms, of a molecule of gas. Notes about the component, vx 1 Even if the molecule were to collide with other walls of the container, the magnitude of the component vx would not change, since the x component is not affected by collision forces that take place in the y or z directions. As pointed out earlier, the net velocity of gas in a stationary container must be zero. Since molecules of gas are moving, they must be doing so with a random distribution of speeds and in random directions to produce a net gas velocity of zero.

The distribution of molecular speeds, v, in a gas has been determined for various temperatures by statistical considerations, assuming random speeds of molecules, by James Clerk Maxwell. vmost probable Nv 2 If the molecule were to collide with another molecule of the gas, its vx might change. However, the change will be shared with the other molecule and the total effective vx in the system due to the first molecule will remain unchanged. Getting it right! Make sure you can differentiate between mean-square speed and root-mean-square speed. ■ Mean-square speed is a square speed and is the vaverage vrms Figure 14.4 Typical Maxwellian distribution of random speeds of a gas at a given temperature, obtained from statistical considerations. mean of square speeds of molecules. Unit is m2 s−2. ■ Root-mean-square speed, being a square root of the mean-square speed, is an actual speed. Unit is m s−1. v ITQ 2 Show that for a gas of density ρ p = 1 ρ 3 Chapter 14 Kinetic theory of gases A typical Maxwellian distribution of speeds is shown in Figure 14.4. The mean speed, the most probable speed and the rms speed are indicated for a given temperature. Equation of state for an ideal gas As with gas pressure, the kinetic theory provides an explanation for gas temperature. This explanation makes use of the equation of state for an ideal gas pV = nRT Mole, molarity and molar mass The mole (mol) refers to 'quantity of substance'. It has been established experimentally that there are 6.022 141 × 1023 atoms present in exactly 12 g of carbon-12. This number of atoms is called the Avogadro constant, NA, and its unit is mol−1. The mole: one mole (1 mol) of any substance is defined as NA particles of that substance. Atomic mass and the mole and the equation For a fixed mass of gas, the following relationships have been established experimentally. One mole of carbon-12, according to the definition of the mole, has a mass in grams that is numerically equal to the atomic mass in atomic mass units, u. The atomic mass of carbon-12 is exactly 12 u, hence the mass of 1 mole of carbon-12 is 12 g. It follows from this definition that one mole of atoms of any element has a mass in grams that is numerically equal to the atomic mass in atomic mass units, u, of the element. It is only pure carbon-12 for which molar mass is an exact whole number – the molar masses of other elements in their natural state are slightly different from whole numbers because: ■ The pressure of a fixed mass of gas is inversely ■ they are mixtures of isotopes 1 pV = Nm 3 derived on the basis of the kinetic theory. Before we combine these two equations, we will discuss certain experimental results and concepts associated with the equation of state. The combined gas equation proportional to its volume, provided the temperature is held constant: 1 p∝ V ■ there are slight changes in mass associated with or pV = constant (Boyle's law) The mass of NA particles of a substance is called the molar mass, M, of the substance. Molar mass is therefore the mass of 1 mol of substance and its S.I. unit is kg mol−1. ■ The volume of a fixed mass of gas is directly proportional to its absolute temperature, provided the pressure is kept constant: V∝T or V = constant (Charles's law) T ■ The pressure of a fixed mass of gas is directly proportional to its absolute temperature, provided the volume is kept constant: p∝T or p = constant (pressure law) T These three relationships have been combined into a single equation: pV = constant T ....(14.7) Equation 14.7 is called the combined gas equation. Equation 14.7, rearranged, becomes pV = constant × T...............(14.8) nuclear binding energy. Molar mass Molarity: the molarity of a substance is the number of moles, n, of the substance present in a given quantity of the substance. The S.I. unit for molarity is the mole (mol). Derivation of equation of state for an ideal gas: pV = nRT According to Avogadro's hypothesis, equal volumes of any gas at the same temperature and pressure contain equal numbers of molecules. This hypothesis has been verified experimentally by measuring volumes of gas released (or used up) in chemical reactions. Experimentally, it has also been determined that 1 mole of any gas at 0 °C (273.15 K) and a pressure of ITQ 3 How many moles of carbon dioxide are present in 56 g of the gas? ITQ 4 What is the molar mass of nitrogen gas? The nitrogen molecule in nitrogen gas is diatomic, i.e. consists of two atoms. The relative atomic mass of natural nitrogen = 14.0067 u. 221 222 Unit 1 Module 3 Thermal and mechanical properties of matter Worked example 14.2: Molar mass of carbon dioxide Q Calculate the molar mass of carbon dioxide (CO2). ■ relative atomic mass of natural carbon = 12.011 (this is greater than 12 because natural carbon contains 1% of the isotope carbon-13) ■ relative atomic mass of natural oxygen = 15.9994 A The total atomic mass units, u, in CO2 ≈ 12 + (2 × 16) u ≈ 44 u Therefore the mass of 1 mol of CO2 ≈ 44 g Hence the molar mass of CO2 ≈ 0.044 kg mol−1 More accurately: The total atomic mass units, u, in CO2 = 12.011 u + (2 × 15.9994) u = 44.0098 u ≈ 44.01 u Therefore the mass of 1 mol of CO2 ≈ 44.01 g Hence the molar mass of CO2 ≈ 0.04401 kg mol−1 101.325 kPa (i.e. at standard temperature and pressure, abbreviated 'STP') occupies 22.4 litres or 0.0224 m3. By considering 1 mol of gas at STP, we can calculate the value of the constant in equation 14.8 as follows: pV = constant × T 101.325 Pa × 103 × 0.0224 m3 = constant × 273.15 K constant = 101.325 Pa × 10 × 0.0224 m (for 1 mol of gas) 273.15 K 3 = 8.31 N m K−1 mol−1 = 8.31 J K mol −1 This value of the constant for 1 mol of gas is known as the molar gas constant, R. Equation 14.8 therefore becomes ....(14.9) Equation 14.9 is called the equation of state for an ideal gas. If the total number of molecules in n moles of gas is N, then N = nNA. The Boltzmann constant, k, is defined as Therefore k = (a) We use the equation of state, pV = nRT pV n= RT 2.0 × 105 Pa × 5.0 × 10−3 m3 = 8.31 J K−1 mol−1× (273 + 27) K = 0.40 moles (b) 1 mol has 6.02 × 1023 oxygen molecules, i.e. 2 × 6.02 × 1023 oxygen atoms (since each oxygen molecule consists of 2 atoms). Therefore 0.40 moles of oxygen gas contains 0.40 × 2 × 6.02 × 1023 oxygen atoms = 4.8 × 1023 atoms Equation 14.9 now becomes pV = NkT...............(14.11) Equation 14.11 is another representation of the equation of state for an ideal gas. Gas temperature and the kinetic theory We now derive the equation 1 3 m = kT 2 2 ■ The left-hand side of this equation represents the average translational kinetic energy of a molecule of a gas. temperature. or R NA A ■ The right-hand side shows the relationship with pV = RT for 1 mole of gas k= A flask of volume 5.0 l (5.0 × 10−3 m3) contains oxygen gas at a pressure of 2.0 × 105 Pa and a temperature of 27 °C. Calculate: (a) the number of moles of oxygen in the flask (b) the number of oxygen atoms in the flask. which explains gas temperature on the basis of the kinetic theory.

−1 pV = nRT for n moles of gas Q 3 constant = 8.31 N m−2 m3 K−1 mol−1 Worked example 14.3: Number of moles and atoms in a gas The total translational kinetic energy of n moles of a gas consisting of N molecules is given by 1 1 1 1 Ek = 2 mv12 + 2 mv22 + 2 mv32 + ... + 2 mvN2 1 = 2 m(v12 + v22 + v32 + ... + vN2) 1 mN(v12 + v22 + v32 + ... + vN2) = 2 N ....(14.10) nR N ITQ 5 Show that the value of the Boltzmann constant is 1.38 × 10−23 J K−1. Chapter 14 Kinetic theory of gases Worked example 14.4: Density of a gas at a given temperature and pressure Q A Calculate the density of oxygen (O2) at a temperature of 25 °C and a pressure of 101 kPa. We consider 1 mole of oxygen. ■ molar mass of oxygen, M = 2 × 16 g = 32 g = 0.032 kg To find the volume of 1 mole of oxygen at 25 °C and 101 kPa, we use the equation of state: pV = nRT...............(14.9) to find the volume, V. We get nRT V= p 1 × 8.31 × (273 + 25) = 101 × 103 Worked example 14.5: rms speed of molecules Q (a) What is the rms speed of helium molecules at STP? (b) What would be the rms speed of nitrogen molecules? A ■ = 2.45 × 10−2 m3 density = M V molar mass of helium = 0.004 kg mol−1 ■ molar mass = 0.018 kg mol−1 at STP The rms speed (√2) depends on the temperature. At STP, the temperature is 0 °C. ■ T = 0 + 273 = 273 K By equation 14.15, 1 3 m = kT 2 2 therefore, 3kT = m Since molar mass, M = 0.004 kg mol−1 then mass of 1 molecule 0.004 kg m= 6.02 × 1023 0.032 kg = 2.45 × 10−2 m3 = 3 × 1.38 × 10−23 J K−1 × 273 K × 6.02 × 1023 4.0 × 10−3 kg 2 −2 = 1 700 981 m s Therefore, vrms = √1 700 981 = 1304 m s−1 (b) For nitrogen 3 × 1.38 × 10−23 J K−1 × 273 K × 6.02 × 1023 = 18 × 10−3 kg = 242 997 m2 s−2 Therefore, vrms = √242 997 = 493 m s−1 = 1.3 kg m−3 = Therefore, 1 Ek = 2 mN...............(14.12) since = (v12 + v22 + v32 + ... + vN2) N But, combining equation 14.10 (pV = NkT) with equation 1 14.5 (pV = 3 Nm) we get 1 Nm = NkT 3 ....(14.13) Equation 14.12, combined with equation 14.13, yields 1 Ek = mN 2 = 3 NkT 2 Ek(average) = Therefore the total translational kinetic energy of the gas is given by 3 nRT 2 The average translational kinetic energy per molecule is 1 m. 2 1 Since Nm = NkT ....(14.13) 3 then 3 NR R = (since k =) 2 NAT NA Ek(total) = 4.0 × 10−3 kg 6.02 × 1023 ....(14.14) = 1 m 2 3 kT 2 ....(14.15) Therefore, average translational kinetic energy per molecule = 3 kT. 2 223 224 Unit 1 Module 3 Thermal and mechanical properties of matter Summary ✓ The kinetic theory of an ideal gas assumes that gases are made up of tiny particles, called molecules, with negligible volume and ■■ are identical to each other ■■ do not attract each other ■■ move independently within a gas with random speeds and in random directions when they collide, they do so elastically (i.e. kinetic energy is conserved) ■■ the duration of a molecular collision is small as compared with the time taken between molecular collisions. ✓ The root-mean-square speed, vrms (or √), 2 of the molecules of a gas is the square root of the mean of the squares of the speeds of the molecules.

In equation form √(v12 + v22 + v32 + ... + vN2 √ = vrms = N ✓ The pressure, p, of N molecules of a gas each of mass m enclosed in a volume V is given by ....(14.5) 10−3 m3). ✓ The molarity, m, is the number of moles of substance present in a given quantity of substance. ✓ Molar mass, M, is the mass of 1 mole of substance. ✓ In terms of temperature, the equation of state, pV = nRT becomes pV = NkT...............(14.11) where k is the Boltzmann's constant defined by k= R NA ....(14.10) ✓ The total kinetic energy of a gas is given by 3 Ek(total) = nRT 2 ....(14.14) ✓ The average translational kinetic energy per molecule of a monatomic gas is given by ✓ In terms of density, ρ, of a gas 1 p = ρ 3 This number (symbol, NA) is called Avogadro's constant and is the number of atoms present in exactly 12 g of carbon-12. ✓ At STP, 1 mol of gas occupies 22.4 litres (22.4 × ■■ 1 pV = Nm 3 ✓ A mole consists of 6.022 141 × 1023 particles. ....(14.6) ✓ The equation of state for an ideal gas is given by pV = nRT.............. (14.9) for n moles of gas; R is called the molar gas constant and has a value of 8.31 J K−1 mol−1. Ek(average) = 3 kT 2 ....(14.15) Chapter 14 Kinetic theory of gases Review questions 5 (a) Distinguish between mole, molar mass and molarity.

(b) Using data from the periodic table of elements, complete the table below for the two gases. Basic assumptions of the kinetic theory 1 (a) State four basic assumptions of the kinetic theory of an ideal gas. (b) Discuss each of the above assumptions in relation to a real gas. 2 (a) What is meant by the expression root-mean-square speed? (b) For the following set of speeds: 1.0 m s−1; 2.0 m s−1; 3.0 m s−1; 4.0 m s−1; 5.0 m s−1; 6.0 m s−1, calculate the: (i) the mean speed (ii) the mean-square speed (iii) the root-mean-square speed. Gas pressure and the kinetic theory 3 Consider molecules within a rectangular box of length l and cross-section area A, as shown in Figure 14.5 y l v1 A x z Figure 14.5 (a) Prove that the average force, F1x, exerted on A by a single molecule, moving with speed v1 is given by mv1x2 F1x = l mv 2 (b) Using the equation, F1x = l1x , derive the equation pV = 1 Nm for N molecules moving 3 with a mean-square speed within the box. 2 4 (a) Starting from the equation, pV = 1 3 Nm,1show that the density, ρ, of a gas is given by p = 3 ρ. (b) The density of nitrogen at a pressure of 104 kPa is 1.66 kg m−3. Calculate the rms speed of the nitrogen molecules. Gas No. of particles Molar mass in 1 mole of gas Mass of 1 molecule Molarity of 100 g of gas neon (Ne) methane (CH4) 6 (a) Write down the equation of state for an ideal gas, giving the meaning of each symbol. (b) Calculate the density of hydrogen (H2) at a temperature of 20 °C and a pressure of 102 kPa. (Hint: consider 1 mole of hydrogen gas and find its molar mass. Use the equation of state to find the volume occupied by this 1 mole of gas at 20 °C.) 7 A flask of volume 4.8 l (4.8 × 10−3 m3) contains only carbon dioxide (CO2) at 1.5 × 105 Pa and 27 °C. Calculate: (a) the number of moles of CO2 in the flask (b) the number of CO2 molecules in the flask (c) the number of atoms present in the gas in the flask (d) the average translational kinetic energy of a molecule of CO2 in the flask. 8 (a) What mass of nitrogen (molar mass 14 g mol−1) at pressure 1.5 × 105 Pa is present in a volume of 0.25 m3 at 30 °C? (b) Determine the rms speed of the nitrogen molecules. Gas temperature and the kinetic theory 9 (a) The equation of state for an ideal gas is given as pV = nRT. Give the meaning of each symbol in the equation. 2 (b) Combining the equation pV = 1 3 Nm (derived from the kinetic theory) and the equation of state pV = nRT (obtained experimentally), derive an expression for (i) the total translational kinetic energy, Ek(total), of a monatomic gas (ii) the average translational kinetic energy per molecule, Ek(average).

10 (a) Calculate the rms speed of oxygen molecules (molar mass 32 g mol−1) in air at a temperature of 27 °C. (b) If the pressure due to the oxygen molecules is 102 kPa, determine the density of the oxygen. 225 226 Unit 1 Module 3 Thermal and mechanical properties of matter 11 (a) Calculate the average translational kinetic energy of molecules in an ideal gas at 27 °C. (b) Calculate the total translational kinetic energy of 1 mole of molecules of an ideal gas at the same temperature of 27 °C. (c) With what speed must a 0.140 kg cricket ball be travelling to have the same translational kinetic energy as 1 mole of molecules of an ideal gas at 27 °C? Answers to ITQs Practical activities and challenges 3 Since there are 44.01 g of CO2 in 1 mol, then number of moles of CO2 in 56 g = 56 ≈ 1.3 mol 44.01 4 The total atomic mass units in N2 = 2 × 14.0067 u = 28.0134 u 1 Determination of the value of R Using a simple Boyle's law apparatus (Figure 14.6), determine the value of R, using air.

A suggested procedure is described below. 0 glass tube 10 1 Newton's third law (if A exerts a force on B, then B exerts an equal and opposite force on A). 2 density, ρ = 1 since pV = Nm (equation 14.5) 3 1 Nm 1 = ρ then, p = 3 V 3 Therefore the mass of 1 mol of N2 = 28.0134 g Hence, the molar mass of nitrogen gas is 0.028 0134 kg mol−1 5 k= transparent protective screen 20 dry air Bourdon pressure gauge 30 oil column 50 connection for foot pump kPa 60 cm3 valve oil reservoir Figure 14.6 Boyle's law apparatus. (a) Read the volume, V, of air trapped in the glass tube and the corresponding temperature, T, and pressure, p. (b) Correct the volume, V, to V0, the volume at STP. Hence, determine the number of moles, n, of air trapped. (c) Using measured values of p, V and T, and the equation pV = nRT, determine the mean value of R.

(d) Calculate the percentage difference between your experimental value of R and the accepted value. Compare this percentage difference with an estimate of the percentage error expected based on the precision of your measurements. (e) How can the apparatus be modified so that gases other than air can be trapped in the tube and investigated? R 8.31 = = 1.38 × 10−23 J K−1 NA 6.02 × 1023 Answers to Review questions 5 40 mass Nm = volume V (b) Gas neon (Ne) No. of particles Molar mass Mass of in 1 mole of gas 1 molecule 6.02 × 1023 methane 6.02 × 1023 (CH4) 7 (a) 0.29 mol (b) 1.7 × 1023 (c) 5.2 × 1023 (d) 6.2 × 10−21 J 11 (a) 6.2 × 10−21 J (b) 3.7 kJ (c) 231 m s−1 Molarity of 100 g of gas 20.2 g 3.36 × 10−26 kg 4.95 mol 16.04 g 2.66 × 10−26 kg 6.23 mol 227 Chapter 15 First law of thermodynamics Learning objectives ■ Define thermodynamics. ■ State the first law of thermodynamics in terms of change in internal energy, ΔU, heat, Q, supplied to a system, and work, W, done on the system: ΔU = Q + W. ■ Solve problems involving the equation ΔU = Q + W. ■ Calculate the work, ΔW, done on a gas initially of volume V and at pressure p ■ ■ ■ ■ ■ ■ using the equation ΔW = pΔV. Determine work done from a p-V graph. Differentiate between isothermal and adiabatic processes. Define molar heat capacity. Use the equations Qv = nCvΔθ and Qp = nCpΔθ.

Explain why molar heat capacity of a gas at constant pressure is different from that at constant volume. Determine efficiency of a p-V cycle. First law of thermodynamics Figure 15.1 The first law of thermodynamics in action. Thermodynamics Energy can be transferred as heat and by work. In Chapter 13, we described heat as a transfer of energy between two objects due to a difference in temperature. There is a net flow of heat energy from a hotter to a colder object. In Chapter 4 we described work as representing a process of conversion of one form of energy to another, for example the process of transfer of energy due to mechanical forces acting over a distance. When a weightlifter does work (for example, when lifting a weight through a height), chemical energy of food is converted to gravitational potential energy that is stored in the weight at its final position. (Incidentally, the weightlifter's muscles also get warm in the energy transfer process of work, showing that while useful work is being done, heat energy may also be transferred.) Thermodynamics is a study of processes involving energy transfers by heat and work. to change. In other words, work done on a gas could produce a change in the internal energy of the system, since internal energy represents the sum total of kinetic and potential energies of the molecules of the system. First law of thermodynamics ITQ 2 Suppose when work is done on a system and heat is added to the system, there is a change in kinetic energy (ΔKE) and potential energy (ΔPE) of the system as a whole as well.

How should equation 15.1 be modified? In Chapter 4, we saw that when work is done on a system of objects, the kinetic energy of the system can change; the potential energy of the system can also change. If the system consists of molecules (as in a gas, for example), the work done could cause the total energy of the system In equation form, the first law of thermodynamics states that ΔU = Q + W...............(15.1) where ΔU represents the change in internal energy of a system, Q represents net heat supplied to the system and W represents the net work done on the system. ITQ 1 The statement, 'work done = change in kinetic energy of an object', is given a special name. What is that name? ITQ 3 Should heat, Q, and work, W, be considered as state variables? 228 Unit 1 Module 3 Thermal and mechanical properties of matter Intuitively, the law expresses the concept that if heat is added to a system and positive work is done on a system, the internal energy of the system will increase in each case. The first law of thermodynamics is therefore a statement of the law of conservation of energy, since both heat and work represent energy transfers. piston F Δd –F cylinder Getting it right! The following are the conventions used in equation 15.1: ■ heat supplied to a system is designated as positive gas ■ heat supplied from a system is designated as negative Figure 15.2 An ideal gas undergoing a compression. ■ work done on a system is designated as opposite in sign to work done by a system. What is a system? We consider a system as an object or set of objects. Everything external to the system we designate as the surroundings. In most of our considerations in this chapter, we focus on a fixed mass of gas as our system. We describe a gas system by its state variables. Quantities that describes characteristics of a gas, such as mass (m), or number of moles (n), pressure (p), volume (V), temperature (T) and internal energy (U), are state variables. F , or is force per unit area, then the applied pressure, p = A F = pA. Work done on the gas by the external pressure is therefore positive and is given by ΔW = pAΔd or, since AΔd is a change in volume, ΔV, ΔW = pΔV...............(15.2) By Newton's third law, the force exerted by the gas on the piston at each stage of the compression is −F.

Hence the magnitude of the pressure exerted by the gas is also p.

Therefore the work done by the gas is given by −FΔd, i.e. by −pΔV. We see, then, that: ■ the work done on a gas is the negative of the work done by the gas Worked example 15.1: Change in internal energy ■ the work done on a gas by a compression is positive. Q 3000 J of heat are added to a gas and 1200 J of work is done by the gas. What is the change in the internal energy of the gas? Graphical determination of work done by a gas A ■ heat added to gas, Q = +3000 J work done on gas, W = −1200 J (work done on a gas is negative of work done by a gas) By the first law of thermodynamics, ΔU = Q + W = 3000 J + (−1200) J = 1800 J ■ The first law and thermodynamics processes Work done on a gas; work done by a gas Let us consider as our system a fixed mass of ideal gas enclosed in a cylinder with a light, freely moveable piston of cross-section area A (Figure 15.2). Work, ΔW, done on the gas during a compression is given by FΔd, where F is the force applied to the piston and Δd a tiny distance moved by the piston in the direction of F. Since pressure Consider n moles of a gas at a state A defined by pressure p1, volume V1 and temperature T1 (Figure 15.3). For a tiny decrease in volume, ΔV, the pressure p1 remains practically constant. By equation 15.2, the small amount of work, ΔW, done on the gas is given by ΔW = pΔV. Since, graphically, pΔV is the area of the shaded region on the graph, then work, ΔW, is equal to the area of the shaded portion under the graph. It follows that the total work W done on a gas in changing its state from A to B is equal to the total area under the p-V graph from A to B. p B p2 T2 A T1 p1 V2 ΔV V1 V Figure 15.3 Work done is equal to the area under a p-V graph. Chapter 15 First law of thermodynamics Getting it right! Since the change of state from A to B involves a decrease in volume (i.e. a compression), the work done on the gas is positive. It follows that the work done by the gas in this case is negative. heat reservoir heat Isothermal process gas In an isothermal process, the system (the gas) is kept at a constant temperature. We imagine the gas in thermal contact with a heat reservoir, a massive body whose temperature does not change significantly during heat exchanges into or out of the gas. It is assumed that an isothermal process is carried out very slowly so as to allow all of the gas to stay at the same temperature as the heat reservoir. p B A temperature TB VA VB V Figure 15.4 Two p-V isothermals for the same fixed mass of an ideal gas. Figure 15.4 shows two p-V graphs for a fixed mass of ideal gas at fixed temperatures TA and TB, where TB > TA. These graphs are called isothermals. Graph B is to the right of graph A. This is so since pV = nRT for an isothermal process (equation 14.9). Since n and R are constants, the product, pV, is larger for larger T. Hence, p1VB is always greater than p1VA if TB > TA. Therefore, for a given p1, VB will always be greater than VA, if TB > TA , resulting in isothermal B being to the right of A. For a monatomic ideal gas, the internal energy, U, is only translational and its kinetic energy, Ek(total), is given by 3 U = Ek(total) = nRT 2 Therefore, for an isothermal process, 0 = Q + W. In other words, W = −Q...............(15.3) temperature TA p1 Figure 15.5 In an isothermal process, when work is done on a gas, heat flows from the gas to the reservoir. ....(14.14) At constant temperature, the change in internal energy of the gas is therefore zero. But by the first law of thermodynamics, ΔU = Q + W Hence the work done on the gas is equal to the negative of the heat flowing into the gas. Put another way, work done on the gas during an isothermal process is equal to the heat flowing out of the gas (Figure 15.5). Adiabatic process In an adiabatic process, heat does not enter or leave the system. Hence, Q = 0. A gas enclosed in a well-insulated cylinder will go through an adiabatic process if the pressure of the gas is increased suddenly. Figure 15.6 shows p-V graphs for a gas initially at a state A, taken to a state B isothermally and to a state C adiabatically. p pf B C adiabatic isothermal pav p1 A V3 V2 V1 V Figure 15.6 p-V graphs for a gas going through an isothermal process (red line, AB) and an adiabatic process (black line, AC) for small pressure changes at relatively small pressures. By the first law of thermodynamics, ΔU = Q + W. Therefore, since Q = 0 ΔU = W...............(15.4) ITQ 4 What would the sign of the work done by a gas in going from state B to state A (Figure 15.3)? ITQ 5 What is another interpretation of equation 15.3, i.e. when the equation is expressed as Q = −W? All work done on the system manifests as a change in internal energy of the system. If the work done on the system is positive, an increase in the temperature of the gas will result.

229 230 Unit 1 Module 3 Thermal and mechanical properties of matter p Discussion example 15.2 Q In which process is more work done on a gas during a pressure change from p1 to pf where pf > p1 (Figure 15.6): adiabatic or isothermal?

A The work done during each process is equal to the area under the p-V curve for that process. We can consider two cases here. (a) Pressure changes small at relatively small pressures. We note that the adiabatic curve, AC is generally steeper than the isothermal curve for small changes in pressure at relatively low pressures. By approximating AB and AC as straight lines (Figure 15.6) and multiplying the average pressure, pav, by the respective changes in volume, we see that pav(V3 − V1) > pav(V2 − V1). Therefore, the work done isothermally is greater than the work done adiabatically for the same change in pressure. With reference to Figure 15.7, a compression from a pressure of 0.0625 units, for example, to 1 unit, results likewise in the area under the isothermal graph (red curve) being greater than that under the adiabatic graph (black curve). (b) Pressure changes large. In Figure 15.7, the isothermal graph is practically vertical over a wide range of pressures, especially in the high-pressure region, whereas the adiabatic curve, still to the right of the isothermal curve, has an approximately negative slope. For a pressure change from 0.0625 units to 20 units, the area under black curve is now greater than area under red curve. So adiabatic work > isothermal work. 20 p 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1 0 C isothermal E D adiabatic V Figure 15.8 p-V diagrams for a gas undergoing adiabatic and isothermal expansions from a state C.

first law of thermodynamics to each of the two processes shown in Figure 15.8. For expansion process CE (isothermal): ΔU = Q + (−W) since work done by a gas is the negative of work W done on a gas. There is no change in internal energy of the gas since there is no change in temperature. Hence, Q = W.

Heat flows to the gas from the heat reservoir since Q is positive. This makes sense intuitively since heat must be put into the gas for it to do some work. For expansion process CD (adiabatic), Q = 0 therefore: ΔU = Q + (−W) results in ΔU = −W. In other words, the internal energy of the gas undergoes a negative change. The internal energy (and hence the temperature) drops. Getting it right! We must be very careful when applying the equation ΔU = Q + W so as to get our signs right. In the first law of thermodynamics equation, Q represents heat flowing into a gas and W represents work done on the gas. Also, if a gas is expanding, the gas is doing positive work; negative of this work is being done on the gas, and vice-versa. Isobaric process 0 2 4 6 8 10 12 14 16 V 18 Figure 15.7 p-V graphs for a gas going through an isothermal process (red line) and an adiabatic process (black line) over a wide range of pressures. In an isobaric process, the pressure of the gas does not change. Figure 15.9 depicts an isobaric process from a state A to a state B. p / kPa 300 Figure 15.8 shows a gas expanding from state C to state D adiabatically, and from state C to state E isothermally. From an earlier discussion, the area under each curve represents positive work done by the gas, which translates to negative work done on the gas.

We can now apply the B A 200 100 1 2 3 4 5 V / m3 Figure 15.9 Isobaric expansion of a gas from state A to state B. Chapter 15 First law of thermodynamics Worked example 15.3: Isobaric expansion of a gas Q A A gas expands from state A to state B isobarically, as shown in Figure 15.9. (a) Is positive work being done on the gas or by the gas? (b) How much work is being done during the process? (c) What becomes of the internal energy of the gas? (a) Positive work is being done by the gas since the gas is expanding (the force of the gas and the expansion of the gas are in the same direction and hence the gas does positive work). (b) work done by gas = area under p-V curve = 200 × 103 Pa × (5 − 2) m3 = 6 × 105 J (c) Since work done by the gas is positive, then heat must flow into the gas in order to achieve the expansion at constant pressure. The internal energy of the gas will increase by 6 × 105 J (and therefore the temperature will increase as well). The molar specific heat capacity of a gas, Cv, at constant volume is defined as the quantity of heat, Qv, required to raise the temperature of 1 mole of a gas by 1 K at constant volume. For a temperature change, ΔT, the defining equation for Cv is Qv Cv = ...(15.5) nΔT which gives Qv = nCvΔT...............(15.6) The molar specific heat capacity of a gas, Cp, at constant pressure is defined as the quantity of heat, Qp, required to raise the temperature of 1 mole of a gas by 1 K at constant pressure.

For a temperature change, ΔT, the defining equation for Cp is Qp Cp = ...(15.7) nΔT which gives Isovolumetric process Qp = nCpΔT...............(15.8) In an isovolumetric process, the volume of the gas does not change. The work done on or by the gas in an isovolumetric process is zero (Figure 15.10). p / kPa 300 B 200 Heat required in isovolumetric processes Figure 15.11 shows n moles of gas initially at state A. The path AB represents an isovolumetric process. From the first law of thermodynamics, if Qv is the heat added at constant volume we get, for this process, ΔUv = Qv + Wv 100 C 1 2 3 4 5 V / m3 Figure 15.10 Gas undergoing isovolumetric process from state B to state C. Since ΔV = 0, then Wv = pΔV = 0. Therefore, ΔUv = Qv i.e. Qv = ΔUv...............(15.9) Molar specific heat capacities of a gas Calculations involving flows of heat into or out of a gas involve applications of specific heat capacities of the gas. These specific heat capacities are different in value, depending on the kind of process to which the gas is subjected. All the heat added to the gas at constant volume is used for raising the internal energy of the gas. p B ΔUv = Qv + 0 Molar specific heat capacities The molar specific heat capacity of a gas is the quantity of heat required to raise the temperature of 1 mole of gas by 1 K. The temperature of a gas can be raised by 1 K by several processes. We consider two processes, isovolumetric (constant volume) and isobaric (constant pressure). ITQ 6 Explain why the work done in the isovolumetric path, BC, is zero. A C ΔUp = Qp + (–Wp) V Figure 15.11 Isovolumetric (AB) and isobaric (AC) processes. Heat required in isobaric processes Figure 15.11 also shows the gas taken from state A to a state C is isobarically. Heat is added to the gas; the gas 231 232 Unit 1 Module 3 Thermal and mechanical properties of matter expands and does work. If Qp is the heat added at constant pressure, and Wp is the work done by the gas, applying the first law of thermodynamics, we get, for this process, ΔUp = Qp + (−Wp) i.e. Qp = ΔUp + Wp...............(15.10) Therefore, 3 ΔU = nRΔT 2 3 But since Qv = ΔU (equation 15.8), then Qv = nRΔT 2 Qv 3nRΔT Therefore, Cv = = nΔT 2nΔT which gives or Qp = ΔUp + pΔV...............(15.11) Cv = From equation 15.14 Comparison between Cp and Cv Combining equations 15.9 and 15.11, we get, Qp − Qv = (ΔUp + pΔV) − ΔUv For the same change in temperature, ΔT, in the two 3 processes, ΔUp = ΔUv, since U = RT (assuming an ideal, 2 monatomic gas at a temperature, T). Therefore, Qp − Qv = pΔV...............(15.12) This means that Qp > Qv. It follows that Cp > Cv (i.e. the molar specific heat capacity at constant pressure is greater than the molar specific heat capacity at constant volume) Q Q since Cp = p (equation 15.7) and Cv = v (equation nΔT nΔT 15.5). The molar specific heat capacity of a gas at constant pressure is greater than the molar specific heat capacity of the gas at constant volume. The difference between Cp and Cv is equal to the work done (pΔV) per mole per degree, as the gas at constant pressure expands. 3 R ....(15.15) 2 Qp = Qv + nRΔT Incorporating the definitions of Cp (equation 15.7) and Cv (equation 15.5), we get nCpΔT = nCvΔT + nRΔT which gives Cp = Cv + R or Cp − Cv = R...............(15.16) 3 Substituting Cv = R (equation 15.15) into equation 2 15.16, we get Cp = 3 R+R 2 Cp = 5 R 2 i.e. ....(15.17) For an ideal, diatomic gas, it can be shown that: Relationship between Cp and Cv For an ideal, monatomic gas, pV = nRT...............(14.9) Therefore V= nRT p and therefore ΔV = nRΔT ....(15.13) p Substituting equation 15.12 into equation 15.13, we get Qp − Qv = pnRΔT ....(15.14) p which gives, Qp − Qv = nRΔT...............(15.14) The internal energy, U, of an ideal, monatomic gas is given by U (= Ek) = 3 nRT ....(14.14) 2 Cp = 7 5 R, Cv = R, Cp − Cv = R 2 2 Efficiency of a cycle Let us consider a fixed mass of gas at an initial state p1, V1 and T1, being taken through processes that change its states and then return the gas to its original state. This path taken by the gas from its original state to some other state(s) and back to its original state is called a cycle. Cycles play an important role in heat engines (where heat supplied to a mixture of gases enables the gases to do work, as in a gasoline-driven engine) and heat pumps (where work is done to transfer heat out of a cool environment into a warmer environment, as in a refrigerator). It can be shown that no such cycles are 100% efficient; that is, some heat always escapes from the system to the environment. In the case of fossil fuels as energy sources, inefficiencies (heat escaping from systems) contribute not only to global warming but also to a waste of this precious energy resource which is non-renewable. Chapter 15 First law of thermodynamics Worked example 15.4: Calculations involving molar specific of heat capacities of a monatomic gas Q A A certain mass of ideal, monatomic gas is initially at a state A, as shown in Figure 15.12.

Using the data given in the figure, calculate: (a) the number, n, of moles of the gas (b) the temperature, TB, at state B (c) the temperature, TC, at state C (d) the heat input to the gas during the process AB (e) the heat input to the gas during the process BC. p / kPa 400 Q p = nCpΔT C B 300 Q v = nCv ΔT 200 A 100 (a) Applying the equation of state pAVA = nRTA we get pAVA n= RTA 1.0 × 105 N m−2 × 0.0060 m3 8.31 J mol−1 K−1 × 273 K ≈ 0.26 mol (b) Applying the equation, pAVA pBVB = TA TB (see equations 14.7 and 14.9) we get pBVB TB = × TA pA VA = 4 × 105 N m−2 × 0.0060 m3 × 273 K 1.0 × 105 N m−2 × 0.0060 m3 ≈ 1092 K (c) Applying the equation, pAVA pCVC = TA TC we get pCVC TC = × TA pA VA = 0.0020 0.0040 (TA = 273 K) 0.0060 V / m3 Figure 15.12 A gas undergoing an isovolumetric process followed by an isobaric process. (d) The heat input, QAB, during the isovolumetric process, AB, is given by Qv = nCvΔT...............(15.6) 3 = 0.26 mol × R × (1092 − 273) K 2 3 = 0.26 × × 8.31 J mol−1 K−1 × (1092 − 273) K 2 = 2654 J ≈ 2700 J (e) The heat input, QBC, during the isobaric process, BC, is given by Qp = nCpΔT...............(15.8) 5 = 0.26 mol × R × (364 − 1092) K 2 5 = 0.26 mol × × 8.31 J mol−1 K−1 × (364 − 1092) K 2 = −2359 J ≈ −2400 J The gas actually loses heat in the process, BC.

4 × 105 N m−2 × 0.0020 m3 × 273 K 1.0 × 105 N m−2 × 0.0060 m3 ≈ 364 K = For simplicity, we consider a cycle consisting of isobaric and isovolumetric processes only (Figure 15.13). Efficiency, e, of a such a cycle is defined as: work output e = heat input = Wnet Qin ....(15.18) where Wnet work output is the work done by the gas. We first discuss how to determine the work done by a gas in p / kPa 400 C D B A 300 200 100 1 2 3 V / m3 Figure 15.13 A p-V cycle involving isobaric and isovolumetric processes. 233 234 Unit 1 Module 3 Thermal and mechanical properties of matter one cycle. Then we show how heat input is calculated based on temperature changes within the cycle. These two quantities, used in equation 15.18, enable us to determine efficiency of the cycle. Figure 15.13 shows a p-V diagram illustrating 3 moles of an ideal gas initially at a state A and going through a cycle depicted by ABCDA. Temperature changes in a cycle In order to calculate heat input in a cycle, we need to know the temperature changes during each process occurring within the cycle (Figure 15.14). We therefore need to know the temperature at the beginning of each process as well as that at the end of each process. p / kPa Work done by gas during one cycle 400 ■ Along AB 300 WAB = pΔV = 2.0 × 10 N m (1 − 3) m = −4.0 × 105 J 5 −2 Getting it right! To calculate a change (e.g. ΔV), we always use change = final value − initial value ■ Along BC The net work done by the gas: WBC = pΔV = 0 J, since ΔV = 0 The net work done by the gas: WCD = 4.0 × 105 N m−2 (3 − 1) m3 = 8.0 × 105 J Since the volume is increasing, the direction of the increase is the same as the direction of the gas pressure. The work done by the gas is therefore positive, in agreement with the sign of our calculated value. ■ Along DA The net work done by the gas: WDA = pΔV 1 2 3 V / m3 Figure 15.14 p-V cycle showing directions of heat flows. Starting with temperature TA at state A, we determine the temperature TB at state B as follows: pAVA pBVB = TA TB (refer to equations 14.7 and 14.9) therefore pV TB = B B × TA pAVA = 2 × 105 N m−2 × 1 m3 × TA 2 × 105 N m−2 × 3 m3 = 1 T 3 A Similarly, it can be shown that TC = 2 T 3 A and TD = 2TA Total heat input in a cycle ■ Along AB ■ For the entire cycle ABCDA The net work done by the gas, Wnet, is given by Wnet = −4.0 × 10 J + 8.0 × 10 J A We now discuss how total heat input is calculated, based on temperature changes and specific heat capacities. = 0 J, since ΔV = 0 5 Q out 100 ■ Along CD D Q out B 3 Since the volume is decreasing, positive work is being done on the gas. The work done by the gas is therefore negative, in agreement with the sign of our calculated value. Q in Q in 200 C 5 = 4.0 × 105 J This is equal to the area enclosed by the p-V cycle. The net work done by a fixed mass of gas in one p-V cycle is equal to the area enclosed by the p-V graph for that cycle. ITQ 7 Will the initial temperature and final temperature at state A be the same or different?
Explain.
Qp = nCpΔT 1 = nCp(TA − TA) 3 5 2 = n × R × (− TA) 2 3 5 = − nRTA 3 The negative value of Qp indicates that heat is flowing out of the gas into the surroundings. This direction Chapter 15 First law of thermodynamics of heat flow is consistent with the fact that since the pressure is constant but the volume is decreasing, the gas temperature must drop so that V/T remains constant, since pV/T = constant. A drop in temperature means that heat must be leaving the gas. The heat input in the cycle shown in Figure 15.14 therefore occurs only along BC and CD. The total heat input, Qin, is therefore given by 1 10 Qin = nRTA + nRTA 2 3 ■ Along BC 2 1 = nCv(TA − TA) 3 3 3 1 = n × R × (TA) 2 3 1 = nRTA 2 The positive value of Qv indicates that heat is flowing into the gas from the surroundings.
This direction, again, can be accounted for by applying the combined gas equation for a gas at constant volume, where the pressure is increasing. There will be a rise in temperature, showing that heat is being added to the gas. ■ Along CD Qp = nCpΔT 2 = nCp(2TA − TA) 3 5 4 = n × R × (TA) 2 3 10 = nRTA 3 The positive value of Qp indicates, again, that heat is flowing into the gas from the surroundings. As before, this direction can be accounted for by applying the combined gas equation for a gas at constant pressure, where the volume is increasing. There will be a rise in temperature, showing that heat is being added to the gas.
■ Along DA = Qv = nCvΔT Qv = nCvΔT = nCv(TA − 2TA) 3 = n × R × (−TA) 2 3 = − nRTA 2 The negative value of Qv indicates that heat is flowing out of the gas into the surroundings. This direction, again, can be accounted for by applying the combined gas equation for a gas at constant volume, where the pressure is decreasing. There will be a drop in temperature, showing that heat is leaving the gas. 23 nRTA 6 Getting it right! The heat input in a cycle includes only the heat put into the system. It does not include heat leaving the system. Net work done by a gas in a p-V cycle, however, does include work done, positive and negative, at each part of the cycle. Efficiency of a cycle Worked example 15.5 shows how cycle efficiency is calculated, based on the cycle shown in Figure 15.13. Worked example 15.5: Calculation of cycle efficiency Q Calculate the efficiency of the cycle depicted in Figure 15.13, which involves an ideal, monatomic gas. Assume an initial temperature of 273 K. A At A: ■ P = 2 × 105 Pa ■ V = 3 m3 If T = 273 K then PV n= RT 2 × 105 × 3 = 8.31 × 273 = 264 moles We apply equation 15.18, using the values of heat input and net work from the preceding two sections. 4 × 105 J × 100% e = 23 nRTA 6 = 4 × 105 J × 6 × 100% 23 × 264 mol × 8.31 J mol−1 K−1 × 273 K ≈ 17% 235 236 Unit 1 Module 3 Thermal and mechanical properties of matter Summary ✓ According to the first law of thermodynamics ✓ The efficiency, e, of a cycle is defined by the ΔU = Q + W...............(15.1) equation e= where ΔU is the change in internal energy of a system, Q is the heat flowing into the system and W is the work done on the system. ✓ The conventions used in heat and work processes in equation 15.1 are as follows: ■■ heat into a system is designated positive ■■ heat out of a system is designated negative ■■ work done on a system is the negative of work done by a system. ✓ Both heat and work involve energy transfer.
✓ The magnitude of work, ΔW, done as a gas initially at volume V and pressure p undergoes a volume change, ΔV, is given by ΔW = pΔV...............(15.2) Graphically, the work done is equal to the area under the p-V graph for the gas. ✓ A system is defined by state variables. For an ideal gas, the state variables are: n = the number of moles of the gas m = the mass of gas being considered p = the pressure of the gas e= net work done by the gas in one cycle heat input Wnet Qin ....(15.18) ✓ The molar specific heat capacity, Cv, of a gas at constant volume is defined by the equation Qv Cv = ...(15.5) nΔT where Qv is the heat supplied to n moles of gas at constant volume and resulting in a temperature rise, ΔT. ✓ The molar specific heat capacity, Cp, of a gas at constant pressure is defined by the equation Q Cp = p ....(15.7) nΔT where Qp is the heat supplied to n moles of gas at constant pressure and resulting in a temperature rise, ΔT. ✓ For an ideal, monatomic gas, it can be shown that: Cp = 5 R 2 ....(15.17) Cv = 3 R 2 ....(15.15) V = the volume of the gas Cp − Cv = R...............(15.16) T = the kelvin temperature of the gas ✓ For an ideal, diatomic gas, it can be shown that: U = the internal energy of the gas. ✓ In a p-V cycle, the net work done by a gas is equal to the area enclosed by the graph. Cp = 7 R 2 Cv = 5 R 2 Cp − Cv = R Review questions First law of thermodynamics 1 (a) State the first law of thermodynamics in equation form, with respect to a fixed mass of gas. Explain all the symbols used.
(b) State the sign convention used in the above equation. 2 (a) In describing a gas, what is meant by state variables? (b) Should heat and work be considered as state variables for a gas? Explain your answer.
(c) 1.5 moles of an ideal gas, initially at a pressure of 2.0 × 105 Pa and a temperature of −20 °C, occupy a volume V1.
Using the equation of state for an ideal gas, determine the initial volume, V1, of the gas. Chapter 15 First law of thermodynamics 3 2000 J of heat are removed from 0.60 mole of helium gas initially at 20 °C, and 1200 J of work are done on the gas during the process, resulting in a change in internal energy of the gas. (a) Calculate the change in internal energy of the gas. (b) Write an equation for change in energy, ΔU, of a monatomic gas, in terms of change in temperature, ΔT. Hence, determine the final temperature of the gas. (c) Using information from (b) above, determine the root mean square speed, vrms, of the molecules of the gas at its final temperature. 4 In a diesel engine, 2500 J of work are done on an air/fuel gaseous mixture in compressing the mixture almost adiabatically. (a) How much heat flows into or out of the gaseous mixture? (b) By how much does the internal energy of the mixture change? Is it a decrease or an increase? Explain your answer.
(c) If the mixture had consisted only of 2 moles of an ideal, monatomic gas, estimate the change in temperature following the compression.
(In diesel engines, the rise in temperature produced by the compression is enough to cause the air/ fuel mixture to ignite, releasing energy to run the engine.) 5 (a) Define internal energy of a monatomic, ideal gas. (b) Show that a change in internal energy of a monatomic ideal gas gives rise to a change in temperature of the gas. (c) Calculate the change in root mean square speed, vrms, for helium gas molecules if the gas temperature changes from 0 °C to 100 °C. 6 Heat Q (= mL), supplied to water of volume V1 boils water into steam of volume V2 at 100 °C. The boiling takes place at atmospheric pressure, P0. (a) Using the first law of thermodynamics, deduce an equation for the change in internal energy, ΔU, of the water in terms of Q, P0, V1 and V2. (Do not attempt to calculate ΔU.) (b) Substituting actual values from experimental results into the equation in (a) yields ΔU > 0. Does this violate the principle that ΔU = 0 for a constant temperature process? Explain your answer.
The first law and thermodynamics processes 7 Figure 15.15 shows a gas, initially at pressure p1 and volume V1, contained in a cylinder of cross-section area A. The gas expands at constant pressure such that the piston moves a distance Δx. ΔV Δx freely moveable piston gas (p1, V1) Figure 15.15 Gas undergoing an expansion in a cylinder.
(a) Show that the work done by the gas during the expansion is given by pΔV, where ΔV is the change in volume of the gas. (b) Draw a p-V graph to show the expansion of the gas. (c) What is the name given to an expansion that takes place without a change in pressure? 8 (a) Explain why the work done by a fixed mass of gas as it expands from a state X to a state Y is equal to the area under the p-V graph between X and Y for the gas. (b) A fixed mass of gas is compressed from a state A to a state B, as shown by Figure 15.16. Determine approximately the work done by the gas. p / Pa 1.4 x 10 5 B A 0.6 x 10 5 0.02 0.06 V / m3 Figure 15.16 p-V diagram of a gas undergoing compression. 9 (a) Explain the meaning of each of the following terms that describe thermodynamics processes: (i) isothermal (ii) adiabatic (iii) isobaric (iv) isovolumetric. 237 238 Unit 1 Module 3 Thermal and mechanical properties of matter (b) Sketch and label accordingly, using the same p-V axes, graphs to illustrate a fixed mass of gas at state A undergoing each of the following processes: W isothermal process X adiabatic process Y isobaric process Z isovolumetric process. (c) In which of the above processes is there: (i) no change in internal energy of a gas? (ii) no work done? (iii) no heat entering or leaving the system? Molar specific heat capacities of a gas 10 (a) Define: (i) the mole (ii) molar specific heat capacity of a gas, Cp, at constant pressure (iii) molar specific heat capacity of a gas, Cv, at constant volume. (b) Using the first law of thermodynamics, explain why Cp > Cv. 3 (c) Prove that for an ideal, monatomic gas, Cv = R. 2 11 A certain mass of an ideal, diatomic gas at state A (TA = 273 K) is taken to a state B as shown in Figure 15.17. Efficiency of a cycle 13 Figure 15.18 shows a fixed mass of gas being taken through p-V cycle, ABCA. p C p2 p1 A B V2 V1 V Figure 15.18 p-V diagram of gas taken through a cycle ABCA. (a) Determine expressions for the work done by the gas in each of the processes AB, BC and CA. (b) Show that the net work done by the gas during the cycle is equal to the area of the triangle ABC. 14 2 moles of an ideal, monatomic gas at temperature 27 °C and at state A are taken to a state B, as shown in Figure 15.19. p / kPa 500 400 A 300 200 p / kPa 400 B 100 B 300 0.050 V / m3 200 100 A 0.010 0.020 0.030 V / m3 Figure 15.17 p-V diagram of gas taken from state A to state B. (a) Determine the number of moles of the gas. (b) Determine the temperature, TB. (c) Using the concept of molar specific heat capacity, determine the change in energy, ΔU, as the gas is moved through AB. (d) Determine ΔU using another approach. 12 Given that the molar specific heat capacity at constant volume, Cv, for helium (molar mass = 4.0 g) is 12.5 J mol−1 K−1, show that the specific heat capacity for helium at constant volume, cv, is 3.13 J g−1 K−1 (or 3130 J kg−1 K−1). Figure 15.19 p-V diagram of gas taken from state A to state B. (b) Determine the change in internal energy as the gas moves from state A to state B. (c) Determine the heat flowing into (or out of) the gas along AB. 15 (a) Define efficiency of a cycle.
(b) In a certain 4-stage cycle, the work done by the gas and the heat flowing into the gas are as follows: Stage Work done by gas / J Heat flowing into gas / J AB −2.0 × 10 −4.99 × 105 BC 0 CD 4.0 × 10 DA 0 5 2.99 × 105 5 9.88 × 105 −5.99 × 105 Determine the efficiency of the cycle. Chapter 15 First law of thermodynamics 16 A fixed mass of n moles of a monatomic gas is carried through a p-V cycle as shown in Figure 15.20 starting from state A. Assume the temperature, TA, at state A is 27 °C. p / Pa a b a b plunger piston plunger piston thermometer clamps thermometer clamps beaker C 2.0 x 10 5 beaker water bath ga syr ga sy sti water bath sti air 1.0 x 10 5 air barrel A B barrel nozzle nozzle seal seal 0.0030 a 0.0090 b V / m3 Figure 15.20 p-V diagram of gas taken through a cycle ABCA. plunger piston (a) Determine the number of moles of the gas. (b) Determine: (i) the temperature, TB, at state B (ii) the heat flowing into (or out of) the gas along air AB barrelalong (iii) the heat flowing into (or out of) the gas BC nozzle (iv) the heat flowing into (or out of) the gas along seal CA. (c) Determine the work done by the gas in the p-V cycle, ABCA. (d) Determine the efficiency of the cycle ABCA. Practical activities and challenges 1 Investigating the first law of thermodynamics using air Note: air is not a monatomic gas but is largely a mixture of two diatomic gases, oxygen and nitrogen. 239 Other gases, such as carbon dioxide, form only a tiny percentage. In this investigation, therefore, we will need to use an 'average' value for Cp for this air mixture of essentially diatomic gases (≈ 29.2 J mol−1 K−1). thermometer clamps beaker gas syringe water bath stirrer retort stand Figure 15.21 Suggested apparatus for investigating the first law of thermodynamics using air: (a) the sealed gas syringe; (b) the suggested experimental arrangement. Use the apparatus shown in Figure 15.21, or devise an appropriate apparatus arrangement, to investigate the first law of thermodynamics. A suggested procedure in outline form is as follows. (a) Seal a 20 cm3 gas syringe at the tip with a suitable glue (Figure 15.21a). Slacken the nozzle and pull the plunger to allow a suitable amount of air to enter the barrel. Tighten the nozzle so as to trap air in the barrel. (b) Test the sealed gas syringe for leaks. This is done by pressing and then releasing the plunger. The plunger should return to its original position rather quickly when released if there is a proper seal and if there is little friction between the plunger and the barrel. (c) Set up the apparatus as shown in Figure 15.21b. (d) Having gently stirred the water in the water bath, take the initial temperature, T1, of the gas. An initial temperature of about 10 °C is suggested (so as to be able to obtain a relatively large temperature change). 240 Unit 1 Module 3 Thermal and mechanical properties of matter (e) Heat the water bath, allowing the trapped air to expand at atmospheric pressure (assumed constant) and take the temperature, T2, near the boiling point of water. Record the change in volume, ΔV. Experimental determination of ΔU According to the first law of thermodynamics, ΔU = Q − W (work is being done by, rather than on, the gas, hence, −W). Therefore ΔU = Q − P0ΔV where P0 is the atmospheric pressure. We can use this latter equation to determine ΔU experimentally. There some challenges, however ... ■ How accurately would you be able to find ΔV? ■ How would you obtain the value of P0? ■ How would you determine the value of Q? (Hint: Q = nCpΔT, but you will have to figure out a way of determining n and, since air is largely nitrogen and oxygen, you may have to use an 'average' value for Cp for air, using Cp values for nitrogen and oxygen, ≈ 29.2 J mol−1 K−1.) ■ Both the cylindrical barrel and the contained air expand on heating. What effect will this have on the measured value of ΔV? Analysis and discussion of results (a) Make an estimate of ΔU based on the equation U = 5 nRT for a diatomic gas. 2 (b) Estimate the experimental value of ΔU using the equation, ΔU = Q − P0ΔV. (c) Comment on the extent to which the first law of thermodynamics has been verified by your investigation. 2 Investigating the human body as a thermodynamic system There is a net flow of heat from the human body to the environment. The heat is provided from internal energy obtained from food through tissue respiration. (a) Make an estimate of the net flow of heat radiated to the environment during 1 hour from the human body at rest.
(Hint: you may find equation 13.5 helpful, but some challenges will include estimating e and the surface area, A, of the human body.) Hence, estimate the heat radiated from the body at rest in 1 day.
(b) Using the first law of thermodynamics, and the value of the net heat flow by radiation, estimate the change in internal energy of the body for 1 day, assuming the body is at rest. Give this estimate in joules, calories and kilocalories (1 cal = 4.186 J; 1 kcal = 4186 J). (c) Compare the result obtained in step (b) above with the recommended daily intake of food Calories for the average adult human body at rest, bearing in mind other methods of transfer of heat between the body and the environment. (Note: food energy values are usually specified in Calories, written with a capital C. 1 Calorie = 1000 cal = 1 kcal.) (d) Does the body at rest really do no mechanical work? Explain your answer.
Answers to ITQs 1 The work–energy principle (see Chapter 4). 2 ΔU + ΔKE + ΔPE = Q + W 3 No. Although Q and W are variables, they do not describe the state of a system. Work and heat are processes involved in changing the state of a system. 4 The direction of the gas pressure (and hence the force of the gas) on the piston is the same as the direction the piston moves as the gas expands. Hence, work = force × distance moved by the force, is positive. Hence work done by gas is positive. 5 Heat put into a gas results in work being done by the gas. 6 Work done = pΔV, i.e. area under the p-V graph.
Since ΔV = 0, the work done = 0 J. 7 They would be the same since pV = nRT and p, V, n and R are the same in both cases. Answers to Review questions 3 (a) −800 J (b) −87 °C (c) 1080 m s−1 5 (c) 221 m s−1 11 (a) 0.44 mol (b) 1090 K (c) 4.5 kJ 15 (b) 15% 241 Chapter 16 Mechanical properties of matter Learning objectives ■ Describe a simple kinetic model of matter. ■ Account for the structure and behaviour of solids, liquids and gases using the simple kinetic model. ■ Describe the structures of crystalline and non-crystalline solids, including metals, polymers and glasses. ■ Relate densities of materials to the simple kinetic model. ■ Derive and apply the equation Δp = ρgΔh (for pressure difference in a liquid). ■ Discuss the load–extension behaviour of springs and wires. ■ Distinguish between elastic and inelastic deformation. ■ Sketch force–extension graphs for ductile, brittle and polymeric materials. ■ Define and use the Young's modulus of elasticity, E = stress strain ■ Describe experiments to determine the Young's modulus of a metal in the form of a wire. ■ Determine strain energy in a deformed material from a force–extension graph. ■ Discuss the importance of elasticity in materials and human-made structures. A simple kinetic model of matter In Chapter 14 we outlined seven basic assumptions of the kinetic theory of an ideal gas. In this chapter, we focus on three aspects of this theory as they relate to mechanical properties of matter. In addition, the model explains structural differences in, and properties of, crystalline and non-crystalline solids, such as metals, polymers and glasses, as well as elastic behaviours in solids, in particular. ■ Matter is made up of numerous tiny particles The graph in Figure 16.1 shows that for separation distances x < x0, the force between molecules is positive (repulsive) and increases sharply with decreasing x.
For x > x0, the force is increasingly negative (attractive) up to x2.
Beyond x2, the intermolecular force is still attractive but approaches zero fairly rapidly with increasing x. In the vicinity of x0, the force, F, varies approximately linearly with x. (particulate nature). In most cases, we shall be considering particles in the form of molecules; in some cases, however, we shall be referring to atoms and ions as these particles. ■ There are electromagnetic forces between particles (e.g. the existence of intermolecular forces). ■ Particles of matter are always moving (kinetic An example of how important the structure of materials is can be seen in the Boeing 787 Dreamliner.
This plane, which made its debut in 2009, uses strong materials that are lightweight: aluminium, titanium and carbon fibre behaviour). This simple kinetic model can be used to account for the common states of matter, namely solids, liquids and gases. a b F motion of molecule motion of molecule x0 x2 x1 x0 x3 x Figure 16.1 Force between two adjacent molecules. (a) The 'spring model' of the force between molecules of mean separation distance, x0. (b) Graph of force, F, against mean separation distance, x, between two molecules that are very close to each other. 242 Unit 1 Module 3 Thermal and mechanical properties of matter reinforced polymers.
The strength-to-weight properties achieved made the Dreamliner 50% lighter than conventional aircraft of its size and therefore about 20% more economical in fuel usage. PE repulsion attraction E3 Solids, liquids and gases Solids A solid has a fixed shape. How does the simple kinetic model explain this? The intermolecular forces of attraction and repulsion must be so strong that each molecule stays in its position relative to the rest. Since the molecules are in motion, according to the kinetic model, their motion must be restricted to vibrations about a mean separation distance, x0. The kinetic model can also explain the solid state of matter in terms of energy. A system consisting of two molecules would have both potential energy (PE) and vibrational kinetic energy (KE) since the molecules influence each other by 'spring-like' forces.
Note: a mass–spring system oscillates with simple harmonic motion. Each mass has kinetic energy associated with its motion; the spring has potential energy due to its extension. The total energy of the system is the sum of the potential and kinetic energies (see Chapter 7). Figure 16.2 shows how the potential energy of a two-molecule system varies with separation distance, x. For a total energy (PE + KE), such as E0, the system of molecules is bound since the total energy is negative. Hence, the particles do not move away from each other. The mean separation distance is x0. Liquids A liquid takes the shape of the container into which it is put. A liquid also evaporates. How does the kinetic model account for these behaviours? Figure 16.2 shows that as total energy increases, to E2 for example, the mean separation distance between the vibrating molecules increases from x0 to somewhere between x0 and x2. This is because the force of attraction departs from linearity. The mean force of intermolecular attraction decreases. However, the liquid remains as ITQ 1 Refer to Figure 16.1. Suggest, in terms of x, the range over which solids are likely to exist. Briefly explain your answer. ITQ 2 Referring to Figure 16.2, for which of the three values of total energy marked (E0, E1, E2) is the system most likely to be liquid? Briefly explain your answer. x0 r (separation) x E2 E1 E0 Figure 16.2 Potential energy (PE) versus separation distance, x, between two molecules. The black dots show how the mean separation increases from x0 with increasing energy. a bulk since the total energy is negative, resulting in a bound system. Liquids easily take the shape of any container into which they are poured since, with the reduced intermolecular forces, molecules can move about somewhat relative to each other. A few molecules near the surface may gain enough kinetic energy due to collisions from other molecules beneath to escape from the surface of the liquid (a phenomenon called evaporation). Gas (or vapour) A gas (or vapour) completely 'fills' any container into which it is put. When a liquid is heated sufficiently, the total energy of the system of molecules increases. The mean separation distance between molecules, for energies such as E3, becomes so large that the mean forces of attraction are almost negligible. The molecules move about independently and therefore bounce continually against the inner walls of the container in which they are put. This is what we describe as 'filling' the container. In terms of energy (Figure 16.2), the molecules possess both potential energy and vibrational kinetic energy when they are bound.
They can escape from the bound state when their average kinetic energy (positive) exceeds the magnitude of their average potential energy (negative) and so their total energy (KE + PE) becomes positive. Note: other states of matter besides solid, liquid and gas do exist, but require more complex models to account for them. A plasma state occurs in gases at very high temperatures or voltages. Electrons and nuclei of atoms move about independently in this state. At very low temperatures (near absolute zero), a Bose–Einstein condensate state is believed to exist. Quantum mechanics is required to account for such states. Chapter 16 Mechanical properties of matter Density of materials Density and state of materials Low density is often a desirable property in engineering applications. Aircraft and modern energy-efficient cars must be made of lightweight, yet strong, materials. Density is defined as mass per unit volume of a substance. In equation form, ρ= M V ....(16.1) where ρ represents density, M is mass and V is the volume of a material. The density of materials can be explained by the simple molecular model of matter. Table 16.1 shows densities of selected materials.
Table 16.1 Densities of selected materials at 0 °C and standard pressure (1 atm) (unless otherwise stated) Material Density / kg m−3 pure ice 0.92 × 103 pure water (at 4 °C) 1.00 × 103 sea water (average) 1.03 × 103 steam (at 100 °C) 0.60 air 1.29 aluminium 2.70 × 103 carbon (graphite; diamond) 2.2 × 103; 3.5 × 103 copper 8.9 × 103 gold 19.3 × 103 iron 7.8 × 103 lead 11.3 × 103 steel 7.8 × 103 platinum 2.1 × 103 polythene (solid) = 920 (variable) rubber = 1100 (variable) wax (solid) = 900 (variable) a As a solid, when heated, usually expands a little to form a liquid. This is because the separation distance of molecules increases slightly in a solid to liquid conversion. Since the mass of a substance remains constant upon heating, the small increases in volume when a solid turns to a liquid results in the density of a liquid being less that the density of the solid from which it was formed. Similarly, the large increase in mean separation of the particles between a liquid and its corresponding gas means that gases would generally have a considerably lower density than liquids or solids. Density and arrangement of molecules Figure 16.3a shows that water expands rather than contracts as it is cooled below 4 °C. This occurs because water molecules, originally attached weakly to each other, begin to reorganize themselves into larger units which occupy a disproportionately large space (see Figure 16.3b). This results in the anomalous situation of the density of water decreasing rather than increasing with decreasing temperature. Polymers (e.g. plastics) generally consist of giant molecules each composed of hundreds or thousands of atoms. These giant molecules usually form tangled masses and therefore occupy much space. This is one reason why solid polymers usually have low densities as compared with other types of solids. Density and relative atomic mass The relative atomic mass of an element is the number of times the mass of one atom of the element (on the average) is as great as the mass of 1/12 that of carbon-12. b Volume ITQ 3 How does the graph in Figure 16.3a imply that the maximum density of water occurs at 4 °C? ice ice and water water molecule water maximum density −5 0 5 regular ice structure 10 15 Temperature / °C Figure 16.3 The anomalous expansion of water. (a) Variation of volume of water with temperature in the vicinity of 4 °C. (b) Model showing water molecules organizing themselves into large units during formation of ice. 243 244 Unit 1 Module 3 Thermal and mechanical properties of matter The atoms in a solid element are usually closely packed. Hence elements of large relative atomic mass (e.g. lead = 207) tend to have a much greater density than those of small relative atomic mass (e.g. aluminium = 27). However, the packing factor of atoms in a crystalline arrangement (e.g. in metals) would also affect the density of a material. Worked example 16.1: Pressure on a submarine Q Estimate the pressure, p, that a submarine (Figure 16.4) must be able to withstand at a depth 100 m below the surface of the sea (mean density of ocean 1.025 × 103 kg m−3). The packing factor is defined as the actual volume of atoms in a repeating cell of a crystal per unit volume of the cell. A large packing factor therefore tends also to contribute towards a large density. Various types of packing are discussed later, in the section on crystalline solids (especially in Figure 16.11, page 247). Δh We see, therefore, that relative atomic mass (and hence molecular mass), packing factor of atoms and molecular arrangement all contribute to the overall density of a material. Effect of density on fluid pressure In Chapter 3, we showed that the pressure, p1, due to a head, h1, of fluid is given by p1 = h1ρg, for a fluid of density ρ (equation 3.16). The pressure, p2, due to a head, h2, is similarly given by p2 = h2ρg. It follows that the pressure difference, Δp, due to a head, h2 − h1, is given by Figure 16.4 A submarine at a depth, Δh, below the ocean surface. A ■ atmospheric pressure, p0 = 1.01 × 105 Pa ■ depth of sea water, Δh = 100 m ■ mean density of sea water, ρ = 1.025 × 103 kg m−3 p = p0 + Δhρg = 1.01 × 105 N m−2 + (100 m × 1025 kg m−3 × 9.80 N kg−1) = 11 × 105 N m−2 = 1100 kPa Δp = (h2 − h1)ρg Crystalline solids or Δp = Δhρg...............(16.2) where Δh is the difference between two levels in a liquid. At the surface of the Earth, the pressure, p0, due to the height of the atmosphere above the Earth is approximately 1.01 × 105 N m−2 (or 101 kPa, since 1 Pa = 1 N m−2).
Crystalline and amorphous solids The arrangement of atoms and molecules and the bonding forces among these give rise to classes of solid materials with special properties that are particularly useful to engineers.
An understanding of these arrangements and forces has led to processes and modifications during manufacture which result in materials of desirable, and even exceptional, properties. Properties such as hardness, tensile strength, elasticity, malleability, ductility, plasticity and even chemical unreactivity, are sought after in various engineering applications.
We consider two broad classes of solids: crystalline and amorphous. Figure 16.5 Diamond, a crystalline solid. Crystalline solids are characterized by flat, smooth surfaces and regular shapes (Figure 16.5). These properties are due to unit cell structures made up of atoms, repeated in a regular patterns based on cubes, hexagons and other simple regular shapes. Many substances in nature, such as metals, might not give the appearance of being crystalline. However, their microstructures are definitely crystalline, and their properties are related to the unit cell type of which each is composed. Each unit cell contains the basic arrangement of atoms we identify as the building block of the crystal.
Chapter 16 Mechanical properties of matter amount of energy is typically supplied by high temperatures (801 °C in the case of sodium chloride) to change the solid to liquid, i.e. to give the ions some mobility. 801 °C is the melting point of sodium chloride. When an atom (originally neutral) loses one or more electrons, the result is an ion that has an excess of positive charge. A sodium ion has an excess of one positive charge. When an atom (originally neutral) gains one or more electrons, the result is an ion that has an excess of negative charge. A chlorine ion has an excess of one negative charge. Figure 16.6 Models showing the arrangement of sodium and chlorine ions in solid sodium chloride. There is a strong electrostatic bonding force between a pair of oppositely charged ions (Figure 16.7). Ionic solids A good example of a crystalline solid is sodium chloride (better known as common salt). If you look at a crystal of sodium chloride you will see it has a regular shape, smooth faces and straight edges. The crystal consists of sodium ions (Na+) and chlorine ions (Cl−) arranged alternately. The models in Figure 16.6 show the arrangement of ions of sodium and chlorine in a salt crystal.
The layers in the arrangement of ions form planes, causing the crystal to have flat surfaces and straight edges. Between a pair of adjacent planes is a space called a plane of cleavage. A crystal, subjected to a large force, e.g. from the blow of a hammer, would tend to break along cleavage planes, resulting in the fragments also having smooth surfaces and straight edges. An ionic bond is formed between each pair of Na+ and Cl− ions. This bond is very strong since electrostatic forces between charged particles are very strong. Ionic bonds are so strong that it takes about 7.8 eV of energy to separate the sodium ions from the chlorine ions in an ion-pair. (The electron volt, eV, as a unit of energy is discussed in Chapter 25. 1 eV is equivalent to 1.6 × 10−19 J.) This is outer electron b − − − − − − − − − − − − sodium ion (Na+) 11 protons 10 electrons − − − − − − − − − − chlorine ion (Cl) 17 protons 18 electrons Figure 16.7 The formation of an ionic bond between a pair of sodium and chlorine ions. – A sodium atom easily loses one electron since its outermost shell has only one electron and the next shell is full with 8 electrons. A shell full of 8 electrons is a stable state. Atoms try to achieve stable states, i.e. to have shells full of electrons by gaining or losing electrons. Properties that are typical of ionic solids, as exemplified by sodium chloride, are: ■ they have high melting points, since the ionic bonding forces are very large ■ they are very hard – due to the strong bonding forces, it is difficult to move particles relative to each other ■ they are not malleable (i.e. cannot be beaten into sheets) but rather fracture along cleavage planes when subjected to large mechanical forces ■ they are generally transparent to visible light because electrons are so tightly bound around each ion that the energy of visible radiation cannot be absorbed to move electrons from one energy level to the next ■ they absorb infrared radiation readily by resonance since the frequency of infrared radiation closely matches the natural frequencies of vibrations of the ions. Covalent solids The unit cell of diamond consists only of carbon atoms. The bonding between carbon atoms is covalent – that is, is brought about by sharing of electrons, as opposed to gaining and losing electrons. The outermost shell of a carbon atom consists of 4 electrons. A stable state consists of 8 electrons in the outermost shell. By sharing pairs ITQ 4 The outermost shell of a chlorine atom has 7 electrons. Explain how such an atom becomes an ion and achieves stability? ITQ 5 Why are solid ionic compounds generally poor conductors of electricity but molten ionic compounds are electrical conductors? 245 246 Unit 1 Module 3 Thermal and mechanical properties of matter carbon atom achieves the stable state of 8 electrons in its outermost shell of electrons with four adjacent atoms (Figure 16.8), a carbon atom achieves the stable state of 8 electrons in its outermost shell. A covalent bond is formed by sharing of electrons in pairs by adjacent atoms. The cohesive strength in a covalent bond is relatively large. Between a pair of carbon atoms, the cohesive bonding energy is of the order of 7.4 eV. Figure 16.9a shows a unit cell of diamond. Carbon is an extremely hard material. ■ The bonds between the atoms are strong. ■ The atoms are not arranged in sheets with cleavage planes between them (as in graphite – see Figure 16.9b).
■ The atoms are positioned such that rigid, interlocking tetrahedral three-dimensional arrangements are formed by the bonds (see Figure 16.9c). Figure 16.8 Formation of covalent bonds between pairs of adjacent carbon atoms.
a b c Metallic solids Metals may not appear to be crystalline. However, their microstructures do display unit cells, which are characteristic of crystalline substances. Typically, the outermost shell of an atom of a metal contains one or two, or at most three, electrons weakly attached to it. An atom, having lost its outermost electrons, becomes an ion with a net positive charge. The positive ions are arranged in a crystalline lattice structure. Because the outermost electrons are free, they wander through the lattice structure and are not attached to any particular atom (Figure 16.10). The presence of free electrons makes metals good conductors both of electricity and heat. The bonding that holds the charged ions and the sea of electrons in a metal is called metallic bonding and is explained by quantum mechanics and statistical mechanics, which are outside the scope of this book. Metallic bonds are not as strong as ionic or covalent bonds. They are of the order of 1 to 3 eV per atom and occur between the positively charged ionic lattice as a whole and the sea of electrons moving throughout it. + Figure 16.9 (a) A unit cell of diamond (carbon) and the bonds between atoms. (b) Unit cells of graphite (carbon) and the bonds between atoms and layers. The black arrows show the strong force due to covalent bonding; the red arrow shows weaker forces due to van der Waals bonds between unit layer arrangement of carbon atoms in graphite. (c) The interlocking tetrahedral shapes formed by carbon atoms in diamond. One atom lies above each triangle, forming the four-sided shape. + sea of delocalized electrons Figure 16.10 The ionic lattice structure of a metal with a swarm (or 'sea') of freely moving electrons permeating it. Chapter 16 Mechanical properties of matter a b c Figure 16.11 Unit cell structures in metals: (a) bcc packing; (b) fcc packing; (c) hcp packing. Metallic bonds occur between a positively charged ionic lattice as a whole and the sea of electrons moving throughout it. The relatively weak bonding in metals explains why metals are so easily deformed (bent or stretched). Because the bonding is between an entire lattice and a sea of electrons, there are no cleavage planes where metals would fracture, as are found in ionic crystals and many non-metallic substances. These types of materials are therefore very ductile (they can be stretched into wires). Bearing in mind the lattice-and-electrons microstructure of metals, we now examine three unit cell structures common in metals. Body-centred cubic (bcc) cell structure The body-centred cubic (bcc) unit cell is a cube with an atom at each corner plus one atom in the middle (Figure 16.11a). This arrangement does not allow layers of atoms to slide easily over each other. Hence metals with bcc unit cell structure tend to be relatively hard and not very malleable (i.e. not easily beaten into sheets). Examples of metals with bcc packing are chromium, vanadium, alpha iron and tungsten. Face-centred cubic (fcc) cell structure In addition to having one atom at each corner, a unit fcc cell also has one atom in the middle of each face of the cell (Figure 16.11b). The fcc structure is more densely packed than the bcc structure.
Along with the hcp structure (see below) it is the most efficient way of packing identical spheres to occupy the smallest possible volume, and therefore has the highest possible packing factor. Examples of metals with fcc structures are aluminium, copper, gold, lead and platinum.
When forces are applied to these structures, the planes of atoms ITQ 6 Suggest one application of carbon based on its hardness. ITQ 7 A close packing of atoms should result in a high density. Why then are platinum, gold and lead so dense, but aluminium is so light? slip more easily over each other than in the bcc packing case.
Hence, such materials tend to be more malleable than materials with bcc packing. Impurities put into fcc structures can, however, make plane slippage more difficult and therefore the material becomes hardened.
Hexagonal close packing (hcp) The hexagonal close packed (hcp) structure, shown in Figure 16.11c, is an alternative close packed arrangement of identical spheres with the same, maximum possible, packing factor as the fcc structure. Adjacent atomic planes in the hcp arrangement do not slip as readily as planes in the fcc structure and so hcp metals tend to be harder and more brittle than fcc metals. Examples of metals with the hcp structure include cobalt, magnesium, titanium and zinc.
Amorphous solids Amorphous solids usually do not have flat smooth faces like 'ionic' crystalline solids. Unlike metals, which display a crystalline microstructure, the units of amorphous solids are not repeated in rows and columns. We consider two classes of amorphous solids: polymers and glasses. Polymers A polymer is a material that is made up of many simple molecules of the same kind called 'mers'. Atoms within each 'mer' are bonded covalently to each other. A polymer is formed when mers bond covalently to form large molecules. Secondary bonds and mechanical entanglement hold these giant molecules together in the solid polymer. Let us consider polythene, a polymer made from ethene. Figure 16.12 shows how a polythene chain is formed from ethene, C2H4, which is made in turn from ethane, C2H6. Each line between the C and H atoms represents a covalent bond, i.e. a bond formed when two atoms share a pair of electrons (one from each atom). A carbon atom has four such electrons to share (see Figure 16.8, which shows the covalent bonding of carbon atoms) and each hydrogen atom has one electron to share. 247 248 Unit 1 Module 3 Thermal and mechanical properties of matter a c b H H H H H H H H C C = C C C C C C C H H H H H H H H carbon Figure 16.12 (a) Two ethene 'mer' units. (b) Polymerization involves opening of the double bonds. (c) A polyethene chain shown in three dimensions. hydrogen Polymerization Polymerization involves the opening up of the double bonds between adjacent carbon atoms. For each mer, a single bond is then available which can link a carbon atom of one mer to another until a very long chain is formed. Chain formation is stopped by the addition of an atom, called the terminator, to each of the bonds available at the ends of the chain. Properties of solid polythene Elasticity Figure 16.12c shows how a chain of carbon and hydrogen are arranged three-dimensionally in a polythene giant molecule. The angle between each adjacent carbon bond is 109°, causing a zig-zag chain. As a result of rotations about the bonds, the chain may also be curled, like a spring. When tensile forces are applied to polythene, uncurling takes place and these chains straighten between the carbon–carbon bonds, resulting in elongations of thousands of times larger than in crystalline substances. Polymers, like polythene, are therefore classified as elastomers since, upon removal of the applied forces, they return practically to their original shapes and lengths. This accounts for the enormous tensile strength of polythene, and polymers in general. The polymer Kevlar is so strong (yet lightweight) that it is used in bulletproof vests and in making frames stronger than steel. Thermoplasticity When polythene is heated, the van der Waals bonds between molecules are easily broken.
Polymers can also move relative to each other fairly easily. The material becomes soft and can easily be moulded using pressure. As the soft moulded material cools, van der Waals bonds re-form and the material takes the shape of the mould. This plastic behaviour, brought about by first softening the solid by heating, is called thermoplastic.
Thermoplastics can be reheated and remoulded. Getting it right! Not to be confused with thermoplastics are thermosetting plastics. The mers of thermosetting plastics have several places where bonding takes place when the raw materials are heated and pressed together.
The chains formed are not linear but very branched and become tangled as the polymer cools. Once set, it is difficult to separate individual molecules by heating – hence the name thermosetting.
Tensile strength Polymer chains usually form a tangled mass of molecules. These molecules are held together by mechanical entanglement and van der Waals forces. The latter are relatively weak, thus allowing some mobility of the molecules. As the number of carbon atoms within each molecule increases, from about 100 to about 1000, the net forces between molecules increases.
This results in the polymer moving from a liquid to a solid phase. As the carbon atoms in each molecule are numbered in thousands, the net forces between molecules become Reactivity and conductivity ITQ 8 Would the density of a polymer solid like polythene be large or small? Explain your answer. Glasses belong to a broad group of materials called ceramics. Ceramics are formed by chemical combination of metallic and non-metallic elements. Alumina (bauxite), Al2O3, and silicon dioxide (sand), SiO2, are ceramic compounds.
Ceramics are typically crystalline, but may ITQ 9 If glasses are so hard, then how come they can be 'cut' using diamond? Because there are no free electrons available in polymer chains (all are involved in covalent bonding), polymers are poor conductors of both heat and electricity. The lack of free electrons also means that visible light would easily pass through the solid since the energy in light is not enough to be absorbed by tightly bound electrons. Thus polymers also tend to be transparent. Glasses and ceramics Chapter 16 Mechanical properties of matter also be formed with an amorphous (random) structure by rapid cooling from the molten state.
Solids with an amorphous atomic structure are described as glasses. In alumina, the crystalline microstructures are formed from ionic bonds between positive aluminium ions and negative oxygen ions. In silicon dioxide (silica), the bonds are covalent. Both types of bonds are very strong, giving these substances their characteristic properties: high melting point, high compressive strength, great hardness and low ductility (due to the fact that the bonded atoms cannot be made to move about relative to each other). Because there are cleavage planes between adjacent unit cells (though these are not arranged in sheets), ceramics shatter easily (i.e. are very brittle) and their tensile strength is relatively low.
Glasses show all these characteristics. The absence of free electrons contribute to their poor conductivity and chemical inertness. Silicate glasses are formed by heating sand to very high temperatures and cooling the melt fairly rapidly. The crystalline microstructure units orient themselves randomly as solidification takes place, and the molten liquid sets in the shape of the mould being used; the molten liquid can also be blown into desired shapes. Elasticity in solids Hooke's investigations Robert Hooke (1635–1703) performed investigations on the stretching of metal wires. He found that the extension, ΔL, due to a force, F, on a wire of length L is directly proportional to F for a small range of extensions. Figure 16.13 shows F, L and ΔL for a wire under test. Close inspection of the graph shows a linear region followed by a non-linear region. Force, F elastic limit UTS yield breaking stress proportional limit 0 Extension, ΔL Figure 16.13 A typical F versus ΔL graph for a wire subject to a tension, F. Hooke's law states that the force causing a wire to stretch is directly proportional to the extension produced, provided the proportional limit is not passed. Figure 16.13 illustrates typical features of a F versus ΔL graph for a metal wire.
■ F varies linearly with ΔL up to a point called the proportional limit. ■ The wire behaves like an elastic up to the elastic limit, i.e. when loads are taken off, the wire returns to its original length. ■ Beyond the elastic limit the material no longer returns to its original length but becomes permanently deformed by the stretching force. The wire is now said to be exhibiting plastic behaviour. metre rule steel spring The relationship found by Hooke can be summarized, in the linear portion, as follows: ΔL ∝ F therefore, F ∝ ΔL F = kΔL...............(16.3) The constant, k, is called the stiffness constant of the wire F represents the force required to being tested since k = ΔL extend the wire by 1 unit of length.
The larger the value of k, the greater the force required to stretch the wire by 1 unit of length, i.e. the 'stiffer' the wire. Arising out of this type of investigation, Hooke found a relationship which has been called Hooke's law. set square used to measure position weight Figure 16.14 Simple laboratory arrangement that can be used to investigate Hooke's law for a spring and to determine the spring constant for that spring. 249 Unit 1 Module 3 Thermal and mechanical properties of matter when a load is taken off in the plastic region is called the permanent set of the wire. In the plastic region, when the wire passes the yield point, a small change in force produces a large extension and the wire soon breaks.
The ultimate strength of the wire is the maximum force it withstands; the ultimate tensile stress (UTS) is the maximum force per unit cross-section area that the wire can take. The breaking stress is the force per unit cross-section area that appears in the wire when it snaps. The breaking stress is smaller than the ultimate tensile stress.
Worked example 16.2: Spring stiffness constant Q A 150 g mass, hung from a spring of length 3.5 cm, causes the spring to lengthen to 5.7 cm. Find the spring stiffness constant in S.I. units. A extension of spring, ΔL = 5.7 – 3.5 cm = 2.2 cm = 0.022 m force, F = 0.150 kg × 9.8 N kg−1 = 1.47 N We assume that the spring has not passed the proportional limit. So we can apply the Hooke's law equation. F = kΔL where k = spring stiffness constant F k = ΔL Table 16.2 Typical ultimate tensile stress values for various materials Material Ultimate tensile stress / N m−2 aluminium 200 × 106 iron 120 × 106 bone (limb) 130 × 106 concrete 2 × 106 glass 50 × 106 (compressive) nylon 75 × 106 polythene ~50 × 106 steel 500 × 106 wood ~40 × 106 Figure 16.14 shows a simple laboratory arrangement that can be used to investigate Hooke's law for a spring and to determine the spring constant for that spring. The springs can be replaced by a polythene thread or opened out rubber band. Note safety first! Impact safety goggles should be worn and other safety procedures should be followed in this kind of experiment, since materials can snap or become loose under the applied forces. Force versus extension graphs Typical force/extension curves for rubber (a natural polymer) and polythene (an artificial polymer) are shown in Figure 16.15. In loading, the force versus extension behaviour is curved for both substances. Progressively smaller changes in force are needed per unit extension. This occurs due to the uncurling of molecular chains and straightening of carbon-carbon bonds. Beyond a certain point, however, both materials get very stiff, i.e. a very large change in force is required to produce unit extension. This is the point where the twisted chains of the molecules have been straightened. During unloading, the rubber (Figure 16.15a) shows a curved graph, indicating that the straightened chains are curling back to their original state. The rubber reaches its = 1.47 N 0.022 m = 69 N m−1 to 2 s.f. a b Tension ■ The amount of permanent deformation in the wire Tension 250 loading loading unloading unloading 0 Extension 0 Extension Figure 16.15 Force versus extension graphs for (a) a rubber band and (b) a polythene thread. original length when all the load has been taken off. The polythene, however, shows a linear force versus extension behaviour upon unloading (Figure 16.15b) and remains permanently stretched after all the load has been taken off. This suggests that only the zig-zag carbon bonds are returning to their original angular orientation but no coiling is taking place once the uncoiling has occurred. The loading–unloading loop shown by the rubber is called a hysteresis loop. Hysteresis means 'lag'. For any given load within the loop, the unloading extension is larger than the corresponding loading extension, i.e. the former lags the latter. The work done in stretching the rubber during loading is the area under the loading curve and is positive since applied force and extension are in the same direction. The work done during unloading is the area under the unloading curve and is negative since the force Chapter 16 Mechanical properties of matter and extension are in opposite directions. The net work done during the hysteresis cycle is therefore positive and can be noticed as heat appearing in the rubber. Springs in series and in parallel Consider two springs of spring stiffness constants, k1 and k2, respectively. When the springs are connected in series (i.e. along a single line), the effective spring constant, ks, is given by 1 1 1 = + ks k1 k2 In equation form, Young's modulus, E, is given by E = stress F/A = strain ΔL/L ....(16.4) where F is the force applied to the material of length L and cross-section area A, and ΔL is the extension produced. The value of E for a given material is practically the same whether the force is tensile or compressive. Table 16.3 shows Young's moduli for various materials. Getting it right! When the springs are connected in parallel (i.e. parallel to each other), the effective spring constant, kp, is given by kp = k1 + k2 The derivation of these two equations is shown in Chapter 17. The modulus of elasticity is defined in the elastic region of the material. The proportional limit is usually close to the elastic limit in metals, the Young's modulus is defined in the region up to the proportional limit in metals. Table 16.3 Typical Young's moduli for various materials Young's modulus of elasticity Material Young's modulus / N m−2 The constant, k, mentioned above, is a measure of the stiffness of a given object, such as a particular spring or a wire of a given cross-section area. A value for the stiffness constant of a material, regardless of its dimensions, is its tensile modulus, also called its Young's modulus. aluminium 70 × 109 bone (limb) 15 × 109 concrete 20 × 109 glass 50–90 × 109 nylon 2–4 × 109 Young's modulus is defined as the ratio of stress on a material to the strain produced in the material. The stress is the force applied per unit cross-section area of the material and the strain is the extension per unit length. polythene 1–4 × 109 rubber 0.01–0.1 × 109 steel 200 × 109 wood 1–10 × 109 Experimental determination of Young's modulus using a wire Practical determination of Young's modulus test wire observer's view Figure 16.16 shows an experimental arrangement that can be used to determine the Young's modulus of a material in the form of a wire.
crossbar rests on micrometer screw spirit level main scale micrometer hinged crossbar hinge horizontal micrometer (screw is behind the weight) Figure 16.16 Apparatus for determining Young's modulus. Note: safety first! Impact safety goggles should be worn and other safety procedures should be followed in this kind of experiment, since materials can snap or become loose under the applied forces. Note the following experimental details. 1 A control wire of the same material and dimensions is used alongside the wire being tested. This is to compensate for either the sagging of the support or temperature changes that may occur during the experiment. It is crucial that precautions like these are taken, since the extensions for wires are usually very small. 2 Just enough weight is added to each wire to ensure the wires are straight before any measurements are taken. 3 The length of the test wire is measured using a metre rule. 251 252 Unit 1 Module 3 Thermal and mechanical properties of matter 4 A micrometer is used to measure the extension, because the latter is so tiny. There is a cross-bar with a spirit-level to enable measurement of extension. The micrometer screw is adjusted to just touch the bottom of the levelled cross-bar when taking each measurement of the extension. A micrometer is also used for measuring the diameter of the wire. Several measurements of the diameter are taken and the average calculated. Worked example 16.3: Young's modulus for cotton thread Q A cotton thread of unstretched length 1.05 m and diameter 0.44 mm, stretches 4.2 cm when a mass of 300 g is hung from it. Estimate the elastic modulus of the cotton material used. A ■ 5 Weights are added gently so that larger impulsive forces are not given to the test wire. Stress versus strain graphs In industrial practice, a tensometer (Figure 16.17a) is used to measure force and extension and plot force versus extension automatically. Stress (force/area) versus strain (extension/length) graphs are shown in Figure 16.17b for ductile, brittle and polymeric materials. ■ Ductile materials, e.g. metal wires, show much extension after the yield point (which follows closely the elastic limit). ■ Brittle materials (e.g. glass) break just after the elastic limit before plastic deformation can take place. ■ diameter of thread, d = 0.44 mm = 4.4 × 10−4 m ■ cross-section area, A = πd2/4 = π × (4.4 × 10−4)2 m2 / 4 = 15.2 × 10−8 m2 ■ extension, ΔL = 4.2 cm = 4.2 × 10−2 m ■ force producing extension, F = mg = 0.300 kg × 9.8 N kg−1 = 2.94 N ■ stretched length of thread, L = 1.05 m By equation 16.4 E = F/A ΔL/L = FL AΔL = 2.94 N × 1.05 m 15.2 × 10−8 m2 × 4.2 × 10−2 m = 4.8 × 108 N m−2 (to 2 s.f.) ■ Polymeric materials extend much more than metals until [illegible]

The presence of cleavage planes, however, results in such materials shattering very easily (i.e. in being brittle). ✓ In covalent solids (e.g. diamond), unit cells are also quite rigid due to the strong bonding forces between atoms, but the cleavage planes are not usually arranged in large sheets. Covalent solids are therefore generally hard and brittle. In the case of diamond, the atoms form rigid tetrahedral arrangements in which bonds from adjacent tetrahedra interlock three-dimensionally within a unit cell, resulting in diamond being extremely hard and not easy to shatter. ✓ In metallic solids (e.g. copper), bonding is not directional and so atomic layers can slide past each other. This makes pure metals easy to be beaten into shapes without breaking and also generally ductile rather than brittle. Metallic bonds are strong, however, giving rise to the large tensile strength generally associated with metals. ✓ The bonding in amorphous substances is generally covalent and, therefore, is also strong. Silicate glasses are brittle because they are covalently bonded; they do not have regular layers of atoms that can slide over each other. A true glass is amorphous – it does not have a crystalline microstructure, although made up of ions.

Polymers are formed by covalent bonding within and between 'mers' to form long molecular chains. Rather weak van der Waals bonds hold these molecules and chains together. Hence these solids do not generally have a crystalline microstructure. ✓ According to Hooke's law, the force, F, exerted on a material is directly proportional to the extension, ΔL, produced, provided that the proportional limit of the material is not exceeded. In terms of an equation, F = kΔL...............(16.3) The constant, k, is called the spring stiffness constant of the object being tested. ✓ The Young's modulus of elasticity, E, of a material is defined by the equation E= stress F/A = strain ΔL/L ....(16.4) ✓ It can be shown that the energy, W, stored in a material stretched an amount ΔL by a force F is given by 1 W = 2 k(ΔL)2...............(16.6) W= 1 2 2 EA(ΔL) L ....(16.7) ✓ The energy stored per unit volume in a wire can be shown to be given by W 1 = stress × strain V 2 ....(16.8) 255 256 Unit 1 Module 3 Thermal and mechanical properties of matter Review questions Kinetic model of matter; Solids, liquids and gases 1 (a) Describe, with the aid of a graph, how the force, F, between two simple molecules varies with separation distance, x, of the molecules. (b) With the help of the kinetic theory of matter, explain the following: (i) a solid has a fixed shape (ii) a liquid takes the shape of the container into which it is put, but generally the particles stay together (iii) a liquid shows evaporation at its surface (iv) a liquid cools as it evaporates (v) a gas completely fills a container into which it is put. 2 (a) Sketch a graph of potential energy versus separation distance for a system of two adjacent molecules. (b) Indicate on the graph, typical total energies, ES, EL and EG, which correspond to solid, liquid and gases states, respectively. (c) Explain why the energies chosen correspond to the states named. Density of materials 3 (a) Define density.

(b) Both aluminium and gold have a high packing factor of atoms per unit volume. Why, then, does aluminium have a far lower density than gold? (c) A certain briefcase of linear dimensions 50 cm × 30 cm × 10 cm is made of a strong, lightweight material. Would you be able to lift such a briefcase if it was tightly packed with bars of aluminium or gold? Explain your answers quantitatively. 4 (a) Explain why substances generally expand on heating. (b) The thermal expansion of water shows an anomaly. Employing a suitable graph, describe the anomaly that occurs during heating from 0 °C to about 10 °C. (c) Explain the anomaly in terms of the arrangement of water molecules. 5 Why do polymers generally have lower densities than metals? Give reasons based on relative atomic masses and arrangement of atoms. 6 A U-tube water manometer (Figure 16.19) is used for measuring the pressure of the gas supply in the lab. open to atmosphere gas H water in U-tube Figure 16.19 (a) If the length H = 30 cm, determine: (i) the excess pressure of the gas in kPa (ii) the actual pressure of the gas in kPa. (b) State, with a reason, one safety precaution that should be exercised when taking a measurement of this kind. 7 (a) Some fish are able to exist 100 m below the sea. Determine the pressure on such fish (atmospheric pressure = 101 kPa). (b) The density of sea water increases with depth of the sea. For objects to rise or sink to desired levels, it must be possible to adjust their overall density (see Chapter 6). How is overall density adjusted in the case of: (i) submarines? (ii) bony fish? Amorphous and crystalline solids 8 How do amorphous and crystalline substances generally differ in overall external appearance? Give an explanation for this difference. 9 In terms of atomic structure explain why: (a) metals are generally malleable (b) metals are generally ductile (c) ceramic substances (e.g. glass) are very hard (d) ceramic substances are very brittle (e) solid ceramic substances can not generally be beaten into sheets. 10 (a) (i) What is meant by the term 'polymer'? (ii) Name one natural polymer and one humanmade polymer. (b) Why are polymeric solids generally very strong? Give two explanations for your answer. (c) Explain why some polymers are thermoplastic whereas others are thermosetting. Chapter 16 Mechanical properties of matter 11 In bio-medical applications, ceramics are used in artificial ball and socket joints (e.g. at the hip) whereas polymers are used in making artificial blood vessels. Suggest two reasons for each of the two applications. Elasticity in materials 12 (a) State Hooke's law. (b) Figure 16.20 shows the force (F) versus extension (ΔL) behaviour for a metal wire.

Name, and give a brief description of, the points labelled A to E. (c) Using the dimensions given in Figure 16.20, determine the stiffness constant, k, for the wire, by graphical means. D F/N B E A 0 5.0 16 Suggest one reason why the value of the elastic modulus of a material is useful to engineers. Explain your answer by referring to a specific application. Practical activities and challenges Note: Safety first! Impact safety goggles should be worn and other safety procedures should be followed in these kinds of experiments, since materials can snap or become loose under the applied forces. 1 You are given a box of chalk. Each stick of chalk is approximately cylindrical. Describe an experiment to determine an accurate value of the density of this chalk. Indicate the instruments to be used and why you have chosen these instruments. Estimate the percentage error in the value you obtain. C 50 15 (a) Describe, with the aid of a labelled diagram, an experiment to determine the Young's modulus of a material in the form of a wire. Give reasons for the instruments used and the steps taken so as to obtain accurate results. (b) Describe and give a reason for one safety precaution that must be taken in this experiment. ΔL / mm Figure 16.20 13 (a) Define: (i) tensile stress (ii) tensile strain (iii) ultimate tensile stress (iv) tensile strength (v) breaking stress (vi) Young's modulus. (b) If the wire in Figure 16.20 had an original (unstretched) length of 1.20 m and diameter of 0.32 mm, determine, for the wire: (i) the Young's modulus (ii) the ultimate tensile stress (iii) the breaking stress. 14 (a) Prove that the work, W, done when a force, F, causes a wire to stretch an amount ΔL is given by 1 1 W = 2 FΔL = 2 k(ΔL)2 where k represents the spring stiffness constant of the wire. (b) (i) What feature of the graph in Figure 16.20 represents work done in stretching the wire? (ii) Using the feature described in (b)(i), estimate the work done in stretching the wire by 2.0 mm. 2 For a similar box of chalk as described in Activity 1 above, describe an experiment to determine an accurate value of the breaking tensile stress of this chalk. Indicate the measurements to be taken, the instruments to be used and why you have chosen these instruments. Estimate the percentage error in the value you obtain. 3 Design and carry out an investigation to compare the elastic moduli of two samples of human hair from two different individuals. By taking percentage errors into consideration, discuss whether the values are the same or different.

4 Investigate the behaviour of rubber (e.g. an open rubber band) and polythene (e.g. a thread) when loaded and then unloaded. 5 Design and carry out an investigation to compare the breaking tensile stress of rubber with that of glass. Answers to ITQs 1 Between x1 and x2. This is because as x approaches x1, the force of repulsion between molecules get larger; as x approaches x2, the force of attraction becomes larger. 257 258 Unit 1 Module 3 Thermal and mechanical properties of matter 2 E2. This is because E2 is negative (the system is therefore bound) but the mean intermolecular separation has increased and so the average attractive force between the molecules is less, allowing them to move freely over each other. 3 The volume is a minimum at 4 °C. The density would be a maximum at that temperature since density = mass and the mass of the water is volume constant. 4 The outermost shell of the chlorine atom has 7 electrons. By gaining 1 electron, the atom achieves the stable arrangement of 8 electrons. 5 Free electrons are not available for conduction in the solid or the molten state. In the solid state, the ions (the charge carriers) are held together by strong electrostatic forces and therefore are not mobile. In the molten state, the ions can move relative to each other and thus electrical conduction is possible. 6 Any practical suggestion, e.g. to make tools that can cut other hard materials. 7 The mass of an aluminium atom is far less than that of an atom of lead, gold or platinum. Hence the density of aluminium would be less, since the packing volumes are not very different. 8 Small, on two accounts: (i) the molecules are coiled and physically tangled, contributing to a large volume and (ii) the relative atomic masses of carbon and hydrogen, which make up the molecule, are small. 9 Diamond is much harder than glass. Also the tiny tip of diamond glass cutters enables tremendous pressures to be directed on the glass in desired lines, to cause breakage of bonds at those sites. Answers to Review questions 3 (c) aluminium 40.5 kg (possibly); gold 290 kg (very unlikely!) 7 (a) 1081 kPa 13 (b) (i) 1.87 × 1011 N m−2 (ii) 9.6 × 108 N m−2 (iii) 7.8 × 108 N m−2 Unit 2 Electricity and magnetism, a.c. theory and electronics, atomic and nuclear physics 260 Module 1 Electricity and magnetism Chapter 17 Electric charges and fields Learning objectives ■ Explain the difference between electrical conductors and insulators using a simple electron model. ■ Explain charging and discharging in terms of transfer of electrons. ■ Use the equation Q = ne to calculate charge. ■ Explain the results obtained when objects are charged by friction, induction and contact. ■ State and apply the fundamental law of electrostatics: 'like charges repel, unlike charges attract'. ■ Discuss simple applications of electrostatics, such as agricultural chemical spraying, paint spraying, dust extraction, photocopying, laser printing and lightning rod protection. ■ Discuss hazards associated with charging by friction. ■ State and apply Coulomb's law for the force between combinations of charges in free space. ■ Define and use the equation for electric field strength: Q 4πε0r2 V ●● between two charged parallel plates: E = d Describe the motion of charged particles: ●● entering perpendicular to an electric field ●● parallel to an electric field. Compare the motion of charged particles in a uniform electric field with that of a projectile in a gravitational field. Define electric potential in terms of work and charge. Use the fact that field strength at a point is numerically equal to the potential gradient at that point (E = − ΔV Δ x). Q Use the equation V = 4πε r for potential due to a point charge. ●● ■ ■ ■ ■ ■ due to a point charge Q: E = 0 ■ Compute the potential at a point due to several point charges. Static electricity ■ Static electricity involves electrical charges at rest or in motion for a brief duration. ■ Current electricity involves charges in motion. In this chapter, the focus is on static electricity. Electrical charges are of two types, positive and negative. Charges similar in behaviour to the net charge on an ebonite rod rubbed with fur (or a polythene rod rubbed with wool) are designated negative. Those that behave like the net charge on a glass (or Perspex) rod rubbed with silk are designated positive. According to the fundamental law of electrostatics: like charges repel; unlike charges attract. Thus a negative charge repels a negative charge but attracts a positive charge. A positive charge repels a positive charge but attracts a negative charge. Chapter 17 Electric charges and fields a nucleus c b n=1 n=2 n=3 n=4 Shell (n) number of electrons 1234 2881 Shell (n) number of electrons 1234 2882 Shell (n) number of electrons 1234 2892 Figure 17.1 Shell arrangements of electrons in selected atoms: (a) potassium atom (19 electrons), (b) calcium atom (20 electrons), (c) scandium atom (21 electrons). Static electrical charges are easily observed on insulators and on insulated conductors. Conductors and insulators In a neutral atom, the number of positive charges in the nucleus (from the protons) is equal to the number of electrons orbiting the nucleus. According to a simple model, electrons of an atom are held in circular orbits due to attraction by the positively charged nucleus. However, the electrons themselves repel each other. The result is that electrons settle in stable energy states. Each state corresponds to an orbit or 'shell' of a different radius and is designated by a quantum number, n = 1, 2, 3 .... According to the model, the maximum number of electrons allowable in each shell is 2n2. Thus, the first (i.e. innermost) shell can accommodate 2 electrons, the second, 8, the third, 18, and so on. It is difficult to add electrons to a stable state or remove electrons from a stable state. It is far easier to do this in other states, i.e. when a shell does not have its full complement of electrons. This model explains why some substances are conductors of electricity and others are not.

8 electrons (the 'octet') in the 3rd shell and 1 electron in the 4th shell. An atom of the metal calcium (with 20 electrons), rather than having 10 electrons in the 3rd shell, consists of 8 electrons in the 3rd shell and 2 in the 4th shell. However, an atom of the metal scandium (with 21 electrons) has 9 electrons in the 3rd shell and 2 in the 4th shell. You are not required to know the actual shell arrangements of electrons at this level. The main point being established here is that atoms of conductors have 1, 2 or 3 electrons in their outermost shells and that these shells do not have their full complement of electrons. Outermost electrons are attached very weakly to atoms of conductors and often wander away from the parent atoms, though remaining within the conductor. When an atom loses an electron, a positive ion is left. Positive ions that result in conductors are arranged in a regular array called a lattice. The electrons given up by each atom meander fairly freely throughout the lattice (Figure 17.2). Such substances are able to conduct electricity very easily by means of these 'free' electrons. (However, electrons may also collide with ions of the lattice, producing an electrical effect called 'resistance'.) This simple shell model, however, does not apply for states n = 3 or greater, as far as stable orbits are concerned. For these cases, the model is modified: stable energy states occur, not when an outermost shell is full, but rather when an outermost shell has 8 electrons (called an 'octet'). In a conductor, the outermost atomic shell is far from being filled to capacity. This type of atomic structure is characteristic of metals.

Typically, metals have 1, 2 or 3 electrons in their outermost atomic shells. Figure 17.1 shows that an atom of the metal potassium (with 19 electrons), rather than having 9 electrons in the 3rd shell, which can take up to 18 electrons, consists of Figure 17.2 Free electrons (coloured red) meandering in random directions through an ionic lattice of a conductor (coloured green). 261 262 Unit 2 Module 1 Electricity and magnetism The ions in the ionic lattice are constantly vibrating, even at room temperature. At higher temperatures, the vibrations are more rapid. This results in a greater probability of the 'free' electrons colliding with the lattice, thereby impeding the rate of flow of charge, i.e. increasing the 'resistance' of the conductor to the flow of electricity. Elements that are classified as insulators, for example sulfur, possess almost no 'free' electrons. As a result, they are not good conductors of electricity. Insulators can be made to conduct a significant amount of electricity if their outermost electrons can be ripped off their outermost atomic shells, as for example, by a large electric field or high voltage. (A third class of elements, called semiconductors, conduct electricity more easily than insulators but not as well as conductors. Unlike conductors, semiconductors have both negative and positive charge carriers. Charge carriers in semiconductors are discussed in Chapter 22.)

Many compounds are insulators due to bonding between atoms. Covalent bonding, such as occurs in rubber, involves mutual sharing of electrons between atoms such that the electron states in their outermost atomic shells are stable ones, making free electrons practically unavailable for conduction. (See Chapter 16 for a discussion of covalent bonding.) Solid ionic crystals, for example sodium chloride (common salt), are also non-conducting. However, this is because the positive and negative ions forming the crystal are held together by very strong electrostatic forces. In ionic bonding in sodium chloride, a sodium atom (with 11 electrons) donates an electron. In the process it becomes a positive sodium ion with 10 electrons, 2 in the n = 1 shell and 8 in the n = 2 shell. A chlorine atom (with 17 electrons) receives the electron.

It becomes a negative chlorine ion with 18 electrons, 2, 8 and 8 electrons respectively in the n = 1, n = 2 and n = 3 shells. Free electrons are unavailable. In the sodium ion, the outermost (n = 2) shell becomes full (8 electrons). In the n = 3 shell of the chlorine ion, the number of electrons becomes 8 (the 'octet'). In the molten state, however, the ions have enough energy to become mobile and current is carried by both positive and negative ions as charge carriers. In the aqueous state, both types of ions are also mobile (Figure 17.3).

ITQ 1 Explain, in terms of electron transfer, how a Perspex (or glass) rod acquires a positive charge when rubbed with silk. a + – c b + – – – – + + + + + – – – – + + – + + – + + + – + – + + + + + – + – + – + + – + + + + + – – + + + – – + + – + + + – – sodium ions (Na+) –+ + – + + chlorine ions (Cl –) water molecules Figure 17.3 Ions in sodium chloride: (a) crystalline solid, (b) molten liquid and (c) aqueous solution. Charging and discharging Charging and discharging involve transfer of electrons. It is far easier to transfer orbital electrons than transfer the protons present in the nucleus of an atom. There are three methods of charging: ■ charging by friction ■ charging by contact ■ charging by induction. Charging by friction Charging by friction works well with insulators (or insulated conductors). When a polythene rod is rubbed with wool, electrons are pulled off atoms of the wool and are transferred to the rod (Figure 17.4). The rod, originally neutral, acquires a net negative charge.

The wool, having lost electrons, is left with a net positive charge. In charging by friction the pair of objects become oppositely charged. – – – 4e + + – + + polythene rod wool Figure 17.4 A polythene rod rubbed with wool acquires a net negative charge through transfer of electrons (e−). Experiments show that both a polythene rod rubbed with wool or a Perspex rod rubbed with silk will attract tiny, neutral pieces of paper. How do we explain this? When a positively charged Perspex rod is bought near to the neutral paper the orbits of electrons of atoms in the ITQ 2 Shortly after a neutral piece of paper makes contact with a negatively charged rod, the paper is sometimes repelled from the rod. How can this be explained? Chapter 17 Electric charges and fields paper – + + + + positively charged rod + + + + – When an object is charged by contact with another charged object, both objects end up with the same sign of charge as the charging object. + Discharging – + – + – + – + Figure 17.5 A positively charged rod can attract neutral paper. paper are shifted toward the rod (Figure 17.5). This causes atoms in the paper to be slightly negative at the end nearer to the positively charged rod since the electrons in orbit spend most of their time displaced toward the rod. The atoms in the paper have become polarized. The positively charged rod attracts the pieces of paper since the negative side of the polarized atoms are nearer to the rod than are the positive side of the polarized atoms. Thus a positively charged rod can attract a neutral object! A similar argument can be made for a negatively charged rod. Getting it right! To test the sign of the charge on an object, attraction of the object to a known charge is an inconclusive test since a known charge will attract both an object that is charged with an opposite sign and a neutral object.

The only sure test is a repulsion test, since charges of the same sign repel each other. Charging by contact Consider a positively charged insulator, A, placed upon another insulator, B, that is neutral. Some electrons from B might be pulled over to A by the strong electrostatic force of attraction. Those electrons from B may neutralize some of the positive charge on A. However, as electrons have left B, B will have a net positive charge. charged rod – –– – – + + + + + + – – metal sphere – – – –– – – electron flow insulating stand 1 Charged rod brought near to sphere + + + + + Consider a charged metal sphere, A, suspended by an insulating thread. For discussion purposes, let us assume a net negative charge on the sphere. If an identical but neutral sphere, B, similarly suspended, is made to touch A, it is reasonable to assume that the charge would be shared equally between the two spheres since there are 'free' electrons in both conducting spheres and the spheres are of the same size. Now consider what would happen if sphere B were as large as the Earth. The charges on A will still be shared between A and B, but, for all practical purposes, all of the charge would reside on B. In other words, A would have become 'discharged'. The Earth is a large conductor. Hence, any charged conductor will be discharged if connected by a wire to the Earth.

Even when touched with a finger, a metal object carrying a small amount of charge will be discharged.

A charged conductor can also be discharged if brought into contact with another conductor with an equal number of opposite charges. This is because conductors have free electrons and these easily move between the two conductors. Charging by induction Because metals have 'free' electrons, a metal object is easily charged by induction, i.e. without the charging object touching it. A negatively charged rod is brought near to a metal sphere (Figure 17.6).

Free electrons in the sphere are repelled to the side away from the rod. If the sphere is connected to earth with the rod still in place, negative charge from the sphere rush to earth, due to the influence of the negatively charged rod. If the earth connection is broken, and then the rod is removed, the sphere is left + + with a net positive charge. + Charging by induction works + + well with conductors. The object being charged acquires a charge opposite in sign to the inducing charge. symbol for earth 2 Sphere touched with finger or connected to earth by wire Figure 17.6 Charging a metal sphere by induction. 3 Rod removed 263 264 Unit 2 Module 1 Electricity and magnetism Since charging transfers whole electrons, the net charge, Q, transferred on to an object is given by the equation: Q = ±ne...............(17.1) where n is the number of electrons transferred (a whole number) and e is the magnitude of the charge on an electron. 'e' is called the elementary charge and has been measured to be equal to 1.6 × 10−19 coulombs. The S.I. unit for charge is the coulomb (C). Worked example 17.1: Charge on a Perspex rod Q 4.0 × 1014 electrons are transferred when a Perspex rod is rubbed with silk. What is the charge gained by the Perspex rod? A By equation 17.1, the charge, Q, gained by the Perspex rod is given by Q = +ne = 4.0 × 1014 × 1.6 × 10−19 C = 6.4 × 10−5 C Note the sign of the charge. The Perspex rod is positive since electrons were transferred from the rod. Hazards of static electricity When substances rub against each other, there is a possibility of them being charged by friction, especially if they are insulators. If there is enough positive charge on an object, it can pull electrons off air molecules. These electrons can be further accelerated due to the proximity of the charged object and knock out other electrons from atoms of the air. An avalanche of electrons results, which as the electrons flow, causes the air to be heated quickly – forming a spark. Thus, when clouds are charged sufficiently (due to rubbing against currents of air) a giant spark,

By equation 17.16, there would be a component of the electric field along the conducting surface.
The field would cause electrons to flow. But electric currents are not observed on the surface of conductor. Hence the surface must be an equipotential. Answers to Review questions 9 –4.1 N 11 0.080 m (i.e. 8.0 cm) from the −3Q charge (but not between the two charges) 13 (b) 2.1 × 10−17 m3 (c) (i) 1.82 × 10−13 kg (ii) 0.0026 × 10−13 kg (d) 1.4 × 10−17 C (e) 88 electrons 277 278 Chapter 18 Capacitors Learning objectives ■ Describe a capacitor as a device for storing charge. Q) and farad as the S.I. unit of capacitance. V Q ■ Use the equation C = to solve problems. V A ■ Use the formula C = to solve problems involving parallel-plate capacitors, and also as a guide in the d construction of large or small capacitances within size limitations. ■ Define capacitance (C = ■ Define dielectric constant κ of a material as a ratio of its electric permittivity to that of free space. 1 = 1 + 1 + C1 C2 ...) to solve problems. Q2 ■ Use formulae for the energy stored in a capacitor to solve problems. U = 1 CV2, U = 1 QV and U = 2 C 2 2 ■ Discuss the mechanism of storage of energy in a capacitor. ■ Derive and use formulae for capacitors in parallel (C = C1 + C2 + ...) and series (−t −t ■ Recall and use the equations for capacitor discharge: Q = Q0e(RC); I = I0e(RC) and V = V0e(RC) ■ Define time constant as τ = RC ■ Sketch and use graphs of Q, V or I (or ln Q, ln V or ln I) against t, illustrating capacitor discharge. ■ Recall and use the equation V = ξ[0(1 − e(RC) and graph of V against t, illustrating the charging of a capacitor, where −t ξ0 is the e.m.f. of the charging source. Capacitors A capacitor is a device for storing charge.
A capacitor consists of two conductors with an insulator or an empty space sandwiched between them. The empty space or insulator is called a dielectric. Figure 18.1 shows the standard symbols used in circuit diagrams. To charge a capacitor, charge must be moved from one of the conductors to the other.
It takes work to remove electrons from a conductor. It also takes work to place electrons on to a conductor that has received electrons, since electrostatic repulsion has to be overcome. Thus, in charging a capacitor, work has to be done by an external fixed capacitor + 1000μF 40V variable capacitor agent (e.g. a battery) between the two conductors. For any amount of electrons transferred, both the donor and the recipient conductors attain charge of the same magnitude but opposite in sign. In Chapter 17, we saw that work is done when a charge, Q, is moved across a potential difference. For a fixed potential difference, V, the work done, W, is given by the equation: W = QV...............(17.15) A potential difference between the plates of a capacitor is created by transferring charge from one plate to the other. trimmer capacitor - + electrolytic ceramic Figure 18.1 Symbols for a variety of capacitors. Chapter 18 Capacitors leaving one with a net positive charge and the other with a negative charge. Experiments have shown that the charge on a capacitor is directly proportional to the potential difference between the conductors. Thus: Factors affecting capacitance The parallel-plate capacitor Q ∝ V. This gives: Q = CV...............(18.1) A where C is a constant and Q is the charge that has been transferred. The constant, C, is referred to as the capacitance of the capacitor. By changing the subject of equation 18.1, Q we get C = , from which we can obtain a definition of V capacitance. Capacitance is the charge stored in a capacitor per unit potential difference. C= Q V ....(18.2) The unit for capacitance is the coulomb per volt, and the S.I. unit is called the farad (F). In practical usage, a farad of capacitance is huge. It is more common to find capacitors marked in microfarads (μF, or 10−6 F), nanofarads (nF, or 10−9 F) or picofarads (pF, or 10−12 F). Worked example 18.1: Charge flowing to charge a capacitor Q How much charge flows when a 3.0 V battery is used to charge a 40 μF capacitor? A Charge will flow until the potential difference across the capacitor is equal to 3.0 V, the potential difference supplied by the battery; charge, Q = ?
capacitance, C = 40 μF = 40 × 10−6 F Q since C = V then, Q = CV = 4.0 × 10−6 × 3.0 = 12 × 10−6 C = 12 μC Getting it right! Do not confuse C (in italics), which stands for capacitance, with C (in regular font), which stands for the coulomb, the unit of charge. ITQ 1 How many electrons are transferred between two conductors if the capacitance is 12 F and there is a resulting voltage of 3.0 V between them? ITQ 2 What would be the S.I. unit for dielectric constant? d conducting plates dielectric Figure 18.2 A parallel-plate capacitor. A capacitor (Figure 18.2) consisting of two parallel conducting plates, equal in size, and separated by various dielectrics can be used to investigate factors affecting capacitance.
For a parallel-plate capacitor, experiments have shown the following: ■ capacitance is directly proportional to the common area of overlap, A, between the plates ■ capacitance is inversely proportional to the separation distance, d, between the plates (i.e. to the thickness of the dielectric). These two relationships can be summarised as: C∝ A d C= εA d ....(18.3) where ε is a characteristic of the dielectric called its electrical permittivity. The electrical permittivity, ε0, of free space, has a value of 8.85 × 10−12 C2 N−1 m−2. The permittivity of air is quite close to this value and in calculations involving air as the dielectric, the value of 8.85 × 10−12 C2 N−1 m−2 is often used. The dielectric quality of a dielectric is usually expressed as a ratio of its permittivity to the permittivity of free space. This quality, called the dielectric constant, κ, is given by the equation: κ= ε ε0 ....(18.4) Thus a dielectric with a large dielectric constant has a large electric permittivity. Table 18.1 shows a list of dielectrics and their corresponding dielectric constants. According to equation 18.4, the dielectric constant of free space is exactly 1 (since the permittivity of free space divided by the permittivity of free space is 1). 279 280 Unit 2 Module 1 Electricity and magnetism Table 18.1 Approximate dielectric constants of selected substances Dielectric Approximate dielectric constant (to 1 significant figure), κ vacuum 1 (exactly) air 1 polystyrene 3 vinyl plastic 3 oil 4 glass 5 rubber 7 mica 7 pure water 80 barium titanate 1000 Worked example 18.2: Capacitance of a parallel-plate capacitor Q A (a) Calculate the capacitance of an air-filled 10 cm × 10 cm capacitor whose plates are separated by a distance of 1 mm. (b) If a sheet of rubber 1 mm thick snugly fits between the plates in (a), calculate the new capacitance of the capacitor. area of plates, A = 10 cm × 10 cm = 10−2 m2 distance between plates d = 1 mm = 10−3 m thickness of rubber, dR = 1 mm = 10−3 m permittivity of air = ε0 = 8.85 × 10−12 C2 N−1 m−2 dielectric constant of rubber κR = 7 (from Table 18.1) (a) Capacitance of air-filled capacitor cA CA = 0 d 8.85 × 10−12 × 10−2 = 10−3 = 8.85 × 10−11 F = 90 × 10−12 F = 90 pF (to1 s.f.) (b) Capacitance of rubber-filled capacitor εA CR = R dR = κRε0A dR 7 × 8.85 × 10−12 × 10−2 10−3 = 62 × 10−11 F = 620 × 10−12 F = 600 pF (to 1 s.f.) Worked example 18.3: Effect of inserting a dielectric in a charged air capacitor Q The air in an air capacitor charged to 30 V is replaced with a material of dielectric constant 4 after the charging source has been removed. What will be the new voltage across the capacitor plates? A The charge, Q, will remain the same because the charging source has been removed. When the dielectric material replaces the air there is no charging source present nor any discharge path for the charge. Q, to travel. capacitance of the air capacitor = CA voltage across air capacitor, CA = 30 V when the material of dielectric constant 4 is inserted, let the new capacitance = CM CM =κCA = 4CA (see 'Useful tip' above) For the air capacitor: Q CA = V (by definition of capacitance) VA i.e. VA = Q CA (i) After the insertion of the new material, the new capacitance is Q CM = (where VM is the new voltage across CM) VM i.e. VM = Q CM (ii) Dividing (ii) by (i) gives VM = Q/CM VA = Q/CA CAVA i.e. VM = CM VA = (since CM = 4 × CA, as explained above) = 4 3.0 = 0.75 V (i.e. the voltage drops by a factor of κ) 4 Capacitors in practical usage Capacitors find extensive usage in electronic circuits, on account of the following properties they possess. Capacitors ■ can store electrical energy ■ can block direct current while allowing alternating current to pass ■ can offer varying resistances (called 'reactances') to Useful tip! Replacing the air in an air-filled capacitor with a dielectric of dielectric constant κ increases the capacitance of the capacitor by a factor of κ. alternating currents depending on the a.c. frequencies ■ in conjunction with inductors, can form resonant circuits (a coil of insulated wire forms an inductance, which is also a.c.-frequency dependent; like a capacitor, an inductor can also store electrical energy). Chapter 18 Capacitors Energy stored in a parallel-plate capacitor The work done in charging a capacitor becomes the energy stored in the capacitor. As charge is being transferred from one plate to the next, the electric field between the two plates increases, since electric field, E, increases with charge and for a field between parallel plates, E = V d (equation 17.17). We can derive an equation for the work done, W (= energy stored, U), in charging a capacitor from 0 V to a voltage V as follows: W = U = total charge, Q, transferred × average voltage between the plates Q(V0 + V) where V0 = initial voltage = 0 V Thus, U = 2 1 Or, U = 2 QV...............(18.5) Getting it right! The work done in transferring a small charge, q, across a fixed voltage V is qV. However, the work done in charging a capacitor from 0 V to a final voltage V in which a total charge of Q is transferred, is ½ QV.
C= Q V substituted into equation 18.5 gives 1 U = 2 CV2...............(18.6) + + − + + − + + − − + − + − + + − + polarised molecules charged plate Figure 18.3 Polarisation of molecules of a dielectric due to charging of a capacitor. How energy is stored within a parallel-plate capacitor Let us now consider a parallel-plate capacitor in which the dielectric is free space. (An air capacitor comes close to this type of capacitor, since the permittivity of free space is practically equal to that of air.) The field E within such a capacitor is very nearly uniform and is given by V = Ed (equation 17.17). εA 1 Since U = 2 CV2 and C = 0 d 2 1 ε ε0A(Ed) Then U = 2 0 d 1 i.e. U = 2 ε0E2Ad...............(18.8) and Q 2 U= ....(18.7) 2C Equations 18.5, 18.6 and 18.7 apply not only to parallelplate capacitors but to all capacitors. Worked example 18.4! Cardiac defibrillator Q A cardiac defibrillator contains a capacitor that is used to send an electric shock to a heart that is fibrillating (beating erratically). If the capacitor is to be charged to 500 V and must deliver 1 000 J of energy in the electric shock, what capacitance would be required of such a capacitor? A W 1 V = 500 V, energy, U = 1000 J capacitance, C = ? 1 U = 2 CV2 2U V2 2 × 1000 = 5002 = 0.008 F = 8000 μF (which is a very large capacitance!) C= Equation 18.8 tells us that the energy of the capacitor is stored in the electric field, E, between the plates. The quantity Ad is the volume of the space between the plates. Hence the energy stored per unit volume within the space, i.e. within the field, is given by: U = 1 ε E2 ....(18.9) Ad 2 0 If there is a material dielectric filling the space between the plates, the molecules of the dielectric become polarized during the charging process (Figure 18.3).
Orbital electrons spend most of their time nearer to the positive plate. This leaves the sides of the molecules next to the negative plate in a positive state for most of the time. Since work has to be done to bring about this polarization, electrical energy would be stored in the polarized molecules. Current through a capacitor Figure 18.4a shows a battery, B, connected to a fixed capacitor, C, having parallel plates, P and Q. 281 282 Unit 2 Module 1 Electricity and magnetism a b B P P P Q Q C Figure 18.4 Flow of charge through a capacitor. As soon as the switch is closed, the positive end of the battery attracts electrons from plate P leaving P with a net positive charge.
We say that conventional current is flowing in the circuit, which then is, by convention, from B to C and its direction is represented by an arrow. At the same time there is conventional current from Q to the battery, resulting in Q having a net negative charge. The flow of charge is very brief. The flow stops when the capacitor is charged to the potential difference of the battery, since both potential differences would be equal but opposite to each other. If the battery were reversed (Figure 18.4b) a similar situation would occur, with a momentary conventional current as in Figure 18.4a but in the opposite direction, as shown by the arrows. P becomes negatively charged and Q positively charged. For practical purposes, since the flow of charge is so brief in both of the above cases, we can say that the capacitor effectively blocks direct current (d.c.). If the battery could be reversed repeatedly, the capacitor would be connected to an alternating voltage. (The symbol for alternating voltage is .) Charge would now flow back and forth along the connecting wires, as shown by the arrows in Figure 18.4c. As long as the voltage is alternating, the current would also be alternating.
We therefore say that a capacitor allows a continuous alternating current in a circuit but blocks a direct current.
The 'resistance' (called 'capacitive reactance') offered by a capacitor depends on the frequency f of the alternating current. The reactance, Xc, is given by Xc = 1 2πfC ....(18.10) Thus the magnitude of the alternating current in any branch of an electric circuit can be controlled by using a suitable capacitor and taking the frequency into consideration. ITQ 3 Would the reactance of a capacitor be greater or less for higher frequency currents? Explain your answer. Capacitors and resonant circuits A capacitor connected in parallel to a coil forms a resonant circuit. At a particular frequency f, called the resonant frequency, the circuit current is a maximum. A variable capacitor allows the circuit to be 'tuned' to desired resonant frequencies. By this means, a radio can be 'tuned' to 'select' a particular frequency of electromagnetic waves even though it is receiving waves of various frequencies. The broadcast frequency signal corresponding to the resonant frequency will be heard the loudest. Capacitor construction The equation for a parallel-plate capacitor (C = εA/d) is a useful guide in constructing capacitors.
Generally, to produce a capacitance that is large: ■ the electrical permittivity of the dielectric must be large ■ the common 'area of overlap' must be large ■ the separation of the conductors (thickness of the dielectric) must be small. Parallel-plate capacitors (variable capacitance) Variable capacitances are usually employed in 'tuning' electronic circuits, such as radio receivers circuits, to particular resonant frequencies. A variable air capacitor makes use of parallel plates; one set of plates is fixed and the other set is moved between the spaces of the fixed plates.
The capacitance decreases or increases as the common area of overlap increases. If a smaller physical size of capacitor is needed (as in some portable radios) the same range of capacitances can be obtained by using a suitable dielectric between the plates, or by making the distances between overlapping plates smaller. In certain uses, precise pre-set capacitances that are not easily accessible to users are needed. Small variable 'trimmer' capacitors are employed. The capacitance is adjusted at the factory by placing and turning a non-metallic screwdriver in a slot of the shaft connected to the moveable plates. Chapter 18 Capacitors a c b protective coating electrode connecting terminals key movable plate ceramic disc (dielectric) hold-offkink metallic electrodes dielectric ceramic connecting wire For specialized high-voltage applications, such as high-power broadcasting and military use, a vacuum must be used between the plates.
If there is air present, arcing can take place. Solid dielectrics also cannot withstand very high voltages. This is because, following polarization, electrons can be ripped off atoms, leading to a 'breakdown' current in the dielectric. The capacitance is changed by sliding concentric metal plates up or down a metal cylinder with which they are not in physical contact. A tight seal is needed on a vacuum capacitor. Parallel-plate capacitors (fixed capacitance) Figure 18.5a shows the construction of a single layer ceramic disc capacitor. It can offer much larger capacitance than a similar sized air capacitor due to the very high permittivity of the dielectric, which is often barium titanate. Single layer ceramic capacitors can withstand fairly high voltages (e.g. 100 V) before dielectric breakdown. Figure 18.5b shows the construction of a multilayer ceramic chip capacitor (MLCC). Through dielectric thicknesses of the order of 0.5 μm, and the stacking of a thousand or so capacitor layers in parallel, both miniaturization and high capacitances of the order of 100 μF have been achieved.
The capacitances of the MLCCs are, however, temperature and voltage dependent. The capacitances also change with time (i.e. they 'age'). Figure 18.5c shows a computer key capacitor in the unpressed condition. When the key is depressed, the rather elastic insulator shrinks, changing the capacitance of the capacitor. This change is detected electronically. Electrolytic capacitors Electrolytic capacitors have large capacitances per unit volume. This is because the dielectric can be made a few molecules thick through electrolytic deposition and capacitance is inversely proportional to the thickness of the dielectric. capacitor flexible insulator fixed plate Figure 18.5 Parallel-plate (fixed capacitance) capacitors. An aluminium (or tantalum) foil is coated electrolytically with an oxide that is a few molecular layers thick. The aluminium itself is called the anode and the oxide is the dielectric. A strip of paper, rolled in the electrolyte used during coating of the anode, is sandwiched between the coated anode and another strip of aluminium. Together the paper and the second strip of aluminium form the cathode. The sandwiched assembly is rolled up and placed in a suitable container. Electrolytic capacitors, depending on their construction, can be made to operate to about 500 V, and can produce capacitances up to about a few hundred microfarads.
However, they are easily destroyed if the anode is connected to the negative side of a potential difference across the terminals or the cathode is connected to the positive of a potential difference across the terminals. For this reason, the terminals on such capacitors are marked; usually only the cathode terminal is marked, using a '−' sign. Electrolytic capacitors possess a higher equivalent series resistance (ESR) than other capacitors and therefore dissipate more heat when current is passing through them.
On account of their large capacitance, electrolytic capacitors are used for 'smoothing' d.c. voltage pulses, such as are generated in the outputs of rectifier diodes (Figure 18.6), into small 'ripple voltages'. As current is being used by a device, the output voltage across the capacitor hardly drops during discharge between pulses due to the large capacitance, i.e. due to the large storage of charge, which results in the voltage across the capacitor being held fairly steady. The large capacitances of electrolytic capacitors provide relatively opposing low 'reactances' (see equation 18.10) to a.c. audio frequency signals. Hence, alternating voltage signals are coupled to stages within an amplifier ITQ 4 Why does a vacuum capacitor carry a tight seal? 283 284 Unit 2 Module 1 Electricity and magnetism a smoothing capacitor b V + capacitor voltage V, or a.c. charging load resistor discharging rectified voltage before smoothing – time, t through suitable electrolytic capacitors. If alternating voltages are superimposed on d.c. voltages (a condition called 'biasing'), care is taken to ensure that the stages are properly d.c. biased so that the net voltage across the capacitor is never large enough to cause reverse voltage breakdown. The capacitors will block d.c. bias voltage from stage to stage while allowing a.c. to pass. Parallel arrangement In the parallel arrangement of capacitors (Figure 18.7b) each capacitor receives the same voltage. The charge on each capacitor is different (assuming the capacitances are different). For this parallel arrangement V = V1 = V2 The net charge, Q, on a single capacitor equivalent to the parallel combination of capacitors is given by Arrangements of capacitors Q = Q1 + Q2 Capacitors can be arranged in series or in parallel (Figure 18.7). In the series arrangement of capacitors (Figure 18.7a), each capacitor in series receives the same charge, Q. This happens because charge is transferred only from the outer plate of the first capacitor to the outer plate of the last capacitor (Figure 18.7). Each capacitor plate receives the same magnitude of charge but the voltages V1, across C1, and V2, across C2, are different (if C1 and C2 are different), such that But since, by equation 18.2, Q = CV, then CV = C1V1 + C2V2 i.e. CV = C1V + C2V (since V = V1 = V2) Thus: C = C1 + C2...............(18.12) V = V1 + V2 The effective capacitance, C, of the combination is given by C = Q/V (by definition of capacitance, see equation 18.2). Hence V= Q C Applying V = V1 + V2, we get Which gives us 1 1 1 = + C C1 C2 +Q C1 –Q...............(18.11) +Q V1 b C2 –Q V2 C1 +Q1 –Q1 V for resistors in parallel. 1 = 1 + 1 . R R1 R2 Using actual figures it can be shown that the effective capacitance of capacitances in series is less than each capacitance and that in parallel is more than any of the capacitances. V1 C2 +Q2 - Q2 + Getting it right! Note that the capacitor arrangement equations are opposite to the resistor arrangement equations: 1 = 1 + 1 , whereas for ■ For capacitors in series, C C1 C2 resistors in series, R = R1 + R2. 1 1 ■ For capacitors in parallel, C = C1 + C2, whereas Q Q V = + C2 C C1 Q Figure 18.6 (a) A capacitor being used in a smoothing circuit. (b) d.c. before and after smoothing. ITQ 5 What are the equivalent capacitances of a 10 μF capacitor and a 50 μF capacitor connected (a) in series and (b) in parallel? - V2 + V+ - Figure 18.7 Capacitors connected (a) in series and (b) in parallel. ITQ 6 What are the maximum and minimum capacitances of four 10 μF capacitors connected together? Chapter 18 Capacitors Worked example 18.5: Capacitors in series and parallel combinations Q Three capacitors are connected as shown in Figure 18.8 to a 12 V d.c. supply and are fully charged.
Calculate: (a) the net capacitance, CT, of the combination (b) the net charge, QT, of the combination (c) the charge, Q1, on the 2.0 μF capacitor (d) the charge, Q2, on the 4.0 μF capacitor. 2.0 μF C1 C3 4.0 μF A 6.0 μF 12.0 V C2 Figure 18.8 Capacitors in a series and parallel combination. (a) C1 and C2 are in series. Hence their net capacitance, C12 is given by 1 1 1 = + C12 C1 C2 1 1 1 = 4.0 μF C12 2.0 μF 4.0 μF 3 = 1.3 μF C12 and C3 are in parallel. Hence their net capacitance, CT, is given by CT = C12 + C3 = 1.3 μF + 6.0 μF = 7.3 μF (b) The net charge, QT, of the combination in Figure 18.8 is given by QT = CTV (since, by definition, C = Q/V) = 7.3 × 10−6 F × 12 V = 88 × 10−6 C = 88 μC (c) The charge, Q3, on C3 = C3V Q3 = 6.0 μF × 12 V = 72 μC Therefore the charge, Q12, on the combination, C12, is given by Q12 = QT − Q3 = 88 μC − 72 μC = 16 μC Since C1 and C2 are in series, the charge in this combination will be the charge on each capacitor. Hence, Q1 = 16 μC (d) By the preceding argument, Q2 (the charge on C2) = 16 μC also. C12 = a b charge + + + C discharge R C – – – − R μA μA Figure 18.9 (a) The circuit for charging/discharging a capacitor. (b) RC discharge circuit. Capacitor discharge Figure 18.9a shows a circuit for charging and discharging a capacitor.
Figure 18.9b shows a capacitor having a charge Q0 and an initial voltage V0 in series with a resistance R. When the switch is closed, electrons flow through R and the microammeter to the positive plate, neutralizing the positive charge and causing the capacitor to become discharged. We represent the discharge current at any time t as I, the charge as Q, and the corresponding voltages on C and R as VC and VR, respectively. Since there is no battery in the circuit, the e.m.f. = 0. Hence the sum of the potential drops around the circuit is zero: VC + VR = 0 For a small change, ΔQ, in Q, taking place in a change in time, Δt, and applying the Ohm's law equation to R and the definition of capacitance to VC, we get Q + IR = 0 C since I = ΔQ Δt we get Q ΔQR + =0 C Δt or ΔQ − Q Δt RC or ΔQ −Δt = Q RC ....(18.13) 285 286 Unit 2 Module 1 Electricity and magnetism Equation 18.13 is converted to calculus form and solved, so the solution becomes t −(RC Q = Q0 e ....(18.14) When time t = RC, equation 18.15 becomes RC V = V0 e−(RC) = V0 e−1 i.e. V = V0/e ....(18.16) where e is a special number in mathematics whose natural logarithm is 1 (i.e. ln e = 1). The value of the number, e, is 2.73... Common logarithms use the number 10 as the base while natural logarithms use the number e as the base (see Appendix 3). Equation 18.16 tells us that the voltage across the capacitor drops to 0.37 of its initial value (or drops by 0.63 of its initial value) when the time t = RC. This time is called the time constant, τ, of the circuit: Getting it right! Do not confuse the number e (2.73...) with the symbol e (= 1.6 × 10−19 C, which means the value of the elementary charge, i.e. the magnitude of the charge on an electron).
For each successive time constant, the voltage drops by the same fraction.
Thus, in Figure 18.10, at the end of the first time constant, the voltage has dropped from V0 to V1 = 0.37 V0. At the end of the second time constant, the voltage has dropped from V1 (= 0.37 V0) to V2 = (0.37 V1) , i.e. to 0.37 × (0.37 V0). The curve is exponential, meaning that in equal times each value of V drops by the same fraction. In the syllabus, and in some textbooks, 'exp' is used in the place of the number 'e'. t e−(RC) is equivalent to exp(− t) RC Since Q = CV, i.e. V = Q/C, equation 18.14 becomes Q Q t = e−(RC) C C t V = V0 e−(RC) ....(18.15) Time constant (τ) A sketch of the graph of capacitor voltage, V, against time of discharge is shown in Figure 18.10. The graph shows that not only does the voltage decrease with time of discharge, but also the rate of decrease of capacitor voltage decreases with time.
The rate of decrease of voltage with time is the slope of the V versus t graph.
V0 The graph of Figure 18.10 can be used to determine the value of the capacitance of a capacitor, if the resistance, R, is known. The value of the time constant τ (= RC) is read from the graph at a point corresponding to V = 0.37 V0. τ if R is known. Since τ = RC then C = R In the practical determination of C, a very high impedance voltmeter (e.g. a digital voltmeter) is placed across the capacitor, so as not to draw too much of the discharge current during measurement. Several values of C are obtained by using several resistances, R, and plotting V versus t for each. The time constants for each R can be obtained from the graph, from which an average value of C can be determined, since the various values of R are known. Alternatively, for a particular resistance used, the average values of several time constants can be obtained from a single graph, from the several values of V read off, using the relationship given by equation 18.15. This yields 0 0 6 t ln V= ln V0−(RC) ln V V1 = V1 = 0.37V0 0.4 V2 = 0.37V1 = 0.14V0 τ = RC...............(18.17) Since, by two rules of logarithms, ln (AB) = ln A + ln B and ln ect = ct (where c is a constant), we can rewrite the above equation as 0.2 t 0 RC t 2RC Figure 18.10 Exponential graph of discharge voltage versus time for a capacitor. ln V = ln V0 + ln e−(RC) = ln V0 + − ln V = − 1 t which, rearranged, gives RC 1 t + ln V0 RC ....(18.18) See Appendix 3 (Logarithms) for more information. ITQ 7 Why is τ called a time constant? Chapter 18 Capacitors Charging of a capacitor ln V 0 Figure 18.12a shows a circuit that can be used for charging a capacitor. R and C represent the resistance and capacitance, respectively, in the circuit; ξ, VR and VC represent the battery e.m.f., the voltage across the resistor and the voltage across the capacitor, respectively.
When the switch is closed, charging of the capacitor begins. For the circuit, 1 slope = – RC ln V ξ = VR + VC i.e. ξ = IR + t Figure 18.11 Graph of ln V against t for equation 18.18. Equation 18.18 is of the form y = mx + c and a sketch of the graph of ln V against time, t, is therefore a straight line, −1 as shown in Figure 18.11. The slope of the line is RC and can be used to determine C if R is known. If equation 18.15 is divided by R we get i.e. ξ = ΔQ R Q + Δt C i.e. ξ − ΔQ R Q − =0 Δt C ....(18.21) Equation 18.21 can be converted to calculus form and solved in terms of Q. The solution is t Q = Qmax [1 − e−(RC)]...............(18.22) t V V0 e−(RC) = R R Since I = V/R, by Ohm's law, this equation becomes t I = I0 e−(RC)...............(18.19) where I represents the discharge current in the circuit shown in Figure 18.9b. As with the equation for t voltage (V = V0 e−(RC)) we can plot I against t (using equation 18.19) to determine C, using the time constant corresponding to I = 0.37 I0 and the known value of R. Taking natural logarithms on both sides of equation 18.19, we get ln I = ln I0 − Q C 1 t RC ....(18.20) As with equation 18.18 (ln V = −(1/RC)t + ln V0), we can use the slope of equation 18.20 to compute C, if R is known. Review questions 13–15 give some practice in using capacitor discharge equations graphically, including the use of natural logarithms. Some practical exercises in determining capacitance by capacitor discharge curves are also given in the 'Practical activities and challenges' section at the end of the chapter. ITQ 8 Estimate the value of the capacitance used in Figure 18.12b, if the resistance, R, = 106 Ω. where Qmax is the maximum value obtained for Q. Since Q is directly proportional to VC, we can rewrite equation 18.22 as t VC = VCmax [1 − e−(RC)]...............(18.23) Or, since the maximum value of VC is the e.m.f., ξ, VC = ξ [1 − e−(RC)]...............(18.24) t A graph of VC against t is shown in Figure 18.12b. When t = τ, the time constant (= RC), equation 18.24 becomes VC = ξ [1 − e−(RC)] RC i.e. VC = ξ [1 − e−1] i.e. VC = ξ [1 - 0.37] which gives: VC = 0.63 ξ...............(18.25) The time constant during charging is defined as the time taken for the capacitor voltage to reach 0.63 of its maximum, the maximum being the e.m.f. of the battery. From the graph of VC against t, based on equation 18.24, the time constant, τ (= RC) can be determined using the value of t when VC = 0.63 ξ. If R is known, C can be calculated using C = τ/R. 287 288 Unit 2 Module 1 Electricity and magnetism a ζ = 10 V b 10 8 6 Vc / V R C VR VC 4 2 0 0 5 Figure 18.12 (a) The circuit for charging a capacitor. (b) The corresponding graph of VC versus t during charging. 10 t/s Summary ✓ A capacitor is a device for storing charge. A capacitor consists of two conductors with a non-conductor (either an empty space or an insulator) between the conductors. The non-conductor is called a dielectric. ✓ The equation, C = εA d (equation 18.3), can be used as a guide in the construction of capacitors with high capacitance but small physical dimensions. ✓ The net capacitance of capacitors connected in series is given by: ✓ The capacitance of a capacitor is defined as the charge it stores per unit potential difference between its two conductors: Q C= ....(18.2) V ✓ The capacitance C of a parallel-plate capacitor with common area of overlap of plates A and dielectric of thickness d and permittivity ε is given by C= εA d ....(18.3) ✓ The dielectric constant κ of a dielectric of ✓ Equations for capacitor discharge in a circuit in the absence of a battery are given by: t Q = Q0 e−(RC)...............(18.14) t V = V0 e−(RC)...............(18.15) t I = I0 e−(RC)...............(18.19) ....(18.4) ✓ Capacitors can store electrical energy, block d.c. while allowing a.c. to pass, offer varying resistances ('reactances') depending on the frequency of the a.c., and can form part of a resonant circuit. U = U = U= 1 2 QV...............(18.5) 1 2 2 CV ....(18.6) Q 2C 2 ....(18.7) In natural logarithmic form: 1 ln Q = − t + ln Q0 RC ln V = − 1 t + ln V0 RC ....(18.18) ln I = − 1 t + ln I0 RC ....(18.20) ✓ The time constant τ (= RC) for discharge is the time taken for the capacitor voltage (or circuit current or charge) to drop to 0.37 (or by 0.63) of its original value. ✓ The energy U stored in a capacitor of capacitance C, having a charge Q and voltage V is given by: ....(18.11) C = C1 + C2 + C3...............(18.12) permittivity ε is the ratio of the permittivity of the dielectric to the permittivity ε0 of free space: ε κ= ε 0 1 1 1 1 = + + C C1 C2 C3 and for a parallel arrangement: ✓ For the charging of a capacitor VC = ξ [1 − e−(RC)]...............(18.24) t ✓ The time constant τ (= RC) for charging is the time taken for the capacitor voltage to reach 0.63 of the maximum voltage (the latter being the battery voltage). Chapter 18 Capacitors Review questions Capacitance 1 (a) What is meant by the term capacitor? (b) Define capacitance. (c) When −20 μC of charge are transferred from one uncharged metal rod to another, the potential difference between them is 6.0 V. What is the capacitance of the arrangement? 2 A 3.0 V battery, in series with a resistor, is connected to an uncharged capacitor. The capacitor becomes fully charged in 12.0 s. If the average charging current is 2.0 mA, determine (a) the charge on the capacitor (b) the capacitance of the capacitor [Hint: when the capacitor is fully charged, current stops. Hence there is no potential drop across the resistor and therefore the potential difference across the capacitor is equal to the potential difference across the battery.] Factors affecting capacitance 3 (a) Define dielectric constant, κ, of a material in terms of permittivity. (b) Estimate the capacitance of a parallel-plate capacitor, of size 28.0 cm × 21.8 cm, made using a 0.02 cm thick sheet of vinyl plastic as the dielectric. [Hint: refer to Table 18.1 and the value of ε0, along with the equation for the capacitance of a parallelplate capacitor.] 4 The air in a 0.001 μF parallel-plate air capacitor, charged to 12 V, is replaced with mica of the same thickness, after the charging source has been removed. What will be the new voltage across the plates? Capacitors in practical usage; Capacitor construction 5 (a) Suppose you were asked to make a parallel-plate capacitor that can fit into a small box but must have a very large capacitance.
Explain how this can be done. (b) Why are electrolytic capacitors capable of extremely high capacitances? What are the limitations to using them. 6 (a) When a capacitor is charged, it stores energy. 1 Derive the formula, U = 2 CV2 for energy stored in a capacitor charged to a potential difference V. (b) A camera flash unit contains a capacitor that is charged to 100 V. If the charged capacitor must release 2.0 J of energy to the camera flashbulb in 0.002 s, calculate: (i) the capacitance needed (ii) the power of the flash. 7 Derive the formula for the energy density (energy stored per unit volume) of an air capacitor: 1 energy per unit volume = 2 εE2 (where E is the electric field between the plates of the capacitor). 8 A parallel-plate 0.002 μF air capacitor is charged to 100 V. (a) Estimate the energy stored in the charged capacitor. (b) A sheet of mica of the same thickness as the air between the plates, replaces the air after the charging source is removed.
Estimate: (i) the new capacitance (ii) the energy stored in this new arrangement. (c) Were the answers to (a) and (b)(ii) the same? Give an explanation. 9 A thundercloud with a fairly flat bottom of area 1.2 square km is approximately 1.0 km above the surface of a flat portion of the Earth. The bottom of the cloud is positively charged. Estimate: (a) the capacitance formed by the bottom of the cloud and the Earth arrangement (b) the energy stored when the cloud is charged to 2 000 000 V. (c) the lightning current, if a lightning discharge between the Earth and cloud takes place in 1.0 s. Arrangements of capacitors 10 With the aid of diagrams, derive the formulae for the net capacitance, C, of two capacitors C1 and C2: (a) in series (b) in parallel. 11 Capacitors of capacitances C1, C2 and C3, shown connected to a 6.0 V supply in Figure 18.13, each have the same capacitance 10 μF.
C1 C3 C2 Figure 18.13 6.0 V 289 290 Unit 2 Module 1 Electricity and magnetism Find: (a) the net capacitance, C, of the arrangement (b) the net charge stored in the arrangement (c) the charge, Q1, stored by the capacitance, C1 (d) the energy, U, stored in capacitor, C2. 12 A 30 cm × 30 cm parallel-plate capacitor has a plate separation of 2.0 mm. If a 15 cm × 30 cm sheet of rubber and a 15 cm × 30 cm sheet of polystyrene, each of thickness 2.0 mm, are placed so as to fill the space between the plates (see Figure 18.14). Estimate the capacitance of this arrangement. [Hint: treat the capacitances due to each dielectric as being in parallel.] rubber 15 (a) Complete Table 18.2 to show values of ln I. The first row is done for you. (b) Plot a graph of ln I versus time, t. (c) Use the slope of the graph in part (b) to determine the time constant of the discharging, and hence calculate the value of C.
(Assume R = 2.4 × 104 Ω.) 16 An uninterrupted power supply consists of a capacitance, C, charged normally to 13 V. The effective resistance in the discharging circuit is 1000 Ω. What size of capacitance is needed if the voltage must not drop below 11 V in 60 s? Practical activities and challenges polystyrene Caution! All of these activities should be conducted only in the presence and under the supervision of a suitably qualified person. Figure 18.14 Capacitor charging and discharging (RC circuits) 13 Table 18.2 shows data obtained during an experiment on the discharging of a capacitor through a resistor R, using a circuit similar to Figure 18.9a. Table 18.2 Current, I / μA ln I Time, t / s 90.0 4.50 0.00 70.0 5.81 50.0 13.79 40.0 19.32 30.0 26.01 20.0 35.22 5.0 69.30 Table 18.3 Drop of current / μA 14 (a) Use the graph in question 13(a) to determine the time constant of the discharging.
(b) If the value of the resistance, R, is 2.4 × 104 Ω, use the value obtained for the time constant to determine the capacitance of the capacitor. Time taken for drop / s from 80 μA to 40 μA from 40 μA to 20 μA from 20 μA to 10 μA from 10 μA to 5 μA (a) Plot and draw a graph of current, I, versus time, t, for the discharging. (b) If the graph in part (a) shows an exponential relationship, the same fraction of current will drop in the same time interval. Complete the Table 18.3 to show the times taken for currents to drop by half. Comment on your results as to whether the graph shows an exponential relationship. 1 Make a parallel-plate capacitor, 20.0 cm × 15.0 cm, using metal plates or using aluminium foil and a sheet of paper. (How would you determine the thickness of the paper?) (a) Predict the capacitance of your capacitor. (b) Determine the capacitance of your capacitor either by direct measurement using a 'capacitance meter', or by measuring the charge stored on the capacitor when it is fully charged by a 3.0 V battery. (A circuit such as Figure 18.9a can be used, in which the microammeter is replaced by a charge sensor connected to a data-logger.) 2 Construct a parallel-plate capacitor and use it to determine the permittivity of a substance. 3 Construct parallel-plate capacitors (or use parallel-plate capacitors supplied by manufacturers) and investigate the relationships between capacitance, C, and: (a) common area of overlap, A (b) distance between the plates (thickness of the dielectric), d (c) permittivity of the dielectric, ε. 4 Design a circuit, similar to Figure 18.9b, and carry out relevant measurements to determine the capacitance of a capacitor using the following discharge graphs: (a) voltage, V, versus time, t (b) ln V versus t (c) current, I, versus time, t (d) ln I versus t. Chapter 18 Capacitors 5 It is best for this to be a teacher demonstration rather than a student experiment. Connect a low-voltage a.c. source, e.g. 6.0 V, in series with a suitable resistor (e.g. 1 kΩ). Connect the voltage across the resistor to the Y-input of an oscilloscope and obtain a sinusoidal trace on the screen using a suitable time base. Notice the a.c. voltage displayed on the screen. Connect a bridge rectifier to the output of the a.c. voltage and observe the pulsating d.c. voltage trace on the screen. Connect suitable capacitances (e.g. 20 V, 100 μF) across the rectifier output and observe the 'ripple voltages' as the capacitor 'smooths' the d.c. voltage. Does the peak-to-peak ripple voltage increase, decrease or remain constant as the smoothing capacitance used is increased? a smoothing capacitor V + d.c. a.c. b load resistor peak-to-peak ripple voltage on capacitor output voltage capacitor output voltage – Figure 18.15 (a) The capacitor circuit used in smoothing a pulsating d.c. voltage. (b) The Appearance of a 'ripple' voltage on an oscilloscope screen. Answers to ITQs Answers to Review questions 1 Q = CV = 1 F × 3.0 V = 3.0 C Q 3.0 C = 1.9 × 1019 electrons Q = ne so n = = e 1.6 × 10−19 C (quite a large number!) 1 (c) 3.3 μF 2 No unit.
Since κ = ε , the unit for permittivity would ε0 cancel leaving a pure number. 3 Since XC = 1/(2πfC), then when f is large XC would be small. 4 If the vacuum seal is not tight, then air can enter the capacitor. Although the capacitance would not be changed by much, arcing can occur at moderately high voltages since there is now a material medium (air) from which electrons can be pulled. 5 (a) In series: 8.3 μF (i.e. less than either 10 μF or 50 μF) (b) In parallel: 60 μF (i.e. greater than either 10 μF or 50 μF) 6 Maximum capacitance occurs in a parallel connection: 4 × 10 μF = 40 μF minimum capacitance occurs in a series connection: 1/ [4 × (1/10)] μF = 2.5 μF 7 The physical quantity represented by τ is time. Since τ is the product of two constants (R and C), then τ must also be a constant – hence the name 'time constant'.
8 From the graph, τ = 5 s (since the voltage rises to 0.63 × 10 V in 5 s. Since τ = RC, and R = 106 Ω, then C = τ = 56s = 5 × 10−6 F = 5 μF R 10 Ω 3 (b) 8 nF 9 (a) 11 nF (b) 21 × 103 J (c) 21 × 10−3 A 15 (a) Current, I / μA ln I Time, t / s 90.0 4.50 0.00 70.0 4.25 5.81 50.0 3.91 13.79 40.0 3.69 19.32 30.0 3.40 26.01 20.0 3.00 35.22 5.0 1.61 69.30 (c) 990 μF 291 292 Chapter 19 Current electricity and electrical circuits Learning objectives ■ State Ohm's law and use the equations V = IR, P = IV, P = I2R and P = ρl A Use energy considerations to distinguish between e.m.f. and p.d. Explain drift velocity (v). Derive and use the equation I = nevA for electrons moving in a metal (n = charge carrier density) and also apply such equations to semiconductors and electrolytes.
Sketch the I–V characteristic for a metallic conductor at constant temperature, a semiconductor diode and a filament lamp; explain these characteristics in terms of the variation in resistance of the device. Sketch the variation of resistance with temperature for a thermistor with negative temperature coefficient; examine different types of thermistors and discuss the differences between their R–T characteristics. Solve problems involving terminal p.d. and external load, given that sources of e.m.f. possess internal resistance. Draw and interpret circuit diagrams; consider d.c. circuits involving sources of e.m.f. and resistive circuit elements. Apply Kirchhoff's laws to given circuits. Derive and use the formula for the effective resistance of two or more resistors in series. Derive and use the formula for two or more resistors in parallel. Use the potential divider as a source of variable and fixed p.d. Use the Wheatstone bridge as a means of comparing resistances. ■ Define the term resistivity and use the equation R = ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ V2 to solve problems. R Circuit principles Current electricity powers modern life. Our lighting, heating, transport and communications are all dependent on it. Electricity is used so widely because it is a clean, efficient means for distributing energy almost instantly to wherever it is required. Electric currents in wires can be voltmeter light bulb switch cell variable resistor (rheostat) easily controlled, and numerous devices are available to convert electrical energy into other forms, including light, heat, sound and motion. In previous chapters you studied the properties of static electric charge and saw how electrostatic forces can be interpreted in terms of fields and potentials. In this chapter we are concerned with charge in motion. V ammeter A Figure 19.1 A simple circuit.
Chapter 19 Current electricity and electrical circuits A switch (open) diode lamp switch (closed) resistor fuse thermistor cell variable resistor voltmeter light-dependent resistor battery light-emitting diode V ammeter Figure 19.2 Circuit symbols. Whenever charge is flowing we say there is an electric current; the strength of the current is equal to the rate at which charge flows. A steady current is only possible when there is a complete loop or circuit of a conductor; if the circuit is not complete then any current will rapidly cause charge to build up at the conductor's free ends, until the strong repulsive forces between like charges prevent further charge flow. Figure 19.1 shows a drawing and a circuit diagram for a simple circuit in which a cell is connected to a lamp. The conventional symbols for the cell, the lamp and some other common circuit components are shown in Figure 19.2. The basic quantities that describe the behaviour of the circuit are reviewed in the paragraphs that follow. Electric charge, Q Electrical energy is transferred around the circuit by the movement of electric charge. In a metallic conductor, such as copper wire, free electrons carry the charge. The basic unit of charge is the coulomb (C).
This is the amount of charge that passes a point in the wire in 1 second when the current is 1 ampere (1 A).
Each electron carries a charge of −1.60 × 10−19 C. This means that when there is a current of 1 A in a circuit, more than one billion billion electrons flow by each point in the circuit per second. Electrons are negatively charged and therefore the actual direction of charge carrier flow in the circuit is from the negative terminal of the cell to the positive terminal.
By convention, however, the direction of the current is shown as from positive to negative. This called the direction of conventional current. From the definition of charge above we can see that the quantity of charge, Q, that flows in a time t is equal to the Q current, I, multiplied by the time: Q = It, and so I = t . Electromotive force, E In the circuit in Figure 19.1 the cell is the source of electrical energy. It drives a current through the bulb. The energy given to each coulomb of charge that flows around the circuit is determined by the electromotive force (e.m.f.), E, of the cell. This is measured in volts (V). A power source with an e.m.f. of 1 V transfers 1 J of energy to each coulomb of charge that flows through it. Typical dry cells used to power radios and other portable equipment have e.m.f.s of 1.5 V. They may be connected in series to provide higher voltages.
Current, I The current, I, in the circuit is measured in amperes (A).
As noted above, the current is equal to the rate of flow of electrical charge, I = Q . Typical currents in domestic t electrical appliances range from a few milliamperes (thousandths of an ampere) or less in calculators and clocks, to a few tens of amperes in cookers and heaters. The current flowing in a circuit can be measured with an ammeter.
Potential difference (p.d.), V When electrons flow through circuit components, such as lamps, the electrical energy given to the electrons by the power supply is converted into other forms. A lamp converts electrical energy into heat and light. The energy transferred to other forms per coulomb of charge that flows through a circuit component is equal to the potential difference (p.d.) between its terminals. This can be measured with a voltmeter. Potential difference has the same units as electromotive force and is measured in volts (V). ■ The current in a circuit is measured by connecting an ammeter in series with the circuit. In order not to affect the current measured, the ammeter should have a very low resistance. ■ The potential difference across a circuit component is measured by connecting a voltmeter in parallel with the component. In order not to affect the voltage measured, the voltmeter should have a very high resistance (Figure 19.3). 293 294 Unit 2 Module 1 Electricity and magnetism A V Resistance, R The magnitude of the current that a given e.m.f. produces depends on the resistance of the components in the circuit, including the wires themselves. The ratio of the potential difference across a component to the current that flows through it is defined as the resistance, R, of the component: ....(19.1) or V = IR...............(19.2) Resistance is measured in ohms (Ω). A component which passes a current of 1 A when the potential difference across it is 1 V has a resistance of 1 Ω. The equation V = IR tells us that the higher the total resistance of a component, the lower the current produced by a given potential difference. In many circuits current is controlled by the inclusion of components with specific resistances, these are called resistors. There are many types of resistors.
first digit of value tolerance Tolerances ± 20 % ± 10 % ±5% ±1% gold energy in = energy out number of zeros following the two digits orange Electromotive force and potential difference are both measured in volts, so what is the distinction between them? Both are measures of energy transfer per unit charge as charge moves around the circuit. An e.m.f. is the energy supplied per unit charge to the electric current by an active device in the circuit such as a battery, a solar cell or a generator. The potential difference (sometimes referred to as the potential drop) across a component is the energy transferred per unit charge from the electric current to other forms (heat for example) as charge flows through that component.
Think of e.m.f. as energy in per unit charge and p.d. as energy out. You will see later that the sum of the e.m.f.s in a circuit loop is always equal to the sum of the p.d.s across the components. This is a consequence of the principle of the conservation of energy: black brown red orange yellow green blue purple grey white 0 1 2 3 4 5 6 7 8 9 purple Getting it right! What is the difference between e.m.f. and p.d.?
second digit of value yellow Figure 19.3 Measuring potential difference and current. V =R I Example Values 3 (= three zeros) no band silver gold pink 4 7 000 = 47 000 Ω = 47 kΩ ± 5 % Figure 19.4 The resistor colour code.
sliding contact tube metal bar terminals terminal coil of constantan wire Figure 19.5 A laboratory rheostat. ■ Fixed resistors. Carbon resistors are made from a mixture of carbon, clay and resin. They are used to limit currents and set voltages in electronic circuits. The proportions of the different materials determine the resistance value. The value of the resistor is indicated by a colour code, as explained in Figure 19.4. Wire wound resistors are used when an accurate resistance value is required. They consist of a coil of resistance wire encapsulated in a resin.
■ Variable resistors enable the resistance value to be changed to control the current or voltage. A laboratory rheostat is shown in Figure 19.5. Ohm's law The relationship between the current through a metallic conductor and the potential difference across it was investigated by German physicist George Simon Ohm (1789–1854). A circuit V that could be used to verify the results of Ohm's A experiment is shown in Figure 19.6. resistance wire 2V Figure 19.6 Investigating Ohm's law. Chapter 19 Current electricity and electrical circuits Worked example 19.1: Using V = IR Q A reverse bias forward bias (a) The current in a conductor is 2 A. The conductor has a voltage of 12 V across it. Find the resistance of the conductor. (b) Calculate the voltage across a 1.5 Ω resistor carrying a current of 4 A. (c) Calculate the current in a 5 Ω resistor that has a potential difference of 20 V across it. I –2 V (a) R = I 12 = 2 =6Ω 0 1 V 2 Figure 19.8 I–V characteristic of a semiconductor diode. When the diode is forward biased (p-type material connected to the positive terminal of the voltage source), the current increases rapidly with voltage and so the diode has an increasingly low resistance in the forward direction. If the diode is reverse biased virtually no current flows. The diode has a very high resistance in the reverse direction. (b) V = IR = 4 × 1.5 =6V V R 20 = 5 =4A (c) I = –1 The current, I, in the circuit is controlled with a rheostat and measured with an ammeter. The potential difference, V, across the resistor is measured with a voltmeter. The graph of V against I, shown in Figure 19.7, is a straight line through the origin. This graph demonstrates that voltage is proportional to current for this conductor. Ohm found that, for a metallic conductor kept at constant temperature, and with all other physical conditions constant, the ratio of the voltage to the current is constant. This result is known as Ohm's law. It may be understood as stating that when the temperature and other physical conditions are constant the resistance, R, of a metallic conductor is constant – it is independent of the current through it. It is important to note that Ohm's law is empirical – it is based on experiment.
It applies to most pure conductors under the conditions stated; in particular the temperature must be constant. If the temperature rises, the resistance of metals is found to increase. The resistance of pure semiconductors decreases with increasing temperature. Semiconducting devices, such as diodes and transistors, do not obey Ohm's law at all: a plot of current against voltage for a diode for example (Figure 19.8), is a curve even at a constant temperature. (Semiconductor diodes are discussed further in Chapter 22.) Electrical power equations When the potential difference across a component is V volts, the component converts V joules of electrical energy into other forms for each coulomb of charge that flows through it. The total energy converted, E, when a charge of Q coulombs flows through the component is thus given by: E = VQ...............(19.3) The quantity of charge that flows through the component in time t is given by: V/V Q = It...............(19.4) 0 I/A Figure 19.7 Graph of V against I for a metallic conductor at constant temperature. ITQ 1 The voltage across a forward biased diode with the characteristics shown in Figure 19.8 is increased from 1 V to 2 V. How does the effective resistance of the component change? Does it increase, decrease or remain constant?
295 296 Unit 2 Module 1 Electricity and magnetism Worked example 19.2: Using P = VI Worked example 19.3: Electrical power calculations Q (a) A 6 V flashlight bulb draws a current of 2 A. Calculate the power of the bulb. (b) An electric light bulb is rated at a power of 60 W. What is the current in the bulb when it is connected in a 240 V circuit? How much energy does the bulb convert into heat and light in 60 seconds? Q A (a) P = VI =6×2 = 12 W (b) Rearrange P = VI to find I P I = V 60 = 240 = 0.25 A E = Pt = 60 × 60 = 3600 J Therefore, E = VQ = VIt Power is the rate at which energy is transferred. The power of a process, P, can be calculated from the equation: power = So, P = energy time taken E VIt = = VI t t P = VI The unit of power is the watt (W), where 1 W is a rate of energy transfer of 1 J s−1. The power of an electrical component is the rate at which it converts electrical energy into other energy forms. Power and resistance We have shown above that P = VI.
Combining this relationship with V = IR we obtain two useful expressions for the power of a component in a circuit in terms of its resistance, R. P = VI = IR × I = I2R P = VI = V × V V2 = R R ITQ 2 You wish to increase the heat output from a heating element connected to a fixed voltage supply. Should the element be replaced by one with a higher resistance or a lower resistance? A (a) An electric bulb is marked as '100 W, 20 V'. (i) What is the resistance of its filament at its operating temperature? (ii) What power does the bulb consume if the voltage supplied is 15 V? (Assume the filament resistance does not change.) (b) Calculate the power of a 250 V motor which takes a current of 0.4 A from the supply.
(c) A 120 V television is switched on for 5 minutes. The set takes a current of 0.50 A. How much energy is supplied to it? V 2 R V 2 so R = P 20 2 = 100 400 = 100 = 4.0 Ω V 2 (ii) P= R 15 2 = 4.0 = 56 W (b) P = IV = 0.4 × 250 = 100 W (c) 5 minutes = 5 × 60 s total energy supplied = Pt = VIt = 120 × 0.50 × 5 × 60 = 18 000 J = 18 kJ (a) (i) P= We thus have three equivalent equations for calculating the power of a component with resistance R when the p.d. is V and the current I, P = VI...............(19.5) P = I2R...............(19.6) P= V2 R ....(19.7) Equation 19.6 shows that: (a) for a fixed resistance the heating effect of a current increases as the current increases – in fact the heating effect is proportional to I2, so doubling the current increases the heating effect by ×4 (22) (b) for a fixed current the heating effect is proportional to the resistance. Chapter 19 Current electricity and electrical circuits Resistivity In his experiments, Ohm investigated how the dimensions of a conductor affect its resistance.
A = 1 m2 Ohm found that the resistance is proportional to the length l of the conductor and inversely proportional to the cross-sectional area A: R∝ l A R= ρl A = ρ x1m 1 m2 =ρΩ l=1m Thus: Figure 19.10 The resistance of a unit cube. ρl R= A ....(19.8) The constant of proportionality, ρ, is known as the resistivity of the material from which the wire is composed. The resistivity is a property only of the material, and does not depend on the dimensions of the sample. Resistivity values for a range of materials are listed in Table 19.1. Table 19.1 Resistivities of selected materials at room temperature (20 °C) Material Resistivity, ρ / Ω m aluminium 2.8 × 10−8 carbon (graphite) 3.5 × 10−5 constantan (copper–nickel alloy commonly used as resistance wire) 4.9 × 10−7 copper 1.7 × 10−8 iron 1.0 × 10−7 silicon 6.4 × 102 silver 1.6 × 10−8 tungsten 5.6 × 10−8 The resistivity of a material is equal in magnitude to the resistance between opposite faces of a unit cube of the material (l = 1 m, A = 1 m2) (Figure 19.10). Theory of electrical conduction in a metal Free electrons carry the electric current in a metallic conductor.
These are the valence electrons that have sufficient energy to escape from individual metal atoms and move freely through the metal structure. The free electrons behave like a gas, moving in random directions. When a conductor is connected to a source of e.m.f. an electric field is created within it. This field accelerates the electrons and they gain a component of velocity parallel to the field. Collisions with the vibrating metal ions transfer the additional kinetic energy the electrons have gained to the metal as heat, and it is this mechanism that gives rise to the resistance of the metal. As the electrons flow the energy lost as heat must be constantly replaced by energy from the e.m.f. in order to maintain the current. As the electrons continuously gain kinetic energy from the field, then lose it again through collisions, they acquire an average 'drift velocity' along the length of the conductor. We can use the diagram in Figure 19.11 to estimate this drift velocity. Worked example 19.4 contains some resistivity calculations. Consider a conductor with cross-sectional area A. Let n be the number of electrons per unit volume, e the charge per electron, and v the average drift velocity of the electrons. ITQ 3 Which of the wires shown in Figure 19.9 do you predict would have the greater resistance? Compare the flow of electricity through the wires to the flow of water through a pipe. Is it easier for water to flow through a long thin pipe or a short broad one? In 1 second an electron travels an average distance v in the direction of the current, so all the electrons within the shaded volume will pass the point P. The volume of the shaded region is vA. number of electrons passing P per second = n × volume = nAv Figure 19.9 Different gauge copper wires. ITQ 4 (a) Show that the unit of resistivity is the ohm metre (Ω m). (b) Calculate the ratio of the resistance of an iron wire to that of copper wire with the same dimensions. Since each electron carries a charge e the charge per second = neAv but the charge per second is just the current (I = Q/t) and so I = neAv...............(19.9) 297 298 Unit 2 Module 1 Electricity and magnetism Worked example 19.4: Resistivity calculations Q An experimenter winds a coil from a 3.0 m length of 1.0 mm diameter copper wire. (a) What is the resistance of the coil at room temperature? (b) What length of 0.50 mm diameter constantan wire would have the same resistance? A (a) The resistance is given by: ρl R= A l = 3.0 m wire diameter, d = 1.0 mm = 10−3 m πd 2 A= 4 3.14 × (10−3)2 = 4 = 7.85 × 10−7 m2 from Table 19.1, ρ = 1.7 × 10−8 Ω m Therefore, ρl R = A 1.7 × 10−8 × 3 = 7.85 × 10−7 = 0.065 Ω (b) For the constantan wire: πd 2 A = 4 3.14 × (0.50 × 10−3)2 = 4 = 1.96 × 10−7 m2 ρl Rearranging R = gives us: A RA l = ρ 0.065 × 1.96 × 10−7 = 4.9 × 10−7 = 0.026 m = 2.6 cm An alternative solution to part (b), based on ratios, is as follows: R2A2 R A ρ ρ2 l2 = = 2 2 1 R A R1A1 ρ2 l1 1 1 ρ1 Now, R1 = R2 l A ρ Therefore, l2 = 1 2 1 A1 ρ2 2 A2 d2 Also, = 2 A1 d1 And so: l d 2 ρ l2 = 1 22 1 d1 ρ2 (3.0 × 0.502 × 1.7 × 10−8) 1.02 × 4.9 × 10−7 = 0.026 m = 2.6 cm This method does not use the previously calculated value of R and is therefore independent of any error made in the calculation in part (a). = Chapter 19 Current electricity and electrical circuits A electron copper wire The effect of temperature As the temperature of a metallic conductor is raised, the metal ions vibrate more vigorously; the frequency of collisions between free electrons and the vibrating ions increases. P I v conventional current electron flow Figure 19.11 Electron flow in a conductor. 0.20 Current / A 0.10 0 0 0.5 1.0 Potential difference / V Figure 19.12 Current–voltage graph for a light bulb filament. The resistance of a metal increases with increasing temperature, so if the current is sufficient to cause a temperature rise (in a light bulb filament for example, Figure 19.12), Ohm's law no longer applies. Metallic conductors only obey Ohm's law if their temperature does not change significantly. The graph of current against voltage for the filament shows that as the voltage is increased, the current does not increase in proportion. This is because, as the filament temperature rises, its resistance increases, and therefore the increase in current is less than would be the case for a conductor that obeyed Ohm's law. In contrast to metals, the resistance of semiconductors (silicon and germanium, for example) decreases with increasing temperature (Figure 19.13). This is because as the temperature rises more electrons have sufficient energy to cross the gap from the valance band to the conduction band (see Chapter 22) and so there are more charge carriers (electrons and holes) available to carry current. Thermistors Thermistors are temperature-sensitive resistors.
A variety of materials are used to make thermistors, including metals, ceramics and semiconductors.
There are two main classes: negative temperature coefficient (NTC) and positive temperature coefficient (PTC). Resistance ■ The resistance of a NTC thermistor decreases with increasing temperature (Figure 19.14a). Temperature Figure 19.13 The variation of resistance with temperature of pure silicon. a ■ The resistance of a PTC thermistor increases with increasing temperature (Figure 19.14b). b Resistance ITQ 6 Advanced theory based on quantum mechanics shows that the average speed of the free electrons in a metal due to their random thermal motion is of the order of 106 m s−1. If it were possible to make a movie of the free electrons in a metal would you be able to see whether

or not there is an electric current? Resistance Temperature Resistance Temperature Figure 19.14 Resistance–temperature graphs for thermistors: (a) negative temperature coefficient (NTC); (b) positive temperature coefficient (PTC).
ITQ 5 Given that for copper n = 8.46 × 1028 m−3, calculate the average drift velocity of the electrons in a copper wire with cross-sectional area A = 1.0 mm2, when the wire is carrying a current of 10 A. (e = −1.60 × 10−19 C) ITQ 7 How does the increase in frequency of electron–ion collisions with temperature change the resistance of a conductor? ITQ 8 An electronics design engineer plans to include a thermistor in a circuit to prevent overheating. The thermistor will reduce the current if the circuit temperature rises. Should the thermistor be NTC or PTC? 299 300 Unit 2 Module 1 Electricity and magnetism Circuit calculations and Kirchhoff's laws ideal cell of e.m.f. E terminal p.d. V internal resistance r Figure 19.15 Internal resistance: the internal resistance of a power source is shown as a resistance r in series with the e.m.f., E. The terminal voltage V is less than E. Internal resistance When there is a current then V is less than E due to the potential drop, ir, across the internal resistance. Figure 19.16 shows a circuit consisting of a power source with two resistors wired across its terminals. When analysing a circuit such as this, the usual goal is to predict the current in each component and the potential drop (the voltage) across them. The current and p.d. are linked by the equation V = IR. With this information we can also calculate the power, VI, dissipated by each component. i P E Internal resistance As we have seen, to maintain a current in a circuit there must be a source of electrical energy (a power source). This may be an electrochemical cell, a generator, a solar cell or some other energy source.
The power source provides the electromotive force (e.m.f.) that drives the current around the circuit. The current passes through the power source as well as the other circuit components. Every power source has an internal resistance, r, to the current through it. A fraction of the e.m.f. is needed to overcome this internal resistance. This produces a drop in the voltage available at the terminals of the cell (or other source) when there is a current.
From Ohm's law, the potential drop is equal to ir where i is the current. For the circuit in Figure 19.15, the potential difference across the external resistor is V and is equal to the terminal voltage. This is less than the e.m.f., E, by the amount ir because of the potential drop across the internal resistance of the cell: Thus: E = V + ir...............(19.10)
Worked example 19.5: Internal resistance calculation Q A cell has an internal resistance of 0.30 Ω. Its terminal voltage is 1.48 V when a current of 0.20 A is being drawn from it. What is its e.m.f.? A E = V + ir = 1.48 + 0.20 × 0.30 = 1.48 + 0.06 = 1.54 V i iB iA RB RA Figure 19.16 A circuit for analysis. Circuit calculations are greatly aided by two laws formulated by Gustav Kirchhoff (1824–1887). ■ Kirchhoff's first law states that at any junction in a circuit, the current into the junction must be equal to the current out. The first law is a consequence of the conservation of charge; if the current out did not equal the current in, charge would accumulate at the junction. Applied to junction P in the circuit in Figure 19.16, Kirchhoff's first law tells us that: i = iA + iB ■ Kirchhoff's second law states that in any closed loop of a circuit the sum of the e.m.f.s must be equal to the sum of the potential drops across the components.
The second law follows from the principle of conservation of energy. When the current is steady, the energy input per unit charge moved around that circuit loop (the sum of e.m.f.s) must be equal to the energy output (the sum of the potential drops). Applied to the circuit loop through the cell and RA in Figure 19.16, Kirchhoff's second law is: E = potential drop across internal resistance r + potential drop across RA Since V = IR we have: E = ir + iARA Similarly for the loop through RB we have: E = ir + iBRB Chapter 19 Current electricity and electrical circuits Worked example 19.6: Applying Kirchhoff's laws Q Use Kirchhoff's laws to find the currents i1, i2 and i3 in the circuit shown in Figure 19.17. Assume that the batteries have negligible internal resistance. i1 R1 6 Ω i3 P i2 E1 = 12 V + – A B R3 = 2 Ω R2 = 4 Ω Figure 19.17 Circuit for analysis. – + E2 = –3 V A Applying Kirchhoff's first law at point P gives, i1 = i2 + i3 (i) Let the clockwise direction around a circuit loop be positive. Kirchhoff's second law applied to loop A gives, E1 = i1R1 + i2R2 12 = 6i1 + 4i2 (ii) The second law applied to loop B gives, E2 = i2R2 + i3R3 −3 = −4i2 + 2i3 (iii) Note that the e.m.f. E2 is negative because this battery is connected so as to drive a current in the anticlockwise direction around the loop – it is working against the current flow produced by E1. The p.d. across R2 is negative since it results from a current flowing anticlockwise around the loop. Substituting for i1 from (i) into (ii) we have, 12 = 6i2 + 6i3 + 4i2 = 10i2 + 6i3 (iv) Multiplying (iii) by 3 and subtracting from (iv) we have, 12 − (−9) = 21 = 10i2 + 12i2 = 22i2 Thus, 21 i2 = A 22 Hence from (iii), 21 −3 = −4 × + 2i3 22 i3 = 84 − 66 18 9 = = A 44 44 22 And, 21 9 30 + = A 22 22 22 A useful check is to apply the second law to the outside loop of the circuit, E1 + E2 = i1R1 + i3R3 Now E1 + E2 = 12 + −3 = 9 V (note that E2 is negative because it is connected in the reverse direction to E1).
Using the values calculated above, 30 9 i1R1 + i3R3 = × 6 + × 2 22 22 180 + 18 = 22 198 = 22 =9V which is consistent. i1 = i2 + i3 = ( ) ( ) 301 302 Unit 2 Module 1 Electricity and magnetism Note: it is clear from these equations that the potential drop across the two resistors is equal (iARA = iBRB). This is the case because they are both connected directly across the terminals of the cell. It follows that when resistors are connected across the same p.d. the ratio of the currents in the resistors is inversely proportional to the ratio of R their resistance values: iA = B . iB RA In general, we can summarize Kirchhoff's two laws with the following statements: V A B R1 I V1 I1 A I2 I I3 Worked example 19.6 shows how this convention is applied in practice. Resistors in series and parallel Resistors and other components are frequently connected in series or in parallel in practical applications. It is often necessary to calculate the effective resistance of a series or a parallel combination. Resistors in series The current flowing through the resistors in Figure 19.18 passes from one resistor to the next.
The current, I, is the same in each resistor. Let the combined resistance be R.
The total potential difference V between A and B is given by, V = V1 + V2 + V3 Applying Ohm's law, V = IR, we have, IR = IR1 + IR2 + IR3 IR = I(R1 + R2 + R3) R2 V2 V3 V ■ Kirchhoff's second law: around any closed loop in the Getting it right! Kirchhoff's first law is straightforward, but the second law is more subtle. To apply it correctly the relative directions of all e.m.f.s and p.d.s must be properly accounted for, as well as their size. A reliable method is to define one direction around a circuit loop as positive. If an e.m.f. in the loop drives a conventional current in this direction the e.m.f. is taken as positive; if it drives a current in the opposite direction it is negative. If a p.d. results from a current in the positive direction it is positive, and negative if the current is in the opposite direction. R3 Figure 19.18 Resistors in series. ■ Kirchhoff's first law: at any point in the circuit: Iin = Iout current: E1 + E2 + ... = I1R1 + I2R2 B I R3 Figure 19.19 Resistors in parallel. Thus, R = R1 + R2 + R3 When resistors are connected in series the total resistance to the current is thus the sum of the individual resistance values. See Worked example 19.7. Resistors in parallel In the parallel circuit in Figure 19.19, the total current, I, divides at A and recombines at B.
Since there is more than one path for the current to take, the combined resistance, R, is less than the resistance of any of the individual resistors. The total current is equal to the sum of the currents in the separate paths: I = I1 + I2 + I3 The potential difference V is the same for all the resistors since they are connected in parallel between the same points. Applying Ohm's law, I = V, we have, R V V V V = + + R R2 R3 R R1 And so, 1 1 1 1 = + + R R2 R3 R1 R1 Worked example 19.8 V E = 1.5 V, i = 0.30 A, RA = 6.0 Ω, RB = 12 Ω Given these values for the quantities in the circuit of Figure 19.16, calculate: (a) the currents in RA and RB (b) the internal resistance of the cell (c) the power dissipated as heat by each component (d) the total power delivered to the circuit by the cell. Chapter 19 Current electricity and electrical circuits Worked example 19.7: Series resistance calculation Q Resistors of 2.0 Ω, 4.0 Ω and 6.0 Ω are connected in series and a voltage of 24 V applied across them (Figure 19.20). Find: (a) the resistance of the combination (b) the current (c) the voltage across each resistor. V1 V3 2Ω 4Ω 6Ω Worked example 19.8: Parallel resistance calculation Q Two resistors of 6.0 Ω and 3.0 Ω are connected in parallel, and an e.m.f. of 12 V is applied across them by a battery (Figure 19.21). Find: (a) the total resistance (b) the main circuit (total current) (c) the current through each resistor.
(Assume that the battery has zero internal resistance.) Figure 19.21 6Ω I1 I2 3Ω Figure 19.20 A 24 V (a) R = R1 + R2 + R3 = 2.0 + 4.0 + 6.0 = 12 Ω V (b) I = R 24 = 12 = 2.0 A (c) V1 = IR1 = 2.0 × 2.0 = 4.0 V V2 = IR2 = 2.0 × 4.0 = 8.0 V V3 = IR3 = 2.0 × 6.0 = 12 V As a check we can see that (4.0 + 8.0 + 12) V does equal 24 V. When resistors are connected in parallel, the reciprocal of the total resistance to the current is thus the sum of the reciprocals of individual resistance values. The combined resistance is always less than any of the individual resistance values. See Worked example 19.8. I A 12 V 1 1 1 1+2 3 = + = = = = + = = 6.0 R1 R2 6.0 3.0 6.0 6 Hence R = 3 = 2.0 Ω V (b) I = R 12 = 2.0 = 6.0 A (c) I1 = V R1 12 = 6.0 = 2.0 A V I2 = R2 = 12 = 3.0 = 4.0 A (a) 1 Potential dividers The potential divider Suppose you have a 12 V d.c. voltage supply but you require only 5 V to power an electronic component. How do you obtain 5 V from 12 V? A simple circuit to achieve this is shown in Figure 19.22. ITQ 10 In Figure 19.22, if R1 = 100 Ω, calculate the value of R2 required to set V1 at 5.0 V. R2 V1 V R1 Figure 19.22 Potential divider. V2 303 304 Unit 2 Module 1 Electricity and magnetism In this circuit: V = I(R1 + R2)(i) And V1 = IR1 (ii) Dividing (ii) by (i) we have, V1 R1 = V R1 + R2 So, V1 = R1 + R2 V Therefore, V1 R1 V − V1 = R1 + R2 Figure 19.23 A potential divider circuit used to provide a variable voltage. Similarly, V2 = VR2 = V − V1 R1 + R2 By selecting the resistance values R1 and R2 appropriately, V1 or V2 can be set to any value between 0 V and the supply voltage V.
This simple arrangement is known as a potential divider – the supply voltage is divided between the two resistors connected in series with it. Note: the simple analysis above assumes that the supply has negligible internal resistance and that the resistance of any load connected across V1 is large compared to R1. If this is not the case, the potential drop across the internal resistance and the current drawn by the load will both affect the voltage V1, and must be accounted for in the analysis. In practice it is always wise to check voltage drops across power supplies and components with a voltmeter and adjust the components accordingly. Figure 19.23 shows a potential divider circuit constructed with a laboratory rheostat (a variable resistor). By moving the slider along the rheostat the output potential can be adjusted continuously between 0 V and the supply value.
This circuit could be used, for example, to investigate the V–I characteristics of a light bulb.
Potential dividers in electronic circuits One mode of operation of a p-n junction transistor is as an electronic switch that is turned on by applying a voltage to the base, and turned off by removing the voltage. (The junction transistor is discussed in Chapter 22.) In the case of the silicon transistor shown in Figure 19.24, there is no current in the collector circuit if the voltage between the base and the emitter is less than about 0.7 V. The base - emitter voltage, Vbe, can be set above or below this value with a potential divider, as illustrated. If one of the resistors in the potential divider circuit is a light-dependent resistor or a thermistor, and the value of Vbe Figure 19.24 Transistor switched by a potential divider circuit. Vbe > 0.7 V, light on; Vbe < 0.7 V, light off. aa bb circuit circuit off in daylight daylight and and on when it is dark is dark Figure 19.25 Potential divider-based alarm circuits. B Q P I1 A C G I2 R S D Figure 19.26 Wheatstone bridge circuit. ITQ 11 The designer of circuit shown in Figure 19.25b requires the warning lamp to switch on when the temperature falls below a pre-set value. Should the thermistor have a positive or a negative temperature coefficient? Chapter 19 Current electricity and electrical circuits changes according to light or temperature, the transistor may be switched on or off by a change in light level or temperature. For example, the collector current could then be used to operate an alarm circuit or security light, as shown in Figure 19.25. The Wheatstone bridge The Wheatstone bridge circuit shown in Figure 19.26 provides a simple but accurate method for finding the value of an unknown resistance by comparing it to a standard resistor. High precision standard resistors are wound from Manganin (an alloy of copper, manganese and nickel) resistance wire. A number of resistance coils may be mounted in a 'resistance box'. Switches or brass plugs connect selected coils in series to give a wide range of accurate resistance values.
Four resistances P, Q, R and S are connected in a diamond arrangement as shown. One of these, P, is the unknown, Q is the standard resistance. The ratio of the resistances R and S must also be known and be adjustable. The supply, with e.m.f. E, is connected and R and S are adjusted until there is no deflection on the galvanometer G when the switch is closed. When this null condition is found then, because no current flows through G, we can conclude that the potential at B must be identical to that at D. As shown this means that Q = S and hence P R P=Q×R S thick copper strips P Q B G A D R S In the null condition no current flows through G so points B and D must be at the same potential. This means that the potential drop across P is equal to the potential drop across R. Now P and Q together constitute a potential divider, and so, EP potential drop across R = R+S Thus, P R = P+Q R+S Taking reciprocals, P+Q R+S = P R So, Q S 1+ = 1+ P R And therefore, Q S = P R The metre bridge The metre bridge is a simple practical arrangement for the Wheatstone bridge circuit (Figure 19.27). The resistances R and S are formed by a uniform constantan resistance wire AC, which is 1 metre long. This wire is stretched between brass terminals adjacent to a metre rule. The unknown resistor P and the standard resistor Q are connected between the terminals and a resistance box. The galvanometer is connected between this brass strip and a sliding contact, which can move along the metre wire. The position of the sliding contact is moved until there is no current indicated by the galvanometer. A protective resistor is usually included in the circuit to prevent damage to the galvanometer when the initial attempts to find a balance point are made. It can be by-passed by closing the switch to increase the galvanometer sensitivity, as the balance point is approached. G At the balance point, protective resistor D Figure 19.27 A metre bridge.
C Proof P R length AD = Q S length CD Therefore, P=Q× AD CD 305 306 Unit 2 Module 1 Electricity and magnetism Worked example 19.9: Metre bridge calculation Q In the metre bridge circuit shown in Figure 19.27, a 10.0 Ω standard resistor is wired in position Q. The circuit is found to be balanced with AD = 79.5 cm. Find the unknown resistance P. (b) A second unknown resistor is now wired in parallel with P.
The balance point moves to AD = 29.2 cm. Find the value of X. A (a) When the circuit is balanced P AD = Q AC 79.5 = 100 − 79.5 = 3.88 Therefore, P = 3.88 × Q = 38.8 Ω (b) The combined resistance, Y, of P and X is found from the new balance point, 29.2 Y = 10.0 × 100 − 29.2 = 4.12 Ω Combined resistance Y of resistors P and X wired in parallel is given by, 1 1 1 = + Y P X Thus, 1 1 1 = − X Y P Therefore, YP X = P−Y 4.12 × 38.8 = 38.8 − 4.12 = 4.61 Ω Summary ✓ An electromotive force (e.m.f.) from a power source such as a chemical cell, a dynamo or a solar cell is required to drive an electric current around a circuit. ✓ The e.m.f. of a source is the energy it provides per coulomb of charge that flows through it; e.m.f. is measured in volts (V). 1 V = 1 J C−1. ✓ The current is the rate of flow of charge, I = Qt . ✓ The potential difference (in volts) across a component is the energy converted to heat, light and/or other forms per coulomb of charge that flows through it. ✓ The resistance of a component is the ratio of the potential difference across it to the current: R = V ; V = IR. I ✓ Ohm's law states that for a metallic conductor at constant temperature and with all other physical conditions constant, the ratio of the voltage to the current is constant.
✓ The power (rate of energy conversion) in a component is given by P = VI. ✓ In terms of the component resistance, R, the power is given by P = I2R = V . R ✓ The resistance of a sample of material of length l and cross-sectional area A is given by R = ρl , A where ρ is the resistivity of the material. Resistivity is a property only of the material, and does not depend on the shape or size of the sample. ✓ The current (I) in a metallic conductor is carried by free electrons.
I is related to the number density n of electrons, the electronic charge e, the cross-sectional area of the conductor A and the drift velocity v of the electrons, through the equation I = neAv. ✓ Drift velocities in conventional conductors are of the order of millimetres per second. They are negligible in comparison to typical electron velocities associated with random thermal motion (≈ 106 m s−1). Chapter 19 Current electricity and electrical circuits ✓ The resistance of a metal increases with increasing temperature; the resistance of a semiconductor decreases with temperature. ✓ A thermistor is a temperature-dependent resistor. The resistance of an NTC thermistor decreases with temperature; the resistance of a PTC thermistor increases with temperature. ✓ Power sources have internal resistance. When current is drawn from a source the terminal voltage, V, is less than e.m.f., E, due to the potential drop across the internal resistance r, so that E = V + ir. ✓ Kirchhoff's first law states that at any junction in a circuit the current into the junction must equal the current out. ✓ Kirchhoff's second law states that in any closed loop of a circuit the sum of the e.m.f.s must be equal to the sum of the potential drops: E1 + E2 + ... = I1R1 + I2R2 + ... ✓ When resistors are connected in series the combined resistance is given by: R = R1 + R2 + R3 + ... ✓ When resistors are connected in parallel the combined resistance is given by: 1 1 1 1 = + + R R2 R3 R1 ✓ Two resistors connected in series across a voltage source V constitute a potential divider. The voltage is 'divided' between the resistors such that: VR1 V1 = R1 + R2 and V2 = VR2 R1 + R2 ✓ A Wheatstone bridge is a diamond-shaped arrangement of four resistors. A cell is connected between one pair of opposite corners of the bridge and a galvanometer between the second pair of opposite corners. The bridge is balanced (the galvanometer is not deflected) when the two pairs of adjacent resistors have the same ratio. Review questions Circuit principles, Ohm's law, Electrical power equations 1 Write down the equations that give the relationships between these quantities: (a) current in amperes, p.d. in volts, resistance in ohms (b) energy in joules, time in seconds, power in watts (c) charge in coulombs, current in amperes, time in seconds (d) power in watts, current in amperes, resistance in ohms 2 State Ohm's law. Draw and describe a circuit that could be used to test its validity for a length of resistance wire. Which law may be regarded as more fundamental: Ohm's law or Newton's first law of motion?
Discuss briefly the difference in the status of these laws. 3 An electric bell has a resistance of 4.0 Ω. When it rings, the current through it is 3.0 A. (a) What voltage is being used to operate the bell? (b) What power does the bell consume as it rings? 4 Calculate the resistance of a toy car motor that has a 6 V battery and a current of 0.2 A in its circuit. Suggest two design changes that could be made to increase the motor's power. 5 An electric radiator has a resistance of 50 Ω. (a) What is the current when it is connected to a 200 V outlet? (b) What is the power of the radiator? (c) Calculate the quantity of electrical energy transformed into heat energy when the radiator is operated for 10 minutes. 307 308 Unit 2 Module 1 Electricity and magnetism 6 The capacity of phone batteries is often quoted in milliampere hours (mAh). A battery rated at 1 mAh can supply a current of 1 mA for 1 h, or a current of 1 X mA for X h. (a) Write a general expression for the time (in hours) that a battery rated at P mAh can supply a current of X mA. (b) A 3.2 V battery rated at 1500 mAh powers a phone that draws an average current of 100 mA.
How long will the phone operate without recharging the battery? (c) For the battery in part (b) calculate: (i) the total charge (in coulombs) that passes through the battery as it discharges (ii) the useful energy (in joules) stored by the battery when it is fully charged. Resistivity 7 A 1.0 kg bar of copper is drawn into 0.50 mm diameter wire. Calculate: (a) the length of the wire (b) the electrical resistance of the wire. ■ density of copper = 8.96 × 103 kg m−3 ■ resistivity of copper = 1.68 × 10−8 Ω m 8 A copper cable and an aluminium cable are manufactured with the same resistance per unit length. By considering a 1 m length of each cable (or more generally a length l), calculate the ratio of the mass per unit length of the two cables. Discuss which metal might be the better choice for overhead power lines. Use the internet to check if your recommendation is used in practice for the manufacture of such lines. ■ density of copper = 8.96 × 103 kg m−3 ■ resistivity of copper = 1.68 × 10−8 Ω m ■ density of aluminium = 2.70 × 103 kg m−3 ■ resistivity of aluminium = 2.65 × 10−8 Ω m 9 (a) A length of fine copper wire is connected in series with a length of iron wire of the same dimensions. The two free ends are connected to a 12 V car battery. (Note: this is very dangerous thing to do.) One section of the wire rapidly becomes red hot and melts. Is it the iron or the copper? Explain your conclusion. (b) Discuss what would happen if the two wires described in part (a) were connected in parallel across the battery terminals.
Which wire would heat up more rapidly? Would either or both melt?
V2 will help Hint: the equations P = I2R and P = R you answer this question. ■ resistivity of iron = 1.0 × 10−7 Ω m ■ resistivity of copper = 1.68 × 10−8 Ω m Theory of electrical conduction 10 (a) Show that the current in a metal wire of crosssectional area A is given by I = neAv where n is the number density of free electrons, e the electronic charge and v the mean drift velocity of the electrons.
(b) The number density of electrons in a metal wire of cross-sectional area 2.0 mm2 is 5.0 × 1022 cm−3. The electrons move along the wire with a mean drift velocity of 0.5 mm s−1. Find the current in the wire.
11 When a potential difference is applied to a semiconductor such as pure silicon, current is carried by both electrons and positively charged holes. The rates at which electrons and holes lose energy through collisions differ, and so they have different drift velocities, ve and vh, as the charges flow. In pure silicon the number density of electrons and holes is the same, and at 20 °C is equal to 1.5 × 1010 cm−3. (a) Show that in a sample of pure silicon with cross-sectional area A the current is given by I = neA(ve + vh). (b) Experiment shows that in silicon ve = 3vh. A silicon wafer with cross-section area 1.0 cm2 and thickness 1.0 mm carries a current of 1.0 mA between its faces. Calculate the drift velocities of the electrons and the holes. (c) Given that the resistivity of silicon = 6.4 × 102 Ω m, calculate the potential difference between opposite faces of the silicon sample. (d) Would the resistivity of the sample increase or decrease if the temperature rose? Explain your answer in terms of the temperature dependence of one of the quantities in the equation quoted in part (a). Chapter 19 Current electricity and electrical circuits 12 The thermistor in the circuit shown is used to measure the temperature. The graph shows how the resistance of a thermistor depends on temperature. 4000 18 State Kirchhoff's laws. In the circuit shown in Figure 19.29, two cells of negligible internal resistance are connected in series. Two resistors are connected across the cells as shown. 3000 Resistance / Ω 17 You are provided with three resistors, with values 10 Ω, 20 Ω and 30 Ω. How many different resistance values can be made by combining all three resistors? Calculate their values. 2000 12 Ω 32 1000 0 20 40 Temperature / °C 60 80 (i) i3 A B i2 Figure 19.28 Temperature dependence of the resistance of a thermistor. (a) What is the resistance at 50 °C? (b) Is this an NTC or a PTC device? (c) Is the device likely to be constructed from a semiconductor or a metal? Explain your conclusion. Internal resistance 13 A battery has an internal resistance of 0.30 Ω. The potential difference between its terminals is 1.48 V when a current of 0.20 A is being drawn from it. What is its e.m.f.? 14 If you accidentally short-circuit a 12 V dry battery (as used in a flashlight, for example) there is a relatively small spark between the conductor and the battery terminal. But if you accidentally short-circuit a 12 V car battery, with a spanner for example, there is a much more powerful and potentially dangerous spark, even though the voltage is the same.
Explain this difference. Circuit calculations and Kirchhoff's laws, Resistors in series and parallel 15 A 15 Ω resistor is connected in series with an unknown resistor. The current in the circuit is found to be 0.20 A when a voltage of 40 V is applied. Find the value of the unknown resistance. 16 Two resistors, each 10 Ω, are connected in parallel to a 20 V battery. Find: (a) the total resistance of the circuit (b) the total current flowing through each resistor (c) the voltage across each resistor. State any assumptions you make.
6V 3V Figure 19.29 Circuit diagram for question 18. (a) Use Kirchhoff's first law to write down the relationship between the three current values in the circuit. (b) By applying Kirchhoff's second law to loops A and B of the circuit, find the three current values. (c) How could the resistance values be adjusted to make the current i3 zero? The potential divider, the Wheatstone bridge 19 A variable resistor with a maximum resistance of 200 Ω is connected as a potential divider to control the p.d. in part of an electronic circuit. The current through the resistor is 25 mA. (a) What is the maximum p.d. the potential divider can provide? (b) A p.d. of 1.5 V is required. Calculate the resistance of the section of the resistor that is 'tapped-off' by moving the slider to provide this p.d. 20 Describe the Wheatstone bridge circuit and explain how it can be used to find an unknown resistance. Derive the formula connecting the resistances when a balance point is obtained.
309 310 Unit 2 Module 1 Electricity and magnetism 21 A Wheatstone bridge circuit is constructed with two 100 Ω resistors, a 200 Ω resistor and a 201 Ω resistor. A 2 V cell with negligible internal resistance and a voltmeter with very high resistance are connected as shown in Figure 19.30. 200 Ω 201 Ω V 100 Ω 100 Ω Figure 19.30 Circuit for question 21. (a) Calculate the reading on the voltmeter. (b) Find the value of a resistor that reduces the voltmeter reading to zero when it is connected in parallel with the 201 Ω resistor. Practical exercises and challenges 1 Figure 19.31 shows apparatus that may be used to investigate the temperature variation of the resistance of a wire or a thermistor. resistance box r G Figure 19.31 Apparatus to investigate the variation of resistance with temperature. A metre bridge is used to find the resistance of the sample at different temperatures by comparing it to a standard resistance selected from a resistance box. To obtain an accurate measurement, the resistance box should be adjusted to give a balance point near the centre of the wire. (Can you explain why this is the case?) The unknown resistance is then given by: l R = 1 × resistance box value l2 The temperature of the sample is raised slowly from room temperature to 100 °C by gently heating the test tube in a water bath. The balance point is found and recorded at regular intervals and the corresponding temperatures immediately noted.
Use this method to investigate the temperature dependence of the resistance of various samples.
Record your data in suitable tables and use it to plot graphs showing how the resistance varies with temperature. 2 (a) Construct and test a circuit to turn on a night light when the light level drops at dusk. Refer back to Figure 19.25. (b) Construct a similar circuit to trigger an alarm if the temperature in the room rises above a pre-set limit. Answers to ITQs 1 The effective resistance is given by VI = R. This ratio is not constant for a diode (the V–I characteristic is non-linear) and so the resistance depends on the voltage. The graph shows that doubling the voltage from 1 V to 2 V will more than double the current since R = VI (the resistance) decreases. 2 2 P = VR . If V is fixed the power may be increased by replacing the element by one with lower resistance R. 3 The longer, thinner wire will have greater resistance. RA ρl 4 (a) R = A , so ρ = l . 2 The units of ρ are therefore Ω m = Ω m m (b) Ratio of resistances of samples with the same dimensions is equal to the ratio of resistivities. ρiron 1.0 × 10−7 = = 5.9 ρcopper 1.7 × 10−8 5 v = I neA 10 = 8.46 × 1028 × −1.60 × 10−19 × 1.0 × 10−6 = −0.00074 m s−1 = −0.74 mm s−1 The minus sign shows that the average drift velocity the electrons, which carry a negative charge, is in the opposite direction to conventional current flow. 6 The drift velocity is a factor of 109 smaller than the average thermal velocity of the electrons. It would not be possible to detect this tiny difference by observing the motion of individual electrons. Chapter 19 Current electricity and electrical circuits 7 An increase in the collision rate will reduce the average drift velocity and so reduce the current for a given voltage.
This corresponds to an increase in resistance. 8 To reduce the current and protect the circuit, the thermistor resistance must increase with temperature. A PTC thermistor is required. 9 (a) i = iA + iB (Kirchhoff's first law) E = ir + iARA = ir + iBRB (Kirchhoff's second law) Therefore iARA = iBRB and so iR iA = B B RA Substituting for iA in (i) gives: iR i = B B + iB RA 0.30 = iB × 12 + iB = 3.0iB 6.0 Therefore iB = 0.10 A and iA = 0.20 A (b) E = ir + iARA 1.5 = 0.30r + 0.2 × 6.0 1.5 − 1.2 r= 0.30 = 1.0 Ω (c) Power dissipated = I2R In RA, power = 0.202 × 6.0 = 0.24 W In RB, power = 0.102 × 12 = 0.12 W In r, power = 0.302 × 1.0 = 0.090 W (d) Total power delivered by cell = Ei = 1.5 × 0.30 = 0.45 W (Check 0.24 + 0.12 + 0.090 = 0.45 W: power delivered by cell equals power dissipated in components, which is correct) VR1 10 V1 = R1 + R2 Therefore V1R1 + V1R2 = VR1 R2 = = VR1 − V1R1 V1 12 × 100 − 5.0 × 100 5.0 = 140 Ω (i) 11 The potential drop across the thermistor must increase to switch the transistor on as the temperature falls. This means that the thermistor resistance must increase as the temperature decreases: it must therefore be NTC. Answers to Review questions 3 (a) 12 V (b) 36 W 4 30 Ω 5 (a) 4 A (b) 800 W (c) 4.8 × 105 J 6 (b) 15 h (c) (i) 5400 C (ii) 1.7 × 104 J 7 (a) 570 m (b) 49 Ω 8 mass per unit length Cu : mass per unit length Al = 2.1 : 1 (the mass of the copper wire is 2.1× the mass of the aluminium wire) 10 8 A 11 (b) ve = 3.1 × 103 m s−1; vh = 1.0 × 103 m s−1 (c) 6.4 V 12 (a) 1500 Ω 13 1.54 V 15 185 Ω 16 (a) 5 Ω (b) 2 A (c) 20 V (assuming the battery has negligible internal resistance) 17 5.5 Ω, 8.3 Ω, 13 Ω, 15 Ω, 22 Ω, 28 Ω, 37 Ω, 60 Ω 18 (b) i1 = +0.5 A; i2 = −1.0 A; i3 = +0.5 A (the negative sign for i2 indicates that the current is in the opposite direction to that of the arrow on the diagram) 19 (a) 5 V (b) 60 Ω 21 (a) 2.5 mV (b) 40 kΩ 311 312 Chapter 20 Magnetic fields and their effects Learning objectives ■ Explain 'magnetic flux density' and the 'tesla'. ■ Sketch magnetic flux patterns due to a long straight wire, a flat circular coil and a long solenoid. ■ Use the expressions for the magnetic flux density at a distance r from a long straight wire, at the ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ centre of a flat circular coil and near the centre of a long solenoid Use Fleming's left-hand rule to predict the direction of the force on a current-carrying conductor. Use the equation F = BIL sin θ to solve problems. Explain how the force on a current-carrying conductor can be used to measure the flux density of a magnetic field by means of a current balance. Predict the direction of the force on a charge moving in a magnetic field; use Fleming's left-hand rule and treat the moving charge as an electric current. Describe the trapping of charged particles by magnetic fields, with specific mention of Earth's magnetic field and the Van Allen radiation belt.
Use the expression F = BQv sin θ to solve problems.
Solve problems involving charged particles moving in mutually perpendicular electric and magnetic fields. Describe the effect of a soft iron core on the magnetic field due to a solenoid; compare this effect with that of the dielectric in a capacitor. Explain the principle of the electromagnet and discuss its uses in door locks, switches and other applications. Explain the origin of the forces between current-carrying conductors and predict the direction of the forces. Explain the Hall effect and use the Hall probe to measure flux density. Magnetic fields The most powerful steady-field magnet yet constructed is the 45 tesla hybrid research magnet at the National High Magnetic Field Laboratory in Tallahassee, Florida. The tesla (T) is the unit of 'flux density', the quantity that indicates the strength of a magnetic field. ■ The Earth's magnetic field has a flux density of around 0.000 05 T. ■ A typical fridge magnet produces a flux density of about 0.01 T. ■ A scrapyard magnet capable of lifting a car might have a field of 2 T. ■ The powerful superconducting magnets in a hospital MRI body scanner produce flux densities of around 5 T. When operating, the 45 T magnet consumes 33 MW of electrical power, which is sufficient to run a small city! Permanent magnets The properties of permanent magnets can be described in terms of poles and fields. The law of magnetism is that unlike poles attract, like poles repel. The region surrounding a magnet in which it exerts a magnetic force is called its magnetic field. The magnetic field is shown by drawing lines of force called field lines or lines of magnetic flux (Figure 20.1). These lines have the following properties: ■ a flux line shows the direction of the force on a magnetic north pole placed at that point in the field ■ the strength of the force is indicated by the density of the flux lines; where they are close together the force is strong, where they are widely spaced the force is weaker ■ the flux lines of permanent magnets always start on a magnetic north pole and end on a magnetic south pole Chapter 20 Magnetic fields and their effects current ammeter A + – N S rheostat plotting compass N S cardboard sprinkled with iron filings Figure 20.3 Investigating the field of a straight wire. field Figure 20.1 The magnetic field (flux pattern) of a bar magnet. ■ flux lines never cross, since the magnetic force cannot act in two directions at the same place current ■ when the fields of two magnets overlap, they combine thumb – current fingers – field by vector addition to produce a resultant field. In nature magnetic poles always come in pairs – a north pole and a south pole together. Although a theoretical possibility, natural magnetic monopoles (isolated north or south poles) have not been discovered. The direction of a magnetic field is thus the direction of the force on a hypothetical north monopole placed in the field. Magnetic field due to a current The forces of electricity and magnetism have been known since ancient times. The two forces have many similarities, and for centuries scientists speculated that they were connected in some way. But it was not until the 19th century that the link between electricity and magnetism was established. wire I N S I compass needle no current I needle below wire The study of electromagnetism began in 1819 when the Danish scientist Hans Christian Ørsted (1777–1851) discovered that an electric current creates a magnetic field. Ørsted observed that a magnetic compass needle placed adjacent to a wire carrying a current is deflected. The results of his observations are summarized in Figure 20.2. Figure 20.3 shows an apparatus to investigate the form of the magnetic field around a long straight wire carrying a current.
Flux lines are plotted with the aid of a plotting compass. The flux lines are always at right angles to the wire and form concentric circles around it. The direction of the flux lines relative to a current may be found with the 'right-hand grip rule' illustrated in Figure 20.4. I needle above wire Figure 20.4 The right-hand grip rule. Point the thumb of your right hand in the direction of the current and curl your fingers as if to grip the wire. The fingers indicate the direction of the field around the wire. needle below wire needle above wire Figure 20.2 Ørsted's observations. The deflection of the compass needle depends on the direction of the current and on whether the wire is above or below the needle. The direction of deflection can be predicted with Ampère's swimming rule, which states: if the observer imagines that he or she is swimming along the wire in the direction of the current and is facing the needle, then the north pole of the needle is deflected towards the left hand. ITQ 1 Compare the forces and fields created by electrostatic charges and by the poles of permanent magnets. What are the similarities? Are there any differences? Figure 20.5 shows the magnet fields of currents travelling towards and away from the observer. The symbols and indicate the direction of the current relative to the plane of the page.
Imagine that the first symbol shows the point of an arrow coming towards you whilst the second symbol shows the feathers of the arrow travelling away from you. ITQ 2 How do the flux lines surrounding a current differ from those produced by a permanent magnet? 313 314 Unit 2 Module 1 Electricity and magnetism a b Figure 20.5 The flux patterns of currents travelling (a) towards and (b) away from the observer. parallel with the solenoid axis and uniformly spaced. The direction of the field inside the solenoid can be deduced by applying the right-hand grip rule to the current in one loop of the coil. Near the ends of the solenoid the lines of flux diverge. Outside of the solenoid the flux pattern is similar to that of a bar magnet. One end of the solenoid behaves as a north pole, while the other end behaves as a south pole. The polarity of the coil can be remembered by attaching arrows to the letters N and S as indicated. If the current appears to flow clockwise when the coil is viewed end on, you are looking at the S pole; if the current flow is anticlockwise, you are looking at the N pole. The strength of a magnetic field: flux density The strength of the magnetic field at a point is indicated by the density (the spacing) of the lines of flux at that point. Where the flux density is high the lines of force are closely spaced and the field strength is high. Flux density is given the symbol B and has units tesla (T). As explained below, the tesla is defined in terms of the force on a current-carrying conductor placed in a magnetic field. Figure 20.6 The flux pattern of a flat circular coil. Figure 20.7 The flux pattern of a long solenoid. Figure 20.6 shows the magnetic field due to a flat circular coil. Note that the field lines at the centre of the coil are perpendicular to the plane of the coil. Figure 20.7 shows the magnetic field produced by a solenoid. A solenoid is a long coil made up of many turns of wire.
Inside the solenoid the lines of flux are ITQ 3 Use the right-hand grip rule to confirm that the directions of the field lines shown for the currents in Figure 20.5 are correct. Getting it right! Flux density and magnetic field strength In more advanced treatments of magnetism, the terms 'flux density' and 'magnetic field strength' are reserved for distinct aspects of the magnetic field, in particular when discussing the fields inside magnetic materials. In free space, however, the flux density and field strength are directly proportional to each other, and the two terms are often used interchangeably when discussing fields outside magnetic materials. In this chapter the terms 'flux density' and 'field strength' are taken as equivalent and used interchangeably, as are the terms 'lines of flux' and 'field lines'. Definition of flux density, B Ørsted's experiments showed that a wire carrying an electric current exerts a force on a nearby magnet (the compass needle). From Newton's third law we can deduce that the wire experiences an equal and opposite force. Experiments show that the force on a wire carrying a current in a magnetic field is: ■ proportional to the current I ■ proportional to the length L of the conductor Chapter 20 Magnetic fields and their effects ■ is a maximum when the conductor is at right angles to the field θ ■ increases when the flux density increases.
P r δB δl Based on these results we can define the flux density B with the equation: I F = BIL where F is the force on a wire of length L carrying a current of I amperes placed at right angles to a magnetic field with flux density B. Rearranging we have, B= F IL The flux density (strength of the field) is thus the force per unit current per unit length of conductor at right angles to the field.
This definition of the strength of the field can be compared to the definitions of gravitational field strength and electric field strength you have met previously: gravitational field strength, g, is the force per unit mass (unit N kg−1); electric field strength, E, is force per unit charge (unit N C−1). The unit of flux density, the tesla is defined as follows: The flux density is 1 tesla when the force on a wire carrying a current of 1 ampere at right angles to the field is 1 newton per metre of length (1 T = 1 N A−1 m−1). Quantitative expressions for the flux density of wires and coils The electric field strength E at a distance r from a point charge Q in free space is given from Coulomb's law as: Figure 20.8 The Biot–Savart law.
The flux density δB produced by the current element IδI is a vector. Its direction is at right angles to the plane containing δI and r, and may be found with the right-hand grip rule. In this case you can use the right-hand rule to confirm that δB is into the plane of the page. Getting it right! Permeability and permittivity Comparing the Biot–Savart and Coulomb laws we see that both are inverse square laws (vary as r−2); the Biot–Savart law contains the factor 'sin θ' because a current element has a direction, but the point charge Q that appears in the Coulomb law does not. The μ constants of proportionality are respectively 4π0 and 1 4πε0 , where μ0 is the permeability of free space and ε0 is the permittivity of free space. When you first meet these laws it may seem odd to write these constants in this complicated way; why not use a single letter for the constant of proportionality as we do with the gravitational constant G? The reason lies partly in the historical development of electromagnetism theory and partly in convenience, as including the factor 4π at this stage simplifies many useful equations derived from the Coulomb and Biot–Savart laws. a b δB 1 Q E= 4πε0 r2 A similar law proposed by Biot and Savart gives the flux density δB due to a short length δl of conductor carrying a current I. Figure 20.8 shows the current element Iδl. The flux density at P due to this current element is given by: μ Iδl sin θ δB = 0 4π r2 In this expression θ is the angle indicated between δl and μ r, and 4π0 is a constant of proportionality. The factor μ0 is called the permeability of free space. We shall see later in this chapter that its value (4π × 10−7 N A−2) is fixed by the definition of the ampere. ITQ 4 Express the tesla in terms of base units of the SI system. r I r μ I B= 0 2πr B= μ0 NI 2r I c B = μ 0 nI Figure 20.9 (a) Magnetic flux density at a distance r from a long μ0I straight wire: B = 2πr . (b) Magnetic flux density at the centre of μ0 a flat circular coil of radius r: B = 2r NI, where N is the number of turns in the coil. (c) Magnetic flux density near the centre of a long solenoid: B = μ0nI where n is the number of turns per metre of length of the solenoid (n = N L where N is the total number of turns and L is the length of the solenoid). 315 316 Unit 2 Module 1 Electricity and magnetism Worked example 20.1 Q A (a) A flat circular coil wound from 50 turns with diameter 20 cm carries a current of 0.4 A. Calculate the flux density at the centre of the coil.
(b) The needle of a small compass placed at the centre of the coil in part (a) turns so that its north pole points away from an observer. With respect to the observer, is the current direction in the coil clockwise or anticlockwise? (c) The flux density measured at the centre of a 0.48 m long air-cored solenoid is 0.003 14 T. If the current is 0.6 A, calculate the total number of turns in the solenoid. μ0 = 4π × 10−7 N A−2 (a) At the centre of a flat circular coil, μ0 B= NI 2r r = 10 cm = 0.1 m; N = 50; I = 0.4 A Therefore, μ0 B= NI 2r 4π × 10−7 = × 50 × 0.4 2 × 0.1 = 0.0013 T B = 1.3 × 10−3 T (b) The field direction at the centre of the coil is away from the observer. The right-hand grip rule applied to any part of the coil will show that the direction of the current is clockwise with respect to the observer. (c) At the centre of a long air-cored solenoid, B = μ0 nI And so, B n= μ0I Now the total number of turns, N, is given by, N = nL where L is the length of the coil. Therefore, BL N= μ0I are always in agreement with experiment, the Biot–Savart law has been confirmed indirectly. To use the law, the contributions of all the current elements that make up the circuit to the flux density at point P are added (by integration) to find the resultant field. This involves some lengthy mathematics that is beyond the scope of this text, but some important results are quoted in Figure 20.9. Force on a current-carrying conductor Fleming's left-hand rule We have seen that the force on a straight conductor of length L carrying a current I placed at right angles to a field with flux density B is given by, F = BIL But what is the direction of the force? Experiment shows that the force is always perpendicular to both the current direction and the field direction. The direction may be found by applying Fleming's left-hand rule, as shown Figure 20.10. Since this force may be used to produce motion of the conductor (the principle of the electric motor) it is also known as the motor rule.
motion field current Figure 20.10 Fleming's lefthand rule (the motor rule).
The thumb, first finger and second finger are extended mutually at right angles. If the First finger shows the direction of the Field and the seCond finger the direction of the Current, then the thuMb shows the direction of Motion (the direction of the force on the conductor). ITQ 5 Figure 20.11 shows the rolling rod experiment. The brass rod experiences an electromagnetic force that makes it roll along the parallel conducting rails.
Use Fleming's left-hand rule to predict the direction of motion. 0.00314 × 0.48 4π × 10−7 × 0.6 = 1999 N = 1999 = Unlike a point charge, a small current element Iδl cannot exist in isolation: it must always be part of a complete circuit. For this reason the Biot–Savart law cannot be proved directly; but it can be used to predict the variation in the strength of the magnetic field with distance from a straight wire, and the flux densities inside and outside coils, solenoids or any other circuit. Since the predictions S brass rod N Figure 20.11 Rolling rod experiment. Chapter 20 Magnetic fields and their effects Worked example 20.2 L I θ B Q Calculate the magnitude and the direction of the force on each side of the triangular loop of wire shown in Figure 20.14. F B F = BIL sin θ B = 0.2 T 1 cm Figure 20.13 Force on a current-carrying conductor at angle θ to a field. In this example F is into the plane of the page. 30˚ A Conductor not at a right angle to a field What if a current-carrying conductor is not perpendicular to the field? What force does it experience then? Experiments (first performed by Ampère) show that as the angle θ in Figure 20.13 is decreased from 90°, the force decreases in proportion to sin θ, becoming zero when the conductor is parallel to the flux lines. I=3A Figure 20.14 A Thus in general, F = BIL sin θ The direction of the force is always perpendicular to the field and the current, and may still be found with Fleming's left-hand rule, where now the second finger shows the direction of the component of the current that is perpendicular to the field (Figure 20.13). Measurement of B Figure 20.15 shows how a sensitive laboratory balance may be used to investigate the strength of a magnetic field by measuring the force on a current-carrying conductor of known length.
The effective length, L, of the conductor can be chosen by connecting the external circuit between different pairs of the five vertical wires. The same apparatus could be used to investigate how the force depends on I and θ. If the flux density in the region between the magnet poles is B, then the force on the length L of wire between the poles is given by F = BIL (assuming that the wire is placed perpendicular to the field). Through Newton's third law, the magnet experiences an equal but opposite force to the force on the wire and the reading on the digital balance changes accordingly. ITQ 6 Faraday imagined that the lines of magnetic flux in Figure 20.12 are in tension and tend to shorten to produce a force on the wire. Does the motion predicted by this explanation agree with Fleming's left-hand rule? C N For side AB: L = 1 cm = 0.01 m; θ = 90°; sin θ = sin 90° = 1 Therefore, F = BIL sin θ = 0.2 × 3 × 0.01 × 1 = 0.006 N Fleming's left-hand rule shows that the force is perpendicular to B and I and is directed into the plane of the page. For side BC: θ = 30°; sin θ = sin 30 = 0.5 1 cm 1 cm L = = = 2 cm = 0.02 m 0.5 sin θ F = BIL sin θ = 0.2 × 3 × 0.02 × 0.5 = 0.006 N The current direction is from B to C and thus the component of the current perpendicular to the field is directed downwards. Fleming's left-hand rule applied to this component shows that the force is perpendicular to B and I and is directed out of the plane of the page. For side AC: θ = 180°; sin θ = sin 180° = 0 Therefore, F = BIL sin θ = 0 The electromagnetic force on a current that is parallel or antiparallel (opposite) to the field direction is zero. S Figure 20.12 The resultant field when a current-carrying conductor is placed in an (initially) uniform field. The resultant field is the vector sum of the fields produced by the magnet and the conductor. 317 318 Unit 2 Module 1 Electricity and magnetism Force on a moving charged particle Figure 20.16 shows a particle with charge Q travelling at right angles to a magnetic field B. The particle has velocity v. As shown below, the particle experiences a force perpendicular to its motion and the field given by: A B C copper wire frame F = BQv A v D Magnadur magnet B into page F = BQv Q iron yoke digital balance Figure 20.15 A simple current balance. (Magnadur magnets are ceramic bar magnets with poles on their faces rather than at their ends.) Figure 20.16 Force on a charged particle travelling at right angles to a magnetic field. Proof Worked example 20.3 Q In an experiment with the apparatus shown in Figure 20.15, the balance reading changed from 271.30 g with no current to 271.71 g with a current of 2 A in the circuit. Given that L = 4 cm, calculate the flux density between the magnetic poles. A L = 0.04 m; I = 2 A; weight = mg where g = 9.8 m s−2 Apparent change in m = 271.71 g − 271.30 g = 0.41 g = 0.41×10−3 kg This corresponds to a downward force given by, F = mg = 0.41 × 10−3 kg × 9.8 m s−2 = 4.02 × 10−3 N Now, F = BIL So, F B= IL 4.02 × 10−3 = 2 × 0.04 = 0.05 T ITQ 7 (a) Use Fleming's left-hand rule to predict the direction of the force on the wire in Figure 20.15.
(b) Why can the forces on lengths AB and CD be neglected? A moving charge is an electric current.
In the previous chapter we showed that the current I in a conductor is related to the drift velocity of the charge carriers by the equation I = nQvA, where n is the charge carrier number density, Q is the charge per carrier, v is the velocity and A is the cross-sectional area of the conductor. We have seen that for a current-carrying conductor of length L, Fwire = BIL And so substituting for I we have, Fwire = BnQvAL Since n is the number density and AL is the volume of the conductor, nAL = N, the total number of charge carried in the conductor. Thus, Fwire = NBQv This is the total force on all N charge carriers in the wire and so the force F on a single particle is given by, Fwire F = N NBQv = N = BQv More generally, if the particle is travelling at an angle θ to the field the force is given by, F = BQv sin θ ITQ 8 Use Fleming's left-hand rule to confirm that the direction of the force shown in Figure 20.16 is correct. Don't forget that current direction (shown by your second finger) is the direction of motion of a positive charge. If Q is negative (an electron for example) the conventional current direction is in the opposite direction to the motion. Chapter 20 Magnetic fields and their effects The trajectories of charged particles in magnetic fields How does the electromagnetic force on a charged particle moving in a uniform magnetic field effect its motion? If the field direction is perpendicular to the particle's velocity, the force is perpendicular to the velocity and to the field. As we saw in Chapter 6, the condition for circular motion is a centripetal force that is always perpendicular to the velocity of the moving particle. In these circumstances the particle will follow a circular trajectory, with the plane of its orbit perpendicular to the field lines. An example is an electron beam in a cathode ray tube, which follows a circular path in a uniform magnetic field which is perpendicular to the beam. field converges there is a horizontal component of the electromagnetic force, which decelerates the particle's horizontal motion. If the field becomes strong enough, the particle is reflected and spirals back (still spinning in the same direction) to A where it will be reflected back to B, and so on. The net effect is that the particle is trapped inside a 'magnetic bottle'.
Worked example 20.4 Q An electron is accelerated by a potential difference of 3 kV in a cathode ray tube. It then enters a uniform magnetic field with strength B = 5 mT, and direction perpendicular to the direction of the electron's motion. Calculate: (a) the speed of the electron (b) the force on the electron due to the magnetic field (c) the resultant acceleration of the electron (d) the radius of the circular arc the electron follows in the magnetic field. ■ electron charge, e = −1.60 × 10−19 C ■ rest mass of electron, me = 9.11 × 10−31 kg A (a) In accelerating through a p.d. V the electron gains kinetic 1 energy Ek = 2 mv 2 = eV. Therefore the speed of the electron is given by 2eV v = m Worked example 20.4 provides more details. More generally, a charged particle will have velocity components both perpendicular and parallel to the field lines. In this case the particle's trajectory in a uniform field will be a spiral, with the axis of the spiral along the flux lines (Figure 20.17). There is no force component parallel to the field direction and so the component of the velocity in this direction is constant – each loop of the orbit is displaced from the previous one as the orbiting particle travels with a constant velocity component in the direction of the field. Now consider the particle moving in the non-uniform field shown in Figure 20.18. The force on the particle is greater at either end where the flux density is greater. As the particle approaches the converging (stronger) field at B it spins in tighter loops. Where the magnetic B Figure 20.17 In general, moving charged particles follow a helical (spiral) path in a uniform magnetic field. 2 × 1.60 × 10−19 × 3 × 103 9.11 × 10−31 = 3.25 × 107 m s−1 (b) The force on the electron in the magnetic field is given by F = BQv = 0.005 × 1.60 × 10−19 × 3.25 × 107 N = 2.60 × 10−14 N (c) F = ma, therefore, F a = m 2.60 × 10−14 = m s−2 9.11 × 10−31 = 2.85 × 1016 m s−2 (d) For circular motion, the centripetal acceleration is related to the speed and the radius of the circle by v 2 a = r Therefore, v 2 r = a (3.25 × 107)2 = m 2.85 × 1016 = 3.71 × 10−2 m = 3.71 cm v = 319 320 Unit 2 Module 1 Electricity and magnetism The magnet bottle effect is responsible for the creation the Van Allen radiation belts that extend from 1000 km to 60 000 km above the Earth's surface. Electrons, protons and other charged particles emitted by the Sun (the solar wind) or formed by cosmic radiation are trapped by the Earth's magnetic field (Figure 20.19). Lines of flux converging towards the poles create magnetic bottles. Q F A F F B Trapped particles constantly spiral back and forth between the poles in a matter of seconds. This energetic plasma of charged particles is a hazard to artificial satellites and astronauts. Exposure to the radiation could damage sensitive electronics and is harmful to life. The aurora borealis (Figure 20.20) is a spectacular natural display of shimmering light that may be seen in high northern latitudes. The equivalent lights in the southern hemisphere are called the aurora australis. These lights are created by energetic charged particles from the solar wind spiralling down magnetic field lines into the Earth's atmosphere. Nitrogen atoms ionized by these particles emit light as they recapture electrons. The detailed mechanism of aurora production is still under investigation. Force on a particle moving in perpendicular electric and magnetic fields Figure 20.21a shows a beam of electrons traveling with velocity v between two parallel metal plates. The upper plate is maintained at a potential V with respect to the lower one. The electric field E between the plates is uniform and the field strength is given by Figure 20.18 A magnetic bottle. North pole protons E= V d where d is the plate separation. Each electron experiences a force F = eE = eV . Since d electrons have a negative charge the beam is deflected towards the upper plate, as shown. electrons South pole Figure 20.19 The Van Allen belts. Figure 20.21b shows a similar electron beam entering a region in which there is a uniform magnetic field B directed into the plane of the page. As we have seen, the force on a moving electron in a magnetic field is given by F = Bev. The electron beam is deflected into the arc of a circle, as shown. In Figure 20.21c the electron beam travels through perpendicular electric and magnetic field simultaneously. By adjusting B and/or E the electric and magnetic forces can be balanced so the beam is not deflected. This occurs when the forces are oppositely directed but equal in magnitude, eV = Bev d Thus V = Bvd Figure 20.20 The aurora borealis. Chapter 20 Magnetic fields and their effects a c b +V F = eE +V B E ee- E= V d B d e- F = Bev zero deflection eE = Bev Figure 20.21 Electrons moving in electric and magnetic fields: (a) electric field, (b) magnetic field, (c) perpendicular electric and magnetic fields. Worked example 20.5 Worked example 20.6 Q Electrons in the beam shown in Figure 20.21c have speed v = 1.5 × 107 m s−1. The plate separation d = 4 cm. The magnetic field in the region between the plates is adjusted to have flux density B = 0.02 T. (a) Calculate the potential difference V between the plates if the beam is undeflected. (b) This p.d. is maintained but the plate separation d is halved. How must the magnetic field B be adjusted to keep the beam undeflected? Q A (a) The beam will be undeflected when the electric and magnetic forces balance, eV = Bev d Thus, V = Bvd = 0.02 × 1.5 × 107 × 4 × 10−2 V = 1.2 × 104 V = 12 kV (b) In the condition for zero deflection (eV = Bev), B is d inversely proportional to d. So if d is halved, B must be doubled to maintain this condition.
Calculate the magnitude and sign of the Hall voltages produced by a current I = 1.0 mA for the following samples in a magnetic field with flux density B = 0.4 T. All the samples are in the form of thin strips, thickness t = 0.1 mm, placed with t parallel to the field direction. (a) A strip of copper foil (charge carrier density n = 1.0 × 1029 m−3) (b) A p-type semiconductor (charge carrier density n = 6.0 × 1020 m−3) (c) An n-type semiconductor (charge carrier density n = 1.0 × 1021 m−3) A (a) The charge carriers in copper are electrons with Q = −1.60 × 10−19 C. The Hall voltage is given by BI VH = nQt 0.4 × 1.0 × 10−3 VH = 1.0 × 1029 × −1.6 × 10−19 × 0.1 × 10−3 = −2.5 × 10−10 V = −0.25 nV conductor t d (b) The charge carriers in a p-type semiconductor are holes with Q = +1.60 × 10−19 C. The Hall voltage is given by BI VH = nQt 0.4 × 1.0 × 10−3 VH = 20 6.0 × 10 × 1.6 × 10−19 × 0.1 × 10−3 = +4.2 × 10−2 V = +42 mV + B BQv + + + + + E I VH + - - - - - EQ = v QVH d – Figure 20.22 The Hall effect. ITQ 9 To increase the Hall voltage should you use a thick sample or a thin sample? (c) The charge carriers in an n-type semiconductor are electrons with Q = −1.60 × 10−19 C. The Hall voltage is given by BI VH = nQt 0.4 × 1.0 × 10−3 VH = 21 1.0 × 10 × −1.6 × 10−19 × 0.1 × 10−3 = −2.5 × 10−2 V = −25 mV 321 322 Unit 2 Module 1 Electricity and magnetism The Hall effect The Hall effect was first noted by Edwin Hall in 1879. When a current-carrying conductor is placed is a magnetic field a small p.d. develops between its sides, as shown in Figure 20.22. The Hall voltage, VH, is proportional to the current I and to the magnetic flux density B. Today semiconductor Hall probes are a convenient tool for the measurement of magnetic flux density. Theory In a magnetic field with flux density B, charge carriers with charge Q drifting with velocity v through a conductor experience a deflection force given by BQv. Initially this will cause a deflection of the charge carriers perpendicular to their flow. Opposite sides of the conductor gain opposite charges – the side towards which the carriers are deflected becomes charged with the same sign as the carriers and the opposite side is left with the opposite charge. The opposite charges on either side of the conductor create an electric field E, which opposes further deflection by the magnetic field (the theory is thus similar to the theory of electron beams in perpendicular electric and magnetic fields in the previous section). In equilibrium, QE = BQv V Now E = dH where VH is the Hall voltage and d is the thickness of the conductor perpendicular to the field. Thus, QVH = BQv d So, VH = Bv d Therefore, VH = Bvd...............(20.1) In Chapter 19 we saw that the current I in a conductor is related to the drift velocity through the equation I = nQvA where n is the number density of charge carriers, v their drift velocity and A is the cross-sectional area of the conductor. Thus, I v= nQA Substituting for v in equation 20.1 we have, BId VH = nQA Now d is just the thickness t of the conductor in the direction of the field, thus, BI VH = nQt Worked example 20.6 contains examples of this type of calculation. ITQ 10 Why are semiconductors more suitable than metals for making Hall probes? Discussion The Hall voltage is inversely proportional to the density of charge carriers. This means that the Hall voltage for a metal is much smaller than that for a semiconductor.
The sign of the Hall voltage depends on the sign of the charge carriers. The Hall voltage of a p-type semiconductor is positive whilst that of a metal or an n-type semiconductor, in which the majority charge carriers are electrons, is negative. Measurement of the Hall voltage can thus be used to determine if an unknown sample of semiconductor is an

n-type or a p-type. The sensor of a Hall probe contains a thin film of semiconductor, which forms part of an integrated circuit. The circuit amplifies the Hall voltage and the output voltage of the probe is proportional to the magnetic flux density. The probe can be used without calibration to plot the flux patterns of coils and permanent magnets. If quantitative flux density measurements are required, the probe must first be calibrated by placing it in a field of known strength – at the centre of a solenoid for example.
Worked example 20.7 Q When placed at the centre of a long air-cored solenoid, the output voltage of a Hall probe is 0.45 V. The solenoid is wound with 20 turns per cm and carries a current of 0.2 A.
When the probe is removed from the solenoid and placed 1 cm from a long wire the output voltage is 0.05 V. What is the current, Iw, in the wire? A The output voltage V of the probe is proportional to the flux density. Inside the solenoid V = 0.45 V; 1 cm from the wire V = 0.05 V. The flux density adjacent to the wire is thus 1/9 of the flux density in the solenoid. For the solenoid: Bs = μ0 nIs where n = 2000 turns per metre and Is = 0.2 A. For the wire: μ I Bw = 0 w 2πr where r = 0.01 m. Now, Bs = 9Bw So, μ I μ0 nIs = 9 × 0 w 2πr Rearranging, 2πnIs Iw = 9 2π × 0.01 × 2000 × 0.2 = 9 = 2.8 A Chapter 20 Magnetic fields and their effects The force between current-carrying conductors Figure 20.23 shows a simple experiment to demonstrate that two parallel wires carrying a current exert a force on each other. This force can be understood in terms of the current in one wire responding to the field created by the other. The force between two parallel wires is the basis for the definition of the ampere. The ampere is the strength of that constant current which when carried by two parallel, straight and very long conductors of negligible cross section and placed 1 m apart in a vacuum produces between these conductors a force of 2 × 10−7 newton per metre of their length. This definition fixes the value of the constant μ0 (the permeability of free space), which appears in expressions for magnetic field strength. foil under terminal 5A 5A For a long straight wire carrying a current I the flux density at a distance r is, μ0 B= I 2πr where B is 2πr which in wooden block (1 cm apart) A Say the force on a parallel wire at the distance r also carrying a current I is given by, μ I2l, F = BIL = 0 2πr B strips of aluminium cooking foil (1 m x 1 cm) And so the force per unit length is given by, F μ I2 = 0 L 2πr Figure 20.23 Currentcarrying wires will attract or repel depending on the directions of the currents. When I = A and r = 1 m then from the definition of the ampere, F μ × (1 A)2 = 0 = 2 × 10−7 N m−1 L 2π × 1 m Thus, Worked example 20.8 Q A As a result of a short circuit in a power station, two parallel power lines 10 cm apart briefly carry currents of 5000 A in the same direction. Calculate the force between them per unit length. Do the cables attract or repel? μ0 = 4π × 10−7 T m A−2 soft iron pole pieces S P For parallel conductors each carrying current I, F μ I2 = 0 2πr L current 4π × 10−7 × 50002 N m−1 2π × 0.1 = 50 N m−1 Currents in the same direction attract. ITQ 11 With the help of the right-hand grip rule, predict the direction of the field produced by wire A in the region of wire B (Figure 20.23). Then use Fleming's left-hand rule to predict the direction of the force on wire B when the two wires carry currents in the same direction. Do the wires attract or repel? Repeat this exercise for oppositely directed currents. ▪ In Figure 20.24 Faraday's electromagnet. 323 324 Unit 2 Module 1 Electricity and magnetism Getting it right! Units of permeability The derivation above shows that μ0 has units N A−2. You will also see it written with units 'Henry per meter' (H m−1). The Henry is the unit of inductance, which you will meet in the next chapter. If N A−2 and H m−1 are both expressed in base units of the SI system they are found to be equivalent. a S unmagnetized b S Electromagnets N partially magnetized An electromagnet is made by winding a solenoid around a solid core of soft ferromagnetic material, iron for example. When current flows in the coil the core becomes strongly magnetized, producing a field that is much stronger than that of the coil alone. Making the core into an almost complete ring concentrates the magnetic field into the region between the poles. Faraday's magnet, shown in Figure 20.24, produces a very strong field in the gap at P between the poles; the field outside the gap is relatively weak. Domain theory The effect of the core is explained by the domain theory of magnetism. Magnetism is created by moving electric charges. One model for the structure of an atom (Figure 20.25) shows it as having a tiny central nucleus, which is positively charged, surrounded by negatively charged electrons – electron + proton – neutron – – N direction of magnetization + + + + + – – Figure 20.25 The nuclear atom. c S N saturated Figure 20.26 Magnetic domains in ferromagnetic materials: (a) unmagnetized; (b) partially magnetized; (c) saturated magnetization. that simultaneously spin and orbit the nucleus, like planets around the Sun. The spin and orbital motions of the charged electrons produce magnetic fields – just like current loops in a coil. The resultant field of the atom depends on the number and relative orientations of its electrons – some elements have permanently magnetic atoms, others do not. The atoms in a sample of a ferromagnetic material (a material which may be strongly magnetized) each behave as a tiny permanent magnet. Neighbouring atomic magnets tend to line up to form magnetic regions called domains (Figure 20.26a). But the whole piece of iron may not itself be a magnet.
This is because the magnetism of different domains point in different directions, and on average, their fields cancel.
To magnetize the material, the magnetization of the different domains must be made to align, so their fields reinforce each other. The magnetic field of a solenoid causes the magnetization of the domains in the core to align, magnetizing the material (Figure 20.26b). The field produced is much stronger than the magnetizing field of the coils alone. If the material is soft magnetically (easily magnetized and demagnetized), then when the current is switched off, the domains revert to their original orientations and the magnetism disappears.
The strength of the field that magnetizes the core is increased by increasing the number of turns and/or the current in the solenoid. The magnetization of the core Chapter 20 Magnetic fields and their effects increases up to a maximum value, at which point it is saturated (Figure 20.26c), and the magnetization of all the domains is aligned. Increasing the current beyond the saturation value does not increase the strength of the magnet significantly. Relative permeability, μr If B0 is the magnetic flux density inside the solenoid without the core and B is the flux density with the core, then the ratio B/B0 is called the relative permeability of the core material. μr = B B0 Non-magnetic or very weakly magnetic materials have relative permeability values close to 1. Ferromagnetic materials are materials with a high relative permeability. The relative permeabilities of selected materials are given in Table 20.1. Table 20.1 Relative permeabilities of selected materials Material Relative permeability, μr air 1 aluminium 1 copper 1 zinc 1 cobalt 70–250 nickel 100–600 carbon steel 50–100 pure iron 150–5000 mumetal (a nickel, chromium, copper, iron alloy) 20 000 Note that the relative permeability is not constant. It varies with the strength of the magnetizing field and depends on the past history of the material – whether it has been magnetized previously and how strongly. Comparing μr and εr
The effect of a ferromagnetic core in an electromagnet can be compared to that of a dielectric material used to increase the capacitance of a parallel plate capacitor. The core increases the flux density by a factor of μr; the dielectric increases the capacitance by a factor εr (the relative permittivity or dielectric constant of the material). But the physical processes are quite different. The magnet core increases the strength of the magnetic field as the solenoid's field aligns its domains. In contrast the dielectric material decreases the electric field strength between the capacitor's plates as its molecules are polarized – the capacitance is increased because a greater charge will now be placed on the plates to produce a given p.d. between them. Applications of electromagnetism The relay (Figure 20.27) and the moving-coil loudspeaker (Figure 20.28) are devices that use electromagnetism to perform useful tasks. A third electromagnetic device – the electromagnetic door lock – is considered in the end of chapter exercises (see Question 20 and Figure 20.34). A relay is an electromagnetic switch operated by turning a small current on or off. The electromagnet makes or breaks a contact to control a much larger current. In a motor vehicle, for example, circuits that require large currents (such as the starter motor and the air conditioning system) are usually switched by relays. When the electromagnet is activated the armature is attracted to the soft iron core, closing the contacts and completing the main circuit. In this way a small current from the vehicle's electronic systems can switch a larger current in the vehicle's power circuits.
springy metal contact terminals contacts pivot soft-iron armature coil terminals coil soft-iron core Figure 20.27 The relay. Inside the loudspeaker a wire coil is fixed to the centre of the paper cone. The coil has flexible leads connected to terminals outside. When a varying electric current passes through the coil, it moves to and fro in the radial magnetic field of the tubular magnet as a result of the force on the current produced by the field. The cone thus sets the surrounding air in vibration and sound is produced. The varying current could come from the output of the amplifier circuits in a radio, a television, a PA system or any similar piece of audio-visual equipment.
frame radial magnet N input signal to coil S N paper cone coils Figure 20.28 The moving-coil loudspeaker. 325 326 Unit 2 Module 1 Electricity and magnetism Summary ✓ Lines of magnetic flux, sometimes called lines of force, surround a straight current-carrying conductor; the flux lines are concentric circles around the current and are perpendicular to the current direction. ✓ The direction of the lines of flux around a current is given by the right-hand grip rule; if the thumb shows the current direction the curled fingers give the direction of the lines of force. ✓ The flux at the centre of a flat circular coil is perpendicular to the plane of the coil. ✓ The flux inside a long solenoid is uniform and parallel to the solenoid axis; the flux pattern outside the solenoid is similar to that of a bar magnet. ✓ If when observed from one end the solenoid current is clockwise, then that end of the solenoid behaves as a south magnetic pole; an anticlockwise current corresponds to a north pole.
✓ Magnetic flux density B is defined as the force per unit current per unit length on a conductor placed at right angles to the field. ✓ The unit of flux density, the tesla (T), is that flux density which produces a force of 1 newton per metre of length on a conductor carrying of 1 ampere at right angles to the field.
✓ Flux density at a distance r from a long (d) straight wire: B = 2πr ✓ Flux density at the centre of a flat coil: μ0 B = 2πN, where N is the number of turns in the coil. ✓ Flux density near the centre of a long solenoid: B = μ0nI, where n is the number of turns per metre. ✓ The constant μ0 is the permeability of free space: μ0 = 4π × 10−7 H m−1 (or N A−2). ✓ The force on a current-carrying conductor of length L at angle θ to a magnetic field is given by F = BIL sin θ, where I is the current. ✓ The force on a charged particle Q travelling at speed v in a magnetic field is given by: F = BQv sin θ. ✓ The direction of the force on a current/moving charge is always perpendicular to both the field direction and the current direction, and may be found with Fleming's left-hand rule. ✓ In general, for non zero θ, the trajectory of a charged particle travelling in a uniform magnetic field is a spiral directed along the field lines. ✓ A charged particle travelling at right angles to a uniform magnetic field is a circle. ✓ A flux pattern which diverges and then converges again (for example between two magnetic poles) forms a 'magnetic bottle' which may trap charge particles. ✓ The Van Allen belts are formed by charged particles trapped in the Earth's magnetic field 1000 to 60 000 km above the Earth's surface. ✓ The condition for zero deflection of a beam of charged particles in crossed electric (E) and magnetic (B) fields is: E = Bv v ✓ The Hall voltage VH between opposite sides of a conductor of thickness t carrying a current I in a Bl , where n is the magnetic field B is given by VH = nQt number density of the charge carriers and Q is their charge. ✓ For a given current and field, the Hall voltage of a semiconductor is greater than that of a metal since n is smaller. ✓ Two current-carrying conductors attract when their currents are in the same direction but repel when their currents are opposite. ✓ The ampere is the strength of that constant current which when carried by two parallel, straight and very long conductors of negligible cross section and placed 1 m apart in a vacuum produces between these conductors a force of 2 × 10−7 newton per metre of their length. ✓ When a ferromagnetic core is placed inside a solenoid the flux density is increased by the factor μr – the relative permeability of the core material. Chapter 20 Magnetic fields and their effects Review questions (a) At the centre of a 200 turn circular coil, radius 0.25 m, carrying a current of 0.5 A (iii) inside a long solenoid with 20 turns per centimetre, carrying a current of 100 mA.
μ0 permeability of free space, μ0 = 4π × 10−7 N A−2 Charge on electron, e = −1.60 × 10−19 C ■ rest mass of electron, me = 9.11 × 10−31 kg Magnetic fields 1 Sketch the shape of the magnetic field in the plane shown in each case in Figure 20.29. a 6 (a) State Fleming's left-hand rule. (b) Find the direction of the forces acting on the conductors in parts a and b of Figure 20.31. a b N b I N S I N S Figure 20.31 Diagrams for question 6. + – 7 Figure 20.32 shows a copper conductor placed between two magnetic poles. Figure 20.29 Diagram for question 1. 2 What is the direction of the magnetic field at P and at O in Figure 20.30?
P current, I O Figure 20.30 Diagram for question 2. 3 A compass needle is placed due west of a vertical wire. The wire is carrying a large current. What is the direction of the current if: (a) the compass needle points south (b) the compass needle points north? 4 (a) Sketch the magnetic flux pattern produced by a solenoid – show the pattern inside and outside the coil. (b) Explain how the direction of the flux lines is related to the direction of the current in the solenoid coils. The strength of a magnetic field – flux density, Force on a current 5 (a) Define flux density and the tesla. (b) Calculate the flux densities at the following points (you may neglect the contribution of the Earth's magnetic field): (i) at a distance of 0.2 m from a long straight wire carrying a current of 10 A N S Figure 20.32 Diagram for question 7. (a) Sketch the magnetic flux pattern between the poles: (i) when there is no current in the conductor (ii) when there is a current directed into the plane of the page as indicated. (b) In what direction is the force on the conductor when it carries the current?
(c) If the direction of the current is reversed, how does the force on the conductor change? (d) How could the strength of the force on the conductor be increased while still using the same length of conductor and the same magnets? (e) When does a current-carrying wire in a magnetic field experience zero force? 8 A l cm length of horizontal conductor is placed perpendicular to the horizontal magnetic field at the centre of a long solenoid. The conductor is connected to a power supply by wires parallel to the field. The current in the conductor is adjusted to I = 5 A. With this current a simple balance mechanism indicates that the magnetic force on the conductor decreases by 20 mg. (a) Calculate the flux density inside the solenoid. (b) If the solenoid is 20 cm long and has 500 turns, calculate the current it carries. 327 328 Unit 2 Module 1 Electricity and magnetism Force on a moving charged particle The Hall effect 9 (a) Write down the expression for the magnetic force on an electron travelling at angle θ to a uniform magnetic field B with speed v. (b) In which direction does this force act? (c) Write expressions for the components of the electron's velocity parallel and perpendicular to the electric field.
(d) By considering the effect of the magnetic force on these two components of the electron's velocity in turn, explain why the electron follows a spiral trajectory along the field lines. 12 Show that the Hall voltage VH between opposite sides of a conductor of thickness t carrying a current I in a Bl , where n is the magnetic field B is given by VH = nQt number density of the charge carriers and Q is their charge. 10 Figure 20.33 shows three types of radiation emitted by a radioactive source. The tracks of the different radiations are recorded with a photographic film or in a cloud chamber (see Chapter 26). When the radiations pass through a magnetic field the tracks of the α and β rays are deflected into arcs of circles. source β γ α Figure 20.33 Diagram for question 10. (a) Given that β rays are electrons, deduce the direction of the magnetic field in the diagram. (b) What is the sign of the charge carried by α rays? (c) What can you deduce about γ rays? 11 A particle with charge Q, mass m and speed v enters a uniform magnetic field B perpendicular to its path. (a) Write down an expression for the force on the particle. m v2 (b) By considering the centripetal force (r) show that the particle follows a circular trajectory of mv . radius r = BQ (c) By finding the time for the particle to complete one orbit in the field, write down an expression for the frequency of the motion. (Note: this frequency is called the cyclotron frequency of the particle.) How does the cyclotron frequency depend on the particle's speed? (d) Calculate the cyclotron frequency of an electron in the Earth's magnetic field (flux density 50 μT). 13 Explain how the Hall effect may be used to determine if a semiconductor is n or p type. 14 A sample of n-type semiconductor in the form of a thin rectangular wafer of thickness t = 0.5 mm is placed with parallel to a magnetic field of flux density B = 0.1 T. With a current of 5 mA between two opposite edges of the wafer, a Hall voltage of −2 mV is measured between the other two edges. (a) Draw a diagram to show the orientation of the wafer relative to the field and the current; label the two edges between which the Hall voltage is measured. (b) Calculate the charge carrier density in the sample. 15 A Hall probe is calibrated by placing it in a solenoid to produce a known field. When placed at the centre of a long solenoid with 2000 turns per metre carrying a current of 0.2 A the output voltage of the probe is 2.51 V. Draw up a calibration chart or table for the probe from which the flux densities corresponding to output voltage readings in the range 0–5 V may be read. State any assumptions you make in producing your chart/table. The force between current-carrying conductors, Electromagnets 16 (a) Define the ampere. (b) Two parallel wires are both carrying currents of 5 A. Calculate the force between the wires. The wires are 10 cm apart and 2 m long. 17 Discuss the effects of following changes on the force F between the wires (consider the changes individually, not in combination): (i) reversing the current direction in one wire (ii) halving the separation of the wires (iii) doubling the current in one wire (iv) doubling the length of the wires. 17 Two long straight parallel wires are placed 0.1 m apart. One carries a current of 4 A and the other a current of 3 A in the opposite direction. (a) Find the force per unit length that acts between the wires. (b) Calculate the magnetic flux density at a point midway between the wires. Chapter 20 Magnetic fields and their effects 18 (a) Define the relative permeability of a material. (b) Explain why a solenoid with an iron core produces a much stronger magnetic field than that of an identical solenoid with an air core. 19 (a) Outline the mechanism of the relay with the help of a labelled diagram. (b) With the help of the internet find three or more practical applications for this device, both historically and in the present day. 20 Figure 20.34 shows a cross-section of an electromagnetic door lock. A lock such as this may be used to secure the door of an apartment building. When a visitor calls, the lock can be operated remotely by pressing a button to let them enter.
door door frame iron bolt (c) investigate the effect of introducing different materials into the core of a solenoid (d) plot a contour map showing lines of equal flux density surrounding a wire. Note: the resultant field around a wire, coil or solenoid is the vector sum of the fields produced by the current and the Earth's magnetic field. Can you identify evidence for the Earth's field in your observations? 3 Use scrap materials to make an electromagnet, a loudspeaker, a buzzer or another electromagnetic device of your choice. Hold a competition to see who can make the most powerful electromagnet powered by a 6 V battery. Test the strength of the magnets by comparing the loads they will lift. Some simple designs for homemade electromagnetic machines are illustrated in Figure 20.35. You may choose to adapt one of these designs, or produce your own. connecting wire solenoid solenoid spring a b nail 50 turn coil of fine copper wire paper cup (a) Explain why passing a current through the connecting wires will withdraw the bolt against the spring. (b) Would the lock work with a brass bolt? Explain your answer with reference to the magnetic a permeability of iron and brass (an alloy of copper and zinc). paper cup metal strip connecting wire Figure 20.34 Electromagnetic door lock. Practical exercises and challenges 329 50 turn coil of fine copper wire aluminium strip cut from drink can 9V switch wooden base cylindrical magnet b metal strip metal strip contact nail coil metal strip 9V magnet 1 Use the apparatus shown in Figure 20.15 to aluminium switch iron nail cut from investigate the force on strip current-carrying conductor drink can in a magnetic field. Adjust or adapt the basic wooden base arrangement to investigate in turn how the force depends on: (a) the current, (b) the length of the conductor, (Answers to ITQs (c) the angle between the conductor and the field, (d) the. 1 Like charges repel, unlike charges attract; like poles 2 Use a Hall probe to investigate the magnetic flux repel, unlike poles attract. The forces of field patterns and flux densities produced by wires, coils created by electrostatic charges and magnetic poles and solenoids. You could: are very similar in character. The most significant (a) plot the variation of flux density with distance in difference is that point charges (positive or negative) different directions can exist in isolation, but magnetic poles cannot; (b) investigate the dependence of flux density on the magnetic fields are created by both north and south poles current in pairs, or by electric currents. 330 Unit 2 Module 1 Electricity and magnetism 2 Unlike the magnet field lines produced by a permanent magnet, the field lines that surround a current are continuous; they do not begin and end on magnetic poles. 3 The right-hand grip rule confirms that the field directions in the diagrams are correct. 4 In terms of base units, 1 N = 1 kg m s−2. Therefore 1 T = 1 N A−1 m−1 = 1 kg m s−2 A−1 m−1 = 1 kg s−2 A−1 5 The current (second finger) is directed from the front (labelled) end of the rod to the back. The field (first finger) is directed upwards from N to S; the motion (thumb) is from left to right. 6 According to Fleming's left-hand rule, the force on the rod will therefore be downwards, in agreement with Fleming's left-hand rule. Faraday's 'tension' picture should not be taken literally – the lines of force are not physical objects, they are just a useful way of visualizing how the magnetic forces vary from place to place. 7 (a) The force on the section AD of the wire is vertically upwards.
(b) Sections AB and CD experience equal but opposite horizontal forces perpendicular to the field. Because the forces balance they can be neglected. (Also the balance is sensitive to vertical forces. Horizontal forces, even if unbalanced, should not affect the reading.) 8 Fleming's left-hand rule confirms that the force direction is correct when Q is a positive charge.
9 The Hall voltage is inversely proportional to the sample thickness in the field direction; a thinner sample will produce a greater voltage. 10 The voltage is inversely proportional to the charge carrier density n. Charge carrier densities in semiconductors are orders of magnitude smaller than in metals; semiconductors therefore produce a much larger (and therefore more easily measured) Hall voltage than metals. 11 Currents in the same direction attract. Currents in opposite directions repel. Answers to Review questions 5 (b) (i) 10−5 T (ii) 2.5 × 10−4 T (iii) 5 × 10−5 T 8 (a) 3.9 × 10−3 T (b) 1.2 A 11 (d) 1.4 MHz 14 (b) 3 × 1020 m−3 17 (a) 2.4 × 10−3 N m−1 (repulsion) (b) 2.8 × 10−5 T 331 Chapter 21 Electromagnetic induction Learning objectives ■ Explain magnetic flux and use the equation Φ = BA to solve problems. ■ Define and explain the weber (Wb). ■ Describe and interpret experiments which demonstrate the relationship between the magnitude ■ ■ ■ ■ and direction of an induced e.m.f. and the change of flux linkage producing the e.m.f. Use Lenz's law to determine the direction of an induced e.m.f. Discuss Lenz's law as an example of the conservation of energy. Use Faraday's law of electromagnetic induction to determine the magnitude of an induced e.m.f. Use the equation ξ = BLv to calculate the induced e.m.f. in a straight conductor moving in a magnetic field. Explain applications of electromagnetic induction, including transformers, a.c. and d.c. motors and generators. N V I Use the relationships N s = V s = Ip for the ideal transformer. p p s Induced e.m.f. At the beginning of the 19th century, Michael Faraday posed the question: 'If electricity produces magnetism – can magnetism produce electricity?' Through careful experiments, Faraday discovered how a current can be induced (generated) by changing the magnetic field through a coil, or by moving a conductor through a magnetic field. This process is called electromagnetic induction. At a demonstration by Faraday of one of his many discoveries, a member of the audience asked, 'But what use is it?' Faraday is said to have replied, 'What use is a new-born baby?' Today, Faraday's inventions have nurtured into the generators, motors, transformers and other electrical machines that provide power to homes, schools, offices and factories throughout the world. Wind turbines, for example, exploit the principle of electromagnetic induction, discovered by Faraday, to transform the kinetic energy of the wind into electrical power. Induced e.m.f. in a coil Figure 21.1 shows a coil of insulated copper wire connected to a galvanometer. When the north pole of a bar magnet is moved into the coil of wire, the needle of the galvanometer is deflected to the right. When the bar magnet is stationary, however, the needle returns to its zero position.
When the bar magnet is removed from the coil, the needle is deflected to the left. The movement of the magnet induces an electromotive force (e.m.f.) in the coil, which causes charge to flow. A similar result can be obtained by moving the coil towards and away from the stationary magnet. 10-turn coil S N sensitive galvanometer Figure 21.1 A magnet moving inside a coil induces an e.m.f. in the coil. Faraday discovered that electromagnetic induction occurs only when the magnetic field through a circuit is changing – either by increasing or by decreasing in strength.
The more rapidly the field changes, the greater the e.m.f. induced in the circuit. Faraday's experiments showed that the strength of the e.m.f., and hence the size of induced current, was increased by: ■ moving the coil or magnet faster ■ increasing the strength of the magnet ■ increasing the size (area) of the coil ■ increasing the number of turns of wire in the coil. 332 Unit 2 Module 1 Electricity and magnetism A a A cos θ θ B Magnetic flux In Chapter 20 we saw that the strength of a magnetic field can be pictured with the help of magnetic field lines. The field is strongest where the lines of flux are closest together (where their density is greatest). We defined the flux density B at a point as the force per unit length per unit current on a conductor at right angles to the field at that point. The flux density is measured in tesla (T), where 1 T = 1 N A−1 m−1. Faraday's experiments showed that induction depends not on the rate the field strength changes at a single point, but on the change over the whole area of the circuit – a bigger coil gives a bigger effect as it surrounds more lines of force. Figure 21.2 shows how the total number of lines of force passing through a loop of conductor in a uniform field with flux density B depends on the area A of the loop, and on the angle θ between the loop and B. Faraday described the total number of lines of force passing through a loop of area A as the magnetic flux through that area. The magnetic flux Φ through an area A at right angles to a magnetic field with flux density B (Figure 21.2a) is given by: Φ = BA..............(21.1) In Figure 21.2b the area A is at an angle to the field direction. Fewer lines of force pass through the area at this angle, and so the total magnetic flux through it is now less than in Figure 21.2a. If θ is the acute angle between the normal to the area A and the field direction, then the projected area perpendicular to the field direction (shown by dashed lines) is A cos θ. Thus Figure 21.2 a b Magnetic flux. Magnetic flux and the lines of force (a) the plane of the loop is perpendicular to B (b) the normal to the loop is at an angle θ to B. general: Φ = BA cos θ..............(21.2) ITQ 1 What is the value of θ when the plane containing the area A is: (a) at right angles to the direction of B (b) parallel to B? Figure 21.2 The flux of a magnetic field (the number of lines of force) through a loop depends on the area of the loop, its angle to the field direction and the flux density. Getting it right! Flux and flux density. Make sure you understand the distinction between flux and flux density, which are often confused. Think of the magnetic field as a crowd of people on a dance floor.
The number of people per unit area at different points on the floor (the people density) is equivalent to the flux density; where they move spread out the density is lower. If the people are evenly spread their density is uniform, in which case we can calculate the total number of people by multiplying the people density by the area of the dance floor. The total number of people that occupy the area is equivalent to the flux. The weber The unit of magnetic flux is the weber (Wb), where 1 Wb = 1 T m2. Thus the flux of a uniform field with flux density B = 5 T through an area A = 0.02 m2 at right angles to the field (θ = 0°) is given by: Φ = BA cos 0° = 5 T × 0.02 m2 × 1 = 0.1 Wb Worked example 21.1 Q A circular loop of conductor of diameter 5 cm is placed at right angles to a uniform magnetic field with B = 0.05 T. (a) Calculate the magnetic flux through the coil. (b) The plane of the loop is now rotated until the angle θ between the normal to the plane and the field is 60°. Calculate the new value of the flux through the loop. Chapter 21 Electromagnetic induction Figure 21.2 shows the relative positions of the coil and the field in this example.) A (a) Φ = BA cos 0 0 BA cos 0°, d2 π × (5 ×10−2 m)2 A= = = 1.96 × 10−3 m2 4 4 Therefore, Φ = 0.05 T × 1.96 × 10−3 m2 × cos 0° = 9.8 × 10−5 Wb (b) Φ = 0.05 T × 1.96 × 10−3 m2 × cos 60° = 4.9 × 10−5 Wb Chapter 21 Electromagnetic induction Flux linkage A coil with N turns wound in a Figure 21.3 The flux linkage of a coil depends on the number of turns. coil with N turns varies with the angle, the area and the number of turns. uniform field B, as shown in Figure 21.3, links the total flux passing through each of the N turns. The flux linked by a coil with N turns is N times the flux through one turn of the coil. Thus, Φ = NBA cos θ..............(21.3) where θ is the angle between the normal to the plane of the coil and B. The magnitude of the flux linkage may be changed by: ■ increasing the area of the coil ■ increasing the number of turns in the coil ■ changing the flux density ■ changing θ to be closer to 0° or 180° (Figure 21.3). Faraday's law Faraday's discoveries about electromagnetic induction may be summarized as follows: an electromotive force, ξ, is induced whenever there is a change in the magnetic flux linked to a coil. The induced e.m.f. is proportional to the rate at which the magnetic flux through the coil changes. An alternative picture, which also fits the experimental observations correctly, is that an e.m.f. is induced in a conductor whenever it cuts through lines of force (cut through). For example, when the coil in Figure 21.4 moves from position A to B, the flux through the loop increases from 1 to 3 lines. Two lines of force have been cut through N the coil as they move from the outside to the inside. The greater the rate at which lines are cut in this way, the greater the e.m.f. Figure 21.4 Lines of force are 'cut through' when the flux linkage changes.
The induced e.m.f. is proportional to the rate of flux linkage or the rate of cutting of lines of force. ξ= −dΦ dt ....(21.4) where dΦ is the change in flux linkage in webers per second. (You should recall that the gradient of a graph of y against x is given by the expression dy/dx. This is the ratio of a small change, dy, in y to the corresponding small change, dx, in x. The gradient of the graph tells us the rate of change of y with x. In the same way, dΦ/dt is the rate of change of Φ with t.) In this equation the constant of proportionality between ξ and dΦ/dt is set at −1. The minus sign is explained by Lenz's law, as described in the following section. The value of one for the magnitude of the constant is confirmed later in this chapter when we calculate the e.m.f. induced in a straight conductor moving through a magnetic field. Lenz's law Faraday's experiments showed that a current induced in one direction when the magnet was moved into a coil, but in the opposite direction when the magnet was pulled out again. The law that gives the direction of an induced current was formulated by Heinrich Lenz in 1834. Lenz's law states that the induced current is always in such a direction as to oppose the change producing it. Consider the experiment in Figure 21.5a where the north pole of a bar magnet is pushed towards the solenoid. Lenz's law tells us that the induced current must be in the direction that makes the end of the solenoid facing the magnet behave as a north pole to repel the magnet's north pole; its motion is thus opposed. When the magnet is pulled out of the solenoid, as in Figure 21.5b, the induced current makes the end of the solenoid facing the north pole of the magnet behave as a south pole, which attracts the north pole and opposes its motion. The direction of the induced current is therefore opposite to that obtained when the magnet moves towards the solenoid. ITQ 2 How can a coil be moved in a magnetic field to make the flux linked with the coil a maximum (b) zero? 333 334 Unit 2 Module 1 Electricity and magnetism a field and why is the direction of the induced current determined by the right-hand grip rule. direction of induced current b magnet moving into coil field set up by induced current magnet moving out of coil N N N Worked example 21.2 Worked example 21.3 Q Q Figure 21.6 shows a pair of flat coils facing each other. Coil A is carrying a current supplied by a cell. The current induced in B. A B Figure 21.5 Lenz's law: (a) a magnet pushed into a coil is repelled; (b) a magnet pulled from a coil is attracted. Figure 21.7 shows a solenoid and a flat coil. The switch S is closed to turn on the current in the solenoid. What is the direction of the current detected in the coil, which rapidly drops to zero again. (a) Explain why the current in the coil only lasts for a short time. (b) Find the direction of the induced current in the coil – is it from A to B or B to A? (c) What is observed when switch S is opened once more? G Figure 21.6 A When coil B is moved towards coil A the current induced must oppose this change (Lenz's law). The coils must therefore repel. In Chapter 20 we saw that wires carrying currents in the same direction attract; wires carrying currents in opposite directions repel. The direction of the induced current in B is therefore in the opposite direction to the current in A. This result may also be arrived at by considering the magnetic poles that appear on the faces of the coils as a result of the currents they carry. Lenz's law predicts that the coils repel as one coil is moved towards the other. We can therefore deduce that like poles appear on the adjacent faces of the coils. The alphabet rule (see Figure 20.7) shows that the right face of A is a south pole. Therefore the induced current in B must make its left face a south pole, and so this current must be in the opposite direction to the current in A. Lenz's law is a consequence of the principle of the conservation of energy. Suppose a magnet is given a small amount of kinetic energy to set it moving towards one end of a solenoid. The current induced in the solenoid makes the end of the coil facing the magnet behave as a magnetic pole. If this pole attracted the magnet then the magnet would be accelerated towards the solenoid, increasing both its kinetic energy and the induced current – energy would be created! With this energy-producing S A B G Figure 21.7 Current induced in a flat coil by the changing field of a solenoid. A (a) According to Faraday's law there is an induced current only when the flux linked with a circuit is changing. A current is detected in the flat coil as the field of the solenoid increases from zero. When the solenoid current reaches a steady value the induced current drops to zero. (b) The alphabet rule (Figure 20.7) shows that the current in the solenoid makes the end facing the flat coil behave as a south pole.
The field produced by the solenoid is therefore directed from right to left in the region inside the flat coil. The strength of this field is increasing. The field produced by the induced current must oppose the increase (Lenz's law), and so must be directed from left to right. The right-hand grip rule (or the alphabet rule) shows that the induced current must flow clockwise around the coil when viewed from the solenoid side – from A to B (in the opposite direction to the increasing current in the solenoid). (c) When switch S is opened a momentary induced current is observed in the flat coil as the field of the solenoid changes. But now the field is decreasing and so, by Lenz's law, the induced current must produce a field in the same direction as the solenoid field to oppose its change.
The current in the flat coil is thus reversed – from B to A. Chapter 21 Electromagnetic induction B v Y X l Figure 21.8 An e.m.f. is induced in a wire moving through a magnetic field.
effect it would be possible to build a perpetual motion machine. But since energy cannot be created, then Lenz's law must apply – the induced current must produce a field that repels the magnet, opposing the change producing it. The e.m.f. induced in a straight conductor moving through a magnetic field Figure 21.8 shows a conducting wire moving at right angles to a magnetic field B.
In the proofs below we show that an e.m.f. ξ is induced in the wire given by the equation: ξ = Blvx..............(21.5) where l is the length of the wire and v its speed. Proof 1 In Chapter 20 we saw that an electron moving at right angles to a magnetic field experiences a magnetic force F = Bev. The direction of the force is given by Fleming's left-hand rule. For the electrons in the wire in Figure 21.8 the force is along the length of the wire, as shown in Figure 21.9. – v This force will initially produce a current as charge flows. One end of the wire gains a net negative charge while the other end is left with a net positive charge. The potential difference V between the ends of the wire grows, creating an electric field E = V which opposes the induced current. After a very short l time the magnetic force, Bev, producing the current will be balanced by an electric force, eE = eV , opposing it. l In equilibrium, Bev = eV l Rearranging we have, V = Blv The potential difference V is equal to the induced e.m.f., ξ, which has transferred energy to the electrons to produce the p.d. (The situation can be compared to a capacitor connected across the terminals of a battery – when the capacitor is fully charged the current is zero and the potential difference between the capacitor plates is equal to the battery e.m.f.) Thus, ξ = Blv Proof 2 In Proof 1 we derived the equation for the e.m.f. induced in a moving wire from the law that gives us the force on a charged particle moving in a magnetic field. In Proof 2 we use Faraday's law to obtain the same result. In Figure 21.10 the straight conducting rod is travelling along conducting rails connected to a voltmeter. The voltmeter detects the e.m.f. induced in the rod due to its motion at speed v through the field. The rod, rails and voltmeter form a complete circuit, which encloses an area A. The field B is uniform throughout this area and so its flux Φ = BA. Now in time dt the rod travels a distance vdt along the rails. This changes the area of the circuit by −lvdt (the change is negative because the area is reduced). The change in the flux linked to the circuit is thus given by dΦ = −Blvdt Therefore, dividing by dt, we have, dΦ = −Blv dt From Faraday's law we have, dΦ ξ = − = Blv dt This is exactly the same result as in Proof 1. F = Bev eF= eV l + B Figure 21.9 The forces experienced by the charge carriers in a wire moving at right angles to a magnetic field. The importance of these two proofs taken together is to show that our earlier assumption that the magnitude of the constant of proportionality in Faraday's law is equal to 1 is correct. 335 336 Unit 2 Module 1 Electricity and magnetism Worked example 21.4 Q A A straight wire of length 20 cm moves sideways with speed v = 20 m s−1 at right angles to a uniform magnetic field of flux density B = 0.5 T. The resistance of the wire is 5 Ω. The ends of the wire are connected with leads of negligible resistance. The motion of the wire is maintained by a force F. Calculate: (a) the e.m.f. induced in the wire (b) the induced current (c) the rate at which energy is being transformed to heat in the wire (d) the force required to maintain the motion. (a) ξ = Blv = 0.5 T × 0.2 m × 20 m s−1 = 2 V (b) I = R 2V = 5Ω = 0.4 A (c) The current produces resistive heating of the wire. rate of energy transformation = power = I 2R = 0.42 × 5 = 0.8 W (d) The force that maintains the motion must provide the energy lost as heat. The power delivered by a force F whose point of application moves at velocity v in the direction of the force is given by P = Fv (the energy transferred by a force F moving distance d in the direction of the force is given by E = Fd; therefore power P = E = Fd = Fv).
t t Thus, Fv = 0.8 W And therefore, 0.8 W F= 20 m s−1 = 0.04 N ITQ 3 Use Fleming's left-hand rule to confirm that the direction of the magnetic force (Bev) on the electron shown in Figure 21.9 is correct. ITQ 4 The magnetic force Bev does work on the electrons, transferring energy to them during the short period that charge flows, to produce the potential difference between the wire ends. What form does this energy take when equilibrium is established and the current becomes zero? ITQ 5 How can you move a conductor in a magnetic field without inducing an e.m.f.? conducting rod dΦ = BA v l V B A = –lvdt rail Figure 21.10 A conducting rod moving along rails in a magnetic field.
Applications of electromagnetic induction Generators By using electromagnetic induction to produce an e.m.f., a generator converts mechanical (movement) energy into electrical energy. A coil of wire is spun between the poles of a powerful magnet. The generator may be powered by gasoline or diesel engines, by the energy of flowing water (hydroelectric power) or the wind, by a steam turbine, or in the case of a bicycle dynamo, by muscle power. A direct current (d.c.) generator produces a current that is always in the same direction, like the current in a batterypowered circuit. An alternating current (a.c.) generator produces a current that constantly changes direction, travelling first one way then the other. Most large generators produce alternating current. A typical domestic electricity supply provides an alternating current that changes direction 60 times each second (has a frequency of 60 Hz or, in some countries, 50 Hz). Alternating current is preferred to direct current for electricity supplies because it is more easily transformed from one voltage to another. The a.c. generator The principle of the a.c generator is shown in Figure 21.12. In a practical generator, the magnetic field would be produced by a powerful electromagnet and the coil would consist of many turns to increase the magnetic flux linkage. The coil is known as the armature winding. In the figure, just one winding of the coil is shown for simplicity. The ends of the coil are fixed to two slip-rings, which rotate with the coil. The current passes to the outside circuit via carbon brushes, which press against the slip-rings. Chapter 21 Electromagnetic induction Worked example 21.5 Q a + b I disc axle S – slip rings N motor with speed control B A Figure 21.11 Disc rotating in the magnetic field at the centre of a solenoid. a (a) Explain why an e.m.f. is detected by the voltmeter. (b) The solenoid has 2000 turns per metre and carries a current I = 0.5 A. Calculate the flux density B at the centre of the solenoid. (c) The disc radius is 30 mm and it spins at a rate of 40 revolutions per second. By considering the motion of a radius of theSdisc, calculate: (i) the magnetic flux cut through per revolution (ii) the rate (dΦ) at which magnetic N flux is cut per second slip rings dt B (iii) the e.m.f. induced between the axle and the rim of the disc.
A carbon brush a.c. + b 1 cycle I0 I magnetic field I rms 0 1 4 1 2 3 4 1 number of turns – magnetic field coil position (a) Consider the disc as a wheel with many spokes. Each spoke is a Figure 21.12 (a) A simple a.c. generator (alternator). a.c. moving conductor cutting through the lines of magnetic flux, which carbon (b) The current output of the a.c. generator. brush are perpendicular to both its length and its direction of movement. From Faraday's law an e.m.f. is induced in a moving conductor as it cuts through lines of flux. This e.m.f. produces a potential difference between the axle and the wheel rim. (b) At the centre of a long solenoid: B = μ0nI = 4π × 10−7 × 2000 × 0.5 = 0.0013 T (c) (i) In one revolution a disc radius sweeps out the disc area A. S Therefore the magnetic flux cut per revolution = BA A = πr2 commutator = π × 0.0302 N = 2.8 × 10−3 m2 Φ = BA = 0.0013 × 2.8 × 10−3 = 3.6 ×10−6 Wb (ii) rate of cutting of magnetic flux = flux cut per revolution × revolutions per second dΦ = 3.6 × 10−6 × 40 dt = 1.44 × 10−4 Wb s−1 (iii) From Faraday's law, dΦ ξ= dt = −0.144 mV I rms 0 long solenoid I 1 cy I0 Figure 21.11 shows a copper disc, which is spun in the magnetic field at the centre of a long solenoid.
A sensitive voltmeter is connected between brushes that make contact with the axle at the centre and the rim of the disc as shown.
mV 337 – + brushes Figure 21.13 A simple d.c. dynamo. The current directions shown in Figure 21.12 may be confirmed with Fleming's right-hand rule (the generator rule). The thumb, first finger and second finger are extended mutually at right angles. If the thumb shows the direction of motion and the first finger the field, then the second finger indicates the direction of the induced current. 1 4 1 2 338 Unit 2 Module 1 Electricity and magnetism As the coil rotates, the magnetic flux through it changes. The rate at which the flux changes is greatest when the coil is horizontal, and is cutting the field lines most rapidly, as shown in Figure 21.12a. The flux through the coil increases until the coil is in the vertical position. When the coil is vertical, its sides are moving parallel to the flux lines, not cutting them, so the induced e.m.f. drops to zero.
As the coil passes the vertical, the flux through it starts to decrease, generating an e.m.f. in the opposite direction. The flux decreases more and more rapidly until the coil is horizontal once more and the e.m.f. reaches its maximum in the negative direction. As the coil continues to rotate, the e.m.f., and the current it produces, alternate at the rotation frequency. Increasing the speed of rotation increases the frequency of the a.c. generated. I = I0 sin ωt ω = 2πf, where f is the rotation frequency, I I0 = peak current, Irms = 0 √2 The d.c. generator In the d.c. generator (Figure 21.13) the armature is connected to a commutator, which contains two half-rings (split rings) mounted on the spindle. The commutator reverses the connections of the coil with the outside circuit every time the coil passes through the vertical position. This ensures that the current in the outside circuit is always in the same direction. Figure 21.14a shows how the current varies with each rotation of the coil. The current is maximum when the coil is horizontal and minimum when it is vertical. The current fluctuates, and so is not 'smooth' like the steady d.c. current of a battery (Figure 21.14b). The induced e.m.f, and hence the induced current, can be increased by: Theory Suppose the area of the generator coil is A and it has N turns. Then, using the equation Φ = BA cos θ derived on page 332, the total flux linked with the coil is given by Φ = BAN cos θ where B is assumed to be uniform and θ is the angle between the normal to the plane of the coil and B.
If the coil is rotating with angular velocity ω (remember ω = 2πf where f is the rotation frequency, see equations 6.6–6.8) then, if at time t = 0 the angle θ = 0, at time t the angle will be given by θ = ωt Now from Faraday's law, −dΦ ξ= dt −d(BAN cos ωt) ξ= dt = BAN ω sin ωt This result comes about since BA and ω are constants and d(cos θ)/dθ = −sin θ) (Students who have studied differentiation in mathematics will recognize that the rate of change of cos θ with θ is equal to −sin θ. If you are not familiar with differentiation, the result can be understood by looking at the shape of the cosine curve. The gradient (rate of change) is zero at θ = 0° and a negative maximum at θ = π/2. This corresponds to the shape of the curve of −sin θ.) Thus the e.m.f. produced by the generator varies sinusoidally with time and is given by ξ = ξ0 sin ωt where ω = 2πf, f is the frequency of rotation and ξ0 = BANω. When the generator is connected to a resistive load (a lamp, for example) the e.m.f. drives an alternating current of a similar sine equation through the load, as shown on Figure 21.12b. ■ using a stronger magnet ■ increasing the number of turns in the coil ■ winding the coil on a soft-iron armature ■ rotating the coil at a higher speed. + a b Current 0 magnetic field 1 4 1 2 3 4 1 Number of turns Current coil position Time Figure 21.14 (a) The current generated by a d.c. dynamo. (b) The current generated by a d.c. cell. Chapter 21 Electromagnetic induction laminated iron core a c b eddy currents induced Vs primary coil secondary coil applied Vp changing flux flux laminations Figure 21.15 (a) The structure of a transformer. (b) The circuit symbol for a transformer. (c) The effect of laminations on eddy currents. Transformers A transformer is a device that makes use of induction to change an alternating voltage – either decreasing or increasing it. The transformer consists of two coils wound on an iron core (Figure 21.15). The coil connected to the a.c. input is called the primary coil. The coil that provides the a.c. output is called the secondary coil. A transformer can only operate on a varying voltage. A constant d.c. voltage connected to the primary coil will not produce any change in the magnetic field. So with d.c., no voltage is induced in the secondary coil. The effect of the iron core is to concentrate the flux produced by the primary coil as the current from the supply flows through it, so that it virtually all the flux passes through the secondary coil as well. As the supply current and therefore the flux it produces are alternating, an alternating e.m.f. is generated in the secondary coil. Consider a transformer with Np primary turns and Ns secondary turns. From Faraday's law, assuming an ideal transformer with no energy losses and in which all the flux of the primary passes through the secondary, we have: ■ flux linkage with primary = NpAB ■ flux linkage with secondary = NsAB NsAB Ns NpAB = Np Now the changing flux passes through both coils, inducing e.m.f.s proportional to the flux linkage (Faraday's law). The ratio of the induced voltages is thus: Vs Ns ....(21.6) Vp = Np ■ ratio of flux linkages = Vs is the output of the secondary coil and Vp is the 'back e.m.f.' in the primary coil induced by its own changing magnetic field. This process is called 'self induction'. From Lenz's law the back e.m.f. must oppose the change producing it (the alternating supply voltage). If the coil has negligible resistance then the back e.m.f. is equal and opposite to the supply voltage. The transformer thus steps the supply voltage up or down in proportion to the turns ratio, secondary voltage secondary turns = primary voltage primary turns Worked example 21.6 Q A 200 V a.c. supply is connected to the primary coil of a transformer. The primary coil has 1000 turns. If the secondary coil has 2500 turns what is the output voltage? A Vs N s = Vp Np 2500 Vs = 200 1000 Therefore, 200 × 2500 Vs = 1000 = 500 V ITQ 6 Why can the output power of a transformer not exceed the input power? 339 340 Unit 2 Module 1 Electricity and magnetism Transformer power ■ Winding the coils with low resistance copper wire – Although the output voltage may be larger than the input voltage, the output power can never be greater than the input power. For an ideal transformer the energy loss is zero and so, input power = output power using the thickest wire and the least number of turns that is practical. ■ Making the core from a soft magnetic material (pure iron).
■ Designing the core and the coils to minimize flux leakage. VpIp = VsIs where Ip is the current in the primary coil and Is the current in the secondary coil. 1 Therefore, Vs Ip Vp = Is V N Now V s = N s and so for an ideal transformer we have, p p Vs Ns Ip Vp = Np = Is 2 3 ....(21.7) This result can be applied to well-designed transformers in which the energy loss is small. In practice, the output power is always smaller than the input power due to energy losses. Energy loss in transformers The sources of energy loss in a transformer are as follows: ■ When an alternating current passes through the primary coil of a transformer, circulating currents, called eddy currents, are induced in the iron core. Part of the input energy is lost as heat energy generated by these eddy currents. ■ Heat is produced in the primary and secondary coils because they have resistance, however small. ■ Some energy is required to repeatedly magnetize and demagnetize the core material. The energy lost in this way is described as hysteresis loss. ■ Some energy is lost if all the flux from the primary coil does not pass through the secondary coil. This is described as flux leakage. These energy losses may be minimized by: ■ Laminating the transformer core to reduce eddy currents (as indicated in Figure 21.15c); instead of a solid piece of iron the core is made from thin iron sheets insulated from each other by varnish or thin paper. ITQ 7 A transformer is used to step up a voltage by 5×. How does the maximum output current compare to the input current? Figure 21.16 A transformer with a centre tapping in the secondary coil. A transformer can be used to provide any a.c. voltage by making it with the correct ratio of Np to Ns. A transformer may have more than one secondary winding on the same iron core to provide different output voltages. Some transformers have a secondary coil with several leads connected to it, as shown in Figure 21.16. These leads are called taps or tappings. d.c. motor A simple d.c. (direct current) motor, illustrated in Figure 21.17, consists of a rectangular coil of wire of many turns, mounted on an axle between two poles of a permanent magnet. Current passes into the coil via two brushes, which press against a split-ring commutator attached to the coil. The commutator reverses the current direction in the coil every half cycle to ensure the coil continues to rotate in the one direction. When there is a current in the coil, forces act on the sides labelled AB and CD (these sides are at right angles to the field). These two forces, equal in magnitude but opposite in direction, produce a torque that causes the loop to rotate in a clockwise direction. When the coil reaches the vertical position, the current and therefore the torque become zero, but the momentum of the coil keeps it rotating. The commutator now causes the current direction to reverse through ITQ 9 If the voltage across terminals 1 and 2 in Figure 21.16 is 12 V, what is the voltage across: (a) terminals 2 and 3 ITQ 8 What happens to the energy lost from a transformer? Most commercial transformers have an energy efficiency of about 90%. (b) terminals 1 and 3? Chapter 21 Electromagnetic induction a c b S B rotation C split ring commutator N D A N A S D N S A D brush Figure 21.17 The rotation of a current-carrying coil in a magnetic field. the coil, so that the torque direction continues to be clockwise. If the current was not reversed each half turn by the commutator, the coil would not rotate continuously, but oscillate to and fro about the vertical position. The simple motor of Figure 21.17 does not have much power, because it only has one coil and the magnetic field is relatively weak. Large commercial motors have a powerful electromagnet (the stator coils) instead of the permanent magnet. They have a number of coils wound around an iron core called the armature. This is shown in Figure 21.18. The carbon brushes, which make contact with the commutator, are held in place by light springs. In time the brushes wear and deposit carbon dust between the segments of the commutator. The motor is made so that the brushes are easy to replace, and the commutator can be cleaned by brushing with a fine brush. Motor speed and power In practice motors are used to overcome loads: to drill a hole, move a vehicle or pump water, for example. But let us first consider a motor that is not loaded and is therefore turning freely. How fast will it turn? armature windings carbon brush shaft armature (laminated iron cylinder) As the current continues to supply energy, the motor speed increases. In the absence of a load or friction you might think that the motor speed will increase indefinitely, but this is not the case.
The ultimate speed of the motor is limited because a back e.m.f. is induced as the coil rotates in the magnetic field. The back e.m.f. opposes the supply voltage, reducing the current. Through Faraday's law the back e.m.f. is proportional to the motor speed; this means that the speed can only increase up to the point at which the back e.m.f. is equal and opposite to the supply voltage. If the motor is frictionless and the coils have negligible resistance, the current is then reduced to zero and the armature continues to rotate at constant speed with no further energy supplied to it.
In practice there will be some friction and the motor coils will have some resistance r. The ultimate speed of the motor will then be a little less, and the back e.m.f. at this speed will not quite balance the supply voltage.
This will allow a current I to pass, supplying the energy needed to overcome friction and the coil resistance. If E is the supply e.m.f and ξ is the back e.m.f. then from Kirchhoff's second law (page 300) we have, stator coils coils in slot When the supply to the motor is switched on, there is a large current in the armature coils. The resulting torque accelerates the armature and it begins to rotate; electrical energy is converted to rotational kinetic energy.
segmented commutator Figure 21.18 The coils, commutator and armature of a practical motor. E − ξ = Ir and so, E = ξ + Ir Multiplying this equation by I we obtain an equation for the power input and outputs of the motor, IE = Iξ + I2r 341 342 Unit 2 Module 1 Electricity and magnetism This equation is a statement of the conservation of energy; power in equals power out. The term on the left, IE, is the power supplied, which is equal to the sum of the power outputs on the right. I2r is power lost as heat from the coils and Iξ is the power needed to rotate the coils against friction in the bearings and between the commutator and the brushes. Worked example 21.7 Q The coil resistance of the d.c. motor in a portable drill is 0.5 Ω. The drill is powered by a battery with e.m.f E = 12 V.
With no load the motor makes 1200 revolutions per minute and takes a current of 0.5 A. When on full load the current rises to 5 A. (a) Calculate the back e.m.f. in each case. (b) Find the power used to overcome friction and wasted as heat in the coils when the drill is running freely. (c) Assuming that the power used to overcome internal friction in the drill is unchanged, find the power wasted as heat in the coils and the useful power delivered when the drill is under full load. (d) Assuming that the back e.m.f. is proportional to the motor speed, calculate the speed of the motor under full load. A (a) E = ξ + Ir Rearranging ξ = E − Ir unloaded: ξ = E − Ir = 12 − 0.5 × 0.5 = 11.75 V loaded: ξ = E − Ir = 12 − 5 × 0.5 = 9.5 V (b) Unloaded power to overcome friction: Iξ = 0.5 × 11.75 = 5.9 W power wasted as heat in coils: 2 I r = 0.52 × 0.5 = 0.13 W (c) Loaded power to overcome load and friction: Iξ = 5 × 9.5 = 47.5 W therefore useful power: 47.5 W − 5.9 W = 41.6 W power wasted as heat in coils: 2 I r = 52 × 0.5 = 12.5 W (d) If speed is proportional to back e.m.f. then, speed under load: 9.5 1200 × = 970 rpm 11.75 Now consider a motor under load, for example an electric drill. As the dill bit is pressed into the material it is cutting the forces opposing the rotation of the motor increase. The motor slows down a little, decreasing the back e.m.f. and so allowing the current to rise. The power supplied, IE, increases to supply the additional energy the motor needs to overcome the load. ITQ 10 Why does a drill motor produce more heat when it is turning under load than when it is spinning freely unloaded? Chapter 21 Electromagnetic induction Summary ✓ Relative motion of a magnet in or out of a coil ✓ A d.c. generator has a split-ring commutator, which reverses the output connections to the coil every half revolution so that the output current is always in the same direction. induces an e.m.f. in the coil. ✓ The magnetic flux linked with a coil is given by Φ = NBA cos θ, where N is the number of turns, B is the flux density, A the area of the coil and θ the acute angle between the normal to the plane of the coil and B.
✓ A transformer makes use of the induction between two coils (the primary and the secondary), wound on an iron core, to step an a.c. voltage up or down. ✓ The unit of magnetic flux is the weber (Wb): 1 Wb = 1 T m2. ✓ For an ideal transformer (one with no energy ✓ Faraday's law states that the induced e.m.f. is proportional to the rate of change of flux linkage or the rate of cutting of lines of force: ξ = − dΦ .
dt ✓ Lenz's law states that the induced current is always in such a direction as to oppose the change producing it. ✓ The e.m.f. induced in a straight conductor moving perpendicular to its length at right angles to a magnetic field B is given by ξ = Blv, where l is the length of the conductor and v its velocity. where Vs and Vp are the secondary and primary voltages, Ns and Np the numbers of turns and Ip and Is the input and output currents. ✓ A commutator reverses the current direction through the coil of a d.c. motor each half turn; the torque is thus always in the same direction and the motor armature rotates continuously. ✓ Rotation of the motor coil generates a back e.m.f. ξ proportional to the motor speed. ✓ When the coil of an a.c. generator rotates in a magnetic field the induced e.m.f. is given by ξ = ξ0 sin ωt, where ξ0 = BANω and ω is the angular frequency of rotation (ω = 2πf). losses and 100% flux linkage between the primary and secondary coils): Vs Ns Ip Vp = Np = Is ✓ The power equation for the motor is: IV = Iξ + I2r, where IV is the power input, Iξ is the power used to overcome the load and any friction, and I2r is the power wasted as heat in the motor coils. Review questions A B N Induced e.m.f., Magnetic flux, Faraday's law 1 Which one of the following actions does not cause an e.m.f. to be induced in a coil of wire? (a) Pushing a magnet into a stationary coil. (b) Moving a coil over a stationary magnet.
(c) Holding a powerful magnet stationary inside the coil. (d) Withdrawing a magnet from inside the coil. 2 Figure 21.19 shows a magnet being pushed into a coil of wire, which is connected to a galvanometer. Which of the following statements is/are correct? S G Figure 21.19 Diagram for question 2. (a) The direction of the induced current is from A to B through the coil.
(b) The direction of the induced current is from B to A through the coil. (c) No induced current will flow.
(d) End B will become a north pole. 343 344 Unit 2 Module 1 Electricity and magnetism 3 A magnet is used to induce a current in a coil of wire. List three things that could be done to increase the current produced. 4 A primary coil AB and a secondary coil CD are arranged as shown in Figure 21.20.
The current in AB can be changed by moving the rheostat slider R to the left or the right.
C D A B G R Figure 21.20 Diagram for question 4. (a) Explain why no current is detected in the secondary coil when the rheostat slider is stationary. (b) When the rheostat slides to the left, in which direction is the current in the secondary coil? (c) Will coil CD be repelled or attracted to coil AB? 5 (a) Explain how the flux of a magnetic field with flux

density B through a coil of area A is calculated. (b) Define the weber. (c) State Faraday's law of electromagnetic induction. (d) State Lenz's law. (e) Calculate the e.m.f induced in a coil of 50 turns when the magnetic flux through it changes at a rate of 25 mWb s−1. If the flux is increasing, and its direction is from left to right through the coil, in which direction (clockwise or anticlockwise) will the induced current be when the coil is viewed from the right? 6 Figure 21.21 shows the jumping ring experiment. The solid aluminium ring is observed to jump up from the top of the solenoid when either a d.c. supply to the solenoid is switched on or off, or the solenoid is connected to an a.c. supply. (There is a video clip of the demonstration at . com/watch?v=Pl7KyVIJ1iE). With reference to both Faraday's law and Lenz's law, explain these observations. If the ring is replaced by one with a small gap in its circumference, how would you expect the outcome of the demonstration to differ? Explain your conclusion. Figure 21.21 The jumping ring experiment. The e.m.f. induced in a straight conductor moving through a magnetic field 7 Figure 21.22 shows a conductor AB in a magnetic field.

On a copy of the diagram mark the direction of the magnetic field. In which direction will a current be induced in the conductor AB when it is moved: (a) in the direction shown by arrow (a) (b) in the direction shown by arrow (b) A N motion (a) S B motion (b) Figure 21.22 Diagram for question 7. 8 A 5 cm length of wire passes through a magnetic field at right angles to the field, with speed 12 m s−1. The flux density is 1.5 T. Calculate the induced e.m.f. 9 An aircraft with wingspan 20 m flies horizontally at a speed of 1000 km/h. The vertical component of the Earth's magnetic field has flux density B = 4.0 × 10−5 T. Calculate the p.d. between the aircraft's wing tips. Chapter 21 Electromagnetic induction 10 A horizontal wire 1 m long pointing east to west is dropped from rest. Sketch a graph showing how the potential difference between its ends varies with time. If the horizontal component of the Earth's magnetic field is 2.0 × 10−5 T, calculate the value of the p.d. after 5 s. State any assumptions you make. 15 The simple a.c. generator shown in Figure 21.24 consists of a 200 turn square coil with edges = 2 cm. The coil rotates with frequency f = 50 Hz in a uniform magnetic field of flux density B = 0.01 T. 11 A wire is moving sideways with speed v at right angles to a magnetic field. By considering the magnetic force on an electron in the wire, show that an e.m.f given by ξ = Blv is induced, where B is the flux density and l is the length of the wire. S Applications of electromagnetic induction 12 Figure 21.23 shows a diagram of a bicycle dynamo. Study the diagram and answer the following questions. soft iron slip rings B a.c. driving wheel coil of wire N cylindrical magnet Figure 21.23 A bicycle dynamo.

(a) What turns the driving wheel of the dynamo? (b) What is connected to the output of the dynamo? (c) Briefly explain how the dynamo produces current. (d) How could the output of the dynamo be increased?

13 Describe the construction of a simple a.c. generator.

Sketch a graph to show how the current from an a.c. generator varies with time. 14 The e.m.f. produced by a simple a.c. generator is given by the equation ξ = ξ0 sin ωt, where ξ0 = BANω. (a) Explain the significance of each term in these expressions. (b) Sketch a graph showing the variation of ξ with time. (c) With reference to the terms in the equation for ξ, explain how the generator output will change if: (i) a coil with twice as many turns is used (ii) the rotation speed of the coil is doubled (iii) the generator magnet is replaced by one with double the field strength. carbon brush Figure 21.24 A simple generator. (a) Calculate the flux linkage with the coil at t = 0 when it is at the position shown. (b) Describe how the flux linkage changes as the coil rotates through 90°. (c) Write down the expression that gives the relationship between the rate of change of flux linkage and the e.m.f induced in the coil. (d) Show that the rate of change of the flux linked to the coil is given by: dΦ = −BANω sin ωt dt where ω = 2πf. Hence calculate the maximum value of the induced e.m.f. (e) Assuming the coil has negligible resistance, calculate the r.m.s. current that flows when the generator is connected to an external load with resistance R = 10 Ω. 16 A transformer has a primary coil with 100 turns and a secondary coil with 250 turns. The primary voltage is 12 V. Is this a step-up or a step-down transformer? What voltage would be obtained from it? 17 A current of 2 A is passed through the primary coil (of 50 turns) of a transformer. The secondary coil has 400 turns. What current would be obtained from this transformer? State any assumptions made. 345 346 Unit 2 Module 1 Electricity and magnetism 18 A step-down transformer gives a current of 5 A at 12 V. If the primary voltage is 240 V, find: (a) the primary current (b) the power input (c) the power output, assuming no power losses. 19 Draw a diagram of a simple d.c. motor. Explain briefly why the motor coil must be connected to the power supply through a split-ring commutator in order for it to rotate continuously. 20 A 240 V motor with armature coil resistance r = 2 Ω takes a current of 0.5 A when running freely with no load. When installed to power a vacuum cleaner the motor takes a current of 5.0 A.

Calculate the back e.m.f. in each case. Estimate the useful power output of the motor when it is running the vacuum cleaner; state any assumptions you make. Extension Use the same apparatus to investigate how the flux linkage depends on: ■ number of turns N (use secondary coils with different numbers of turns) ■ area A (use secondary coils of different diameter) ■ flux density B (vary the current I through the solenoid). 2 Make a simple d.c. motor using one of the designs below. The first design is a moving coil motor with a commutator (Figure 21.26a), as described in the main text of the chapter. coil of varnished copper wire – 20 turns make sure this is well balanced a wooden base bare wire supports bent from paperclips Practical exercises and challenges 1.5 V cell 1 Use the circuit shown in Figure 21.25 to test Faraday's law. signal generator A S bare copper (cleaned with a knife) cathode ray oscilloscope (CRO) cylindrical magnet P varnish Figure 21.25 Circuit for Practical exercise 1. b Switch off the CRO time base. The trace on the CRO will be a vertical line. The length of this trace is proportional to the induced e.m.f. in the secondary coil. Adjust the CRO sensitivity so the trace is 1 cm long. Record the frequency f and the length of the trace l. Increase the frequency in steps of 50 Hz, adjusting the signal generator output at each step to keep the current I in the primary circuit constant. Record corresponding values of f and l in an appropriate table. Plot a graph of f against l. AA 1.5V Use the low impedance output of the signal generator; start with the frequency set at 100 Hz. neodymium button magnet Caution: the wire effectively short-circuits the cell and so a large current may flow. Figure 21.26 (a) Moving coil d.c. motor; (b) homopolar motor. Run your motor for a short period only. If the cell becomes hot disconnect the wire. Use only a non-rechargeable 1.5 V AA or AAA cell for these experiments. The second design is a homopolar motor (Figure 21.26b). Homopolar means that the poles (+ and −) to which the moving wires are connected stay the same – a commutator is not required to constantly reverse the connections. Theory Φ Faraday's law state that ξ = −d dt . The rate of change of the flux produced by the solenoid when it is supplied with an alternating current of a given amplitude is proportional to the signal frequency f – the higher the frequency the higher the average rate of change of flux. If Faraday's law is correct, the plot of trace length l against f should be a straight line passing through the origin.

copper wire Note: clean the varnish from one half only of the two ends of the copper coil. This will provide commutation as the coil turns Chapter 21 Electromagnetic induction With the magnet placed on the axis of rotation as shown, lines of flux diverging from its upper pole have radial components perpendicular to the vertical wires. You can use Fleming's left-hand rule to confirm that the forces on the wires on either side of the cell make the wire loop rotate in the same direction. Faraday's first electric motor was based on this principle. Answers to ITQs Answers to Review questions 5 (e) 1.3 V; the current direction is clockwise. 8 0.9 V 9 0.2 V 10 1 mV 14 (a) 0 Wb (b) 0.25 V (e) 18 mA 1 (a) 0° (b) 90° 16 30 V 2 (a) The plane of the coil should be perpendicular to the field (θ = 0° or 180°). (b) The plane of the coil should be parallel to the field (θ = 90° or 270°). 18 (a) 0.25 A (b) 60 W (c) 60 W 3 In a conducting wire the charge carriers are electrons, which carry a negative charge.

The current, shown by the second finger, is therefore in the opposite direction to the velocity v. If the left hand is rotated so that the first finger shows the field direction, then the thumb indicates the force. The left-hand rule confirms that the force direction shown is correct. 4 Some of the energy will have been transferred to heat in the wire by the current. The remainder of the energy is stored in the electric field created by the p.d. between the wire ends. 5 By moving it parallel to the field direction – so no flux lines are cut. 6 If the output power exceeded the input power the transformer would be creating energy – this would violate the principle of the conservation of energy. Vs Ip Vs Ip 7 For an ideal transformer V = I . If V = 5 then I = p s p s I 5, so Is = 5p . The maximum output current is 1/5 the input current. 8 The energy lost is transferred to heat (internal energy) in the core and windings. When in use transformers generally become hot. 9 (a) 12 V (b) 24 V 10 Under load the rotation speed is reduced which means the back e.m.f. is reduced and the motor takes more current from the supply. The resistive heating in the motor coils is proportional to I2 (P = I2r) and so more energy will be transferred to heat. 17 0.25 A 20 Back e.m.f. unloaded = 239 V back e.m.f. loaded = 230 V useful power = 1 kW 347 348 Module 2 A.c. theory and electronics Chapter 22 Simple semiconductor devices Learning objectives ■ Describe voltages and currents used with semiconductor devices. ■ Distinguish between intrinsic and extrinsic semiconductors.

■ Distinguish between conductors and semiconductors in terms of arrangement of atoms, charge carrier density, charge carrier types and resistivity. ■ Outline the effect of doping on semiconductors in terms of population density of electrons and holes. ■ Distinguish between p-type and n-type semiconductors. ■ Explain the formation of a depletion layer in an unbiased p-n junction. ■ Discuss the I–V characteristics, and applications of these characteristics, for a p-n junction when forward biased and ■ ■ ■ ■ ■ reverse biased. Describe the use of diodes for half-wave and full-wave rectification. Discuss the smoothing of rectified a.c. by capacitors and the significance of the time constant, RC. Draw a junction transistor using the transistor symbol. Explain the use of input transducers (the light-dependent resistor, LDR, the thermistor and the microphone), including usage as potential dividers, in electronic circuits. Describe the operation and use of output transducers (the light-emitting diode, LED, the buzzer and the relay) in electronic circuits. Voltage and current sources Direct and alternating currents and voltages a b The working of semiconductor devices involves the use of d.c. (direct current) and a.c. (alternating current). Y-gain Y-gain V/cm V/cm ■ In d.c., the current is in one direction – as obtained from a battery.

■ In a.c., the current alternates direction several times every second along the conductors of the circuit – as in household mains electricity. Figure 22.1 shows graphs of voltages resulting from direct current and from sinusoidally varying alternating current, with time. Note that a.c. and d.c. are sometimes written using capital letters as 'AC' and 'DC'. ms cm -1 20 10 5 2 ms cm -1 1 20 50 0.5 50 1 0.2 0.1 1 2 20 50 s cm -1 10 5 2 1 0.5 2 20 50 s cm -1 0.2 0.1 Figure 22.1 Oscilloscope display of voltage (V) versus time (t) resulting across a resistor from (a) a steady direct current through the resistor and (b) a sinusoidal alternating current through the resistor. The oscilloscope time base was set at 10 ms per cm and the y-gain was set at 5 V cm−1. The grid shown on the display is 1 cm × 1 cm. Chapter 22 Simple semiconductor devices Sinusoidal a.c. voltages and currents A sinusoidal current, I, as a function of time, t, is represented by the equation I = I0 sin ωt..............(22.1) where ω = 2πf, ω is the angular frequency of the current, f is the frequency of the current (i.e. the number of cycles per second) and I0 is the maximum or peak current. By Ohm's law (see Chapter 19), the voltage, V, across a resistance, R, for a current, I, is given by the equation V = IR. It follows that the voltage across the resistance is given by V = I0R sin ωt = V0 sin ωt..............(22.2) The voltage resulting from the sinusoidal current through R is therefore also sinusoidal. Figure 22.1b shows a graph of equation 22.2, i.e. voltage varying sinusoidally with time. The amplitude, V0, is called the peak voltage; the range between the positive and negative peak values is called the peak-to-peak voltage. The mean voltage per cycle is zero, since the voltage is positive and negative in equal amounts during one cycle. However, the average power per cycle is not zero. This is because power depends on the square of the current (see Chapter 19), and squares are positive. It can be shown that the average power, Pav, is given by the square of the root mean square value of the voltage (Vrms), divided by R. (The concept of rms value is discussed in Chapter 14 (about the kinetic theory of gases). For the average power, we get V 2 Pav = rms ....(22.3) R It can also be shown that, for a sinusoidal variation with time, the rms value is (1/√2) times the peak value. Hence, 1 V0 √2 Vrms = ...(22.4) or V0 = √2 Vrms Similarly Irms = 1 I0 √2 ....(22.5) or I0 = √2 Irms It is common practice to quote a.c. voltages and currents as rms values. Worked example 22.1: Determining frequency from a graph of voltage (V) versus time (t) Q A What is the frequency of the voltage shown in Figure 22.1b? The period, T (the time taken for one complete cycle) = 4 × 10 ms = 40 ms = 0.040 s Frequency, f = 1 (see equations 8.3 and 8.4) T therefore, f = 1 = 25 Hz 0.040 A.c. and d.c. usage A.c. is readily available as a power source since most places are served by mains electricity. Furthermore, a.c. can easily be stepped up or down (see Chapter 21). Stepping up a voltage allows power to be transferred with reduced energy loss. Stepping up or down allows desired voltages to be obtained easily, e.g. for powering industrial, household or communications and entertainment equipment. D.c. is portable, as batteries can be carried almost anywhere. Most electronic equipment requires d.c. power. (When connected to the mains, the a.c. first has to be converted to d.c.) However, d.c. is not as easily stepped up or down as a.c. and usually involves more power loss in the process.

Worked example 22.2: Population density of electrons in copper Q A Determine the population density, n, of free electrons for copper. We assume one free electron per atom of copper. atomic mass of copper = 63.5 amu 63.5 g of copper will contain 6.02 × 1023 free electrons (as the atomic mass of any element contains 1 mole of atoms and 1 mole = 6.02 × 1023 particles).

Therefore, 0.0635 kg of copper contains 6.02 × 1023 free electrons. mass (m) density (ρ) = volume (V) density of copper = 8.9 × 103 kg m−3 m volume of 1 mole of copper = ρ population density, n = n= ITQ 1 For a household a.c. voltage of 120 V, what is the value of the peak voltage? 0.0635 kg 8.9 × 103 kg m−3 = 7.13 × 10−6 m3 = N V 6.02 × 1023 7.13 × 10−6 ≈ 8.4 × 1028 electrons per m3 which is quite a large number of free electrons! 349 350 Unit 2 Module 2 A.c. theory and electronics Conductors, semiconductors and insulators a outermost shell has 1–3 electrons b outermost shell has 8 electrons One measure used for comparing the ability of a substance to conduct electricity is its resistivity (see Chapter 19). Resistivity is defined as the resistance of a material of unit length and unit cross-sectional area. The lower the resistivity of a material, the less opposition it offers to the flow of electric charge – in other words, the greater its electric conduction properties. Semiconductors are elements whose resistivities fall between those of conductors and insulators (see Table 22.1). To account for this wide range of resistivities, we will now examine the atomic structures of these three classes of substances. Atomic structure of conductors Each atom of a conductor typically has one, two or three electrons, called valence electrons, in its outermost shell. This is not a very stable arrangement. These electrons, being furthest from the positively charged nucleus, experience the least force of attraction. Further, the electrons in the inner shells of the atom repel these outermost electrons. Hence, the outermost electrons are only very weakly attached to parent atoms and can move about within the conductor very easily. These electrons are called free electrons and they wander randomly throughout a vibrating lattice of positive ions formed as electrons leave their parent atoms. Free electrons give conductors their characteristically great conducting ability since these electrons are not only relatively free to move but also are numerous, since at least one free electron is Table 22.1 Resistivities of various substances Substance Conductors Semiconductors Insulators Resistivity, ρ / Ω m silver 1.6 × 10−8 copper 1.7 × 10−8 gold 2.4 × 10−8 aluminium 2.7 × 10−8 tungsten 5.6 × 10−8 iron 10 × 10−8 constantan (an alloy) 49 × 10−8 mercury 98 × 10−8 nichrome (an alloy) 100 × 10−8 carbon ~10−4 germanium ~10−1 −8 silicon 50 glass 109 – 1012 rubber 1013 – 1015 nucleus inner shells Figure 22.2 Simplified view of (a) an atom of a conductor and (b) an atom of an insulator. assumed to be available from each atom of a conductor. Metals are very good conductors. The population density, n, of free electrons in a conductor is the number of free electrons, N, per unit volume of the conductor. Atomic structure of insulators Whereas for an atom in a conductor, when there are only a few electrons in the outermost shell, there are eight electrons, the full 'octet' complement, in the case of an insulator (see Figure 22.2). The 'octet' arrangement is discussed in Chapter 17, Conductors and insulators (page 261). The exception is helium, whose full outermost complement is two. This is a stable arrangement and therefore free electrons are not available. It takes a tremendous amount of energy to dislodge an electron from an atom of an insulator. Hence, there are practically no free charge carriers to conduct electricity in an insulator. Group VIII elements (such as helium, neon, argon and krypton) are insulators. Elements (e.g. sulfur) and compounds (e.g. rubber) whose atoms bond together by covalent sharing of electrons also tend to be insulators because their outermost shells become filled or 'octets' are formed by this type of mutual sharing of electrons. Atomic structure of semiconductors Atoms of semiconductors belong to Group IV of the Periodic Table; they have four electrons in their outermost shells.

The elements carbon, germanium and silicon fall in this category. By covalent sharing of these electrons between neighbouring atoms, they achieve the stable configuration of the 'octet' of eight outermost electrons (see Figure 22.3).

At room temperature, however, due to thermal vibrations of the atoms, some electrons gain enough energy to become free from octets. Vacancies, called holes, result. Chapter 22 Simple semiconductor devices a 'hole' is created when an electron is dislodged Comparison between conductors and semiconductors free electron In a conductor, the charge carriers are negative (electrons only); in a semiconductor, the charge carriers are negative (electrons) as well as positive ('holes'). The population density of charge carriers in a conductor is huge when compared to that in a semiconductor. This is because the outermost shell of every atom of a conductor has one or more free electrons; in the case of a semiconductor at room temperature, there are much fewer electrons and 'holes'. outer shell of each silicon atom has 4 electrons with covalent sharing, each atom has a stable outer arrangement of 8 electrons Figure 22.3 Covalent sharing of outermost (valence) electrons within a semiconductor. Because the semiconductor as a whole is neutral, the 'holes' behave as though they are positively charged. Figure 22.4 shows this. Consider a d.c. voltage, V, applied across a semiconductor. Electrons move to the right on account of the negative terminal. Assume an electron (e) moves from one atom to fill a hole in another. The movement creates a 'hole' in the first atom and the 'hole' in the second atom is filled. It is as if the 'hole' (h) in the second atom has moved to the left due to the positive terminal. Hence, 'holes' behave as if they are positively charged. The specialized study of the behaviour of semiconductors is part of an area of study called solid state physics. e – + h Figure 22.4 Movement of electrons (e) and holes (h) in a semiconductor due to an applied voltage, V. a phosphorus atom silicon atom free electron b The resistance (and hence, resistivity) of a conductor increases with temperature. This is because as temperature increases, the vibrations of ions of the lattice within a conductor increases, making it more probable for collisions between free electrons and ions to take place. These collisions impede the flow of electrons and hence increase the resistivity of the material. Although vibrations of atoms increase with temperature in a semiconductor, the increase in vibrational energies causes more electrons to be dislodged from octets, and hence, more electron–hole pairs to be formed. The dramatic increase in electron–hole pairs results in increased conductivity and hence reduced resistivity. The presence of trace impurities hardly affects resistivity of a conductor. Certain trace impurities in an intrinsic semiconductor (i.e. a semiconductor with no impurities) can, however, dramatically increase the number as well as types of charge carries per unit volume.

Figure 22.5a shows the effect of introducing a pentavalent impurity (i.e. an atom whose outermost orbit contains five electrons). Covalent sharing of the phosphorus atom with surrounding silicon atoms results in an electron being 'free'. The semiconductor now has an excess of electrons over holes and is referred to as an n-type semiconductor, since electrons are negatively charged. boron atom silicon atom hole The process of introducing impurities into a semiconductor is called doping.

Doping is done by placing the semiconductor wafer in an oven and Figure 22.5 The effect of doping an intrinsic semiconductor: (a) using a pentavalent dopant; (b) using a trivalent dopant. 351 352 Unit 2 Module 2 A.c. theory and electronics a b gap energy allowing the hot dopant, in vapour form, to diffuse into the material. The population density of charge carriers can be tailored by carefully controlling the temperature of the oven and the time allowed for diffusion. Since the presence of tiny amounts of impurities drastically affects the number of charge carriers produced, much care is taken to eliminate even traces of dust in the chamber in which the doping process takes place. By introducing a trivalent dopant (the outermost orbits of whose atoms contain three electrons), an excess of holes over electrons as charge carriers are produced, as shown in Figure 22.5b. Such doped semiconductors are referred to as p-type semiconductors, since holes behave as if they are positively charged. Semiconductors that are doped are called extrinsic semiconductors, as compared with intrinsic semiconductors which are pure elements. Band theory explanation of conduction in solids An alternative theory to the 'octet' theory, called band theory, has been put forward to explain conduction in solids. Band theory is based on the premises that ■ electrons exist in various energy states or levels in a substance ■ no two electrons can occupy the same energy level ■ in solids, energy levels occur in bands with gaps between bands of allowed energies (forbidden gaps) between (many unoccupied states) electrons energy levels very close to each other energy Figure 22.6 The energy states possible (a) for an outermost valence electron in a single atom; (b) for outermost valence electrons in a crystal consisting of the same type of atoms. ■ movement of an electron within a substance occurs only if an electron can jump from one energy level to another. Figure 22.6a shows the energy levels possible for a valence electron in a single atom. Figure 22.6b shows that in a crystalline material, since the atoms are so closely packed together, energy states of valence electrons are crowded together to form energy levels crowded together (called bands) with gaps between bands. Figure 22.7 shows band structures corresponding to conductors, semiconductors and insulators. In a conductor, when many atoms are brought together to form a solid, the atomic energy levels of the valence electrons split to form a band of energy levels called the valence band. Unoccupied energy levels above the valence band form the conduction band. In a conductor there is no gap between the valence band and the conduction band (Figure 22.7a). Since the difference in energy levels between states is small, electrons move easily from state to state. In fact, at room temperatures, electrons leave parent atoms readily in a conductor and move randomly throughout the vibrating lattice of ions formed by their departures. In a semiconductor, the valence band is almost full of electrons. The energy gap between the valence band and the conduction band is small (~1 eV). Thermal energies at room temperature are, on the average, less than 1 eV, so only the most energetic electrons can gain enough energy to conduction band. Electrons conduction band conduction band b valence band c empty conduction band (partially filled) Figure 22.7 Band structures for (a) a conductor, (b) a semiconductor and (c) an insulator. valence band (nearly full) Chapter 22 Simple semiconductor devices b a p type n n – + – + – + + – – + – + V V V p n n x depletion layer depletion region to jump into the conduction band (Figure 22.7b). Having relatively few electrons in the conduction band at room temperatures results in low conductivity, that is the relatively high resistivity, of a semiconductor. At higher temperatures, thermal energy causes more electrons to be able to move from the valence band into the conduction band where there are very many unoccupied states. Hence the increase in mobile charge carriers and the decrease in resistance with temperature in a semiconductor. In an insulator, the valence band is filled but the gap between this band and the next highest (the conduction band) is large (~10 eV). At room temperature, thermal energy of vibrating atoms are so small that hardly any electrons are able to jump the gap to go into the conduction band, which has plenty of unoccupied states (Figure 22.7c). This explains the extremely low conductivities, that is high resistivities, of insulators. The p-n junction diode The p-n junction A special p-n junction occurs when p-type material meets n-type material during doping. This p-n junction finds many applications in semiconductor devices today. The p-n junction is formed when a semiconductor crystal is doped to produce p- and n-type material from opposite sides. Both before and during doping, both sides are neutral. As electrons from the n-region and holes from the p-region diffuse (Figure 22.8a), a thin depletion layer is formed at the common boundary between the two. Electrons from the n-region fill holes in the p-region and quickly a thin region is formed where there are a b reverse bias no mobile charge carriers.

Instead, on the n-side of the boundary, a layer of positively charged ions is formed due to electrons leaving atoms of that region. We can model the electrons moving into holes as holes in the p-region themselves moving in opposite direction to the electrons. Hence a layer of negative ions is formed on the opposite side of the depletion layer (Figure 22.8b). The net result is an internal electric field, E (of the order of 104 V cm−1), which opposes further diffusion of holes or electrons across the depletion layer. A dynamic equilibrium is set up and there is no net current.

On account of the field, E, a small, internal voltage, ΔV, appears across the depletion layer. A graph of voltage versus position, corresponding to Figure 22.8b, is shown in Figure 22.8c.

p-n junction characteristics An electrical connection made to the p-end is called an anode; one to the n-end, a cathode. Since there are only two electrodes, the arrangement is called a diode. Figure 22.9a shows the positive terminal of a variable voltage source connected to the p end of a junction diode and the negative terminal to the n end. An external electric field, Eext, due to the externally applied voltage, is set up within the diode. For small external voltages, V, the external field is small and not enough to overcome the internal field.

The current in the diode remains practically zero.

For larger voltages, the field is strong enough to maintain a conventional current, I, which increases almost exponentially with voltage. The diode is said to be forward biased in this arrangement, when the positive of the voltage source is connected to the anode and the negative to the cathode. In the forward biasing, current increases significantly with applied voltage. c mA hand (cathode end) −6.0 p n symbol for diode 0 reverse breakdown current Figure 22.8 Formation of a depletion layer at a p-n junction. 0.2 0.4 0.6 forward bias V A B C D longer end (anode) Figure 22.9 Applying an external voltage to a diode. 353 354 Unit 2 Module 2 A.c. theory and electronics Worked example 22.3: Diode protection resistor Q A 9.0 V supply is to be used in series with a junction diode. What value of resistance, R, must be put in series with the circuit so that the current, I, does not exceed 50 mA (Figure 22.10)? 9.0 V Figure 22.10 Diode protection by a series resistor. R A We assume a voltage drop, Vd, of 0.6 V within the diode when the current is 50 mA and neglect the internal resistance of the supply.

We use the equation V = sum of potential drops within the circuit where V = applied voltage, I = current in the circuit and R = resistance of the protective resistor (see Chapter 19). We get V = IR + Vd Therefore, 9.0 V = 0.050 A × R + 0.6 V From which we get R = 168 Ω = 170 Ω Worked example 22.4: Zener diode protection resistor Q Determine the resistance, R, to be used in the circuit shown in Figure 22.11 for a battery supply voltage, Vs, of 9.0 V, a Zener voltage, VZ, of 6.0 V, a Zener current, IZ, of about 20 mA and a load current, IL, of about 60 mA. The voltage supplied is equal to the voltage drop across R plus the voltage drop across the Zener diode/load combination.

Hence, Vs = IR + VZ Therefore VR = Vs − VZ = 9.0 V − 6.0 V = 3.0 V Also, Is = IZ + IL = 0.020 A + 0.060 A = 0.080 A But, VR = IsR (by Ohm's law) 3.0 V Therefore, R = R = Is 0.080 A = 38 Ω The symbol for the diode is the arrow followed by a bar, and is shown in Figure 22.9a. The direction of the arrow is the direction holes flow in a diode, and is the conventional current. The physical appearances of two diodes are shown in Figure 22.9c. In one, a bar is marked near the cathode end; in the other, the longer terminal is the anode. A graph of I versus V for a diode is shown in Figure 22.9b. Such a graph is called the characteristic of the diode. As the forward current increases with applied voltages, the diode gets hot. The thermal energy generated within the diode results in an increase in electron–hole pairs, thus causing the current to increase further, and hence, further overheating. This resulting overheating could destroy the diode. Hence a resistor is usually placed in series with a diode so that the voltage does not exceed about 0.8 V and the current does not therefore exceed a safe level. Under conditions of reverse bias, i.e. the positive of the voltage supply connected to the anode and the positive to the anode, the external electrical field is in the same direction as the internal field. This opposes the flow of holes to the cathode side and electrons to the anode side. A very tiny, fairly constant, reverse current of the order of a few microamperes flows (see left half of Figure 22.9b). However, when a certain reverse voltage, Vb, is reached (called the breakdown voltage), there is a large sudden increase in the reverse current, which could damage the diode through overheating, unless some form of diode protective resistance is used. Zener diode The phenomenon described in the preceding paragraph is called the Zener effect, and is put to good use in a specially constructed Zener diode to stabilize voltages supplied to certain types of circuits in which voltage variation cannot be tolerated. The Zener diode is connected in reverse bias in series with a current-limiting resistor (Figure 22.11) and operates at its breakdown reverse voltage. It stays at this voltage for fairly large variations in current. + R Iz V3 load – symbol for Zener diode Figure 22.11 Maintaining a constant voltage across a load by Zener action. The voltage supplied to the load is equal to the Zener voltage, VZ, because both the diode and load are connected in parallel. For a stable voltage supply to reach the load (and Chapter 22 Simple semiconductor devices IL V reverse current diode present photodiode symbol for photodiode Figure 22.12 Photodiode circuit. The symbol for a photodiode is a diode with arrows pointing toward it. the Zener diode, which is parallel to the load) the supply current, Is, must be kept constant. This is achieved since Is = IZ + IL..............(22.6) If the load current, IL, increases (e.g. to meet varying current demands by a load), then, according to equation 22.6, IZ decreases (still keeping VZ constant). If the load current decreases, more current, IZ, flows through the diode (while the Zener voltage is still maintained) and Is remains constant. A practical application is to keep the voltage, and hence the speed, of a d.c. motor constant.

Photodiode A specially constructed p-n junction, operated in reverse bias at voltages before breakdown is reached, is the photodiode.

A circuit involving a photodiode is shown in Figure 22.12. When light is incident on the photodiode, electrons gain enough energy to jump from the valence band into the conduction band. The resulting reverse conduction current caused by the charge carriers generated varies linearly with applied voltage. Photodiodes are used in converting laser light pulses, reflected off CD and DVD discs, into electrical impulses. This process is called 'reading' the disc. Photodiodes are also used as detectors of laser light sent along optical fibres in telecommunications. Photovoltaic cell (solar cell) Specially constructed p-n junctions give rise to an increase in charge carriers when light is incident on a doped p-n junction semiconductor. The energy imparted to these charge carriers gives rise to a flow of charge (i.e. a current) in a circuit if special means are employed to ensure only one-way flow. Voltage is therefore produced by the cell, typically of the order of 0.6 V. Elements such as silicon, gallium, cadmium and arsenic are used in the manufacture of photovoltaic cells. ITQ 2 What are two main differences between forward bias and reverse bias resistances of a p-n junction diode? load resistance, R VL IL Figure 22.13 Half-wave rectification by a p-n junction diode. though the latter three elements are very toxic – which has implications in manufacturing, usage and disposal. Considerable research is on-going to produce photovoltaic cells of increasingly higher efficiency (currently at a level of about 35%) and higher voltages. Cells are connected in series arrays to achieve the larger voltages. Photovoltaic cells are made into panels that power calculators, homes and even equipment aboard spacecraft.

p-n junction rectification Figure 22.13a shows a p-n junction connected in series with an a.c. power supply and a load resistance. Since a p-n junction offers little resistance in the forward bias mode and very high resistance in the reverse bias mode, current will be basically only in the forward part of the a.c. voltage cycle. Figure 22.13b shows graphs of voltages (VL) across the load versus time (t) in the absence of a diode, and with a diode present. Since, with the diode present, positive, but not negative voltages appear across the load, then current is only in one direction in the circuit, i.e. only direct current. The conversion of alternating current to direct current (by allowing current in one direction only) is called rectification. The diode used in this mode is called a rectifier. Since positive output voltage appears in only half of a cycle, and no voltage in the other half, this type of rectification is called half-wave rectification. Figure 22.14 shows how full-wave rectification can be achieved using a 'bridge' formation of four diodes. Let us consider the half-cycle in which the end A of the a.c. supply is positive (Figure 22.14a).

Conventional current can only take the path through diode D1, as shown by arrows, resulting in the terminal at X being positive. During that time, since B is negative, conventional current is flowing to B, and can only do so via diode D3, resulting in the terminal at Y being negative. Figure 22.14c shows the variation of the output voltage V with respect to time t achieved with bridge rectification. 355 356 Unit 2 Module 2 A.c. theory and electronics a a b A + D4 D1 X B – D2 D3 A – + D4 Y D1 X + 0 B + D3 R V Y D2 – Y Figure 22.14 A diode bridge rectifier circuit and output voltage (V) versus time (t) – This kind of rectification is called full-wave rectification.

The voltage does not change polarity and is therefore a direct (as opposed to an alternating) voltage. Because the d.c. voltage does vary with time, this kind of voltage is called a pulsating d.c. voltage. Most electronic devices require a steady or 'smooth' d.c. voltage to operate properly. The mains supply, however, is a.c. The capacitor of capacitance C (Figure 22.15) connected across the output helps to 'smooth' the rectified, but pulsating, d.c. produced by the bridge. Capacitors store charge (see Chapter 18) and hence maintain a voltage across them. The voltage across a capacitor is given by the equation The p-n junction transistor The p-n junction transistor consists of a lightly doped region, called the base, sandwiched between two regions heavily doped with charged carriers of a sign opposite to that in the base. Figure 22.16 shows the two types of transistors, n-p-n and p-n-p, and their respective electrical symbols. The base, collector and emitter are labelled. The arrow on the emitter shows the direction of conventional current. c −t V = V0 e RC ....(22.7) where V0 is the maximum charging voltage and the product RC is the time constant of the circuit. For large t RC and small times t during a cycle, the voltage V ≈ V0. Hence the RC voltage, V, across the load can be very nearly constant. The small ripples of voltage (Figure 22.15b), which must be removed for certain very voltage-sensitive electronic circuits. a b capacitor X + load resistor R – Y V a c b p n e c b p n p circuit symbol c ebc b n e Figure 22.16 A p-n-p and an n-p-n transistor. V0 = 0, since e0 = 1. However, there is still a slight variation in V, called a ripple voltage (Figure 22.15b). The input voltage is a sinusoidal wave but smoothed, with slight ripple (t) R1 X potential divider R2 Y c b e Figure 22.17 Investigating transistor characteristics output of the transistor. Vb is the input voltage. Vce is the output voltage, Ic is the output current. The potential divider R1 (see Chapter 19) connected to the supply voltage Vs is adjusted to provide a suitable positive potential difference between the base (b) and the emitter (e). The p-n junction, b-e, is therefore forward biased. The potential divider R2 is adjustable and Figure 22.15 Capacitor smoothing of a pulsating d.c. voltage. ITQ 3 Use Figure 22.14b. Consider the half-cycle when the end B is positive and A is negative. (a) Through which diode does conventional current, leaving B, pass? (b) Which of the output terminals, X or Y, becomes positive? ITQ 4 How can the ripple voltage be made even smoother by using a device mentioned earlier in the chapter? Rb R1 c Ib b R2 e Vbe Ic Vce Vs + 6V – Figure 22.17 Circuit for investigating transistor characteristics using an n-p-n transistor. Chapter 22 Simple semiconductor devices b / μA Ic = 0 mA Vce = 6 V 50 5 Ib = 30 μA Ib = 10 μA 0 0 0.5 Vbe / V Figure 22.18 Transistor characteristics: input. provides a potential difference Vce between the collector and the emitter. Input, output and transfer characteristics typical for a junction transistor are shown in Figures 22.18–22.20.

Transistor input characteristic (Ib versus Vbe) The graph of Ib versus Vbe (Figure 22.18) is typical of a p-n junction in forward bias.

As Vbe increases, the base current (Ib) suddenly increases exponentially at about 0.50 V. The protective resistance (Rb) limits the base current so that the transistor does not overheat for voltages beyond 0.50 V. The input resistance (rbe) of the base–emitter circuit is defined as the ratio ΔVbe/ΔIb, i.e. the reciprocal of the slope of the Ib versus Vbe graph. The graph shows that at first the input resistance is very high and then decreases with increasing Vbe. Transistor output characteristic (Ic versus Vce) Figure 22.19 shows a family of Ic (collector current) versus Vce (collector–emitter voltage) curves for various values of base current, Ib. The graphs show the following. ■ For any given value of Ib, at first Ic rises very sharply with increasing Vce to a maximum value (called the saturation current).

■ For a given Vce, of the order of less than 1 V, a tiny base current gives rise to a large collector current. This means that the transistor can function as a switch, where a small current in the base, Ib, switches on a large current, Ic, in the collector.

Practical diagrams illustrating applications of the transistor as a switch are shown in the next section. ■ Note that modern digital electronic circuits involve switching by transistors millions of times per second.

One such 'integrated circuit', consisting of millions of transistors, each requiring a base current for switching operation, even in the order of microamperes, requires considerable total current and so leads to power wastage.

In a cell phone, maximizing talk time on a 0.5 V drive / I A / c Figure 22.19 Transistor characteristics: output. single battery is important. Hence a different type of transistor called a 'field-effect transistor' (FET) is often used. A field rather than a (base) current switches the transistor on or off and this involves hardly any current. ■ Once saturation is reached, Ic increases linearly, but not sharply, with Vce. In this region, Ic is affected mostly by Ib and hardly by Vce. Hence in this region, the transistor is used as an amplifier. The output resistance (rce) is defined as the ratio ΔVce /ΔIc and is of the order of tens of thousands of ohms. The output resistance is equal to the reciprocal of the slope of the Ic versus Vce graph in the region beyond the saturation current. Transistor transfer characteristic (Ic versus Ib) The transfer characteristic, shown for a fixed Vce in Figure 22.20, illustrates the transistor functioning as a current amplifier. For Ib = 0, there is a tiny 'leakage' current in the collector–emitter. Thereafter, the collector–emitter current, Ic, varies almost linearly with Ib. The current gain, hfe, is defined by the equation ΔIc hfe = ΔIc ....(22.8) and is the slope of the Ic versus Ib graph. Since Ib is of the order of μA and Ic is of the order of mA, the transistor current gain can be quite large. By placing a resistor in the collector circuit, the collector current variations are converted to collector voltage Ic / mA 5 0 0 50 Ib / μA Figure 22.20 Transistor characteristics: transfer. 357 358 Unit 2 Module 2 A.c. theory and electronics LDR light relay switch R Vs LDR Vbe e R Vbe e bell Figure 22.21 Transducers used as inputs to transistor circuits: LDR. Figure 22.22 Transducers used as inputs to transistor circuits: variations, resulting in voltage amplification.

A practical application of voltage amplification, using a varying input voltage from a microphone, is discussed in the next section. When light is incident on the LDR, the resistance decreases. The voltage across it drops and hence the base–emitter voltage decreases.

The result is that Ic drops to a very low value and may not be sufficient to light the lamp shown or activate a relay connected to a lamp circuit. Transducer usage in electronic circuits A transducer is an electrical device that can convert one form of energy to another.

Figures 22.21–22.23 show three transducers connected to the inputs of transistors in the common emitter mode. In the common emitter mode, the emitter is connected to the negative of the power supply and this connection is grounded, i.e. connected to earth. Light-dependent resistor Figure 22.21 shows a light-dependent resistor (LDR) and a resistor (R) forming a potential divider. An LDR consists of a light-sensitive semiconductor. When light is incident on an LDR, valence electrons gain sufficient energy to be dislodged from their octets. Electron-hole carriers increase in the semiconductor enabling greater conductivity (less resistance). In the dark, the resistance of the LDR is high, since the charge carriers are few. The voltage across the LDR is large, on account of the potential divider arrangement. This results in a large base–emitter input voltage. This large input voltage, in turn, results in a large collector current which could be used either to light a lamp, or activate a relay switch to a circuit in which there is a lamp. ITQ 5 Describe an application of an LDR connected as shown in Figure 22.21a. ITQ 6 Explain why the thermistor is placed between the positive of the voltage supply and the base of the transistor in the circuit of Figure 22.22, between the negative supply and the base of the transistor.

Thermistor Figure 22.22 shows a thermistor (temperature-dependent resistor) in a potential divider arrangement connected to the supply voltage. A thermistor is a semiconductor that is very sensitive to temperature. Energy due to temperature rises causes electrons to become dislodged from their octets, creating an increase in charge carriers, and hence a decrease in resistance of the thermistor. The circuit shown is designed to produce a large collector current, Ic, when the thermistor gets hot. The output can be connected to a relay switch to turn on a fire alarm circuit or to turn off power going to a motor if the motor gets too hot. When the relay coil switches from the 'on' to the 'off' position, it induces a large 'back' e.m.f. in the collector circuit, which could result in a large current that can damage the transistor. The diode connected in reverse bias, as shown, provides a low resistance path to the power supply for such a current, thus protecting the transistor. (Induced currents are dealt with in Chapter 21.) Since the resistance of a thermistor varies with temperature, a thermistor can be placed in series with a milliammeter, suitably calibrated to read temperature, i.e. the circuit can function as a thermometer. Microphone Figure 22.23 shows a microphone input in which the transistor serves as a voltage amplifier. A microphone converts sound energy to electrical energy. The tiny input voltage, vi, of the microphone adds to the base–emitter input voltage (Vbe). This resulting voltage, in turn, causes a Chapter 22 Simple semiconductor devices Relays and buzzers Ic V0 C1 vi microphone C2 Vbe Vce e earphone Figure 22.23 Transducers used as inputs to transistor circuits: microphone. large varying collector current (Ic), which results in a large varying output voltage V0 = IcRL across the load resistor (RL). From the diagram, we see that, for a relay switch ■ The relay coil shown in Figure 22.22 converts electrical energy (via electromagnetism) and operates a switch mechanically. Electromagnetic buzzers also convert electrical energy to mechanical energy via electromagnetic action. Electromagnetic relays and buzzers are described in Chapter 20. Some crystals, e.g. quartz, undergo mechanical, elastic deformation when a voltage is applied across them. If a varying voltage is applied across such a crystal, the latter vibrates, mechanically, producing sound. Such transducers convert electrical energy to sound. These kinds of buzzers are called piezo-electric buzzers. Vce + V0 = V Therefore, V0 = V − Vce V0 The voltage Vce is a varying d.c. voltage. The capacitor C2 placed on the output, blocks the d.c. component of the output and allows the a.c. component to pass to the earphone. The capacitor C1 placed at the input similarly allows the a.c. component from the microphone to pass but blocks d.c. from going to the microphone. The behaviour of capacitors is dealt with in Chapter 18. Light-emitting diode (LED) Light-emitting diode (LED) converts electrical energy to light energy. Figure 22.24 shows a LED connected in a forward bias arrangement, with a protective resistor. Practically no current flows for very low voltages.

However, for voltages of the order of 0.6 V, a fairly large current flows across the p-n junction. The electrical energy supplied by the voltage source bonds to be broken within the crystal, causing electrons to jump into higher energy levels (electrons jump from the valence band into the conduction band). As electrons return to lower levels, the energy is given off as light. LEDs emitting visible light are used to indicate when power is on in electronic devices. LEDs emitting infrared light are used in remote control devices for operating TVs and other electronic equipment. ITQ 7 What is the function of the resistor R in Figure 22.24? 359 360 Unit 2 Module 2 A.c. theory and electronics Summary ✓ Intrinsic semiconductors belong to a class of elements in Group IV of the Periodic Table. Atoms of these elements have four free electrons, called valence electrons, in the outermost shell. ✓ In a semiconductor crystal, valence electrons form their full complement of eight electrons (called an octet) by this sharing and a stable configuration results. ✓ Due to vibrations of the lattice of atoms in the crystal at temperatures above absolute zero, a few valence electrons gain sufficient energy to break loose from octets, leaving 'holes'. Holes behave as though they are positive charge carriers since, when an electron moves to fill a hole, the hole appears to have moved to the site vacated by the electron, i.e. in the opposite direction. ✓ The charge carriers in conductors are numerous and consist of electrons only. The charge carriers in semiconductors are not as numerous and consist of electrons and holes. ✓ An increase in temperature causes a pronounced increase in the numbers of charge carriers per unit volume in a semiconductor, and hence an increase in conductivity. ✓ In a conductor, an increase in temperature causes an increase in vibrations of the ionic lattice which tends to impede the flow of electrons, and hence a decrease in conductivity. ✓ When an intrinsic semiconductor is doped with a trivalent element, a p-type semiconductor results (having an excess of holes over mobile electrons). Doping with a pentavalent element results in an n-type semiconductor (having an excess of mobile electrons over holes). Semiconductors produced by doping are called extrinsic semiconductors. ✓ During manufacture, when a p-type and an n-type semiconductor meet in a single crystal, a depletion layer is formed at the p-n junction in which there are practically no charge carriers. The layer is bounded on both sides by oppositely charged ions which create an internal electric field (and hence an internal voltage). ✓ When a p-n junction is forward biased, it allows charge to flow only after a certain voltage (of about 0.2 V) is reached. The current then increases exponentially with voltage. In reverse bias, there is practically no current until the breakdown voltage of the order of several volts (the Zener voltage) is reached, at which the current becomes quite large. ✓ A single p-n junction diode can be used as a halfwave rectifier, allowing current in one direction only during half-cycles of alternating current. A bridge of four p-n junction diodes can form a full-wave rectifier, allowing current in one direction during entire a.c. cycles. ✓ Rectified a.c. can be smoothed using a capacitor. ✓ A p-n-p transistor consists of an n-type material sandwiched between p-type material; an n-p-n transistor consists of p-type material sandwiched between n-type material. ✓ A small input current in a transistor can switch on a large output current. ✓ A small input voltage variation of a transistor results in a large output current, and hence the transistor can function as a voltage amplifier. This results in a transistor being used as a very fast electronic switch or as an amplifier. ✓ Transducers such as LDRs, thermistors and microphones are often used in input circuits of transistors. Transducers such as LEDs and buzzers are found in the outputs of transistor circuits. Chapter 22 Simple semiconductor devices Review questions Voltage and current sources 1 (a) Name two sources each of a.c. and d.c. voltages. (b) State two advantages of a.c. over d.c. (c) State two advantages of d.c. over a.c. 2 (a) 240 V mains outlets are provided in some homes. What is the least number of volts an appliance must be able to withstand when plugged in to such a supply? (b) If a 30 Ω heating element is connected to such a supply, what will be the power consumption of the element? (Hint: see equation 22.3) Conductors, semiconductors and insulators 3 (a) Describe, qualitatively, the difference in resistivity between semiconductors and conductors. (b) Account for the difference in resistivity between conductors and semiconductors. (c) State and explain the effect of increasing temperature on resistivity in semiconductors and conductors. 4 (a) What class of substances are 'intrinsic semiconductors'? (b) With the aid of a diagram, explain how charge carriers arise in an intrinsic semiconductor. (c) Describe the effect of applying a d.c. voltage across an intrinsic semiconductor. 5 With the aid of diagrams, describe how n-type and p-type semiconductors are formed. 6 (a) With the aid of a diagram, explain how a depletion layer is formed at a p-n junction. (b) With the aid of a diagram, explain why an internal voltage appears across the depletion layer, and account for the direction of the internal electric field through the layer. 7 (a) What is meant by forward biasing a p-n junction diode? (b) (i) Sketch a typical characteristic (current, I, versus voltage, V) for a junction diode in forward bias.

(ii) Give an explanation for the shape of the graph. (iii) Estimate, by calculation, the value of resistance R needed to protect a diode in forward bias if a 6.0 V supply is used and the current must be limited to 80 mA. (iv) Draw a circuit diagram of the arrangement to be used in part (iii) above. 8 (a) What is meant by the 'Zener voltage' in a diode?

(b) With the aid of a circuit diagram, explain how a Zener diode can be used to stabilize the voltage supplied to a load resistance (RL). Include a current-limiting resistor (R) in your diagram. 9 Draw circuit diagrams and voltage (V) versus time (t) graphs in each case to show how a sinusoidal voltage can be rectified using p-n junction diodes to produce d.c. that is: (a) half-wave (b) full-wave. 10 (a) Why is the rectified d.c. output from junction diodes not suitable for powering electronic circuits? (b) Explain, with the aid of suitable diagrams, how the d.c. output from a p-n junction can be 'smoothed.' The p-n junction transistor 11 (a) Draw diagrams to show the normal electronic symbols for the following. Label the collector, base and emitter in each case. (i) a p-n-p transistor (ii) an n-p-n transistor. (b) How can the circuit of Figure 22.22 be modified such that when it is bright, a large collector current results? Explain your answer. (c) Suggest a practical application of the arrangement in part (b) above. 12 (a) Define 'current gain' of a transistor. (b) Calculate the current gain of the transistor shown in Figure 22.20. 13 A 9.0 V d.c. supply is connected to resistors R1 = 100 Ω and R2 to form a potential divider. What must be the value of R2 so that the voltage across R2 = 1.0 V? (Hint: see the section on the potential divider in Chapter 19.) 14 Draw a circuit diagram and explain how a transistor can be used as an electronic switch. Transducer usage in electronic circuits 15 (a) What is a 'transducer'? (b) Name a transducer in each case that can perform the following energy conversions: (i) light energy to electrical energy (ii) electrical energy to light energy (iii) electrical energy to sound energy (iv) sound energy to electrical energy. 361 362 Unit 2 Module 2 A.c. theory and electronics 16 (a) Describe the structure of a light-emitting-diode (LED) and explain how it works. (b) Draw a diagram to show how a LED can safely be operated from a 6.0 V d.c. power supply. (c) Would the diode above work on a 6.0 V a.c. power supply? Explain your answer.

Practical exercises and challenges Note: students are encouraged to obtain simple electronics 'hobby' books that may be available locally. These will give specifics of components and technically useful information, and the stores might even carry the components needed. Although the experiments suggested involve low voltages only, students should attempt to perform them only under consultation with, and supervision of, a properly trained person. 1 Investigating the I–V characteristic of a junction diode Using the arrangement shown in Figure 22.25, investigate the I–V characteristic of a junction diode. variable resistance R and the voltmeter to ensure that the voltage across the thermistor is kept at a constant value. (Alternatively, a simpler circuit involving a digital resistance meter connected to the thermistor may be used and resistance could be plotted against temperature.) (b) Using the apparatus of Figure 22.26, and a liquid-in-glass laboratory thermometer as your 'standard', obtain a current versus temperature (or resistance versus temperature) graph for the thermistor. Comment on the shape(s) of the graphs obtained. (c) Measure the temperature of the room, making use of each of the graphs. What was the percentage difference between the two readings? (d) Obtain an equation for current (or resistance) versus temperature for each graph. (Using the paired data obtained in each case, a best-fit regression equation that describes the relationship can be obtained using features present in some calculators or computers.) thermometer mA + variable low-voltage d.c. supply – stirrer V R A water protective resistor V Figure 22.25 2 Investigating characteristics of a transistor Using the arrangement shown in Figure 22.17, investigate the base, output and transfer characteristics of a transistor. Note that a high impedance voltmeter, e.g. a digital voltmeter, and a very low resistance ammeter, e.g. a digital ammeter, will be necessary so as not to alter the characteristics of the operating circuit too much. 3 Calibrating a thermistor thermometer (a) Using apparatus similar to that shown in Figure 22.26, and other apparatus that may be necessary, take measurements of current and corresponding temperatures and plot a calibration graph of current against temperature, based on ice and steam points only. Note the presence of the heat Figure 22.26 Taking current–temperature measurements for a thermistor thermometer. 4 Fun (and educational) projects Using manuals and kits from 'hobby' electronics books or other resources, construct and test the following: ■ a switch that automatically turns on a light only when it is dark ■ a switch that turns on an alarm only when light is incident on it ■ a switch that turns off power to a d.c. motor when the motor gets hot ■ a one-transistor microphone amplifier that can power a headphone ■ a beam of infrared light from a LED ■ an alarm that sounds when an infrared beam is interrupted. Chapter 22 Simple semiconductor devices Answers to ITQs Answers to Review questions 1 120 V is the rms voltage. The peak voltage, V0 = √2Vrms = √2 × 120 V ≈ 170 V. 7 This example shows us that although the household mains voltage is 120 V, the voltage in the wires could be as high as 170 V! 2 In forward bias, the resistance is very high for about 0.5 V; thereafter, the resistance decreases exponentially with voltage. In reverse bias, the resistance is extremely high and constant until the breakdown voltage is reached, at which point, the resistance very quickly approaches zero. 3 (a) D2 (b) X (as before) 4 By connecting a suitably protected Zener diode in reverse bias across the load resistance, R. 5 A light-operated switch which turns a street lamp on when the place gets dark. 6 A heated thermistor has a small resistance. In the potential divider arrangement with R, this gives rise to a small voltage across the thermistor and a large voltage across R, and hence across the base input. This large input voltage results in the desired large out put current. If the thermistor is connected the other way described, the input voltage (now across the thermistor) would get smaller when the thermistor heats up. This would give rise to a reduced output current which might not be sufficient to activate the relay switch. 7 R limits the current through the diode and thus protects it from overheating and being damaged. (b) (iii) 75 Ω 13 R2 ≈ 13 Ω 363 364 Chapter 23 Operational amplifiers Learning objectives ■ Describe the properties of an ideal operational amplifier. ■ Compare the properties of a real operational amplifier (op-amp) with those of an ideal op-amp. ■ Use the fact that the magnitude of the output voltage of an op-amp cannot exceed the voltage of ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ the power supply. Define gain and bandwidth of an op-amp.

Interpret gain–frequency curves. Determine bandwidth from a gain–frequency curve. Explain the reason for using logarithmic scales in a gain–frequency curve for a typical op-amp. Describe the use of an op-amp as a comparator. Draw several representations of circuit diagrams for both the inverting and non-inverting amplifier with a single input. Explain and use the concept of virtual earth, and its limitations, in the inverting amplifier.

Derive and use expressions for the gain of an ideal op-amp, both as an inverting amplifier and as a non-inverting amplifier. Discuss the effect of negative feedback on the gain and bandwidth of an inverting and a non-inverting op-amp. Perform calculations related to single-input and 'cascaded' amplifier circuits. Use the fact that a non-inverting amplifier has a very high input impedance. Describe the use of the inverting amplifier as a summing amplifier. Solve problems related to summing amplifier circuits – including the use of the summing amplifier as a digital to analogue converter. Analyse the response of amplifier circuits to input signals, using timing diagrams. Analyse simple operational amplifier circuits. Describe the use and importance of the operational amplifier as a voltage follower. The operational amplifier Operational amplifier representation In Chapter 22 we outlined the functioning of the junction transistor as a simple amplifier. The operational amplifier (op-amp) is a very specialized type of amplifier. Before today's digital computers, analogue computers were used as summers, integrators and for other mathematical functions. The term 'operational amplifier' ('op-amp') has its origin in the computing field where high gain analogue amplifiers were needed to perform the mathematical operations. Today's op-amps have very high gains. Most of the calculations in this chapter on op-amps are based on two very simple concepts – Ohm's law and +Vs a Vin+ input voltage to non-inverting input Vin – input voltage to inverting input + +9 V b Vout – –Vs output voltage Gnd Gnd –9 V Figure 23.1 (a) Symbolic representation of an op-amp with the input, output and supply voltage pins and common ground connection for voltages. (b) A simple dual voltage supply made using two 9 V batteries. Chapter 23 Operational amplifiers the potential divider, which were discussed in Chapter 19. Calculations involving logarithms are also used (logarithms were used in Chapter 10 in defining sound levels in decibels; see also the discussion in Appendix 3). Today's operational amplifier (Figure 23.1) is a silicon integrated circuit amplifier having: ■ a positive or non-inverting input, shown as Vin+ ■ a negative or inverting input, shown as Vin− ■ an output, shown as Vout ■ a positive supply pin, shown as +Vs ■ a negative supply pin shown as −Vs earth (Gnd). For simplicity of analysis, the supply pins are not usually indicated on op-amp diagrams. When a d.c. voltage is applied to the inverting input with the non-inverting input connected to the reference Gnd potential, the output voltage is always of the opposite voltage polarity to the voltage on the inverting input. With a d.c. voltage applied to the non-inverting input, the output voltage will be of the same polarity as the input. The inputs are marked '–' and '+' only to indicate 'inverting' and 'non-inverting', respectively; they should not be confused with the supply voltage polarities. Getting it right! In the case of a positive differential voltage, i.e. a positive (Vin+ – Vin−), a positive output voltage always results. If the inverting input is connected to earth, which is at zero potential, a positive (Vin+ – Vin−) would mean that there is a positive voltage at the non-inverting input, and hence a positive output voltage. If the non-inverting input is connected to earth, a positive (Vin+ – Vin−) would mean that there is a negative voltage at the inverting input, and hence, again, a positive output voltage. Op-amp power supply Op-amp circuits are powered by dual-balanced d.c. supplies, +Vs and −Vs in ranges ±3 V to ±15 V, and referenced to 0 V(see Figure 23.1b). Op-amps require their d.c. operating points to be defined. Inputs may earth Figure 23.2 Symbols in use for ground reference. When only one polarity of d.c. supply is

used as the d.c. input operating reference level is typically one half of the d.c. supply voltage, to which is added the input signal; for example, a 10 V single supply would then have a 5 V d.c. input signal reference level. Op-amp input sources ■ a common reference, typically ground, also called signal / common ground be d.c. or a.c. signal voltages. The 0 V reference level additionally serves as the reference for both the input and output signal voltages and is also described as the 'common' or 'ground' potential. Figure 23.2 shows symbols in use for ground. chassis ground The input sources of an operational amplifier make use of d.c. (direct current) and a.c. (alternating current). Whether it is a d.c. or an a.c. source, the ideal voltage source is defined as a source having no internal resistance or internal 'impedance'.
The symbol Z is used to represent impedance and the unit for impedance is the same as that for resistance, the ohm. The opposition to current by capacitances and inductances is termed 'reactance'; impedance is the generic term for resistance or reactance. Real voltage sources have internal resistances, i.e. Rint > 0 Ω, or impedances, i.e. Zint > 0 Ω. Alkaline d.c. cells which may power op-amps have Rint in the region of tens to hundreds of milli-ohms (mΩ). Figure 23.3 shows symbols for ideal and (b) real voltage sources used in operational amplifiers. Saturation and clipping an important feature of an op-amp is that its output cannot exceed the supply voltages. Typically the output becomes saturated to a approaches its ability to function as a linear amplifier, as the output remains constant and no increase with increasing input voltages (Figure 23.4a). Figure 23.4b shows the clipping of the output from a sinusoidal input voltage. If the amplified input voltage appearing at the output is required to be slightly greater than the supply voltage, saturation of the output signal occurs, resulting in no amplification beyond these voltages. We say that the waveform has been clipped.
365 366 Unit 2 Module 2 A.c. theory and electronics a b Vout V+ V+ Vsat Vout Input ... The ideal operational amplifier The ideal op-amp has infinite voltage gain, AV, so an op-amp in any configuration is defined as the ratio of the output voltage to the differential input voltage (equation 23.1). V Av = ΔVout in V = V −outV in+ in− ....(23.1) The open loop voltage gain, AOL, of an op-amp is defined as the ratio of the output voltage to the differential input voltage (equation 23.1) when in open loop. V AOL = ΔVout in V = V −outV in+ in− Figure 23.4 (a) Saturation; (b) clipping of an output waveform due to saturation, op-amp linear region ....(23.2) Equation 23.2 implies that a huge open loop gain requires a tiny differential input voltage (Vin+ − Vin−) for a fixed output voltage (Vout).
Practical op-amps have very high gains in the hundreds of thousands, as infinite gains do not exist. Note that in an 'open loop' arrangement, there is 'no feedback' applied to the input; in a closed loop arrangement, there is. Open loop and closed loop arrangements will be discussed later in the chapter. The ideal op-amp and its 'virtual' input potential Worked example 23.1 shows that a high gain for a 1 V given output leads to a small differential input voltage (10 mV for a gain of 100) at the op-amp input pins; and a very high gain leads to a very small input voltage (100 μV for a gain of 10 000). An infinite gain would therefore lead to 0 V differential input voltage to realize the required output voltage. In such an ideal case, this would mean that the two input signals are virtually at Worked example 23.1: Open loop gain variation with differential input voltage Q A differential input signal, from a given amplifier, is applied to the input of an op-amp.
What differential input signals (Vin+ − Vin−) are required for the op-amp to produce an output voltage, Vout, of 1 V, if the op-amp has an open loop gain, AOL, of: (a) 100 (b) 10 000? A Use equation 23.1: V AOL = out ΔVin Vout AOL (a) Therefore, for an output voltage, Vout = 1 V and an open loop gain of 100: 1V Vin+ − Vin− = 100 = 10 mV This means that only a 10 mV input signal is required for the output of 1 V. (b) For an open loop gain of 10 000: 1 V Vin+ − Vin− = 10 000 = 100 μV This means that only a 100 μV (0.1 mV) input signal is required for an output of 1 V. Then, Vin+ − Vin− = the 'same' potential (however, they are not connected electrically). Infinite gain is only a theoretical concept, as op-amps with infinite gains cannot be manufactured; however, a gain of, say, 10 000, is achievable. The signal voltage from a microphone is typically greater than tens of millivolts.
With a gain of 10 000, connecting the microphone directly to the input pins of the op-amp would give 1000 V at its output; this is not realistic!! A negative Chapter 23 Operational amplifiers feedback technique (to be discussed later) enables the use of such large input signals with realistic output voltages. Analysing these techniques make use of the concept of the input at virtual earth potential. We shall be making use of a gain of 10 000 as an 'infinite gain approximation' and, consequently, inputs at virtually the same potential in certain calculations later. (Getting it right! A signal voltage appears at the output of a signal source; do not confuse it with the differential input voltage applied to the pins of the op-amp. For linear amplification, signal voltages are not applied directly to op-amp pins, since op-amp gain is unstable (e.g. it varies with temperature). Worked example 23.2: Electret cellular phone microphone and op-amp input impedance Q The majority of cellular phones use electret microphone modules having an output impedance of 2.2 kΩ. The op-amp input resistance is of the order of 1 MΩ. Let us calculate the attenuation of the input signal voltage. A We can use an equivalent circuit showing the microphone voltage, Vmic, the microphone impedance, Zmic, and op-amp impedance, Zin (Figure 23.6). a b Z mic Rmic + Vin+ =2.2 kΩ 2.2 kΩ (Vin+ − Vin −)A v Vmic Zs − Z in Vin − − The ideal op-amp has infinite input impedance Vs − Vout Zout (Vin+ − Vin−)A v represents the voltage generated in the op-amp due to the gain of the amplifier Gnd The integrated circuit op-amp amplifier is made up of a namely, transistors, resistors and capacitors. It consists, Vin + Rmic Z mic takes place Opposition to currents on+account of Z in 2.2 kΩ resistances, as well as the presence of Z as Xtreactances (due to (Vin+ − Vin−)A v Vmic Z s Z in Vmic Zout X to represent reactance.) The combined opposition 1 MΩ Zs − Z in Vin − to current in a circuit is called impedance, and, like V mic − (Vin+ − VZ in − A resistance, is measured in ohms. We use the symbol two represents the voltage generated in 1 Vrms the op-amp due to the gain of the amplifier Gnd capacitances and inductances (We the amplifier to capacitances and inductances within the op-amp (Figure 23.6).
We can extend the modelling of the op-amp Figure 23.6 (a) Microphone connected to input of an op-amp. (b) Equivalent input circuit showing Zin = (Known as) connected in by adding input and output impedances at each of the parallel with (Zin). terminals, as illustrated in Figure 23.5; input signal source impedances are also shown. Using the potential divider equation (see Chapter 19), and the input voltage, Vin− at ground potential of 0 V, and Zmic << Zin we The op-amp can be used in various ways. It may be set: to simplify the signal from temperature sensors, to Zin input voltage, V in = Vmic Zin + Zmic This may be Vmic with a more op-amp voltages so as to realize the function of audio Zmic + Zin A Zs+ Vin+ R Vs+ So, (Vin+ − Vin−) (i.e. there is hardly any attenuation of the source voltage). (Vin+ − Vin−) A v Z V Zin − Vin − Vout Zout Vs − Gnd Figure 23.5 Impedances: source impedances Zs and op-amp input Zin comprising Rin and Cin, and output impedance Zout of the op-amp circuit. 367 mixing desks, to perform digital to analogue conversions, and a host of other applications. The ideal amplifier will not electrically 'load' the signal source (i.e. will not attenuate the signal voltage reaching the inputs), for a wide range of source impedances. Worked example 23.2 shows that with a 2.2 kΩ source impedance, the 1 MΩ op-amp input impedance hardly attenuates the electret microphone's signal. Calculations ITQ 1 What is the S.I. unit for impedance? 1 Vrms Vmic 1 Vrms 368 Unit 2 Module 2 A.c. theory and electronics Worked example 23.3: Cellular phone voltage supply (Vs) and power output at 2 Vrms Q Two op-amps are used in a cellular phone to power the left and right headphones.
Each headphone has an impedance of 30 Ω (assume this to be purely resistive). The output stage of each op-amp has a resistance Zout of 10 Ω (Figure 23.7).
(a) How much power is delivered to each headphone when Vout is 2 Vrms? (b) If the op-amp's power is supplied by a single 800 mAh lithium battery, for how long can one listen to 2 Vrms music with the headphones? (c) How much power is wasted in the op-amp's output resistance? (d) Now assume each headphone has a resistive Zout of 1 Ω (instead of 10 Ω); how much power is wasted in the op-amp's output resistance? 2 out 10 V Va.c. Vout = 2 Vrms Rphone = 30 Ω A Figure 23.7 Schematic circuit showing output voltage of an op-amp connected to a headphone (a) Power delivered to each headphone: Vout22 = W (= 0.133 W) Rphone 30 (d) power lost in the op-amp's output resistance of 1 Ω = 2 × (2phone × Rout) = 2 × (0.0067)2 × 1 (I) = 0.089 W fractional power loss in op-amp resistance 0.0089 = ≈ 3.2% 0.275 Note that, for a 10 Ω output impedance, nearly 30% of the power is wasted in the op-amp's output impedance. However at 1 Ω output impedance, only 3.2% of wasted power occurs in the op-amp. The ideal is to have zero output impedance, that is zero power loss. This illustrates the importance of approaching zero output impedance as per the ideal op-amp. In this example we may conclude that an op-amp with less than 1 Ω output impedance is virtually an ideal op-amp (for this application). similar to the above show that for other applications, with source impedances in the hundreds of kΩ, and at the same input frequencies, Zin will lead to significant attenuation of the signal voltage that is to be amplified on reaching the op-amp's input terminal. Note that a capacitor, placed in parallel with the input impedance (as is sometimes done), can also result in attenuation at high frequencies, since capacitive reactance decreases with frequency. The ideal op-amp has zero output impedance (b) The current, Iphone, through each phone: Vout 2V = = 66.7 mA Rphone 30 Ω The op-amp output may be connected to a headphone socket. Headphones may be specified with impedances in the range of 30 Ω.
The op-amp amplifies the signal of the source. It may then be required to deliver signal power into the 'load' at the output, Vout, of the amplifier. Any impedance in series with the output stage, especially resistive impedance, results in power loss.
A waste of battery power is very important to avoid, for example, in the case of a cellular phone, where increasing talk time per charge is of prime importance. To minimize power losses, the real op-amp must therefore have near zero output impedance.
The left and right channels will therefore together take 2 × 66.7 mA = 133 mA For the 800 mAh battery to be discharged in time, t 800 mAh = 133 mA × t therefore, t = 6 h The 800 mAh battery will supply this current for approximately 6 hours (c) Power lost in op-amp's output resistances (impedances) due to two headphones, each being connected in series with its op-amp output resistance = 2 × (Iphone × Rout) = 2 × (0.0667 A)2 × 10 Ω = 0.089 W total power used by phones and output resistances = 2 × 0.133 W + 0.089 W = 0.315 W therefore fractional power loss in op-amp output impedances 0.089 = = 0.28 0.315 ≈ 0.30 (or nearly 30%) Worked example 23.2 gives more details.
The ideal op-amp has infinite speed When a signal is applied to the input of a non-ideal op-amp, there is a delay before the output reacts (the capacitances within the op-amp require finite times for charging and discharging by the signal voltage). This reaction delay is called the 'propagation' delay through the op-amp. Also, if the input signal commands the output voltage to rise instantly, a non-ideal amplifier will always take a short time to reach the target voltage (Figure 23.8). An ideal op-amp will have zero propagation delay and zero output rise time. Chapter 23 Operational amplifiers the equations that follow, I, P and V represent intensity, power and voltage, respectively, for a signal. propagation delay Vout ideal op-amp Vout, non-ideal op-amp rise time Vout 0 Figure 23.8 Non-ideal op-amp propagation delay and rise time (Chapter 10, equation 10.5) since by definition, power gain, Ap, in dB is given by: P Ap (in dB) = 10 log ΔPout in But since power is directly proportional to the square of voltage, voltage gain can also be expressed in dB: ( ) Vout rise time t (μs) Figure 23.8 Propagation delay and rise time for a real op-amp. An ideal op-amp has zero offset voltage If the voltage at the inputs Vin+ and Vin− of the op-amp are exactly the same, equation 23.2 suggests that the op-amp voltage, Vout, should be 0 V. This is the case for ideal op-amps. In real amplifiers, because of the large voltage gain and component mismatches within the amplifier, the output voltage will not be zero volts; one must bear this in mind when using the amplifier as a comparator (the comparator will be discussed shortly). The offset voltage of an op-amp is defined as that voltage, which is applied to one of the input pins of the op-amp, causes the output to be at 0 V. The ideal op-amp has zero offset voltage. Real operational amplifiers Use of logarithmic scales Voltage gains of op-amps, and corresponding frequencies, vary very widely and are usually expressed on a logarithmic scale. As in the case of sound levels (Chapter 10), the unit for this logarithmic ratio is the decibel (dB). The decibel is a power ratio. Hence, sound power intensity (= area × time) ratios can also be expressed in dB. In Worked example 23.4: Voltage gain in dB (a) What is the voltage gain, in dB, of an op-amp that has a voltage gain of 1 million? A The voltage gain of 1 million: V Av = out = 1 million = 106 ΔVin but, by equation 23.3: V Av (dB) = 20 log out ΔVin = 20 log106 = 20 × 6 = 120 dB See Appendix 3 for a discussion of logarithms. ( ) ( ) Voltage gain (in dB) is defined by: V 2 Av (in dB) = 10 log ΔVout 2 in That is, V 2 Av (in dB) = 10 log ΔVout in V Or, Av (in dB) = 20 log ΔVout ....(23.3) in ( ) ( ) ( ) Worked example 23.4 gives more detail. Voltage transfer characteristics The voltage transfer characteristic of an op-amp is defined as the variation of output voltage, Vout, with differential input voltage, ΔVin. Figure 23.9 shows three curves, each curve corresponding to a specific voltage gain.
Worked example 23.5 considers an op-amp used in a mobile telephone supplied by voltages ±2.0 V. The graphs in Figure 23.9 show that the useful operating range of input signals allowed by an op-amp is limited by the gain of the amplifier. For a gain of 40 dB, the range is only 0.2 mV. Op-amps in open loop mode So far we have been looking at the op-amp without any 'feedback'. The open loop description of an op-amp denotes one without 'feedback' resistors. This 'open loop' usage is not generally used in precision applications, as the voltage gain is very high and varies with temperature and supply voltage, resulting in output distortions (non-linear amplification). Open loop gain and frequency Let us consider the general purpose op-amp, the Texas Instruments OPA211 (for full details see: ti.com/lit/ds/symlink/opa211.pdf).
The simplified open loop gain curve (Figure 23.10) shows how the voltage gain varies with the input signal frequencies. At d.c. and low frequencies up to the turn-point frequency of about 369 Unit 2 Module 2 A.c. theory and electronics Worked example 23.5: Voltage transfer curves of a 40 to 80 dB voltage gain op-amp Q A An amplifier is connected to voltage supplies of ±2.0 V. Sketch the voltage transfer curves (with x-axis in mV and y-axis in volts) when the voltage gain, Av, of the amplifier is set from 40 dB to 80 dB in 20 dB steps. The supply voltage of ±2.0 V implies that the output voltage, Vout, will be saturated at 2.0 V. V For Av = 40 dB, the gain (out) = 100 (see equation 23.3 and ΔVin ITQ 1). Hence, by equation 23.1: 2V ΔVin = 100 = 20 mV Similarly, for Av = 60 dB: 2V ΔVin = 1000 = 2 mV and for Av = 80 dB: 2V ΔVin = 10 000 = 0.2 mV By equation 23.1, the graphs of Vout versus ΔVin are straight lines. The graphs corresponding to the voltage gains of 40 dB, 60 dB and 80 dB are shown in Figure 23.9. output voltage 'saturated', i.e. limited by the supply voltage +Vs (given = 2 V Vout / V +2 V Av = 80 dB Av = 60 dB Av = 40 dB −20 ΔVin / mV output voltage 'saturated', i.e. limited by the supply voltage −Vs (given = −2 V) Figure 23.9 Voltage transfer characteristics of 40, 60 and 80 dB op-amps having supply voltages Vs+ = +2 V and Vs− = −2.0 V. 120 Open loop gain, A v / dB 370 100 A v slope = 20 dB / decade 80 60 40 20 dominant 1.0E+00 1.0E+02 1.0E+04 1.0E+06 Frequency / Hz Figure 23.10 Open loop voltage gain versus input frequency. Note that 1.0E+x is the same as 1.0 × 10x.
200 Hz (described as the dominant pole frequency) the gain is flat with a value of 112 dB, i.e. × 400 000; thereafter the gain decreases with increasing frequencies. Plotting the gain using a linear scale is somewhat impractical due to the very large gain values.
As shown in Figure 23.10, the voltage gain is plotted on the y-axis, with values in dB (i.e. values of 20 logVout/Vin). This is therefore a log scale. Likewise, the frequency on the x-axis is plotted on a logarithmic scale, to cater for the large range of frequencies. (See Appendix 3 for a discussion on logarithms and logarithmic scales.) Open loop bandwidth The bandwidth of an op-amp is the range of frequencies over which the gain is constant within an attenuation of 3 dB. In Figure 23.10, the 3 dB attenuation occurs at the intersection of the flat and slope projections of the curves. Following the dominant pole frequency of 200 Hz, the open loop gain falls off with a slope of −20 dB per decade (i.e. per power of 10 frequency). The graph in Figure 23.10 shows that at a voltage gain of 0 dB, i.e. for an amplifier gain of 1, the frequency is 100 MHz (written as 1.0E+08 or 1.0 × 108 Hz).
The frequency of 100 MHz is referred to as 'unity gain frequency' or 'the unity gain bandwidth' (UGBW) of the amplifier. Before we continue with practical circuits, let's make a few observations. For d.c. and low frequencies, the OPA211's ITQ 4 Explain why a voltage gain of 0 dB corresponds to an amplifier gain of 1. ITQ 2 How much is the op-amp gain that corresponds to a voltage gain of 40 dB? ITQ 3 What do the slopes of the graphs in Figure 23.9 represent? ITQ 5 Figure 23.10 What is the maximum voltage gain from the op-amp at 1 MHz is 40 dB. What is the maximum voltage gain, at 10 kHz? Chapter 23 Operational amplifiers open loop gain is 112 dB (≈400 000) Now this very large low frequency open loop gain is not stable; it changes with temperature, supply voltage and output load impedance (or resistance). For this reason the op-amp OPA211, although rated at a typical gain of 114 dB, specifies the maximum practical Av as 110 dB with load at RL = 600 Ω connected to the output. Applications requiring precise gains will not use the open loop mode. Open loop circuit applications The open loop gain of an op-amp varies with temperature, supply voltage and its output load. Care must therefore be taken when using this configuration. Comparator The comparator circuit (Figure 23.11) is used to indicate at its output when a signal on one of its inputs is less or greater than a set reference voltage at its other input. The reference voltage can be 0 V, or any other value that can be tolerated by the input of the op-amp. Since the op-amp in the comparator configuration is only comparing voltages, the supply voltage and the amplifier selected must give good stability with input voltage variations. Vin V Vs+ +Vs V+ Vout V− Vs− −Vs Figure 23.11 Comparator circuit Recalling equation 23.1, voltage gain, Av, for a given output voltage, Vout, is given by V Av = ΔVout in V = V −outV in+ in− ....(23.1) Hence, Vout = Av(Vin+ − Vin−) ....(23.4) ■ If Vin+ > Vin−, then ΔVin is positive and, therefore, by equation 23.4, Vout is positive.
■ If Vin < Vin−, then ΔVin is negative and, therefore, by equation 23.4, Vout is negative. For a change in the sign of (Vin+ − Vin−), therefore, Vout changes sign.
Recall also, that for a very large voltage gain, a small difference in (Vin+ − Vin−) gives rise to a large Vout. The output voltage, Vout, thus ■ indicates if there is even a tiny difference between the two input voltages (by Vout being large and non-zero), and ■ indicates which of the input voltages is the greater (by the sign of Vout). Vs+ Vin+ Vin − V1 V2 Vs − Vout RV Figure 23.11 Op-amp comparator circuit. For these reasons, the circuit of Figure 23.11 is called a comparator circuit, since it serves to compare the sizes of two input voltages. The above equations are true only when there is no input offset voltage or if the input offset voltage (125 μV max for the OPA211) has been nullified. Worked example 23.6: Op-amps as a comparator Q Assume there is zero input offset voltage. For a supply voltage Vs = ±5 V and an op-amp gain Av = 106: (a) what minimal differential input voltages, ΔVin, will result in saturation with a negative voltage on the output? (b) which input would be at the greater voltage? A (a) By rearranging equation 23.1, we get ΔVin = out Av 5 V (at saturation, Vout = −Vs for a negative 106 output voltage) = −5 μV Therefore, for Vout = −5 V, we have ΔVin ≤ −5 μV. (b) Since Vout is negative, then by equation 23.4, ΔVin is also negative. Therefore, Vin− > Vin+ (see discussion immediately preceding this worked example). ■ Sine-to-square wave converter (zero-crossing detector) The sine-to-square wave converter is another application of the open loop arrangement (Figure 23.12). The sine-to-square wave converter is, in effect, a comparator with a zero reference voltage. The input signal is applied to the non-inverting input Vin+ with Vin− connected to the zero voltage ground reference. The differential voltage to the op-amp is thus ΔVin = Vin+ − Vin− = Vin+. For low frequency operation, the open loop gain is very high (see Figure 23.10). Therefore, the output voltage quickly reaches positive saturation voltage when Vin+ is positive, and negative saturation voltage when Vin+ is negative. Worked example 23.7: Light-operated alarm circuit Q Figure 23.12a shows an op-amp used in a low frequency sinusoidal voltage, V, connected to a potential divider so as to provide an input voltage at Vin+. Vin− is connected to the output of a potential divider that provides a reference voltage Vref. The op-amp supply voltages are ±5 V. (b) Figure 23.12 ... 371 372 Unit 2 Module 2 A.c. theory and electronics a b R2 space V+ +5 V Vout Vin R1 4 Vs+ Vout VDC = 112 dB R2 Vout Vs+ Vin− +V Vref 0 V Vin = + Vout 5 V Vs− V1 R1 +5 V Vout VDC = 112 dB R2 0V Space Vin+ +5 V Vout mark Vs +5 V V ref Vin+ V −5 V ref 0V Vin − a G Vin− = Vref > 0 V Vin− sine wave = V x R1 R1 + R2 Vout = wide pulses (mark > space) Figure 23.12 Sine to square wave converter using a zero crossing detector. (a) Zero-crossing voltage detector (Vin− = 0 V) with sinusoidal input applied to Vin+. (b) Square wave output produced (mark = space) from a sine wave input when Vin− = (reference voltage) = 0 V. (c) Vin+ set as Vref > 0 V and sine wave applied to Vin−. (d) Pulsed output produced (mark > space) from sine wave Vin+ = (reference voltage) > 0 V. input voltage crosses 0 V. The output remains in this state for the first half-cycle (Figure 23.12b). As the input voltage crosses 0 V to begin the second half-cycle, the output voltage quickly reaches negative supply voltage, and the amplifier remains in this state for the entire second half-cycle (Figure 23.12b).
This gives the 'mark:space' ratio of 1:1, resulting in a 'square wave'. The circuit is therefore likened to a 'zero crossing detector'; a positive slope crossing gives a positive (saturated) output voltage and a negative slope crossing gives the opposite output voltage polarity. In Figure 23.12c, Vin+ is set at a reference voltage Vref > 0 V. Let us suppose that in Figure 23.12d the Vin+ line intersects the sine wave graph for input Vin− at times a, b, c and d, as shown in the diagram. Very shortly after time a, ΔVin (= Vin+ − Vin−) < 0; hence the output saturates to −5 V. Very shortly after time b, ΔVin > 0 V, and the output saturates to +5 V. The output remains at this positive value until time c, where saturation to −5 V occurs very shortly thereafter, since ΔVin < 0 V again. Very shortly after time d, ΔVin > 0 V, and the output saturates to +5 V, and the cycle repeats with the next sine wave. Introducing the positive Vref results in a longer duration of high positive output voltage than in the previous square wave example. Hence, the 'mark-to-space' ratio, is >1:1, resulting in a train of positive pulses of a fairly ITQ 6 What would happen to the mark-to-space ratio if the input pins connections were reversed, i.e. the positive reference voltage now appears on the Vin− input pin and the divider resistors sine voltage on Vin+? large duration for each cycle of the sine wave voltage input.
When the reference voltage is set positive, zero or negative, a comparator can therefore produce from a low frequency sinusoidal voltage input a train of positive pulses, each train with different mark-to-space ratio dependent on the magnitude of Vref. A note of caution: when the signal source exceeds the input voltage range of the op-amp, a potential divider R2, R1 can be used to attenuate the signal as it is always prudent to connect any sources to the op-amp inputs using a resistor. This is to avoid destructive high-voltage signals appearing at the input pins of the op-amp. Temperature alarm The thermistor is a negative temperature coefficient (NTC) resistor made of semiconductor material – that is, its resistance decreases with increasing temperature. Its characteristic can be approximated by the equation: change in resistance ΔR = k ΔT where k = constant of resistance temperature coefficient and ΔT = change in absolute temperature. A comparator circuit, involving a thermistor, can be used as a temperature alarm (see Discussion example 23.7). Analogue to digital converter (ADC) Comparator circuits, with certain additions, can be used to convert analogue voltages to digital binary codes (groupings of 'bits' – each bit represents a 0 or a 1). A discussion of such circuits lies outside the scope of this book. Chapter 23 Operational amplifiers Discussion example 23.7: Car coolant temperature warning light Q Let us look at the use of the thermistor, and in terms of linear temperature measurement, but as an alarm (Figure 23.13). A red warning light on the dashboard of a car will illuminate when the car engine's coolant temperature exceeds a safe level. A The circuit uses two potential dividers to define the Vin+ and Vin− voltages.
The NTC resistor (R2) and R1 are selected such that below the selected temperature, Vin+ is lower than Vin−, causing ΔVin to be negative. Vout becomes approximately equal to Vs−. Because Vout is negative, the LED warning light will be OFF, since the LED would be reverse-biased. As the temperature rises, the resistance of the NTC (R2) falls causing Vin+ to increase, making ΔVin positive.
The output, Vout, now switches to Vs+ when the coolant or oil temperature is exceeded. This occurs for a tiny difference in voltage between Vin+ and Vin−. The tiny input voltage required for saturation makes the op-amp more sensitive than the transistor in detecting temperature difference. R4 R 2 (NTC) V Vin R1 A vDC = 112 dB Vout = +/- 5 V R3 Vin− Vs −5 V 0V Figure 23.13 Comparator used as an oil temperature alarm in a car. Vs1 Vs3 Vmic / V 0.1 0.2 0.3 Time / ms Figure 23.15 Analogue signal from a microphone sampled at a particular time interval by an ADC. The analogue voltage is converted and stored as digital data. For an N-bit ADC, there are 2N codes. For an ideal 3-bit ADC, there would be 8 codes. More bits will be used about the binary notation later in this chapter and in the following chapter. Figure 23.14 shows digital output in bits for an ideal 3-bit ADC as a function of analogue input voltage. Figure 23.15 shows a sampling of an analogue voltage produced by a microphone. Vs +5 V Vin+ Vs2 Figure 23.16 analogue voltages at a particular sampling frequency. If the sampling frequency is 10 kHz, the analogue input voltage would be sampled each 0.1 ms and each voltage value stored as a 4 bits of digital data (for a 4-bit ADC). Figure 23.15 a microphone signal sampled at 1 ms intervals and stored as a 2-bit digital code. The digital code from the output of the ADC, sequenced at the input of a digital to analogue converter (DAC), recreates the (red) analogue signal. (DACs will be discussed shortly.) Using 4-bit samples, a more faithful reproduction of the mic signal at Vout is obtained.
111 Op-amp in closed loop modes 110 Inverting and non-inverting modes Digital output 101 1 LSB Closed loop configurations use the op-amp with feedback (Figure 23.17). In the inverting mode, the inverting input (Vin−) is grounded (or at a lower voltage than the positive input, Vin+) and the signal, V1, is applied to the non-inverting input (Vin+). The output is therefore 180° out of phase with the input signal.
In the non-inverting mode, the inverting input (Vin−) is grounded (or at a lower voltage than the positive input, Vin+) and the signal, V1, is applied to the non-inverting input (Vin+). The output is in phase with the input signal. 100 011 010 001 000 0 1 8 1 4 3 8 1 2 5 8 3 4 7 8 Analogue input Figure 23.14 Digital output for an ideal ADC as a function of analogue input in volts. ITQ 7 For a 4-bit ADC, how many digital codes would there be? 373 Unit 2 Module 2 A.c. theory and electronics a b 0.3 2 bit DAC Vout / V 0.3 Mic voltage / V 374 0.2 0.1 0 0.2 Figure 23.16 (a) Analogue signal from a microphone sampled and stored as a 2-bit digital code. (b) The digital code is sequenced at the input of the DAC which gives the (red) analogue signal. Using 4-bit samples would give a more faithful reproduction of the mic signal at Vout. 0.1 0 0 1 2 3 4 5 Time / ms 01 10 10 11 11 10 2 bit digital code corresponding to the mic voltage A-D, analogue to digital conversion (2 bit words) ADC 'coder' 0 1 2 3 4 5 Time / ms 01 10 10 11 11 10 2 bit digital code corresponding to the mic voltage D-A, digital to analogue conversion (2 bit words) DAC 'coder' Op-amp inverter (closed loop) So far we have looked at cases where the negative input Vin− has been grounded, or at a lower voltage than the positive input, Vin+. Let us now consider grounding the positive terminal Vin+. R1 is connected to Vin− and a feedback resistor, RF , is connected between the output Vout and Vin−. The amplifier is described as being connected in 'inverting mode' with negative feedback through RF and R1 (Figure 23.17a). The feedback is called negative when a portion of the output voltage is fed back to the inverting input of the op-amp; this leads to a more precise and stable output voltage gain. When feedback is applied, the amplifier is said to be operating in the closed loop mode. I1 = IF, and therefore, by Ohm's law: V1 Vout R1 = − RF Therefore, closed loop gain: V R ACL(inv) = − Vout = − RF 1 1 The negative sign in equation 23.5 indicates that the input, V1, is out of phase with the output, Vout. Equation 23.5 tells us that the gain of any op-amp in the inverting mode depends only on the input resistor, R1, and the feedback resistor, RF; the gain in this mode does NOT depend on what kind of op-amp it is, so long as its open loop gain is high enough! The OPA211 acts as a real op-amp on account of its very high input impedance. If we choose R1 = 1 kΩ and RF = 100 kΩ, all the current (I1) through R1 would go through RF (as less than 100 nA is taken by the op-amp). Therefore we can say that I1 = IF. For an output of 1.0 V, the differential input voltage is ΔVin− = 1.0 V/Av ≈ 2.5 μV which is negligible (since Av ≈ 400 000). With Vin+ at ground this voltage is small enough to be ignored and Vin− is considered to be virtually at ground potential (this concept was discussed in this chapter, page 366). So: a RF Vin+ Vin− Av Vin+ Vin Vout time 0V Av ΔVin Vin− ΔVin 0V Vout 0V R1 V Getting it right! Virtual op-amp potential denotes potentials that are almost, that is virtually, the same. In the inverting mode, the positive input is usually connected to ground; on account of the very small value of ΔVin, the negative input is described as being at a virtual ground potential. In an op-amp, although the potentials are virtually the same, large input impedance separates them making ΔV the virtual ground current (Z in) negligible. Practically in all the current due to V1 passes through the input and feedback resistors, R1 and RF1. b Vout virtual ground ....(23.5) RF Vout V Vout = inverted sinusoid Vs = input sinusoid Figure 23.17 Op-amp in closed loop (a) inverting mode and (b) non-inverting mode. R1 0V Vin time Vout = in phase sinusoid Vs = input sinusoid Chapter 23 Operational amplifiers Discussion example 23.8: Gain of an op-amp in the closed loop inverting mode Consider the op-amp shown in Figure 23.18, with RF = 100 kΩ and R1 = 1 kΩ. (a) What will be the closed loop gain and bandwidth of such an arrangement? (b) What would be the closed loop gain and bandwidth when RF = R1? a b RF = 100 kΩ R1 = 1 kΩ V virtual ground Vin− Rin Av Vin+ Vout = 1 V Open loop gain, A v / dB Q 120 100 60 40 20 0 0V A v slope = 20 dB / decade 80 closed loop gain UGBW dominant pole 1.0E+00 1.0E+02 40 dB BW 1.0E+04 1.0E+06 Frequency / Hz 1.0E+08 Figure 23.18 (a) Op-amp in closed loop inverting amplifier mode. (b) Op-amp open loop and closed loop gains compared. Note the larger bandwidth at 40 dB. A (a) Using equation 23.5, we get the closed loop gain 100 kΩ ACL(inv) = − 1 kΩ therefore, magnitude of gain = 100 (or 40 dB). The open loop gain (the upper curve) and the closed loop gain (the lower curve) are compared in Figure 23.17b. When the negative feedback resistor is applied, the closed loop gain is reduced to ×100 (or 40 dB). At this gain the bandwidth is 1 MHz. R (b) When RF = R1, ACL(inv) = − F = −1 (unity gain) R1 magnitude of gain = 1 (or 0 dB) At this gain the bandwidth is 100 MHz. If a single op-amp is used in the inverting mode, the output voltage is inverted (180° out of phase) with respect to the input signal. In audio and some other applications this inversion can be problematic.
One can therefore use a second op-amp in the inverting mode to revert the output in phase with the input signal. To sum up, some advantages of negative feedback in the closed loop inverting mode are: ■ a more stable output voltage, since the gain is less and therefore the output voltage is further from saturation voltage ■ a more predictable gain, since gain depends only on the input resistor (R1) and the feedback resistor (RF) and not on what kind of op-amp it is ■ a more linear amplification (i.e. distortion of the amplified signal is lessened) ■ the gain is constant over a wider range of frequencies (large bandwidth). Closed loop (inverting) applications The summing amplifier With the op-amp in the inverting mode, additional resistors can be added to the inverting input and connected to other voltage sources. Vin− is a virtual ground pin, typically within 100 μV from Vin+ which is at ground potential (Figure 23.19). As all the currents through R1, R2, R3 and R4 flow into RF (the op-amp typically takes less than 100 nA), a mathematical expression can therefore be written, as follows, for input signals, V1, V2, V3 and V4, of the order of mV: V1 V2 V3 V4 Vout ....(23.6) R1 + R2 + R3 + R4 = − RF Now consider the case where R1 = R2 = R3 = R4 = R. The expression becomes V1 V2 V3 V4 Vout R + R + R + R = − RF 375 376 Unit 2 Module 2 A.c. theory and electronics a b MSB Q3 R1 Q2 R1 R2 RF R2 R3 V2 V3 0V R3 Vin− V1 R4 V4 RF Q1 Av Vin+ LSB Q0 Vin− Av R4 Vout = 1 V Vin+ Vref 0V virtual ground virtual ground Q3 Q2 Q1 Q0 input pulses Vout = 1 V up-down counter An up-down counter is used to provide the signal codes at Q 3 , Q 2 , Q1 and Q0 for the above 4 bit DAC Figure 23.19 (a) The op-amp in inverting mode as a summing amplifier. (b) VREF connected via switches to input resistors. Thus R Vout = −(V1 + V2 + V3 + V4) × RF ....(23.7) This is the equation of a summer, where the output voltage is the resistor ratio RF/R multiplied by the sum of the input voltages V1, V2, V3 and V4. Worked example 23.9: Output voltage from an inverter summing amplifier Q In Figure 23.19a, an ideal op-amp is used in the inverting configuration with Vin+ connected to ground. The input resistors to the negative input Vin− are R1 = 2R, R2 = 4R, R3 = 8R and R4 = 16R.
The negative feedback resistor RF = R.
Four voltage sources each of value VREF are connected to each of the input resistors. Derive an expression for the output voltage, Vout. A Using equation 23.6, we get V V1 V2 V3 V4 + + + = − out R1 R2 R3 R4 RF Summing amplifier as an audio mixer The technique described above can be used to design an audio mixing desk in which several microphones, CD players and other audio signals are to be mixed. Note that the output signal is of the opposite phase to that of the input signals.
An inverting stage with input Rin and output RF resistances of equal magnitude can therefore be cascaded (connected in series) with the output of the summing op-amp to restore the audio signal to its true phase. Typically the line output voltage from a portable music player is much higher than that of a microphone. For an RF = 100 kΩ, Rin(mic) may be say 1 kΩ (giving a gain RF/Rin(mic) of 100); whereas Rin(music) may be 10 kΩ (giving a gain of 10). The music amplification is therefore smaller than that of the microphone. Equation 23.7 shows that it is the resistor ratio, RF/R, that determines the accuracy of the summer. The integrated circuit (IC) process permits accurate matching of resistors (and other devices), so accurate summers are realizable in the IC process. Summing amplifier as a digital to analogue converter (DAC) First, look at Worked example 23.9. Let us consider four switches that can connect a voltage (VREF) to each of the input resistances (Figure 23.19b). VREF VREF VREF VREF V + + + = − out i.e. 2R 4R 8R 16R R VREF VREF VREF VREF Vout = −R × + + + 2R 4R 8R 16R (= −VREF × (21 + 41 + 18 + 161)) ....(23.8) When a switch is in the 'on' position (e.g. the switch shown associated with Q0) VREF is connected to the associated input resistance (R4 in this case). When a switch is in the 'off' position, the voltage at the associated input resistance is 0 V. Signals Q0, Q1, Q2 and Q3 action the switches between their 'on' and 'off' states. For signal Q0, we can call the 'on' position Q0 at logic 1 and the 'off' position Q0 at logic 0. The four switches, Q0 to Q3, then represent four binary 'bits' of 'on'/'off' logic. The bits may have been produced by an analogue to digital 4-bit converter (ADC). (See earlier discussion about ADCs.) Four 'bits' of digital data from an ADC are called a 'word' or 'byte'. ITQ 8 What would be the output voltage corresponding to the 4-bit word 0011 (equation 23.8)? Chapter 23 Operational amplifiers Worked example 23.10: DAC with weighted resistors: counting and timing diagrams Q A four-bit DAC uses the weighted resistor arrangement and an ideal high gain, high input impedance, op-amp in the inverting configuration with Vin+ connected to ground (see Figure 23.19 and Worked example 23.9 above). The switches in Figure 23.19b are activated by a 4-bit binary up-down counting digital circuit. The activation of the switches is being done by outputs Q3, Q2, Q1, Q0 in a binary counting sequence (4-bit in this case). This sequencing is repeated until the clock CLK is stopped. (a) Write in a table the digital states for one count-up cycle starting at zero (and finishing at 15). (b) Sketch a timing diagram of the output voltage (Vout) versus the clock pulses and the digital outputs Q3, Q2, Q1, Q0 of the counter.
A (a) Table 23.1 shows the up-down count for 15 clock pulses. Table 23.1 # clock pulses Q3 Q2 Q1 Q0 reset 0 0 0 0 1 0 0 0 1 2 0 0 1 0 3 0 0 1 1 4 0 1 0 0 5 0 1 0 1 6 0 1 1 0 7 0 1 1 1 8 1 0 0 0 9 1 0 0 1 10 1 0 1 0 11 1 0 1 1 12 1 1 0 0 13 1 1 0 1 14 1 1 1 0 15 1 1 1 1 V1 R1 0 16 1 16 2 16 3 16 4 16 5 16 6 16 7 16 8 16 9 16 10 16 11 16 12 16 13 16 14 16 15 16 Equation 23.8 suggests that if VREF = 1 V, we can 1 reconstruct, in steps of 16 V, any voltage from 0 V to 15 16 V by appropriate combinations of the four bits.
For example, the 4-bit combination of switches, 1000, starting from R1 and proceeding in sequence to R4, would give 1 1 an output voltage of (2 + 0 + 0 + 0) V, i.e. 2 V. The 4-bit 1 combination of 1001 would result in an output of (2 + 0 1 9 + 0 + 16) V, i.e. 16 V. In this case, the 4 bits are referred to as 4 bits of digital data (or a 4-bit digital 'word') and the op-amp output as an analogue representation of the 4 digital bits. We have digital words in, analogue levels out. The reconstruction of an analogue voltage from digital bits is called 'digital-to-analogue' conversion. The circuit (b) Timing diagram The outputs of the counter are connected to the digital inputs of the 4-bit DAC; this gives the voltage 'staircase'. Vout the DAC's resolution is Vout_max /2 n where n is the maximum number of input bits of the DAC Each state of the counter corresponds to an analogue voltage at the output of this single supply voltage DAC. For this 4-bit DAC there will be 2 4 output voltage levels. Q3 Q2 Q1 Q0 CLK 01 08 16 No. of CLK pulses The clock pulses (CLK) are counted by the 4-bit digital counter circuit using 'falling edge triggering' Time / ms Figure 23.20 Timing diagram: voltage output (Vout) and counter output (Q0, Q1, Q2, Q3) as a function of clock pulse (CLK) number. Q0 to Q3 switch between logic 1 and logic 0. in Figure 23.19b is therefore a crude type of digital-toanalogue converter (DAC) where the digital data bits Q0 to Q3 are converted to an analogue representation. CDs, flash drives, etc., can store musical digital bits; DACs re-convert these bits to the original analogue audio music. The timing diagram (Worked example 23.10 and Figure 23.20) shows the reconstruction of (analogue) voltages in steps during one cycle involving a train of 15 clock pulses (in the reset state, Q0 to Q3 are at 0000; clock pulse 1 initiates the count to 0001, which is the binary count 1 for 16 V). The initiation of each pulse makes use of 'falling edge triggering' (see Chapter 24, page 394.) 377 378 Unit 2 Module 2 A.c. theory and electronics R3 = R A R5 = R B R7 = R C RF = 2R R9 = 2R D Vin− R1 = 2R R2 = 2R R4 = 2R R6 = 2R Vin+ R8 = 2R MSB LSB Vref Q0 Q1 Q2 Q3 For a DAC, more precise reconstruction of voltages using bits can theoretically be done if more input resistances and corresponding bits are used, since the voltage increments (bits) would then be smaller. However, there are two major disadvantages to this approach: ■ obtaining precision resistors of various values is not an easy requirement to meet and is expensive ■ the different input resistors will load the op-amp inputs differently, resulting in errors in the calculated values of Vin, and hence of Vout. R-2R Digital-to-analogue converter (DAC) The inverting configuration along with a R-2R 'ladder' resistor arrangement is another technique used to realize a digital-to-analogue converter.
Figure 23.21 shows a 4-bit R-2R DAC. It can be shown (though it is outside the scope of this book) that the R-2R ladder network in conjunction with the op-amp (as shown in Figure 23.21) encounters the same input resistance (2R) regardless of which combination of resistances is switched ON by Q0 to Q3. Thus the problem of input loading variations for various resistance combinations (that exists in the weighted resistor DAC approach) is eliminated. Further, there are only two precision resistances involved (R and 2R) and this is neither difficult nor expensive to manufacture. Hence, this method of reconstructing analogue voltages from digital words is far superior to the 'weighted' method depicted in Figure 23.19. Discussion: high-quality ADCs and DACs The sounds we hear are analogue, i.e. variations in amplitude are continuous and smooth (Figure 23.16a). To obtain a high-quality reproduction, low cost, small sizes and for other reasons, most recordings today are – Av + Vout Figure 23.21 The R-2R resistorladder network used with inverting op-amp for a digitalto-analogue converter (DAC). The MSB control bit Q3 enables switching of R8 to Vref. The digital outputs Q0 to Q3 switch between LO and HI voltage levels. The reference voltage for the R-2R ladder uses the independent voltage reference Vref; Q0 to Q3 actuate its use. transformed from analogue to digital words or bytes (using analogue to digital converters, ADCs) immediately after capture (e.g. after the microphone). Processing the sound (adding reverberation, etc.) is done by computer manipulation of the digital bytes, then the resulting bytes are stored on a hard drive, flash drive or written to a CD. CDs use 16-bits to represent each analogue level; this gives over 65 000 (i.e. 216) levels of voltages for each analogue conversion; for a 1 V analogue signal this represents a voltage step of ~15 microvolts! Such a difference is not perceived by the human ear. On playback, the 16 bits are read from the storage device and sent to a digital-to-analogue converter (DAC) to reconstruct the signal back to the analogue form. Worked example 23.9 above showed a simple 4-bit system (which allows for only 24, i.e. 16, combinations of bits permitting 16 analogue levels) to explain how this is done. DACs are very important devices, especially when used in 16- or 32-bit communications. Hopefully, you have gained some insight into how analogue voltages can be sampled, converted to groups of digital bits ('digital words', called 'bytes'), then stored on a HDD or CD in digital form. The stored digital words can be read and directed to the input of a DAC. For each digital word, the DAC reconstructs the voltage that had been sampled from an analogue signal. This procedure occurs when we play a CD or listen to digitally stored music. Real operational amplifiers: closed loop (non-inverting) configurations The non-inverting amplifier For this course, only negative feedback around the op-amp is being considered – as positive feedback is used for complex functions. The summing op-amp previously described applies negative feedback in an inverting mode.
Chapter 23 Operational amplifiers Vs +5 V Vin+ + Vout A vDC = 112 dB ΔVin Vin– Vs +5 V Vin+ – + ΔVin RF Vin– Vs –5 V Vs Vout A vDC = 112 dB – RF Vs –5 V Vs R1 = ∞, i.e., open circuit R1 Figure 23.22 Op-amp in closed loop non-inverting mode.
Figure 23.23 High gain op-amp as unity-gain buffer. We are now considering non-inverting configurations (Figure 23.22).
The op-amp unity-gain buffer (voltage follower) The portion of the output, Vout, that is fed back to the negative input, Vin−, is found using the potential divider equation, that is R Vin− = Vout × R +1 R 1 F = Vout β R where β = R +1 R 1 F ....(23.9) 1 = R 1 + RF 1 β is called the feedback factor in this closed loop, non-inverting mode and represents the portion of the output voltage that is feedback to the negative input. Now for very large open loop gain, Av, of the OPA211, Vin+ is virtually the same as Vin−, so, the voltage, V, of the source can be expressed as R V = Vin+ = Vout × R +1 R ....(23.10) 1 F = Vout β Therefore, Vout = V β V Closed loop gain, ACL(ninv) = Vout = In this case, closed loop gain is given by V R ACL(ninv) = Vout = 1 + RF = 1 ....(23.12) 1 A resistor of infinitely large value is in effect an open circuit. Therefore, according to equation 23.12, removing R1 gives a gain of 1.
This is the unity-gain buffer (Figure 23.23). The circuit is also called a voltage follower, since the output voltage 'follows' (i.e. is equal to) the input voltage and is in the same phase. We had assumed an ideal op-amp. The ideal op-amp has infinite input impedance, so the buffer circuit above has infinite input impedance. The ideal op-amp has zero output impedance, so the buffer above has very low output impedance. Unity gain buffers are useful when high impedance sources needs be interfaced with low impedances; using a unity gain buffer the loading of the source is therefore very reduced (i.e. very little current from the source flows through the buffer op-amp input resistance).
Some types of microphones and sensors require high input impedance amplifier stages to avoid loading effects. See Worked example 23.11. Impedance matching 1 β V 1 R i.e. ACL(ninv) = Vout = or 1 + RF β 1 V R From equation 23.10, closed loop gain Vout = 1 + RF , it is 1 seen that when R1 increases in value towards infinity, RF ∞ = 0. .... (23.11) Therefore the closed loop gain with a very high open loop gain op-amp, is equal to (1 + the resistor ratio RF). R1 We have discussed the fact that in order that a source voltage (Vs) is not loaded by the amplifier input stage, the input impedance of the amplifier must be large. Figure 23.25 shows an equivalent circuit of a signal source of voltage V and impedance Z connected to the input of an op-amp of input impedance Zin.
The input voltage at the op-amp is ΔVin. An input current I flows in the input circuit. 379 380 Unit 2 Module 2 A.c. theory and electronics Worked example 23.11: Error from unity gain in using inverting op-amp configuration with real op-amps Q An non-ideal op-amp is used in unity gain non- inverting configuration (Figure 23.24). Compute the error from unity gain when the open loop voltage gain Av is: (a) 60 dB (b) 20 dB Vs +5 V Vin+ + ΔVin Vin– Vout Av – Vs –5 V Vs Figure 23.24 Non-ideal op-amp in unity-gain buffer configuration. A (a) For a non-ideal amplifier, Vin+ and Vin− cannot be considered to be very close to the same potential.
There is a significant potential drop, ΔVin, between the two. By definition, open loop gain, AOL, is given by V AOL = Av = out ΔVin Therefore, Vout = Av(Vin+ − Vin−) But Vin− is connected to Vout Therefore, Vout = Av(Vin+ − Vout) i.e. Vout = AvVin+ − AvVout or Vout + AvVout = AvVin or, Vout(1 + Av) = AvVin Rearranging gives the unity gain as Vout Av = ...(23.13) 1 + Av Vin V2 Av in dB is given by the expression 20 log(V) (see equation 23.3). 1 V Hence, 60 = 20 log(2) V1 V i.e. 3 = log (2) V V 1 so, 103 = 2 = Av V1 For 60 dB gain, we therefore get a numerical gain Av = 1000 (Similarly, for 20 dB gain, Av = 10.) Substituting Av = 1000 into equation 23.13, we get Vout 1000 1000 = = ΔVin 1 + 1000 1001 Gain error is therefore 1 − 1000 1 = or ~0.1% 1001 1001 (b) For Av = 20 dB, numerical gain is = 10. Hence Vout 10 10 = = 1 + 10 11 Vin Gain error is therefore 1 − 10 1 = or ~ 9.09% 11 11 ITQ 9 Would low and high impedance microphones work well using the unity gain buffer? ITQ 10 What do you notice about the percentage error in unity gain as the open loop gain of the op-amp decreases? Chapter 23 Operational amplifiers Zout Z Vin Zin a.c. signal source V Vout ZL eout V op-amp input Vin op-amp output eout = Av*Vin load Figure 23.25 Equivalent circuit diagram for an a.c. source connected to input of op-amp and op-amp output connected to load. For an input current I we get V = I(Zs + Zin) But, ΔVin = I(Zin) ΔV I(Z) Therefore, V in = I(Z +inZ) in (Zin)V Or, ΔVin = Z + Z in ....(23.14) Equation 23.14 shows that to obtain the largest voltage across the amplifier input, i.e. for best voltage matching, Zin must be much larger than Z. In other words, the input impedance of the amplifier must be much larger than the impedance of the source. It can also be shown that that for maximum power to be delivered to the load, i.e. for best transfer of power from source to load (e.g. from the output of an amplifier to a loudspeaker) the source impedance (output impedance of the amplifier) must be equal to the impedance of the load (impedance of the loudspeaker). The equation for best power matching is therefore Zout = ZL...............(23.15) 381 382 Unit 2 Module 2 A.c. theory and electronics Summary ✓ Table 23.2 compares the properties of an ideal operation amplifier (op-amp) with those of a real op-amp. One must bear in mind, however, that one op-amp that might be 'ideal' for a certain application might not be ideal for another. Table 23.2 Ideal op-amp Real op-amp Infinite open loop voltage gain, Av Finite, open loop voltage gain. For practical purposes, gains, Av >105 ≈ 100 dB can be considered as infinite in most calculations, which is constant from d.c. to about 100 Hz and then falls rapidly with increasing frequency. Small bandwidth (≈100 Hz) at very high gains. Large bandwidth (≈10 MHz) at Av ≈ 1 ≈ 0 dB Bandwidth increases with decreasing gain. Infinite input impedance, Zin Finite input impedance. For large Zin say ≈ 106 Ω, negligible op-amp input current is taken from the signal source therefore permitting most of the signal to be at the pin for amplification. Differential input voltage very small, so the inputs are at virtually the same potential. Since real op-amps are manufactured with very high voltage gains, the resulting differential input voltages are small enough for the inputs to be considered at virtually the same potential. Zero output impedance, Zout Have output impedances related to its usage Zout < 102 Ω may be 'ideal' for some applications. Infinite speed (zero propagation Finite speed (non-zero propagation delay and delay and zero output rise time). non-zero output rise time). Zero offset voltage. Non-zero offset voltage. ✓ The gain, Av, of an op-amp for a differential input voltage, ΔVin, and output, Vout, is given by the equation: V Av = out ....(23.1) ΔVin ✓ In the op-amp open loop mode there is no feedback resistor to permit a portion of Vout to be fed back to the input. In the closed loop mode there is a feedback resistor. ✓ Graphs of Vout versus ΔVin in the open loop mode are called voltage transfer characteristics of the op-amp. ✓ The bandwidth of an op-amp operating at a certain gain is the range of frequencies over which the gain is constant.
At very high gains, the bandwidth of an op-amp is only a few hundred Hz. As its gain decreases, its bandwidth increases. ✓ In the open loop mode, a real op-amp can be used as a comparator (with applications to alarm circuits) since the op-amp responds to tiny differences in voltages between the input terminals. The open loop mode can also used in the sine-to-square-wave converter (also called a 'zero-crossing' converter). ✓ Table 23.3 compares features of the op-amp in the closed loop non-inverting mode with those of the closed loop inverting mode.
Table 23.3 Op-amp closed loop inverting mode Op-amp closed loop non-inverting mode The feedback resistor from Vout is connected to the inverting input, Vin−. The feedback resistor from Vout is connected to the inverting input, Vin−. The op-amp can be used as a summing amplifier or as a digital-toanalogue converter (DAC). The op-amp can be used as a unity-gain buffer (also called a 'voltage follower'). The output is given by the equation The output of a summing amplifier consisting of four input resistors, R, in RF parallel and a feedback resistor, RF, is Vout = Vs (1 + R) given by the equation ....(23.11) R Vout = −(V1 + V2 + V3 + V4) × F R ....(23.7) The inverting gain, ACL(inv), is given by the equation ACL(inv) = − Vout R =− F Vin Rin The non-inverting gain, ACL(non-inv), is given by the equation ACL(non-inv) = − ....(23.5) Unity gain occurs if RF = Rin. In the unity-gain inverting mode, the op-amp inverts the polarity of a signal, and can therefore be cascaded in series with the output of another op-amp to re-invert the phase of the output of the latter. Vout R =1+ F Vin Rin ....(23.12) Unity gain occurs if Rin = infinity. This results in a 'voltage follower' or described as 'unity-gain buffer' which can be used in impedance transformation. ✓ To obtain the largest voltage transfer across the amplifier input, i.e. for best voltage matching, the input impedance of the amplifier must be much larger than the impedance of the source. For maximum power to be delivered to the load (e.g. from the output of an amplifier to a loudspeaker) the source impedance (output impedance of the amplifier) must be equal to the impedance of the load (impedance of the loudspeaker). Chapter 23 Operational amplifiers Review questions Voltage and current sources; The operational amplifier; The ideal operational amplifier 1 State the difference between an ideal and a real voltage source. 2 State four properties of an ideal op-amp. For each property, discuss how a real op-amp compares with an ideal op-amp. 3 (a) What is meant by 'voltage characteristic curves' of an op-amp? (b) What is meant by 'saturation' of an op-amp, and under what condition does it occur? (c) On the same graph paper and using the same scales, sketch voltage transfer curves for an op-amp in the open loop mode when operating from a supply voltage of ±5 V at gains of (i) 40 dB (ii) 80 dB (Hint: See Worked example 23.3) 4 (a) Define voltage gain in 'decibels' (dB). (b) (i) Why are graphs of voltage gains usually scaled in dB? (ii) An op-amp has an open loop voltage gain of 300 000. Express this gain in dB. (c) Frequency scales are usually 'logarithmic'. (i) What does logarithm of a number mean? (ii) Why are logarithmic scales used?
5 An electret microphone of impedance 2.2 Ω and output 0.50 Vrms. is connected to an op-amp of input impedance 800 000 kΩ and output impedance 8.0 Ω. The output of the op-amp is fed into a 30 Ω headphone (i.e. into one side only). (a) Calculate the input current to the op-amp. (b) Calculate the power delivered to the headphone if the output voltage of the op-amp is 3.0 Vrms. (c) Calculate the power lost in the op-amp due to the op-amp's output impedance. 7 An op-amp is used in a cellular phone to power a headphone of impedance of 30 Ω (assume this to be purely resistive). The output stage of the op-amp has a resistance Zout of 10 Ω. (a) What is the minimum supply voltage, Vcc, needed before saturation (or clipping) of Vout occurs when Vout is 2 Vrms? (Hint: consider the open loop situation with a load.) (b) Draw a schematic circuit to represent the information. 8 Vin+ Av Vin– Q Vout V s– 0V Figure 23.26 Figure 23.26 shows an op-amp in open loop mode, powered by supply voltages Vs+ and Vs− = +10 V and −10 V, respectively. A sinusoidal voltage, Vin = ±1.0 V peak-to-peak, is applied to the input. (a) Sketch and label, on the same axes, a graph of Vout versus Vin. (b) Draw another diagram to show how the op-amp may be made to produce a square wave output with mark-to-space ratio of (i) 1 (ii) less than 1. Explain your answer in each case. (Hint: see the section on 'Sine-to-square wave converter'.) 9 Figure 23.27 shows a light-operated alarm circuit. Explain how the circuit works. Real op-amps in open loop mode 6 (a) Draw an op-amp comparator circuit in which the input is protected from large voltages. (b) Why is the circuit in (a) called a comparator circuit? (c) For a supply voltage Vs = 10 V and an op-amp gain Av = 4 × 105 Ω (i) what differential input voltages, ΔVin, will result in saturation with a positive voltage on the output? (ii) which input would be at the greater voltage? (Hint: see Worked example 23.6) V s+ P R3 LDR V1 Vs +6 V Vin+ Av V2 R1 R2 Vs –6 V 0V Figure 23.27 Vin– RD 383 384 Unit 2 Module 2 A.c. theory and electronics Real op-amps in closed loop mode; Impedance matching 10 (a) Re-draw the circuit of Figure 23.26 with a 1.0 MΩ feedback resistor, and a 200 kΩ resistor connected between P and earth. Remove the a.c. input and place it between Q and earth. (b) Name this type of closed loop circuit configuration. (Hint: see Figure 23.17a) (c) Determine the voltage gain of the circuit. (d) Calculate the maximum input voltage, Vin, for no saturation to take place at the output (if saturation takes place at ±6 V). 11 (a) What is meant by the term 'bandwidth'? (b) Referring to Figure 23.18b, determine the bandwidth for an open loop gain of 60 dB. 12 (a) Draw the diagram of a real op-amp in the inverting mode, used as a summing amplifier, with input voltages V1 and V2 connected to input resistances of value R and a feedback resistor, RF. (Hint: see Figure 23.19) (b) Derive a formula for the output voltage, Vout, in terms of V1, V2, R and RF. (c) Explain how the 'virtual ground' concept is used in the derivation of the formula in (b). 13 (a) Derive the formula ACL(inv) = −RF/R, for the closed loop inverting amplifier with input resistor R and feedback resistor RF. State two assumptions made in arriving at the formula. (b) An experiment was designed to determine the variation of gain with input resistance R, RF being kept constant at 100 kΩ. How can a suitable graph be plotted involving values of R and corresponding values of ACL(inv) to check the nature of the variation between the two? 14 (a) Draw a diagram of an op-amp as (i) a unity-gain inverter and (ii) a unity-gain follower. (b) Describe, with explanation, one use of each type of op-amp circuit. Practical exercises and challenges Students are advised to carry out these exercises only under proper supervision of a teacher who will be able to reinforce safety practices and observances of safety codes. 1 The syllabus recommends, and gives guidance on, the following practical exercises: (a) Plotting the transfer characteristic of an op-amp in the inverting mode. (b) Determining the theoretical and experimental gain of an op-amp in the inverting mode. (Closed loop experiments are suggested, as open loop gain is very difficult to obtain experimentally.) (c) Investigating the gain versus frequency behaviour of an op-amp. 2 Construct the light-operated alarm shown in Figure 23.27 for various levels of light intensity that trigger the alarm. Obtain a graph showing triggering light intensity versus resistance, R1. 3 Design and carry out an investigation to determine the theoretical and experimental closed loop gains of two different op-amps in the non-inverting unity gain buffering modes. Answers to ITQs 1 The ohm. V V 2 40 = 20 log ΔVout . Therefore, log ΔVout = 2. in in V Therefore, ΔVout = 100 in V 3 Slope = ΔVout = open loop voltage gain, Av. in V V 4 Av = 20 log ΔVout . Therefore, 0 = 20 log ΔVout . in in Vout Vout Therefore, 0 = log ΔV . Therefore ΔV = 1 in in 5 80 dB ( ) ( ) ( ) ( ) ( ) 6 Using a similar reasoning, you will find that the 'mark-to-space' ratio over one time period would be 8 16 V 9 High impedances, yes, as discussed in the section.
The buffer will convert the high impedance of the microphone to a low impedance with unity gain. Low impedances, yes, too; but since the buffer's gain is 1 and the output is of low impedance, it is a waste to connect the microphone to the buffer since this microphone already has a low output impedance. 10 The percentage error in unity gain increases as the open loop gain of the op-amp decreases. Answers to Review questions 5 (a) 0.63 μA (b) 300 mW (c) 80 mW 7 (a) +VCC = 4.0 V; −VCC = −4.0 V 11 (b) 100 kHz 385 Chapter 24 Logic gates Learning objectives ■ Describe the functioning of the following logic gates: NOT, AND, NAND, OR, NOR, EXOR, EXNOR. ■ Use truth tables to represent logic states of logic gates with not more than two inputs.
■ Form equivalent gates from combinations of different gates. ■ Form equivalent gates from combinations of only NOR or only NAND gates.
■ Design and analyse combinations of logic gates to perform control functions. ■ Use timing diagrams to represent the response of digital circuits to different input signals. ■ Draw a circuit to show the construction of a half-adder and explain its operations.
■ Construct a full-adder using two half-adders and an OR gate. ■ Explain the operation and use of a flip-flop consisting of two NAND gates or two NOR gates. ■ Describe the operation of the triggered bistable. ■ Combine triggered bistables (T flip-flops) to make a 3-bit binary counter. ■ Discuss everyday applications of digital systems in the home and industry. Logic gates Discussion example 24.1 Logic gates as electronic switches Q A An electric logic gate is an electronic switch or a combination of electronic switches. Electronic switches are operated by voltages and function millions of times faster than mechanical switches (hence their use in high-speed computer circuits). Electronic switches are combined to form logic gates. Logic gates are based on the binary system. Voltages applied to the input(s) of a gate are either low (logic 0) or high (logic 1), and quickly transition the in-between voltages for the gate to function properly. Figure 24.1 shows, for a 5 V supply, logic '0' (low) is defined for the voltage range 0 V to 1.5 V and logic '1' is defined as the voltage range 3.5 V to 5 V; these ranges are used typically CMOS logic 1 3.5 V to 5 V not used 1.5 V to 3.5 V logic level V/V 0 1 0 1 logic 0 0 V to 1.5 V t/s Figure 24.1 CMOS logic 0 and logic 1 voltage ranges.
0 What are CMOS gates? In Chapter 22 we saw that the junction transistor can act as a switch or 'gate' (see Figure 22.19). The input base current (resulting from an input base–emitter voltage) controls the output collector current, which is converted to an output voltage as it passes through a suitable collector output resistance.
One of the main disadvantages of using junction transistors to provide logic systems lies in the fact that junction transistors require base (input) currents and collector (output) currents to maintain their logic states. The use of junction transistors for digital systems is called transistor-transistorlogic or TTL. in 5 V CMOS (complementary metal-oxide silicon) logic circuits. Most modern CMOS logic circuits operate with a supply voltage of 1.5 V, where logic '0' corresponds to the voltage range 0 V–0.53 V and logic '1' to the voltage range 0.98 V–1.5 V; the lower supply voltages permits speedier logic operation and lower power consumption and are predominantly used in portable hand-held devices. CMOS (complementary metal-oxide silicon) logic has now largely replaced the older TTL logic systems. CMOS logic uses p and n channel MOSFETs (metal-oxide silicon field-effect transistors); pMOSFETs and nMOSFETs are considered complementary as in the static state they are Unit 2 Module 2 A.c. theory and electronics b V supply ICMOS gate Vin 5 source 4 drain Vout 3 Vout CMOS inverter The output transitions from the logic 1 (5 V) to logic 0 (0 V) Vin (CMOS inverter's input voltage) 2 I CMOS (power supply current through the CMOS inverter 1 HI LO ICMOS / nA to μA a Vin Vout / V 386 GND 0 t / ns range switched ON together such that current passes through them. Figure 24.2a

illustrates a NOT gate formed by a complementary pair. CMOS transistor logic circuits reduce current consumption significantly. The important advantage of the complementary arrangement is that once the CMOS has transitioned to its logic state, there is zero current through the logic transistors, which is great for cellular phones, computers, iPods, iPads and other portable devices (Figure 24.2b). Some portable devices (smart phones, laptop computers, etc.) contain billions of gates so even tiny idle currents from each gate can add up to significant amounts of current, which will continuously discharge the battery; this results in reduced usage times. Figure 24.2 shows a CMOS inverter circuit (also described as a NOT gate) and the transient current. At all logic '0' to logic '1' and logic '1' to logic '0' input signal transitions there is a spike of current during the transition times. Note that there is no current after the transitions occur (corresponding to static states). The current amplitude per transistor varies with the size of the MOSFET (it is in the range nA to µA) and the transition times vary with time taken for a LO to HI or HI to LO logic level transition (it is in the ns range). The oxide in a CMOSFET is an insulator that prevents an input gate current. Rather a field is set up (by the gate voltage) at the drain to source channel which controls the drain to source current. Hence the name field-effect transistor (FET).
Today's basic logic element is the CMOS inverter, called a NOT gate (to be described shortly). When the input voltage, Vi, is low, the output voltage, V0, is high, and vice-versa. ITQ 1 Look at Figure 24.3. (a) What would the voltage at A be if R1 = 500 Ω and R2 = 100 Ω? (b) What would the voltage at A be if R1 were very high and R2 remained at 500 Ω? Figure 24.2 (a) A CMOS 'NOT' gate. (The symbolic representation of CMOSFETs is not required in this course; the two MOSFETs are shown in the diagram just as a matter of interest.) (b) Current consumption of a CMOS 'NOT' gate (also referred to as a CMOS INVERTER) as a function of Vin, Vout and time. t. Controlling input logic states Care must be taken to prevent the inputs of logic functions from exceeding the voltage supply used by the logic circuit. Before connecting a signal of a different voltage range to the logic system, it may be necessary to use a potential divider to ensure voltage compatibility. Figure 24.3 shows a potential divider providing 5 V a gate input A, from a 6 V supply. The voltage drop across resistor R2 is R 500 Ω × 6 V = 5 V; given by R + R2 × 6 V = 100 Ω + 500 Ω 1 2 hence the voltage drop across R1 is 1 V. (The potential divider formula was discussed in Chapter 19.) +6 V R1 100 Ω A +5 V R2 to gate input 500 Ω 0V Figure 24.3 A potential divider providing a gate input voltage of 5 V from a 6 V supply. Truth tables A truth table shows gate output states for all possible combinations of gate input states. Truth tables are extremely useful as design aids in designing digital circuits to perform certain functions, to analyse digital circuits and to predict their functioning.
Truth tables are presented for all gates described in this chapter. A, B and C will usually represent input connections; Y will usually represent the output connection. Chapter 24 Logic gates a +6 V b NOT A Y Input Output A Y 1 0 0 1 LDR R Worked example 24.2: AND gate circuit Q Using Figure 24.5c, explain why the LED would light only if S is closed and the LDR is in the dark. A For the LED to light, both A and B must be at high voltage (logic 1) – see truth table in Figure 24.5a). If S is open, the LED will not light since the voltage at B would be 0 V. When S is closed, the voltage at B would be 5 V, i.e. high. In the dark, the LDR will have very high resistance as compared to R2, and hence the voltage at A would be high (close to 5 V), on account of R2 and the LDR forming a potential divider. The output at Y will therefore be high when S is closed (input B is high) and the LDR is in the dark (input A is high), causing the LED to glow off light. In bright light, the LDR will have low resistance. Most of the supply voltage would drop across R2. A low voltage results at A.
Hence the LED will not light. Y LED 0V Figure 24.4 The NOT gate: (a) symbol and truth table and (b) controlling an LED. Types of logic gates Note: In this chapter we are using the American system of representing gate symbols. NOT gate (inverter) ■ When a low voltage is applied to the input, A, of a NOT ■ An AND gate is one in which the output voltage at Y is gate, a high voltage output is obtained at the output, Y. high (logic 1 = 5 V) only when the input voltage at A is high AND the input voltage at B is high as well (Figure 24.5a). ■ When a high voltage is applied to the input, A, of a NOT gate, a low voltage output is obtained at the output, Y. The NOT gate inverts the input logic and is therefore also referred to as an inverter. Figure 24.4a shows the gate symbol and truth table for a NOT gate. The LDR (light-dependent resistor) and variable resistor, R, of the order of a few hundred ohms, in Figure 24.4b form a potential divider.
In the dark, the resistance of the LDR is high; hence there is very little current through the LDR and R. A large voltage drop therefore appears across the LDR, and a small drop across R, resulting in a low voltage at input terminal A.
A high output appears at the output Y, causing the LED (light-emitting diode) to light. (The functioning of LDRs and LEDs was discussed in Chapter 22.) AND and NAND gates Figure 24.5a and 24.5b show the symbols and truth tables for an AND gate and an OR gate. Unlike the NOT gate, each of these gates has two input terminals. The input terminals are labelled A and B and the output terminal is labelled Y. ITQ 2 In Figure 24.4, what is the function of the resistor in series with Y and the LED? ITQ 3 What is a practical application of the circuit shown in Figure 24.5c? ■ An OR gate is one in which the output voltage at Y is high only when the input voltage at A OR the input voltage at B (or both) is high (Figure 24.5b). Figure 24.5c shows an AND gate operating an LED (note the presence of the protective resistor in series with the LED). NAND and NOR gates The output of an AND gate is high only when both inputs are high.
The NAND gate behaves in an opposite manner (Figure 24.6a). ■ The output of a NAND gate is low (i.e. not high) only when both inputs are high. NAND stands for NOT-AND. The output of an OR gate is high when either input is high or both inputs are high. The NOR gate behaves in an opposite manner (Figure 24.6b). ■ The output of a NOR gate is low when either input is high or both inputs are high. NOR stands for NOT-OR. A NOR gate can be made from a combination of an OR gate and a NOT gate (Figure 24.6c). ITQ 4 In what way does the truth table for a NAND gate differ from that for a NOR gate? 387 388 Unit 2 Module 2 A.c. theory and electronics a AND A Y B b OR A Y B Input a A Output AND b c OR A Y Y A S A B Y A B Y A B Y B 0 1 0 1 0 0 0 1 0 1 0 0 1 0 0 1 1 1 1 0 1 0 0 1 1 1 0 1 1 1 A 0 R1 1 1 Y 0 1 0 1 +5 V R2 S A Y LED (lightdependent resistor) +5 V R2 S A Y LED (lightdependent resistor) LDR (lightdependent 0 resistor) R R 1 1 1 1 0 1 0 1 +5 V A Y Output Input Y B A Y R 2 S B Input Output Input Output Output Input Output OR c R2 B 0 1 0 0 1 0 1 1 Y LED (lightdependent resistor) 0V LED (lightdependent resistor) 0V Figure 24.5 Logic gates: (a) AND gate symbol and truth table; (b) OR gate symbol and truth table; (c) AND gate operating an LED. a A NANDa A NAND A Y B B B b C NOR Y Y A cAND A B c NOT AND NOT AND NOT NOT A Y1 Y2 Y1 B Output A A Y B A Y1 YB1 Y2 Y1 Y2 A AB AB Y B Y Y A AB ABY B Y Y 0 1 0 1 0 1 0 1 0 0 1 0 1 0 0 1 1 1 1 0 0 0 0 1 1 1 1 1 1 0 0 1 0 1 0 1 0 0 1 0 1 1 1 0 0 0 0 0 1 1 1 0 0 0 0 1 1 0 1 0 1 0 0 1 0 1 0 1 0 1 0 1 1 1 0 1 0 1 0 1 0 1 0 1 1 1 0 1 0 1 1 1 1 0 1 0 1 0 1 0 1 0 1 1 1 0 1 0 1 0 1 0 1 0 1 1 Output Input Input Output Input Input Output Output Input 1 1 0 Input 0 0 1 1 Output Figure 24.6 (a) NAND gate symbol and truth table; (b) NOR gate symbol and truth table. (c) NAND gate from AND-NOT gates. +5 V R1 T thermistor placed inside water tank R B EXNOR gates to water tank EXOR and EXNOR gates EXOR stands for 'Exclusive OR'. EXNOR stands for 'Exclusive NOR'. relay ■ In an EXOR gate, the output is high when either of the two inputs are high but the output is low if the two inputs are different. EXOR gate, the output is low if either of the inputs are the same. R2 ■ In an EXNOR gate, the output is high only when the inputs are the same. Figure 24.7 shows a NOR gate which turns a heater off if the water level in the tank drops below the level of two conducting probes or the water is too hot. The thermistor is placed in the water tank along with the probes. Figure 24.8 shows the symbols and corresponding truth tables for an EXOR and an EXNOR gate. A common usage of EXOR gates is in binary number addition. The EXOR gate is used in combination with an AND gate to add two binary numbers (to be discussed later in the chapter). EXOR A A Q A Explain how the NOR gate shown in Figure 24.7 turns the heater off if either the water level falls below the level of the probes or the water in the tank is too hot. When the water drops below the level of the probes, the resistance between the probes is very high as compared to R1. A high voltage (logic 1) appears at B. When the water gets too hot, the resistance of the thermistor, T, becomes small as compared with R2. A high voltage (logic 1) therefore appears at A. Either situation results in a low output voltage (logic 0) which is unable to activate the 'normally off' relay shown. Input A 0 0 1 1 A Y B Worked example 24.3: NOR gate operating an automatic water heater EXNOR B Input B 0 1 0 1 Output Y 0 1 1 0 Y B Input A 0 0 1 1 Input B 0 1 0 1 Output Y 1 0 0 1 Figure 24.8 (a) EXOR gate symbol and truth table. (b) EXNOR gate and truth table. ITQ 5 An EXOR gate could be used in the circuit shown in Figure 24.7 to produce the same automatic control effect if a 'normally on' switch replaces the 'normally off' switch.
Why is this not a desirable circuit to use?
Chapter 24 Logic gates Equivalents of combinations of gates of different types a b pin 1 notch NOT gate placed at a gate output A NOT gate placed at the output of an AND gate results in a NAND gate (see Figure 24.6c). Similarly, a NOT gate placed at the output of an OR gate results in a NOR gate. The reader can verify this. The reader can also verify, using Figure 24.9 (see ITQ 6 at the bottom of this page), that a NOT gate placed at the output of an EXOR gate results in an EXNOR gate. It is again left as an exercise to the reader to show, by using circuit symbols and completion of the corresponding truth table, that a NOT gate placed at the output of an EXNOR gate results in an EXOR gate. The effect of placing a NOT gate at the output of a gate is to invert the logic of the gate. NOT gates placed at the inputs of an AND gate Figure 24.10 shows two NOT gates placed at the inputs, C and D, of an AND gate. Complete the truth table to show the logic states at the AND inputs, C and D, and at the output, Y.
Verify, by completing the truth table, that the gate combination results in a NOR gate.
As this example shows, a truth table is very useful in determining the output logic corresponding to various gate input logic types. A C Y D B A 0 0 1 1 B 0 1 0 1 C 1 D 1 Y 1 1 2 4 3 7 6 5 Figure 24.11 (a) Top view showing a CMOS quad, 2-input NAND gates. (b) IC with connection pins. notch NAND gate, IC (integrated circuit) chip. All logic gates can be constructed using only NAND (or only NOR) gates. NOT from single input NAND or single input NOR Figure 24.12a shows a NOT gate made from a single input NAND gate.
Figure 24.12b shows that a NOT gate can also be made from a single input NOR gate. A Y A b Y A A Y A Y 1 0 0 1 0 0 1 1 Figure 24.12 (a) NOT from single input NAND. (b) NOT from single input NOR. Useful tip! A NOT gate (Figure 24.4) can be made from a single input NAND gate or from a single input NOR gate (Figure 24.12) by connecting the input pins of each together. Figure 24.10 NOR gate obtained from two NOT gates at inputs of an AND gate. AND from NORs only Equivalents of combinations of gates of the same type By completing the truth table for the circuit in Figure 24.13, the reader can verify that the NOR gates circuit shown results in an AND gate. In the manufacturing process, a large number of gates can be realized on a single chip. Figure 24.11 shows the internal and external connections in a quad, 2-input A Y B ITQ 6 What would be the effect of placing a NOT gate on the output of an EXOR gate?
Complete the truth table in Figure 24.9 to show this. EXOR A H NOT Y1 Input A 0 0 1 1 Input B 0 1 0 1 Y1 Output Y Figure 24.9 A NOT gate placed at the output of an EXOR gate. A1 B1 A 0 0 1 1 B 0 1 0 1 A1 B1 Y 0 Figure 24.13 AND from NORs only. OR from NANDs only The outputs of two single-input NAND gates, connected to the inputs of another NAND gate (as shown in Figure 24.14) result in an OR gate. To show this, the reader can complete the truth table in Figure 24.14. The reader can verify this. A1 A A 0 0 1 1 Y1 B1 B B 0 1 0 1 A1 1 1 A1 B1 Y Figure 24.14 OR from NANDs only. Note that the switch input checks if the relay and the battery circuit are working. You would want to make this check if, for example, it is dark and the mains is OFF but the battery circuit fails to operate. Y 0 1 Figure 24.17 This shows the truth table corresponding to Figure 24.16. Figure 24.14 OR from NANDs only. Binary addition using logic gates EXNOR from NANDs only Binary and decimal numbers The arrangement of NANDs shown in Figure 24.15 produces an EXNOR gate. The reader can complete the corresponding truth table to verify this. Binary adders make use of binary numbers. Y2 A Y1 Y4 A 0 0 1 1 Y B Y3 Y4 A 0 0 1 1 Y B Y3 Y2 B1 B ■ The columns in a binary number can be headed 20, 21, B 0 1 ■ 0 1 Y4 Y 1 × 101 + 0 × 100 = 10 + 0 = 10 The number 14 in the decimal system means: 1 × 101 + 4 × 100 = 10 + 4 = 14 The number 10 in the binary system means: 1 × 21 + 0 × 20 = 2 + 0 = 2 (in the decimal system) Control by combinations of logic gates The number 110 in the binary system means: Logic gates can be used for making 'decisions'.
Figure 24.16 shows an example. A mains power sensor and daylight sensor are connected to the inputs of two NOT gates. The outputs of these gates connect to the inputs of an AND gate. A high output of the AND gate activates a relay which switches on a battery power circuit. The truth table shows that only when both the mains power is off and it is not daylight that the battery circuit will be switched on automatically.
1 × 22 + 1 × 21 + 0 × 20 = 4 + 2 + 0 = 6 (in the decimal system) Converting decimal numbers to binary is easy. The decimal number is divided by 2 and the remainders at each step, taken in the specific order shown by the arrows, represent the binary number. Figure 24.18 shows the conversion of the decimal number 13 into binary.
There are electronic circuits that can convert decimal numbers into binary. The OR gate is included so that a test switch input can be incorporated to check if the relay and the battery circuit A NOT gate Y1 Y3 AND gate daylight sensor B Y2 NOT gate OR gate test switch B 0 1 0 1 Y C C 0 0 0 0 Y1 Y2 normally OFF relay battery circuit Figure 24.16 'Decision making' by two NOTs and one AND gate. test switch OFF A 0 0 1 1 In decimal numbers, the columns are headed 100, 101, 102, and so on from right to left (note that 100 = 1). The columns in binary are headed 20 (= 1), 21 (= 2), 22 (= 4), 23 (= 8), and so on from right to left (note that 20 = 1).
= 1), 22Y, 1and Yso 2 on, Y3 Y 1 0 0 0 A 0 0 1 1 B 0 1 0 1 C 1 1 1 1 Y1 Y2 Y3 Y 1 1 1 1 Figure 24.17 Truth tables corresponding to Figure 24.16.
Chapter 24 Logic gates 2 2 2 2 13 6l 3r0 1r1 0r1 check: 1101 1 1 0 1 = 1 × 23 + 1 × 22 + 0 × 21 + 1 × 20 = 8 + 4 + 0 + 1 = 13 a b A A 0 0 1 1 S B A C 1 1 0 1 halfadder B S C B 0 1 0 1 Sum 0 1 1 0 Carry 0 0 0 1 Figure 24.18 Converting a decimal number (13) into binary (1101). Figure 24.20 (a) The half-adder symbol and block representation. (b) The half-adder truth table. Decimal and binary addition compared The (single-bit) half-adder The following rules apply in adding (single) binary digits: The half-adder cannot handle the addition of binary numbers with more than two digits. This is because it does not allow the input of a carry bit from the addition of two previous digits. A circuit that can handle three inputs forms the basis for adding any two binary numbers. Such a circuit is called a full-adder. 0+0=0 0+1=1 1+0=1 1 + 1 = 0, carry 1 (In decimal this reads: 1 + 1 = 2 (i.e. 1 × 21 + 0 × 20)). The full-adder In adding additional of two numbers, we add digits from right to left, column by column, in a column with one digit on the left. In like manner, in binary addition digits are added from right to left. Whenever the total in a column equals or exceeds 10, we 'carry' over 1 and add it to the next column on the left. In like manner, in binary addition, when the total in a column equals or exceeds 2, we put 0 as the sum below the column (in this case) and then carry over 1 to the next column on the left. Figure 24.19 shows a comparison between decimal and binary addition. a b carry 1 1 1 1 1 4 9 2 2 carry carry 1 1 1 1 1 + 8 7 1 3 0 carry + a The (single-bit) half-adder The half-adder is a digital circuit that can add two bits at a time (a 'bit' is either a 0 or a 1, i.e. a binary digit). The binary half-adder has two inputs, A and B, and two outputs. The EXOR output is the 'sum' of the two bits (S) and the AND output is the 'carry' (C). The inputs to the EXOR gate are also the inputs to an AND gate (see Figure 24.20a). Figure 24.20 shows a half-adder and its corresponding truth table. The truth table shows that when inputs A and B are parallel. Figure 24.19 Addition using: a three-digit, single-bit, full-adder formed using two half-adders, HA1 and HA2.
Two half-adders, HA1 and HA2, connected as shown, provide these three inputs. The 'sum' of each half-adder is the output of each EXOR. The 'carry' of each half-adder comes from the AND outputs. The final 'carry' comes from an OR output.
Figure 24.21 shows a full-adder and the corresponding truth table. 1 1 0 0 1 Figure 24.19 (a) Adding decimal numbers 13 and 9, and 43 and 87. (b) Adding binary numbers 111 and 1, and 1001 and 101. Construction of a full-adder from two halfadders S b truth table D E F 0 0 0 1 0 1 1 0 0 1 0 1 0 1 1 0 1 0 0 1 1 0 1 C1 S1 C2 C S 0 0 0 0 0 1 1 1 0 0 1 1 0 0 0 1 0 0 0 1 0 1 1 1 0 Decimal number 0 1 1 1 2 0 0 0 1 1 0 1 2 3 Figure 24.21 (a) Full-adder from two half-adders. (b) Corresponding truth table (incomplete). ITQ 8 Convert 15 (decimal) into a binary number. ITQ 9 How many single bits can: (a) a half-adder add (b) a full-adder add? 391 392 Unit 2 Module 2 A.c. theory and electronics Worked example 24.4: Truth table of a full-adder Q Complete row 6 of a full-adder in Figure 24.21b to show the addition of 1 + 0 + 1.
A D(1) + E(0) = S1(1) + C1(0) S1(1) + F(1) = S(0) + C2(1) C2(1) + C1(0) = C(1) (from OR gate) Hence, the final result is S = 0 and C = 1 (i.e. 1 0 in binary, or 2 in decimal) Addition of two binary numbers using a fulladder The full-adder in Figure 24.21a has three single-bit inputs (D, E, F) and two single-bit outputs (S, C).
We can represent the full-adder as shown in Figure 24.22a, where F is replaced by Cin, the carry from a previous full-adder. a A B b full-adder (FA) Cout C in S A B Cin S Cout 0 0 0 0 1 0 1 1 1 0 0 1 0 1 0 1 1 0 1 1 0 0 1 0 1 1 0 1 1 1 0 0 1 0 0 1 1 1 1 Figure 24.22 (a) Full-adder block representation. (b) Corresponding truth table. Let us suppose we wish to add two binary numbers, A2A1 (e.g. 1 1) and B2B1 (e.g. 0 1). Binary addition gives us: carry 1 1 A2A1 B2B1 1 1 0 0 1 + 1 a sum, S2 = 0 carry C2 = 1 ... Hence our final sum is 1 0 0. The reader can verify the result from each full-adder, FA1 and FA2, by checking the full-adder truth table in Figure 24.22b. Timing diagrams and logic gates A timing diagram shows the variation of output voltage with time for a corresponding variation of input voltage with time. Timing diagrams employ 'square' wave voltages, i.e. voltages whose rise and fall times are very fast and whose maximum and minimum voltages are very constant. Figure 24.24 shows two NOT gates connected in series and the corresponding timing diagram. We shall be making use of timing diagrams in the sections on bistable circuits that follow.
a b Vin Vin V1 Time Figure 24.24 (a) Two NOT gates connected in series. (b) Corresponding timing diagram.
Worked example 24.5: Timing diagram involving two inputs Q Figure 24.25a shows a timing diagram associated with two inputs, A and B, and output D. Complete the truth table for the circuit in Figure 24.25b and name the logic gate that would give rise to such a timing diagram. time sequence Two full-adders, arranged in parallel as shown in Figure 24.23, will give us this sum.
The addition of A1 and B1 in FA1 (i.e. 1 + 1) gives us a sum, S1 = 0 and a carry C1 = 1. The addition of C1, A2 and B2 in FA2 (i.e. 1 + 1 + 0) gives a 2) full-adder (FA) 2 C2 (1) S2 (0) B1 (1) C1 (1) A 1 (1) full-adder (FA) 1 V1 Vout A B2 (0) Vout B D b 1 2 3 4 5 6 7 8 Time sequence 1 2 3 4 5 6 7 8 A B D 0 0 1 1 0 1 1 0 Figure 24.25 (a) Timing diagram associated with two inputs.
(b) Truth table (incomplete). A first three rows of the truth table have been completed. It is left as an exercise for the reader to complete the remaining rows. The completed table will show that the output, D, is 1 only when both inputs are 0 – in other words the gate is a NOR gate. S1 (0) Figure 24.23 Full-adder block representation showing addition of two binary numbers. ITQ 10 In Figure 24.24, what do you notice about the phases of Vin, V1 and Vout? Chapter 24 Logic gates S1 switched 'on' a +5 V → Y1 (logic 1) – because of action of NAND gate 1. A1 1 Y1 B1 A2 S1 (set) Y2 1 B2 S2 (reset) 0V b → A1 (logic 0) – since A1 is now connected directly to 0 V A1 B1 Y1 A2 B2 Y2 circuit S2 'switched 'on' S1 depressed 1 0 1 0 1 0 1 1 1 0 S1 released 1 0 1 1 1 1 0 S2 depressed 1 1 1 0 0 0 1 1 1 1 Figure 24.26 (a) A bistable circuit when S1 is depressed (switched 'on'). (b) Corresponding truth table. Bistable circuits The Set-Reset (SR) flip-flop Bistable circuits The NAND gates combination shown in Figure 24.26a is called a flip-flop circuit and is the basis of computer 'memory'. (Two NOR gates can also be combined to form a flip-flop circuit.) When the power is switched on, the circuit remains in one of two stable states (hence the name 'bistable'). Either: ■ the output Y1 is at logic 0 and Y2 at logic 1, or ■ the output Y1 is at logic 1 and Y2 at logic 0. The output state is unpredictable when the power is switched on: this circuit is also known as a metastable state (or a quasi-stable state). The logic at each output may oscillate between '1' and '0' until at the end of metastable state, the flip-flop settles down to either of the logic states described above. We now examine the logic states for the circuit, assuming that the initial state when powered. Let us assume that, when power is switched on, Y1 settles at logic 0 and Y2 at logic 1. In this state, A1 and B2 are at logic 1 (since S1 and S2 are open). Since Y1 is at logic 0, then B1 must be at logic 1. A2 must be at logic 0, which is in keeping with Y1 at logic 0, since A2 is connected to Y1. The first row of the truth table in Figure 24.26b summarizes this state. ITQ 11 Why is it that, according to the text above, 'since Y1 is at logic 0, then B1 must be at logic 1'?
→ A2 (logic 1) – since Y1 is connected directly to A2 → Y2 (logic 0) – since B2 is at logic 1 (since S2 is open) and A2 is at logic 1 into NAND gate 2 → B1 (logic 0) – since Y2 is connected to B1.
The net result of switching S1 'on' is that Y1 and Y2 have 'flipped': Y1 is now at logic 1 and Y2 at logic 0. S1 switched 'off' after being switched 'on' → A1 (logic 1) – since S1 is now open → B1 (logic 0) – from previous state → Y1 (logic 1) – because of NAND gate 1 → A2 (logic 1) – because Y1 is connected directly to A2 → B2 (logic 1) – since S2 is open → Y2 (logic 0) – because of action of NAND gate 2. The net result of switching S1 'off' after being switched 'on' is that Y1 and Y2 remain in the 'flipped' state, i.e. Y1 at logic 1 and Y2 at logic 0. Even if S1 is then switched off again, the output logic levels, Y1 and Y2, remain 'latched' at 1 and 0, respectively. S1 is referred to as the 'set' switch. S2 switched 'on' after S1 is switched 'off' A similar analysis as the above shows that the logic states of Y1 and Y2 become interchanged when S2 is switched 'on' after S1 is switched 'off'. The circuit has 'flipped' again ('flopped'). Even if S2 is then switched 'off', analysis shows that the circuit remains latched in the 'flopped' state, which is the original state. Hence S2 is referred to as a 'reset' switch. Figure 24.27 summarizes the setting, latching and resetting of a bistable circuit. Bistable alarm circuit On account of its latching feature, a bistable can be used as one type of burglar alarm circuit. If a burglar steps on a mat switch, S1, a buzzer, connected to Y1, will sound an alarm. The alarm remains on even if the burglar steps off the mat switch. The switch S2 is used to reset the alarm. Triggered bistable (T flip-flop) A T flip-flop is a bistable whose output 'toggles' between logic 0 and logic 1 when 'triggered' by a train of square wave pulses. Figure 24.28a shows the symbol for a T 393 394 Unit 2 Module 2 A.c. theory and electronics Switching sequence invalid state at power 'on' A1 B1 Y1 A2 B2 Y2 Comments 1 X X X 1 X metastable (oscillating states) 1 1 0 0 1 1 settled state (assumed) S1 'on' 0 0 1 1 1 0 set: Y2 = 1 (Y2 ⇒ Y1) S1 'off' 1 0 1 1 1 0 latch S2 'on' 1 1 0 0 0 1 reset: Y2 ⇒ 0 (Y2 ⇒ Y1) S2 'off' 1 1 0 1 1 0 1 latch a T0 b T1 T2 T3 T4 Triggered bistables are 3-bit binary counter T5 Three T flip-flops, FF0, FF1 and FF2, connected in sequence (as shown in Figure 24.29a) can count pulses applied to the first clock (CK) input. Note that the input CK0 enters the first flip-flop (FF0) and the output Q0 from the flip-flop feeds the clock input in this way. It is a triggered flip-flop. Figure 24.28 (a) T flip-flop symbol. (b) Timing diagram. T Q T Q Q Q ■ The input to FF0 is from the clock pulses. The output Q of FF0 is fed to the clock input of FF1. flip-flop. Figure 24.28b shows a train of square wave triggering pulses (T) as a function of time, as well as corresponding outputs Q and Q.
We observe that the output Q is at logic 1 when the input T is high. (Q and Q are 0 and 1, respectively), and that the flip-flop will toggle on an input LOW-to-HIGH (rising-edge) transition (e.g. the transitions taking place at T1, the outputs will be as shown in the timing diagram (Figure 24.28b). At T1, there is a rise in the input voltage and therefore Q will toggle to 1. At T3, there is again a rise in the input, and therefore Q toggles back to 0. The timing diagram (Figure 24.28b) shows that the output frequency is half the input frequency. Hence, this arrangement can be used to divide the input frequency by 2. a Q2 (m.s.b.) Q1 Q2 Q0 (l.s.b.) Q1 CK 2 b Q0 FF1 ? 6 5 FF0 g h 1 0 1 0 clock pulses CK 0 Q1 FF2 c Q C H T U 4 e 3 f c 2 d F E a C The outputs of the three flip-flops, from left to right, gives the total count of the clock pulses in binary. The right to left digit on the right is called the least significant bit (l.s.b.) of the output Q0 while, on the extreme left, is the most significant bit (m.s.b.) The output of the circuit is given in the truth table, with comments, given in Figure 24.29c, explains how the counting in binary occurs. The truth table for the 3-bit counter shows that: ■ the l.s.b. (least significant bit) toggles with each clock pulse ■ the next bit (to the left) toggles with each two clock pulses (or each pulse from the preceding T flip-flop; and Clock pulse no. clock pulses Outputs = binary count Q2 Q1 Q0 0 0 0 All outputs at 0 1 0 0 1 Q0 : 0 → 1 due to falling edge ab from clock pulse 1 2 0 1 0 Q0 : 1 → 0 due to falling edge of from clock pulse 2. Q1 : 0 1 due to falling edge CD from Q0 . Q0 : 0 1 due to falling edge gh from clock pulse 3. Q1 : 0 → 1 due to falling edge GH from Q0 .A 1 0 1 1 0 1 0 B D M A L Q0 (l.s.b.) Q1 Q2 (m.s.b.) Comment 0 3 1 Q ■ The input to FF2 is from the output of FF1. Q2 : 0 → 1 due to falling edge TU from Q1
5 1 0 1 Q0 : 0 1 6 1 1 0 Q0 : 1 0 Q1 : 0 1 7 1 1 1 Q0 : 1 1 0 Figure 24.27 Bistable state: metastable, set, latch and reset. Note that the output Y1 is often designated as Q and Y2 as Q̄. 7 1 1 1 Figure 24.29 (a) 3-bit binary counter using T flip-flops. (b) Timing diagram. (c) Clock pulses and corresponding binary count. Chapter 24 Logic gates ■ the m.s.b. (or extreme left) toggles with each four clock pulses (or the toggles taking place in pulses 0, 4 and 7. An examination of the output bits shows that they correspond to the number of clock pulses expressed in binary (e.g. 5 = 1 0 1 in binary). Summary ✓ Logic gate functions can be summarized by the following table: Gate Summary of function NOT Inverts the logic state of its input. A single-input NAND gate (i.e. inputs connected together) or a single-input NOR gate functions as a NOT gate. AND Output is 1 only when both inputs are 1. NAND Output is 0 only when both inputs are 1. OR Output is 1 if either or both inputs is 1. NOR Output is 1 only if both inputs are 0. EXOR Output is 1 if the inputs are different; output is 1 if the inputs are both 0 or both 1.
EXNOR Output is 1 if the inputs are the same; output is 0 if the inputs are different. ✓ A truth table shows logic function output states for all possible combinations of its input states. ✓ All logic functions can be constructed using only NAND (or only NOR) gates. ✓ Logic gates can be used for making 'decisions'. A block diagram consisting of sensors and gates can illustrate control by combinations of logic gates. ✓ A bit is a single digit – either a 1 or a 0. Binary addition involves bits only. ✓ A half-adder can add only two single-bit numbers. The result of the addition is a sum and a carry bit. ✓ A full-adder can add three single-bit numbers. The result is a sum and a carry bit. A full-adder can be combined to form a full-adder.
✓ Full-adders can be combined in parallel to add multi-bit numbers. For example, two full-adders thus combined can add two 2-bit numbers. The result is a 3-bit number where the digit on the right is the 'least significant bit' (l.s.b.) and the leftmost bit the 'most significant bit' (m.s.b.). ✓ A timing diagram shows the variation of output voltage (or logic state) with time for a corresponding variation of input voltage (or logic state) with time. Functional timing diagrams employ 'square' wave voltages, i.e. voltages whose rise and fall times are very fast and whose maximum and minimum voltages are very constant. ✓ Two NAND gates (or two NOR gates) can be combined to form a flip-flop circuit. When the power is switched on, the output states are unpredictable and oscillate for a brief period, the metastable state, the circuit settles down into one of two stable states (hence the term bistable).
One of the two outputs settles to logic 1 and the other to logic 0 or vice-versa. ✓ On account of its latching feature, a bistable can be used as one type of burglar alarm circuit. If a burglar steps on a mat switch, S1, a buzzer, connected to Y1, will sound an alarm. The alarm remains on even if the burglar steps off the mat switch. The switch S2 is used to reset the alarm. ✓ A T flip-flop is a bistable whose output 'toggles' between logic 0 and logic 1 when 'triggered' by a train of square wave pulses. The timing diagram shows that the output frequency of a T flip-flop is half the input frequency. Hence, this arrangement can therefore be used to divide the input frequency by 2. ✓ Three T flip-flops, connected in series, can form a binary counter. 395 396 Unit 2 Module 2 A.c. theory and electronics Review questions a b Logic gates; Types of logic gates R2 1 (a) Draw the symbol and truth table for (i) an AND gate (ii) a NAND gate (iii) a NOT gate. (b) Show, by drawing and a truth table, that a NAND gate can be formed by combining an AND gate and a NOT gate. a 2 By labelling inputs and outputs suitably, construct a truth table for the following combination of safe kind of gate does the gates (Figure 24.30). What combination represent? A safe S mat switch +5 V b R2 R1 combination lock buzzer LDR S mat switch 0V B Figure 24.30 3 A quad chip consists of four NAND gates (see Figure 24.11). Show how the following gates can be made by making connections to these NAND gates only. Use the minimum number of NAND gates in each case and verify each case using an appropriate truth table. (a) NOT (b) AND (c) OR (d) NOR (e) EXOR (f) EXNOR.
4 (a) What kind of gate results when the two inputs of a NOR gate are connected together?
Illustrate your answer with a gate diagram and the corresponding truth table. (b) Would the same result be achieved if a NAND gate were used instead? Illustrate your answer with a gate diagram and the corresponding truth table. Control by combinations of logic gates 5 Figure 24.31 shows a burglar alarm circuit protecting a safe in the dark.
The alarm sounds if either someone steps on the mat switch, S, or the LDR, located on the combination lock, is illuminated. Figure 24.31 (a) Explain how the circuit works. (b) How can the circuit be modified, using an OR gate, so that a 'test' switch can be used to check the functioning of the buzzer?
(c) Draw a decision-making block diagram to show the modified circuit in (b). 6 Draw a decision-making block diagram to show that if the water level in a car cooling system is low (logic 0) or the oil pressure is low (logic 0), or both are low, a warning light will come on. 7 Draw a decision-making block diagram to show that if a pot on a stove is empty and the stove is 'on' or if boiling liquid is spilling over from a the pot on the stove, a relay turns on a buzzer. Binary addition using logic gates 8 Figure 24.32 shows a half-adder. S A halfadder B R1 combination lock C Figure 24.32 (a) What do A, B, C and S represent? (b) Draw a truth table corresponding to this half-adder. (c) By using two such half-adders, show how a fulladder can be constructed. Complete the truth table for this full-adder. (d) In terms of single bits, what is the difference between a half-adder and a full-adder? LDR Chapter 24 Logic gates 9 (a) Convert 41 into a binary number. (b) Express 12 and 7 each as a binary number. (c) Add 12 and 7 using bits (binary digits). (d) Verify that the result from binary addition in (c) is compatible with the result from decimal addition. 10 Using gate diagram symbols, draw (a) a half-adder circuit (b) a full-adder circuit.
11 Using block diagrams to represent full-adders, draw a diagram to show a combination of two full-adders that can add two 2-bit numbers. Draw a corresponding truth table.
(Hint: see Figure 24.23.) Timing diagrams and logic gates 12 Figure 24.33 shows bistable outputs, Q and Q, due to triggering pulses, T. T0 T1 T2 T3 T4 T5 T (c) Explain, using the results from your truth table, the terms 'latch' and 'reset'. (d) Why is the circuit called a 'flip-flop'?
15 A quad chip consists of four NAND gates (see Figure 24.11). (a) Show how a S-R flip-flop can be obtained using a minimum number of NAND gates from this chip. (b) Verify your diagram using a truth table. Practical exercises and challenges Caution! CMOS circuits have a defined range of supply voltages V+. Too low a voltage will result in faulty operation and too high a voltage will result in the destruction of the transistors, so always follow the manufacturers' recommendations found in the products datasheets. All electronic circuit experiments should be carried out only under direct supervision of a suitably qualified instructor. It is necessary to handle MOSFET devices with great care; due to their extremely high input impedances, static electricity introduced to the gate through touching can destroy the device. Q Q Figure 24.33 (a) If a NOT gate were placed between the trigger output and the bistable input, draw the timing diagram produced for all three sets of pulses. (b) How does the frequencies of the outputs compare with the input frequency? (c) Is the circuit a falling-edge or a rising-edge response circuit? (d) Draw a truth table showing the outputs as a function of the sequence of the input pulses. (e) How many triggered bistables are being used in the circuit to produce the results shown? 13 (a) Draw a binary counter circuit consisting of two triggered T flip-flops that make use of falling-edge triggering. (Hint: see Figure 24.28.) (b) Draw a timing diagram showing four input pulses and the corresponding output pulses. (c) Draw the corresponding truth table. Include the binary and decimal outputs in the truth table. Bistable circuits 14 (a) Draw a diagram to show how two NOR gates can be connected to form a flip-flop. (b) Verify your drawing with a corresponding truth table. 1 LED logic probe +V Figure 24.34 Logic probe. Construct and test various designs of LED logic probes. A very simple one is shown in Figure 24.34. Others can be found on the internet. 2 Constructing and testing logic gates Using a quad NAND chip, construct and test combinations to produce various gates. (First draw the circuit and show your teacher.) 3 Constructing and testing S-R flip-flop circuits (a) Using a quad NAND chip, construct and test a S-R flip-flop circuit with LED outputs. (First draw the circuit and show your teacher.) (b) Construct and test a burglar alarm to the following specifications, to be used to guard a safe that is kept in a dark room. (First draw a circuit diagram and discuss your plan with your teacher; you may wish to base your diagram on the flip-flop circuit shown in Figure 24.26a.) 397 398 Unit 2 Module 2 A.c. theory and electronics Specifications: ■ the circuit has a main power switch ■ a 'Test' switch is included to check if the alarm device (e.g. a buzzer or an LED) is working, but the flip-flop output is isolated electrically from the testing circuit during the testing ■ a 'Reset' switch is included to 'set' the alarm initially in the 'off' state ■ the alarm is tripped either by a mat switch or by a light being used to read the dial on the door of the safe ■ once tripped, the alarm stays on even if the burglar steps off the mat or turns off the flashlight ■ management must be able to open the safe, day or night, without the alarm operating. Answers to ITQs R1 1 (a) Voltage drop across R1 = 6 V × R + R = 6 V × 100 600 1 2 = 1 V.
Hence the voltage at A would be 6 V − 1 V = 5 V.
R1 would approach 1, meaning (b) The fraction, R + R 1 2 that the voltage drop across R1 would approach 6 V. Hence the voltage at A would approach 0 V. 2 If the current through an LED exceeds a certain value, the LED overheats and is destroyed. A suitably chosen resistor limits the value of the current reaching the LED and thus protects the LED. 3 The circuit in Figure 24.6c is therefore suitable for turning on a light only when the place is dark and the main switch (S) is on. 4 The output of a NAND gate is low only when both inputs are high; the output of a NOR gate is high only when both inputs are low. Also, the output of a NAND gate is high if any of its logic inputs are low; the output of a NOR gate is low if any of its logic inputs are high. 5 This circuit is not desirable on safety grounds. If a fault occurs such that the gate output voltage is not high enough, the 'normally on' relay switch would remain on – resulting in the water overheating. 6 An EXNOR gate results, as shown by the truth table. Input A Input B Y1 Output Y 0 0 0 1 0 1 1 0 1 0 1 0 1 1 0 1 7 First, make the NOR a NOT gate by connecting the two inputs to form a single input. Then connect the single input to the output of the given circuit. 8 1 1 1 1 9 (a) Two single bits. (b) Three single bits. 10 They are out of phase with each succeeding the other (as one would expect with the input and output of a NOT gate). 11 Because, for a NAND gate, for an output of 0, both inputs have to be at logic 1. Answers to Review questions 9 (a) (b) 41 (decimal) 12 (decimal) 7 (decimal) (c) & (d) 12 (decimal) + 7 (decimal) = 19 (decimal) = = = = = = 25 1 0 0 0 0 0 24 0 0 0 0 0 1 23 1 1 0 1 0 0 22 0 1 1 1 1 0 21 0 0 1 0 1 1 20 1 0 1 0 1 1 (binary) (binary) (binary) (binary) (binary) (binary) [25 + 23 + 20] [23 + 22] [22 + 21 + 20] [24 + 21 + 20] 399 Module 3 Atomic and nuclear physics Chapter 25 Atomic structure Learning objectives ■ Give a brief account of early theories of atomic structure, including those of Thomson, Bohr and Rutherford. ■ Describe the Geiger–Marsden α-particle scattering experiment and discuss the evidence it provides for the nuclear ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ model of the atom. Use the relation A = Z + N to solve problems. Explain the term 'isotope'. Use the standard notation for representing a nuclide; for example, 147 N. Describe Millikan's oil drop experiment. Discuss the evidence in Millikan's oil drop experiment for the quantization of charge.
Explain the process of X-ray production. Explain the origins of line and continuous X-ray spectra. Use the relationship I = I0 exp (−µx) for the attenuation of X-rays in matter, where µ = linear absorption coefficient. Discuss the use of X-rays in radiotherapy and imaging in medicine. Discuss how line spectra provide evidence for discrete energy levels in isolated atoms. Use the relationship hf = E2 − E1 to solve problems. Distinguish between absorption and emission line spectra. Atomic theory In Ancient Greece Democritus and his fellow philosophers imagined cutting a piece of matter in two, and then in two again, until it could be cut no more. They called the pieces that were left 'atomos' which means 'indivisible'. Atoms were considered to be the smallest particles from which all matter is made. by arranging the elements in order of their atomic mass, from hydrogen (the lightest) to uranium (the heaviest known at the time), Dmitri Mendeleyev produced the first periodic table. This revealed that the chemical properties of the elements changed with atomic mass in a systematic way. The development of modern atomic theory began in the 18th and 19th centuries, with the chemical studies of John Dalton and others.
It was established that there were more than 60 chemical elements (today we know of 118), each with distinct chemical properties. It was believed that each element consisted of atoms of a given mass and that every atom of an element was identical. In 1869, Up to this point in the story, atoms were regarded as fundamental particles – particles that are not composed of more basic constituents. Then in 1887 the physicist Joseph John (J. J.) Thomson discovered that cathode rays (beams of charged radiation produced by applying a high voltage between two electrodes in an evacuated tube) are streams of particles nearly two thousand times 400 Unit 2 Module 3 Atomic and nuclear physics metal case insulation light beam oil spray microscope microscope oil drop window plan view metal supports lighter than hydrogen, the lightest known atom. Thomson referred to these particles as 'corpuscles' but we now know them as electrons. The electron was the first known subatomic particle. By measuring the deflection of the cathode ray beam in crossed electric and magnetic fields, Thomson determined the charge to mass ratio e/m of the electron.
By using a method similar to (but less refined than) the Millikan oil drop experiment, described in the next section, Thomson estimated the charge on an electron, and could thus calculate its mass. His original estimate was that the electron has 1/1700 the mass of the hydrogen atom. The accepted figure today is 1/1840. The Millikan oil drop experiment In 1906 Robert A. Millikan made an accurate measurement of the charge on the electron by measuring the force on tiny charged oil drops produced by an electric field E. His apparatus is shown in Figure 25.1. A fine spay of non-evaporating oil drops is created above the top plate so that a few drops fall through the gap. The drops are brightly illuminated and appear as bright dots when viewed through the microscope. Initially the plates are earthed. The speed of fall v1 of one drop is recorded with the aid of a stopwatch. The forces acting on the drop when there is no electric field are shown in Figure 25.2a. a b Fdrag = kv1 Fdrag = kv2 qE mg mg Figure 25.2 The forces on an oil drop in the Millikan experiment: (a) with no electric field between the plates; (b) with an electric field E between the plates. Figure 25.1 The Millikan oil drop experiment. (Refer back to Chapter 3, page 47, for a discussion of the forces that act on a sphere falling through a fluid under the influence of gravity.) The falling drop rapidly acquires a downward terminal velocity v1. Its effective weight, m'g, is then exactly balanced by the drag force, Fdrag, due to its motion through the air. (The effective weight m'g of a drop is a little less than the actual weight mg because of the upthrust from the air the drop displaces – by Archimedes' principle. The method used to calculate the effective weight is explained in Review questions at the end of the chapter.) m'g = Fdrag The drag force is proportional to the drop's velocity and so we can write, Fdrag = kv1 where k is a constant of proportionality.
Thus, m'g = kv1 ....(25.1) A high voltage V is now applied between the plates to produce an electric field, E = V , where d is the plate d separation. The upper plate is connected to the positive terminal of the high-voltage supply. Drops from a spray are usually already charged by friction, but they may also be charged (or their charge changed) by exposing them to a beam of X-rays or radiation from a small radioactive source. If the drop is carrying a negative charge q, it experiences an additional upward force qE in the electric field. This force changes the drop's terminal velocity to a new value v2. The measurement of rate of fall is repeated to find v2, and, as shown in Figure 25.2b, the equation for the balanced forces now becomes, m'g − qE = kv2 ....(25.2) Chapter 25 Atomic structure Assuming the weight and shape of the drop have stayed the same, we can subtract equation 25.2 from equation 25.1 to obtain, qE = k(v1 − v2) and so, q= k (v − v2) E 1 ....(25.3) The value of k can be calculated from v1 and the known properties of the oil and air (as explained in the Review questions at the end of the chapter). So, given that E = V , d we can calculate the charge q from the measured values of v1 and v2. Models of the atom The discovery of the electron showed that atoms were not indivisible, as had been thought, but were themselves made up of even smaller particles, some of which, at least, could be removed from the atom by strong electric fields. Electrons carry a negative a charge (they are emitted from the negative cathode and attracted to the positive anode in a cathode ray tube). But atoms are electrically neutral. The atom must therefore contain positive matter as well as electrons. Thomson proposed the 'plum pudding' model for atomic structure shown in Figure 25.3. In this model the atom consisted of a uniform ball of positive 'pudding' in which the electrons are embedded. Worked example 25.1 Q A In a Millikan oil drop experiment the following values were measured/calculated: V = 1500 V d = 1.5 cm Drop 1 k = 2.1 × 10−10 kg s−1 v1 = 2.5 × 10−4 m s−1 v2 = 1.0 × 10−4 m s−1 Drop 2 k = 2.3 × 10−10 kg s−1 v1 = 4.0 × 10−4 m s−1 v2 = 1.9 × 10−4 m s−1 Calculate the magnitude of the charge on each drop. Drop 1 kd 2.1×10−10×1.5×10−2 q = (v1−v2) = ×(2.5×10−4 − 1.0×10−4) V 1500 = 3.2×10−19 C Drop 2 kd 2.3×10−10×1.5×10−2 q = (v1−v2) = ×(4.0×10−4 − 1.9×10−4) V 1500 = 4.8×10−19 C The charges found in Worked example 25.1 are both greater than the charge of the electron (1.6 × 10−19 C). To modern eyes it is obvious that drop 1 carries two units of charge, while drop 2 has three units. Millikan, however, did not know the unit of charge. He measured the charges of hundreds of different drops. Within the limits of experimental error he found that q was always a whole number multiple of the same basic unit. Millikan concluded that charge is quantized – i.e. cannot take any value, but comes in units of a fixed size – and used his data to find this basic unit of charge, e. You will find some sample oil drop data to analyse in the Review questions at the end of the chapter. electron ('plum') negative charge 'pudding' positive charge Figure 25.3 The Thomson model of the atom. Thomson suggested that an atom consisted of a sphere of positive charge in which electrons are embedded like 'plums in a pudding'. The Geiger–Marsden experiment In order to investigate the distribution of positive and negative charges in atoms Ernest Rutherford suggested that his young colleague Hans Geiger and his student Ernest Marsden measure the deflections of alpha particles (now known to be the positively charged nuclei of helium atoms) as they passed through a thin gold foil. If Thomson's model was correct, the light electrons and uniform distribution of positive matter should produce deflections of no more than a few degrees. The experimental arrangement used by Geiger and Marsden is shown in Figure 25.4. The results of the experiment were unexpected. Most of the alpha particles passed almost straight through the foil with deflections of less than 1 degree. A few alpha particles experienced large deflections, with about 1 in 8000 bouncing back from the foil. Rutherford said that the surprise at finding some alpha particles deflected backwards from the foil was like firing an artillery shell at tissue paper and finding it coming back ITQ 1 The air in the Millikan oil drop apparatus must be maintained at a constant temperature if accurate results are to be obtained. Can you suggest why?
401 402 Unit 2 Module 3 Atomic and nuclear physics a gold foil zinc sulfide screen gold foil source microscope alpha source b A problem with Rutherford's atom θ to vacuum pump vacuum electrons nucleus alpha particles atom Figure 25.4 The Geiger-Marsden experiment: (a) the apparatus; (b) a representation of the results and their interpretation. to hit you. Rutherford realized that these observations meant that most of the atom was empty space, since most alpha particles passed straight through. Almost all the atomic mass must be concentrated in a dense central nucleus, with the occasional collisions of alpha particles with these minute nuclei producing the few very large deflections. Rutherford's nuclear atom is shown in Figure 25.5. The atom consists of a positively charged central nucleus, around which the negatively charged electrons orbit like planets in a solar system. The electrons are held in their orbits by the electrostatic force between the positive and negative charges. Rutherford was able to calculate the size of the nucleus from the Geiger–Marsden data. Again the result was a surprising – the nucleus was approximately one hundredthousandth (10−5) the diameter of the atom. If a football stadium represents an atom, then the nucleus is just the size of a pea on the centre spot. The orbits of the planets are stable; their orbital energy remains constant over time. But charged particles should behave in a different way. According to the theory of electromagnetism developed in the 19th century, when an electric charge accelerates it emits electromagnetic radiation, and so its energy decreases. Thus, according to classical physics, Rutherford's atom should not be stable. The orbiting electrons should radiate energy because of their centripetal acceleration, and spiral into the nucleus. The new theory of quantum mechanics was required to resolve this problem. The Bohr atom In quantum theory electrons and all other 'particles' have a combination of particle characteristics, such as linear momentum (p = mv), and wave characteristics, such as wavelength. (This 'wave particle duality' was first postulated by Louis de Broglie in his PhD thesis in 1924 and is discussed in more detail in Chapter 28.) The momentum p of a particle is connected to its wavelength λ through the de Broglie equation, p= h λ ....(25.4) In this equation h is the Planck constant.
This constant was introduced by Max Planck in 1900, in his theory of the spectrum of electromagnetic radiation emitted by hot objects such as the Sun and light bulb filaments. We now b EM radiation nucleus trajectory of electron ITQ 2 How does the Rutherford atom account for the results observed by Geiger and Marsden? Figure 25.5 (a) The Rutherford atom. (b) Classical physics predicts that the Rutherford atom is not stable. The orbiting electrons should radiate energy and spiral into the nucleus. ITQ 3 Why does a particle moving at constant speed in a circular orbit have acceleration? Chapter 25 Atomic structure ionization n=5 n=4 –0.544 eV –0.85 eV nucleus –1.51 eV second excited state n=3 first excited state n=2 ground state n=1 n=1 –3.4 eV n=2 n=3 Figure 25.6 The Bohr model of the hydrogen atom. For the allowed orbits of the electron nλ = 2πr, where r is the orbital radius and n is a whole number. –13.6 eV know that h is the constant of proportionality between the energy E of a quantum of electromagnetic radiation and the frequency of the radiation, E = hf. Planck's constant is one of the fundamental constants of quantum theory. In a first attempt to combine quantum ideas with the classical picture of the atom, Bohr suggested that the electron orbits in the Rutherford atom are stable only for certain values of angular momentum. The de Broglie equation applied to Bohr's proposal shows that this requires a whole number of electron wavelengths to fit exactly around the circumference of the orbit. This idea is analogous to the standing wave patterns that can be generated on a string or in a pipe. Bohr's model of the hydrogen atom is shown in Figure 25.6. (The angular momentum L of a mass m moving a speed v in a circular orbit radius r is given by L = mvr = pr. Just as linear momentum is conserved when no external forces act on a system, angular momentum is conserved when there are no external torques.) In the Bohr atom the allowed states (orbits) of the electron are labelled by the quantum number n. Each state corresponds to a specific momentum and orbital radius; the energy of the electron (the sum of its KE and electrostatic PE) is therefore also fixed for each state.
Stable orbits do not exist between the allowed sates. The lowest energy state, n = 1, is known as the ground state. Since there are no lower energy states the electron cannot lose energy, and so this orbit is stable. Figure 25.7 The energy levels of the hydrogen atom. Energy level diagrams An energy level diagram shows the allowed energy levels for an electron in an atom as horizontal lines – like unevenly spaced rungs on a ladder. Figure 25.7 is the energy level diagram for the Bohr model of the hydrogen atom. Energy values are in electron volts (eV). In the hydrogen atom there is only one electron, which normally occupies the lowest energy level (the ground state for which n = 1). The energy of a stationary free electron outside the atom is taken as zero. The energy levels inside the atom have negative values, since energy must be provided to an electron in one of those levels in order for it to escape from the atom. The Bohr model predicts that the magnitude of the electron's energy is proportional to 1/n2, and so the energy levels get closer together with increasing n. Getting it right! The electron volt When dealing with energy changes at the atomic scale the joule is a very large unit. It is more convenient to use the much smaller energy unit called the electron volt (eV) for the tiny energy changes involved. The electron volt is defined as the energy transferred when the charge of one electron is moved through a p.d. of 1 volt. 1 eV = 1.602 × 10−19 C × 1 V = 1.602 × 10−19 J To convert from eV to joules multiply by 1.602 × 10−19. To convert from joules to eV divide by 1.602 × 10−19. ITQ 4 The angular momentum L of a particle with momentum p moving in an orbit of radius r is given by L = pr. Use the de Broglie equation to show that the electron in the Bohr hydrogen atom can only take angular momentum values nh , where n is a 2π positive integer. An electron in an atom can be excited to a higher energy level by providing it with energy equal to the energy difference between its current level and the higher level 403 404 Unit 2 Module 3 Atomic and nuclear physics a b 0 n=∞ n=3 E3 n=3 n=2 E2 n=2 0 n=∞ E3 E2 Energy Energy ΔE = E2 – E1 = hf E1 ΔE = E3 – E1 = hf E1 n=1 (Figure 25.8a). When in an excited state the electron may decay to a lower energy level by emitting a photon of electromagnetic radiation with energy equal to the difference (Figure 25.8b).
Every atom has its own characteristic energy level diagram. Although the Bohr model successfully predicts the principal energy levels of the hydrogen atom, it fails to predict the energy levels of multi-electron atoms. In the 1920s quantum mechanical models based on the Schrödinger wave equation replaced the Bohr model. The full quantum theory shows that the allowed energy levels are grouped in 'shells' corresponding to the n = 1, 2, 3 ... energy levels of the Bohr atom. Each shell is further divided into subshells. The shells can hold increasing numbers of electrons (2, 8, 18, etc.) as n increases. The number of electrons occupying the outermost shell determines the chemical properties of the atom. The shell model for atomic structure is covered in depth in chemistry courses. Although not complete, the Bohr picture gives us a fundamental insight into the origin of many atomic aa n=1 Figure 25.8 Transitions between levels. (a) Excitation: the electron must absorb energy in order to be excited from a lower level to a higher one. (b) Decay: when an electron decays from an excited state to a lower energy level it emits energy. characteristics, including atomic energy levels and their connection to the absorption and emission spectra of atoms. Atomic spectra Emission spectra Matter may be induced to emit electromagnetic radiation by raising its temperature or by exciting its atoms with another form energy, for example by passing an electric current through a gas in a glass tube (a discharge tube). Any visible light emitted can be analysed by dispersing it (separating its component wavelengths) into a spectrum with the aid of a prism or a diffraction grating (see Chapter 9). Figure 25.9 shows the optical emission spectra of the Sun and of hydrogen gas. The solar spectrum (Figure 25.9a) is an example of a continuous spectrum – it contains all wavelengths, from the long wavelength red end of the visible spectrum and below, to the short wavelength violet end and above. This type of spectrum is characteristic of matter at high temperatures. bb Figure 25.9 Optical emission spectra of (a) the Sun and (b) hydrogen. ITQ 5 How does the Bohr atom overcome the problem with the Rutherford atom? ITQ 6 Use the energy level diagram of hydrogen (Figure 25.7) to find the energy needed to completely remove an electron in the ground state from the atom. Express your answer in electron volts and in joules. Chapter 25 Atomic structure excited states 0.00 eV ∞ –0.38 eV –0.54 eV n=6 n=5 IR –0.85 eV –1.51 eV Paschen series –3.4 eV ground state Use the values for the energy levels of the hydrogen atom in Figure 25.10 to calculate the frequency and wavelength of the photons emitted by hydrogen when an electron decays from: (a) the first excited state (n = 2) to the ground state (n = 1); (b) from the n = 3 state to the n = 2 state. (c) To which parts of the electromagnetic spectrum do these photons belong? h = 6.63 × 10−34 J s c = 3.0 × 108 m s−1 A (a) E2 = −3.4 eV E1 = −13.6 eV E2 − E1 = −3.4 − (−13.6) eV = 10.2 eV = 10.2 × 1.6 × 10−19 J = 1.63 × 10−18 J Now, E2 − E1 = hf Therefore, E − E1 1.63 × 10−18 f= 2 = = 2.5 × 1015 Hz h 6.63 × 10−34 Also, c 3.0 × 108 = = 1.2 × 10−7 m λ= f 2.5 × 1015 n=4 n=3 Lyman series n=1 −13.6 eV Figure 25.10 Transitions between the energy levels of the hydrogen atom. The hydrogen spectrum (Figure 25.9b) is a line spectrum; only certain well-defined wavelengths are present. Figure 25.10 shows how the line spectrum arises from transitions between allowed electron energy levels in the hydrogen atom. The electric current passed through the hydrogen gas excites electrons from the ground state to the allowed excited states. In the case of hydrogen each atom has only one electron, which can be excited to only one of the possible states, but in a gas of many atoms all possible transitions will take place. As the electron decays from an excited state E2 to a lower energy state E1 a photon of radiation with energy equal to the energy difference E2 − E1 between the two states is emitted. The energy of a photon of electromagnetic radiation is given by E = hf, where h is Planck's constant and f is the frequency of the radiation. Thus, ...(25.5) E2 − E1 = hf Or, since f = c , where c is the velocity of light and λ is its λ wavelength, hc E2 − E1 = λ Q n=2 Balmer series UV Worked example 25.2 ....(25.6) The line spectra of atoms are clear evidence of the existence of discrete energy levels for atomic electrons. Since only certain energy values are allowed, the transitions between these levels result in spectral lines of well-defined wavelengths through the equation E2 − E1 = hc . λ Figure 25.10 shows that the possible transitions between hydrogen energy levels can be grouped into series. Each (b) E3 = −1.51 eV E2 = −3.4 eV E3 − E2 = −1.51 − (−3.4) eV = 1.89 eV = 1.89 × 1.6 × 10−19 J = 3.02 × 10−19 J Now, E3 − E2 = hf Therefore, E − E2 3.02 × 10−19 f= 3 = = 4.6 × 1014 Hz h 6.63 × 10−34 Also, c 3.0 × 108 = = 6.5 × 10−7 m λ= f 4.6 × 1014 (c) The photon in part (a) is in the ultraviolet region of the spectrum; the photon in part (b) is in the visible region (see the diagram of the electromagnetic spectrum on page 152).
series corresponds to a band of lines in the emission spectrum. The Lyman series results from transitions down to the ground state (n = 1); the Balmer series from transitions down to the n = 2 state and so on. The Lyman series lines all lie in the ultraviolet part of spectrum, the Balmer series falls in the visible region (as seen in Figure 25.9b), and the Paschen series in the infrared. Absorption spectra The optical emission spectrum of a sodium lamp (a discharge lamp containing sodium vapour at low 405 406 Unit 2 Module 3 Atomic and nuclear physics 700 600 500 a 400 nm b E3 E2 a b 700 600 500 400 nm Figure 25.11 Emission and absorption spectra of sodium: (a) emission; (b) absorption. pressure) includes a pair of closely spaced yellow lines known as the sodium doublet (Figure 25.11a). These are produced by electron transitions from two closely spaced energy levels to a lower level. When light with a continuous spectrum is shone through a cool sodium vapour, dark lines appear at the position of the sodium doublet (Figure 25.11b). The cool sodium vapour has absorbed photons with identical frequencies to those emitted by a hot vapour. Figure 25.12 shows how the same energy levels may produce either absorption or emission. ■ Emission takes places when a level that has been vacated by an excited electron is filled by an electron from a higher level (Figure 25.12a). ■ Absorption occurs when an electron is excited from the lower level to a vacant higher level (Figure 25.12b). E1 Figure 25.12 Electron transitions producing (a) emission and (b) absorption at the same frequencies. passes through. By measuring wavelengths, and matching them to the known spectral lines of the elements, we can deduce the composition of the stars. The atomic nucleus The name 'proton' was given to the hydrogen nucleus in 1920 by Rutherford. The nucleus of the most common form of hydrogen consists of a single proton with positive charge equal in magnitude, but opposite in sign, to the charge on the electron. The nuclei of heavier elements contain increasing numbers of protons: helium 2, lithium 3, beryllium 4, and so on. The number of protons is the atomic number Z of the element. As well as protons, the nuclei of all atoms (apart from the simplest form of hydrogen), contain neutrons. The neutron is a particle with almost the same mass as the proton but with no charge. Its discovery by Chadwick was discussed in Chapter 5. The number of neutrons in a nucleus is the neutron number N. The total number of neutrons plus protons is equal to the mass number A of that nucleus. Thus, An absorption spectrum thus appears as a series of dark lines in an otherwise continuous spectrum. The absorption spectrum of an element may be regarded as the inverse or negative of its emission spectrum. Note, though, that the energy absorbed is subsequently re-emitted when the excited electrons drop back into the vacated levels.
Because these emissions may be in any direction, energy is 'lost' from the direction of the incident beam.
Thus dark lines are observed in spectrum of the transmitted radiation. The composition of a particular atomic nucleus is represented thus: AZX, where X is the symbol for the element, Z is the atomic number and A the mass number. For example: 126C, or 197 79Au. Spectral analysis Isotopes and nuclides Each element has its own characteristic line

spectrum corresponding to the energy levels of its electrons. Both emission and absorption spectra can therefore be used as 'fingerprints' in chemical analysis to identify which elements are present.

A nuclide is an atomic nucleus with a specific number of 197 protons and neutrons: so 126C, 238 92U and 79Au are different nuclides. All nuclei may therefore be referred to as nuclides. The spectra of the Sun (Figure 25.9a) and other stars contain bands of dark lines known as Fraunhofer lines (after Joseph von Fraunhofer who first observed them). These lines result from absorption by elements in the cooler outer atmosphere of the star as the stellar radiation A = Z + N ....(25.7) Isotopes are atoms of the same element that have different mass numbers because they have different numbers of neutrons in their nuclei. The element carbon, for example, has naturally occurring isotopes 126C, 136C and 146C. ITQ 7 Write down the number of protons, neutrons and electrons 197 in neutral atoms of: 126C, 238 92U and 79Au. Chapter 25 Atomic structure X-rays X-rays were discovered by William Röntgen in 1895. He called them X-rays because at first their nature was unknown. Röntgen discovered that invisible radiation was emitted from the end of a cathode ray tube that the cathode rays (electrons) were striking. This radiation was detected by its effect on a photographic film. Although the radiation could not be seen, it a darkened the film in similar way to light, but, unlike light, was able to pass through the opaque wrapping to affect the film beneath. X-rays are emitted whenever cathode rays are brought to rest by matter. The kinetic energy that the cathode rays gained in accelerating between the cathode and anode of a cathode ray tube is transformed into X-rays when the cathode rays are decelerated back to rest. We now know that X-rays are high-frequency electromagnetic radiation. The frequencies of X-rays are in the range 1016–1021 Hz and their wavelengths are in the range 10−7–10−13 m. X-ray production Figure 25.13 is a cross-section of an X-ray tube. Cathode rays emitted by a hot filament are accelerated towards a copper anode held at a potential difference of several tens of thousand volts with respect to the cathode. Another heavy metal such as lead is sometimes inserted as a target into the anode, at the point where the cathode rays strike, in order to produce X-rays with certain characteristic energies. When cathode rays strike the heavy-metal target most of their energy is turned into heat, but about 0.5% is ultimately converted into electromagnetic radiation in the X-ray waveband.

The excess heat energy produced is carried away by the heavy copper anode, which may have water-cooling coils wrapped around it. Modern high output tubes have a rotating anode. Because the cathode rays are not continuously incident on the same point of the anode, higher currents and therefore more intense X-ray production can thus be achieved without overheating.

A tungsten–rhenium target can withstand the high temperatures produced; a molybdenum core conducts heat away from the target. X-ray spectrum Figure 25.14 shows a typical X-ray spectrum from an X-ray tube. The minimum wavelength λmin is determined by the potential difference V applied between the cathode and the anode, and is given by the expression derived in the following section, λmin = c ch = fmax eV ....(25.8) The spectrum extends from λmin to higher wavelengths and consists of two components.

■ The continuous spectrum is a continuous range of all wavelengths greater than λmin.

This radiation is emitted by the cathode rays as they decelerate on striking the anode. It is sometimes referred to as braking or 'bremsstrahlung' radiation (bremsstrahlung is the German word for braking rays). This part of the spectrum can be compared to the continuous spectrum of light from a high-temperature source such as the Sun. ■ The characteristic spectrum is a series of higher intensity lines of particular wavelengths superimposed on the continuous spectrum. These spectral emission lines are produced by electron transitions in the atoms of the target. Electrons are excited from the energy levels of the inner shells by the incident cathode rays; as electrons from higher energy shells refill these shells, the excess energy is released as photons with well-defined energies and wavelengths. The position of the lines in the line spectrum is a characteristic of the element from which the target is made. In just the same way as the spectral lines of hydrogen are grouped in series, so are the characteristic X-ray lines of heavy metals. The K series lines correspond to transitions to the K (n = 1) shell; the L series to the n = 2 shell and so on. X-rays electron beam (attracted by high voltage on anode) cooling fins low voltage supply copper anode vacuum heavy metal target hot filament (electron source) Figure 25.13 An X-ray tube. 407 408 Unit 2 Module 3 Atomic and nuclear physics electron by the accelerating voltage. X-rays with higher frequencies and shorter wavelengths, which carry more energy per photon, may be generated by increasing the voltage applied to the X-ray tube. Kα Intensity Kβ Intensity Lα K series λ min Lγ Lβ L series Wavelength Figure 25.14 A typical X-ray spectrum. Energy changes in an X-ray tube The cathode rays that excite X-ray emission in the X-ray tube gain kinetic energy as they are accelerated from the cathode to the anode by the large potential difference applied. The potential difference between two points is defined as the energy change per unit charge in moving between those two points. If the potential difference between the cathode and the anode is V, then the energy gained by an electron with charge e in accelerating from one to the other is eV. This is the kinetic energy of the electron as it strikes the anode, The brightness or intensity of the X-ray beam is controlled by the quantity of the X-ray photons produced, not their energy. This quantity depends on the current (the number of electrons passing from the cathode to the anode per second) in the tube. But increasing the current has costs – the anode must be cooled more effectively or it will overheat and melt. As noted above, high-intensity X-rays tubes have rotating anodes so that cathode rays do not constantly strike the same point on the metal surface.

A large moving anode has a higher heat capacity than a small stationary anode. Worked example 25.3 Q An X-ray tube is operated at 40 kV.

Calculate: (a) the kinetic energy of an electron accelerated by this potential (b) the speed to which an electron is accelerated by this potential (c) the maximum frequency of the X-rays produced by the tube (d) the minimum wavelength of the radiation (the 'cut-off' wavelength). ■ velocity of light, c = 3 × 108 m s−1 ■ charge on the electron, e = 1.602 × 10−19 C ■ mass of the electron, me = 9.11 × 10−31 kg ■ Planck's constant, h = 6.63 × 10−34 J s A (a) KE = 12 mv2 = eV = 1.602 × 10−19 × 40 × 103 J = 6.41 × 10−15 J (b) KE = 12 mv2 So: 2KE 2 × 6.41 × 10−15 v = m s−1 = 1.19 × 108 m s−1 = m 9.11 × 10−31 1 eV = 2 mv2 The energy of a photon of electromagnetic radiation is given by E = hf. If one electron gives all its kinetic energy to one X-ray photon, then from the conservation of energy, 1 hf = 2 mv2 = eV This is the maximum possible energy for X-ray photons produced by cathode rays accelerated through a voltage V.

The maximum frequency of the X-rays produced by an X-ray tube operated at V volts is thus: fmax = eV h Using the wave equation, c = fλ, the minimum wavelength ('cut-off' wavelength) of the X-rays emitted is, λmin = c fmax ch = eV X-rays with this frequency and wavelength are the most energetic emitted by the X-ray tube. Most of the radiation emitted will have a lower frequency and a longer wavelength than these limiting values, since the photons will not carry out all the energy originally given to an (c) fmax = eV 1.602 × 10−19 × 40 × 103 = = 9.67 × 1018 Hz h 6.63 × 10−34 (d) λmin = c fmax = 3 × 108 = 3.10 × 10−11 m 9.67 × 1018 X-ray attenuation The intensity of an X-ray beam passing through any material decreases progressively with distance as the beam penetrates into the material; we say that the beam is attenuated by the material. Chapter 25 Atomic structure X-rays penetrate low-density materials significantly. For example, X-rays pass through paper and human flesh with little attenuation. However X-rays are strongly absorbed by metals. The X-rays from a tube operating at 100 kV are almost completely absorbed by a thin layer of lead (a few mm). The energy of the X-ray photons is transferred to electrons in the material as they are absorbed; dense materials contain more electrons per unit volume and so absorb X-rays more strongly. Experiment shows that the fractional change in intensity δI I of an X-ray beam in passing through a small thickness δx of material is proportional to δx. Therefore, δI = − μ δx I where μ is a constant which depends on the material and the energy of the X-rays.

The negative sign indicates that the intensity is decreasing. This equation may be solved by integration to find how the intensity varies with the distance the X-rays have penetrated. (Note: integration is not required by the CAPE Physics syllabus, but the equations are included here for the interest of students who are familiar with integration from their mathematics course.) I δI x = − μ dx I0 I 0 where I0 is the initial intensity of the X-ray beam and I the intensity after passing through thickness x of the material.

∫ 1.0 0.6 I I0 ln I − ln I0 = −μx Therefore, I = exp(− μx) I0 And so, I = I0 exp(−μx) ....(25.9) The constant μ is called the linear absorption coefficient of the material. Values of μ for a selection of materials are given in Table 25.1. Table 25.1 X-ray linear absorption coefficients, μ (unit m−1) Energy / keV Air 20 8.5 × 10−2 40 2.9 × 10 60 80 Water Aluminium Iron 7.4 × 101 8.8 × 102 2.0 × 104 1.1 × 105 2.6 × 10 1.5 × 10 3 2.8 × 10 1.7 × 104 2.2 × 10−2 2.0 × 101 7.2 × 101 9.3 × 102 6.0 × 103 2.0 × 10−2 1.8 × 101 5.3 × 101 4.7 × 102 2.8 × 103 −2 1 2 0.4 iron 0.2 0 lead 0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 Thickness, x / mm Figure 25.15 The attenuation of 60 keV X-rays by aluminium, iron and lead. X-ray attenuation is an example of exponential decay; it follows the same pattern as the discharge of a capacitor through a resistor (page 286) or the decay of a radioactive isotope (page 424). Figure 25.15 compares the attenuation of 60 keV X-rays by aluminium, iron and lead. Worked example 25.4 Q A dental X-ray machine operates at voltage V = 80 kV. To ensure the long-term safety of the staff at a dental practice, the room in which the machine is operated is to be lined with lead sheet. The sheet must absorb 99.9% of any stray X-rays from the machine.

What thickness of lead sheet is required? A In Table 25.1 the linear absorption coefficient of lead for 80 keV X-rays is given as μ = 2.8 × 103 m−1 Now, if 99.9% of the X-rays are to be absorbed, 0.1% will be transmitted. Therefore, I = 0.001 I0 But, I = I0 exp(−μx) Therefore, I = exp(−μx) = 0.001 I0 ∫ Performing the integrations gives, aluminium 0.8 Taking natural logarithms of both sides, −μx = ln (0.001) ln (0.001) 6.9 x = = = 2.5 × 10−3 m = 2.5 mm μ 2.8 × 103 Lead ITQ 8 Suggest why a point source is used to produce a radiograph. (Hint: there are no practical lenses for focusing X-rays). ITQ 9 How would a tooth cavity show up on a dental X-ray?

Would it be lighter or darker than the surrounding enamel? 409 410 Unit 2 Module 3 Atomic and nuclear physics rotating target converging electrons 'focus' of X-rays adjustable diaphragm conical beam definer X-rays Figure 25.17 A dental X-ray. A lead grid placed between the patient and the detector helps to reduce blurring of the image by blocking scattered photons from reaching the film/detector. To stop the grid itself casting a shadow, it is slowly moved across the film as the exposure takes place. scattered X-ray patient film grid A radiograph recorded on a photographic film is a negative (Figure 25.17). The regions of the film exposed to the most X-rays are darkest. Lighter regions occur where bone or other dense tissue have cast shadows by absorbing X-rays. Figure 25.16 The production of a radiograph (an X-ray).

Medical uses of X-rays Radiography The most familiar use of X-rays is the production of radiographs, for example to investigate a fractured bone or to inspect teeth for evidence of decay. The principle of radiography is show in Figure 25.16. A converging beam of electrons is focused to a point on the target in the X-ray tube. A conical beam of X-rays diverges from the point towards the area to be X-rayed. A photographic film or a digital detector is placed on the far side of patient. Dense tissue (bone or tooth enamel) absorbs X-rays more strongly than soft tissue, and so casts a darker shadow on the detector. CAT scan CAT stands for computer axial tomography. A CAT scan provides a far more detailed image of the internal structure of the body than is available from a simple radiograph. In a radiograph, structures that overlay each other cannot be distinguished, and there is very little contrast between different regions of soft tissue. With CAT scanning detailed cross-sections of the body, providing three-dimensional information, can be obtained. The basic components of a CAT scanner are shown in Figure 25.18. The collimator produces a narrow pattern monitor X-ray tube computer collimator detectors Figure 25.18 How CAT scanning works. Chapter 25 Atomic structure front back radiation source moves in circle around patient shadow 2 left incident radiation position 1 right Cancer cells at centre of circle.

The radiation hits these cells all the time.

shadow 1 incident radiation position 2 Figure 25.19 The principle of CAT scanning. Figure 25.20 The radiation source rotates around the patient positioned with the tumour at the centre of the circle.

of X-ray beams, which fan out to pass through a slice of the body before reaching the detectors. The X-ray source and X-rays detectors are rotated around the body, which is placed on the rotation axis. The detectors monitor the changes in intensity of the beams as the scanner turns through 360°. A huge amount of data is recorded, which is processed by computer to produce a cross-section of the body. By moving the patient in steps along the axis of rotation, a complete 3D picture can be produced. The principle underlying the technique is explained in Figure 25.19. X-ray therapy A simple X-ray radiograph taken from one angle is a 2D shadow of a 3D object. In Figure 25.19, shadow 1 of the three objects shows that the cube is to the left of the sphere, and the pyramid to the right, but it does not show which is closer to the observer and which most distant. Shadow 2, formed at 90° to shadow 1, reveals the arrangement front-to-back. From the two images we can deduce the 3D arrangement of the objects. A CAT scan provides images from multiple angles, which are combined to provide detailed 3D information.

TQ 10 Why is the radiation source rotated around the patient during radiotherapy? Megavoltage X-rays produced by a very high-energy X-ray generator called a linear accelerator (LINAC) are used to deliver radiotherapy to cancer patients. The LINAC uses microwave radiation to accelerate a pulsed beam of electrons to energies of the order of 10 MeV. The electrons gain energy from the microwaves in a similar way to a surfer taking energy from water waves. X-rays up to the same energy are produced when the electrons strike a heavy metal target – often tungsten. The purpose of radiotherapy is to destroy malignant cancer cells by 'zapping' them with damaging radiation. There are of course side effects, as the radiation can also damage the healthy cells it passes through. To minimize unwanted damage, the X-ray beam must be carefully focused onto the tumour. Figure 25.20 shows how rotation of the focused source around the patient maximizes the time the cancer cells are exposed to the radiation, while reducing the exposure time of the healthy tissue that surrounds the tumour. 411 412 Unit 2 Module 3 Atomic and nuclear physics Summary ✓ By making measurements of the charge to mass ratio e/m of cathode rays, and the basic unit of charge e, J. J. Thomson showed that cathode rays consist of subatomic particles with mass just 1/1840 that of the hydrogen atom. We now know these particles as electrons. ✓ The Millikan oil drop experiment provides an accurate measurement of the charge on the electron and confirms that charge is quantized. ✓ The Geiger–Marsden alpha particle scattering experiment demonstrates that the atom is mostly empty space with nearly all of its mass concentrated in a tiny positively charged central nucleus.

The diameter of the nucleus is just 10−5 the diameter of the atom. ✓ In Rutherford's model of the atom, electrons orbit the central nucleus in a similar way to planets orbiting a star in a solar system. ✓ According to the classical theory of electromagnetism, the Rutherford atom should not be stable; charged electrons should radiate energy as a result of their centripetal acceleration and so spiral into the nucleus. ✓ The Bohr atom was the first attempt to use quantum ideas to explain atomic structure; in the Bohr atom electrons can only occupy orbits for which a whole number of wavelengths (as calculated from the de Broglie equation p = h) fit λ exactly around the circumference. ✓ The Bohr model successfully predicts the principal energy levels of the hydrogen atom, but fails to predict the energy levels of multielectron atoms.

✓ Full quantum mechanical atomic models based on the Schrödinger wave equation show that atomic energy levels are grouped in a series of 'shells'.

✓ When electrons make transitions between energy levels, atoms emit and absorb radiation of well-defined frequencies to produce line spectra. ✓ The frequency of the radiation absorbed or emitted may be calculated from the energy difference between the two levels involved in a transition: E2 − E1 = hf ✓ Emission spectra appear as bright lines against a dark background. ✓ Absorption spectra appear as dark lines in a continuous spectrum. ✓ The mass number A of an atomic nucleus is related to the atomic number Z (the number of protons) and the neutron number N by A = Z + N ✓ The composition of the nucleus of element X is represented by AZX; for example 238 92U. ✓ Isotopes are atoms of the same element with different mass numbers, for example 126C, 136C, and 146C. ✓ X-rays are high-frequency electromagnetic radiation, which may be produced when cathode rays are brought to rest by striking matter (especially heavy metals). ✓ A typical spectrum from an X-ray tube consists of a continuous spectrum – produced by the deceleration of the cathode rays – plus a higher intensity characteristic spectrum of spectral lines produced by electron transitions in the atoms of the target material. ✓ The minimum (cut-off) wavelength of an X-ray spectrum is given by λmin = c = ch fmax eV ✓ X-rays are absorbed by matter; the attenuation is given by I = I0 exp(−μx), where μ is the linear absorption coefficient of the material. ✓ In medicine X-rays are used to produce radiographs and CAT scans. High-energy (megavolt) X-rays are used in the treatment of cancer by radiotherapy.

Chapter 25 Atomic structure Review questions Table 25.2 Millikan drop experimental data (note: simulated data for illustration only) Atomic theory, The Millikan oil drop experiment 1 Explain briefly how J. J. Thomson's discovery of the electron in 1887 changed the view of atomic structure that had been held since the time of the Ancient Greeks. 2 (a) Describe the principle of the Millikan oil drop experiment to determine the electronic charge e. (b) In the oil drop experiment, the effective weight m'g of a drop is a little less than the actual weight mg because of the upthrust from the air the drop displaces (by Archimedes' principle). Given that, weight of drop = volume × density of oil × g m'g = 4 3 πr (ρo − ρa)g 3 where ρo is the density of the oil and ρa the density of air. (c) The drag force on a spherical drop of radius r moving at velocity v1 through air of viscosity η is given by Stokes' law, Use this result, together with the expression for m'g in part (b) of this question, to show that the radius of a drop in the Millikan experiment is given by, 1 9ηv1 2 r= 2(ρo − ρa)g [ ] 3 Table 25.2 gives data obtained by observing a single drop in the Millikan oil drop experiment.

Between each measurement of v2 the drop was exposed to radiation from a radioactive source in order to change its charge q. 3.0 × 10 2 5.5 × 10 −4 3.67 4.5 × 10−4 3 1.5 × 10 −4 1 6.0 × 10−4 4 3.5 × 10 −4 2.33 7.0 × 10−4 4.67 5.5 × 10 3.67 −4 n, fourth guess 2 If we assume that the charge is a whole number multiple, n, of the charge on the electron, e, then q = ne. From q = k (v1 − v2), we have, E k ne = (v1 − v2) E k A graph of (v1 − v2) against n should thus produce a straight line with gradient e E . Since k can be k E calculated from the experimental parameters (as outlined in the previous question) a value for e can be found. But there is a problem: we don't know the value of n for each drop. However, we can deduce n as follows.

Start by guessing that the smallest value of (v1 – v2) corresponds to n = 1. If (v1 – v2) is proportional to n we can then find the n value for each of the other (v1 – v2) values. These must all be integers, so if one or more turns out to be a fraction, we know our initial guess was wrong, and we can then try n = 2. The first column has been completed in the table. Was this successful? Follow the procedure outlined to find the correct values of n for each (v1 – v2). Plot a graph (v1 – v2) against n and hence calculate a value for e.

Note: in practice many more measurements of (v1 – v2) for a number of different drops are made to determine a reliable value for e. k q = (v1 − v2) E [1 −4 (v1 – v2) = ne where v1 is the drop's terminal velocity in the absence of an electric field. (d) The theory of the oil drop experiment gives the charge on the oil drop q as, Given that from Stokes' law k = 6πηr, show that, 1 9ηv1 6πη 2 q= (v1 − v2) E 2(ρo − ρa)g 1.67 1.5 × 10−4 n, third guess or, Fdrag = 6πηrv1 2.5 × 10−4 n, second guess k = 3.2 × 10−15 C s m−1 E upthrust = volume × density of air × g Show that for a spherical drop of radius r, n, first guess 3.0 × 10−4 and, (v1 − v2) / m s−1 413 414 Unit 2 Module 3 Atomic and nuclear physics Models of the atom, Atomic spectra, The atomic nucleus 4 Describe the Thomson 'plum pudding' model of atomic structure. Explain how the results of the Geiger–Marsden alpha particle scattering experiment demonstrated that this model could not be correct.

How did Rutherford's 'nuclear' model account for the observations that: (i) most alpha particles pass through a thin gold foil with very little deflection; (ii) a small number of alpha particles are deflected through large angles? 5 Why, according to classical theory, are the orbits of the electrons in the Rutherford atom unstable? How does the atomic model suggested by Bohr overcome this difficulty? 6 Explain briefly how the Bohr model of the hydrogen atom accounts for the observation of discrete spectral lines in the emission spectrum of hydrogen gas. 7 Figure 25.21 shows the n =1, 2 and 3 energy levels of the hydrogen atom. 3 –1.51 eV 2 –3.4 eV 1 –13.6 eV Figure 25.21 (a) Find the frequencies of the photons emitted or absorbed in all possible transitions between these levels. (b) Calculate the corresponding wavelengths of the radiation. (c) Which one of these wavelengths can be observed as a line in the optical absorption spectrum of hydrogen gas? Describe the appearance of the line in the absorption spectrum. 8 Given that the ground state of the hydrogen atom has energy −13.6 eV, calculate: (a) the minimum frequency of the electromagnetic radiation which will cause ionization (removal of the electron) from hydrogen (b) the minimum energy (in joules) and speed of an electron that can ionize a hydrogen atom by colliding with it. 9 (a) An isotope of nitrogen can be represented as 147N. What is the significance of each of the numbers 14 and 7? (b) Write down the numbers of protons, neutrons and electrons in neutral atoms of each of the following 138 235 nuclides: 95 36Kr, 56Ba, 92U. X-rays, Medical uses of X-rays 10 (a) Draw a labelled diagram of an X-ray tube. What features of the operation of the tube determine: (i) the intensity of the X-rays (ii) the penetrating power (maximum energy) of the X-rays? (b) Sketch a typical X-ray spectrum produced by an X-ray tube. Label the 'continuous' and 'characteristic' regions of the spectrum. Describe briefly how the two regions of the spectrum are produced.

11 (a) Outline the energy changes that take place in an X-ray tube. (b) An accelerating potential of 25 kV is applied to an X-ray tube. Calculate: (i) the kinetic energy of the electron accelerated by this potential (ii) the speed to which an electron is accelerated by this potential (iii) the maximum frequency of the X-rays produced by the tube (iv) the minimum wavelength of the radiation (the 'cut-off' wavelength). (c) Explain how X-rays with a shorter wavelength could be produced. 12 The attenuation of X-rays by matter is described by the equation I = I0 exp(−μx). Describe briefly the meaning of the symbols I, I0, μ and x in this equation. Sketch a graph to show how I varies with x as X-rays pass through matter. 13 The thickness x 1 of a layer of material required to 2 reduce the intensity of an X-ray beam by half is described as the 'half-value thickness'.

If I = I0/2 when x = x 1 , use the expression I = I0 exp(−μx) to show that: 2 x1 = 2 1 0.693 ln 2 = μ μ Use the data given in Table 25.1 to calculate the half-value thicknesses of air, water, aluminium, iron and lead for 80 kV X-rays. Chapter 25 Atomic structure 14 Compare and contrast the processes of producing a radiograph and a CAT scan for medical diagnosis. Explain why the CAT scan provides more detailed diagnostic information than the radiograph. Practical activities and challenges 1 Use a spectrometer and diffraction grating (see Chapter 9) to determine the wavelengths of the emission lines in the visible spectrum of a hydrogen discharge tube. Compare the wavelengths you measure to the wavelengths predicted for transitions to the n = 2 level of the Bohr atom (the Balmer series). 2 Use the apparatus in Figure 25.22 to determine the linear absorption coefficient and half-value thickness of thin paper for visible light. lamp computer or data logger paper light sensor Answers to ITQs 1 The drag force experienced by the oil drop, and hence its terminal velocity, depends on the viscosity of the air; the upthrust from the air depends on the air density (see Review question 2 above). Both viscosity and density change with temperature. The apparatus must therefore be held at constant temperature to obtain consistent results. 2 Most of the Rutherford atom is empty space and therefore most of the alpha particles will suffer zero or very small deflections as they pass through. Only when an alpha particle passes close to, or makes a direct hit on the tiny nucleus will it be deflected through a large angle; since most of the atomic mass is concentrated in a tiny volume a direct hit on the nucleus can cause an alpha particle to bounce back from the target. 3 A particle moving at constant speed in a circle of radius r has a centripetal acceleration ω2r, where ω is the angular velocity (see Chapter 6).

This acceleration is perpendicular to the particle's instantaneous velocity and produces the continuous change in the direction of the motion required for movement in a circle. 4 Angular momentum L = pr.

Figure 25.22 Adjust the settings of the light sensor so that full-scale reading is obtained with nothing between the light source and the sensor. p= Place one sheet of white paper between the source and the sensor and record the new sensor reading. Repeat, increasing the number of sheets, n, one at a time. Devise a simple method to determine the thickness of a single sheet with a ruler. (Hint: measure the thickness of a stack of sheets?) Plot an appropriate graph to test if the intensity variation is given by I = I0 exp(−μx). (Hint: this is the equation of exponential decay. What form will a plot of ln I against x take if the decay is exponential?) For a particle that obeys the de Broglie relationship, If possible, determine μ and x 1 from your graph. 2 h λ For the Bohr atom, nλ = 2πr And so, λ= 2πr n Therefore L = pr = hr nhr nh = = 2πr 2π λ 5 The energy of the electron in the Bohr atom is quantized. When the electron occupies the ground state there is no lower energy state to which it can move by losing more energy; the ground state is therefore stable. 6 The energy of the electron in the ground state is −13.6 eV. The energy of a stationary electron outside the atom is 0 eV. The minimum energy required to completely remove a ground state electron from the atom is thus +13.6 eV = 13.6 × 1.6 × 10−19 J = 2.18 × 10−18 J 415 416 Unit 2 Module 3 Atomic and nuclear physics 7 The neutral carbon atom has 6 protons, 6 neutrons and 6 electrons. The neutral uranium atom has 92 protons, 146 neutrons and 92 electrons. The neutral gold atom has 79 protons, 118 neutrons and 79 electrons. 8 Because there are no lenses available to focus the X-rays, a point source is required to produce a sharp shadow which shows detail. This may be compared to producing an image with a pinhole camera. If the source/pinhole is extended the shadow/image is blurred. 9 The cavity (hole) is less dense than the surrounding material of the tooth and so will absorb fewer X-rays. The corresponding region of an X-ray film will be exposed to a greater intensity of X-rays and be darker (remember a radiograph is like a photographic negative – areas which are exposed to more radiation are darker). The cavity will therefore show as a dark shadow in the X-ray. 10 To minimize the exposure time of healthy tissue to the radiation beam. Answers to Review questions 3 (v1 − v2) / m s−1 n, first guess n, second guess 2.5 × 10−4 1.67 3.33 5 1.5 × 10−4 1 2 3 3.0 × 10−4 2 4 6 5.5 × 10 3.67 7.33 11 4.5 × 10−4 3 6 9 1.5 × 10 −4 1 2 3 6.0 × 10−4 4 8 12 3.5 × 10 −4 2.33 4.67 7 7.0 × 10−4 4.67 9.33 14 5.5 × 10 3.67 7.33 11 2 4 6 −4 −4 3.0 × 10−4 7 n, third guess n, fourth guess The third guess gives all integers; therefore a fourth guess is not required. From the gradient of a graph of this data e = 1.6 × 10−19 C. Transition f / Hz 3→1 2.9 × 1015 1.0 × 10−7 2→1 15 2.5 × 10 1.2 × 10−7 3→2 4.6 × 1014 6.5 × 10−7 λ/m 8 (a) 3.28 × 1015 Hz (b) 2.18 × 10−18 J; 2.16 × 106 m s−1 95 9 (b) 36Kr – 36 protons, 36 electrons, 59 neutrons 138 56 Ba – 56 protons, 56 electrons, 82 neutrons 235 92 U – 92 protons, 92 electrons, 143 neutrons 13 Material Air half-value 35 thickness / m Water Aluminium Iron Lead 0.038 (3.8 cm) 0.013 (1.3 cm) 2.5 × 10−4 (0.25 mm) 1.5 × 10−3 (1.5 mm) 417 Chapter 26 Radioactivity Learning objectives ■ Relate radioactivity to nuclear instability. ■ Discuss the spontaneous and random nature of nuclear decay. ■ Identify the origins and environmental hazards of background radiation. ■ Describe experiments to distinguish between the three types of emissions from radioactive substances. ■ Write and interpret equations for radioactive decay. ■ Discuss the environmental hazards of radioactive emissions. ■ Discuss the necessary safety precautions for handling and disposal of radioactive material. ■ Explain 'activity', 'decay constant' and 'half-life', and use the relationship A = λN. ■ Use the law of decay dN = −λN and N = N0 exp(−λt) to solve problems. dt 1 ln 2 to solve problems. 2 λ ■ Describe an experiment to determine the half-life of a radioactive isotope with a short half-life. ■ Use the relation T 1 = ■ Discuss uses of radioisotopes as tracers, for carbon dating and in radiotherapy. ■ Describe the operation of simple detectors. Historical introduction In 1896 Henri Becquerel (1852–1908) found, by chance, that when he placed a uranium–potassium sulfate crystal on a photographic plate wrapped in thick black paper, the developed plate showed a blackened image of the crystal. No light had reached the plate, so why was it affected? Becquerel suggested that the uranium salt gave off rays that could pass through the wrapping paper. Becquerel later discovered that similar rays were emitted from other uranium compounds. He concluded that uranium was responsible for the emission of the invisible rays. Becquerel's discovery stimulated an intense period of research into the new phenomenon of radioactivity – both by his students Marie and Pierre Curie, and by other research groups around the world. Marie Curie (1867–1934) was a leading pioneer of research into radioactivity (a term she coined). Her many achievements include: the discovery of two elements (radium and polonium); the award of Nobel Prizes for both Physics and Chemistry (she is the only person to have been awarded prizes for two separate sciences); the first woman to be appointed professor at the University of Paris; the establishment of the first mobile X-ray units (during World War 1); the first research into radiotherapy for the treatment of cancer. We now know that all naturally occurring elements with atomic numbers above 83 are radioactive. A few of the elements with atomic numbers below 83 also have naturally occurring radioactive isotopes. Carbon is an example. Natural carbon is mainly C-12, but carbon also contains small percentage of the isotope C-14, which is radioactive. Nuclear instability Radioactivity is a result of nuclear instability. A radionuclide (radioactive nucleus) is in an unstable state; it can decay to a more stable state by emitting excess energy in the form of radiation. The stability of a nuclide depends on both the number of protons and the number of neutrons it contains. These nucleons are held together by the very short range 'strong nuclear' force that acts between them. This force is about 100 times stronger than the electromagnetic force that acts between charged particles, but it only acts over Unit 2 Module 3 Atomic and nuclear physics Characteristics of nuclear decay 120 The early studies of radioactivity revealed that it has the following characteristics: 110 band of stability 100 ■ The nuclei of radioactive elements spontaneously decay into more stable nuclei by emitting radiation. 90 Number of neutrons, N 418 ■ Radioactive decay is a random process.

The probability 80 N=Z of the decay of a single nucleus in a given time interval is constant, but it is not possible to predict exactly when it will decay. 70 60 ■ For a large number of nuclei of the same type, the 50 fraction that decay in a given time interval is constant. ■ The radiations from radioactive elements produce 40 30 20 10 0 0 10 20 30 40 50 60 70 80 90 Number of protons, Z Figure 26.1 Plot of neutron number N against proton number Z for stable nuclides. short distances – you can imagine it as acting like a strong glue that holds nucleons together inside the nucleus but does not extend outside the nucleus. A nucleus consisting of only two or more protons is not stable.

This is because of the electromagnetic repulsion between the positive charges the protons carry. The presence of neutrons stabilizes the nucleus against this repulsion. Figure 26.1 show a plot of the neutron number against the proton number for the stable nuclides found in nature. In general, heavy nuclides contain more neutrons than protons and so the 'band of stability' curves upwards. Nuclides with neutron to proton ratios on either side of the band of stability are unstable – they are radioactive. These nuclides decay spontaneously, emitting radiation in such a way that their neutron to proton ratio moves closer to the stable region. Details of the different types of decay are discussed in the sections that follow. ITQ 1 Suggest why the neutron to proton ratio is greater than one for heavy nuclides. ITQ 2 Give a simple argument to account for the observation that the heaviest elements have atomic numbers of the order 100 and not say 500 or 1000. bright flashes of light when they strike certain compounds. The compounds are said to fluoresce. For example, rays from radium cause zinc sulfide to glow. A mixture of radium and zinc sulfide was once used to make luminous paint that made watch hands glow in the dark. ■ The radiations released by radioactive substances cause ionization of air molecules. Their energy knocks out electrons from the molecules, leaving them with a positive charge. ■ Radiations from radioactive elements can penetrate the heavy black wrapping around a photographic film. When the film is developed, it appears black where the radiations struck the film. ■ Radiations from radioactive elements can destroy the germinating power of plant seeds, kill bacteria and burn or even kill animals and plants. Radiation can also be used to destroy cancers.

Types of radiations and their properties In 1897, Ernest Rutherford showed that the invisible radiation discovered by Becquerel had at least two components with different properties. Rutherford called them alpha (α) and beta (β) radiation. A third component of this radiation was later discovered by Pierre Curie.

It was called gamma (γ) radiation. The properties of gamma radiation differ from those of alpha and beta radiations. However, all three originate from the nucleus of the atom. Most radioactive substances emit at least two kinds of radiation simultaneously (usually alpha and gamma, or beta and gamma).

When radiation from a source that emits all three kinds is directed into an electric field or a magnetic field, the three radiations behave differently, as shown in Figure 26.2a. Chapter 26 Radioactivity a b + lead container lead container α β γ γ α source of radiation electric field – source of radiation The alpha radiation is attracted towards the negative plate. The beta radiation is attracted towards the positive plate. The gamma radiation is unaffected and passes straight through. This shows that alpha radiation carries a positive charge, beta radiation has a negative charge, whilst gamma radiation has no charge associated with it. The deflections of these radiations in a magnetic field support these observations. In a magnetic field alpha particles and beta particles are deflected in opposite directions; alpha particles are deflected by only a small amount because they are more massive than beta particles. Gamma rays are not deflected at all. Alpha radiation Alpha radiation is made up of positively charged particles emitted from the nucleus of radioactive atoms. Measurement of their charge to mass ratio made by J. J.

Thomson at the beginning of the 20th century suggested that they were helium nuclei. This was confirmed by Rutherford and others. The helium-4 nucleus (42He), which contains two protons and two neutrons, is particularly stable. Since it contains two protons, the alpha particle carries a charge of +2e. Alpha particles typically travel at about 5% of the speed of light. Because of their positive charge, they tend to attract electrons away from nearby atoms and so cause ionization in a gas. Alpha particles have a low power of penetration, travelling only a few centimetres in air. They can be stopped by a thin sheet of paper or the outer layer of skin. Because they do not travel very far, even in a gas, all their ionizing power is concentrated in a small volume. Their effect is therefore intense where it occurs. Alpha particles can be detected on a photographic plate, in a cloud chamber (by straight thick, short tracks, all of about the same length) and by a spark counter. The most energetic alpha particles can be detected by a Geiger–Müller (GM) counter.

The principles of operation of various nuclear radiation detectors are described later in this chapter. Americium-241 is a good laboratory source of alpha radiation. magnetic field into page β Figure 26.2 The passage of alpha, beta and gamma radiation through: (a) an electric field; (b) a magnetic field. Beta radiation Beta radiation is made up of particles emitted from the nuclei of radioactive atoms. Measurement of their charge to mass ratio shows that beta particles are fast-moving electrons, like cathode rays. They are much lighter than alpha particles and have a charge of −1e. Their speed ranges from 30% to 99% of the speed of light. They are much less ionizing than alpha particles but are far more penetrating.

The most energetic beta particles have a range in air of a few metres and are stopped by about 5 mm of aluminium or 1 mm of lead. The term 'beta particle' is reserved for an electron which comes from the nucleus of an atom.

The electrons outside the nucleus are simply called electrons. Beta particles can be detected on a photographic plate, in a cloud chamber (by thin and twisted tracks) and by a GM tube. Strontium-90 is a good laboratory source of beta radiation. Gamma radiation Gamma radiation is high-energy electromagnetic radiation emitted from the nucleus of an atom. It travels at the speed of light and carries no charge. The nature of gamma radiation was not established until 1914. Like X-rays, gamma rays are very penetrating, but their ionizing power is very low. Their intensity can be reduced significantly by several centimetres of lead. They have shorter wavelengths than X-rays. There is no sharp dividing line between ultraviolet radiation and X-rays, or between X-rays and gamma radiation (Figure 26.3). wavelength 1 nm 100 nm gamma X-rays 400 nm UV visible blue frequency 760 nm red Figure 26.3 Part of the electromagnetic spectrum. 1000 nm 419 420 Unit 2 Module 3 Atomic and nuclear physics 238 U α 237 β Mass number, A γ Figure 26.4 The penetrating powers of radiation through the human body. Gamma emission generally occurs after alpha or beta emission. Its effect is to carry away excess energy from the excited nucleus. Gamma rays can be detected on a photographic plate, in a cloud chamber (by straight tracks spreading out from the gamma ray source) and by a GM tube. Cobalt-60 is a good laboratory source of gamma radiation. Summary Table 26.1 A comparison of alpha, beta and gamma radiations Radiation type Ionizing power Penetrating Typical Electric Absorbed by ... power range in air charge /m /e alpha (α) high low 0.05 +2 paper beta (β) medium medium 3 −1 ~5 mm aluminium gamma (γ) low high 100+ 0 ~1 cm lead halves intensity Equations of nuclear decay Alpha decay In an alpha decay, the atomic number, Z, of the nucleus goes down by two, and the mass number, A, goes down by four.

An example of alpha decay is the decay of uranium-238, producing thorium-234. 4 U → 234 90Th + 2α 238 92 In general for alpha decay, X→ 235 Th 234 89 90 91 92 93 Atomic number, Z Figure 26.5 Alpha decay of uranium-238 to thorium-234. Note that both the mass numbers and the atomic numbers balance on either side of the equation, A=A−4+4 Table 26.1 and Figure 26.4 provide a summary and comparison of these three types of radiation. A Z α 236 Y + 42α A−4 Z−2 ITQ 3 An experimenter observes that the radiation from a radioactive source is halved in intensity by a 2 mm aluminium sheet and is deflected towards a positively charged electrode. Which type of radiation is the source emitting? Z=Z−2+2 We can show this change in another way, by drawing a grid with mass number, A, vertically and atomic number, Z, horizontally. The original nucleus loses four units of mass and two units of positive charge (Figure 26.5). Beta decay A beta particle is an electron. Electrons do not exist inside the nucleus, but can be produced if a neutron changes into a proton by emitting an electron. n → 11p + −10e 1 0 The electron is emitted from the nucleus as a beta ray. The atomic number, Z, is increased by one (because a new proton has been created) but the mass number, A, remains unchanged. For example, when strontium-90 undergoes beta decay, the daughter nucleus becomes yttrium-90. 0 Sr → 90 39Y + −1e 90 38 Showing this change on a grid gives the diagram in Figure 26.6. In general for beta decay, X→ A Z Y + −10e A Z+1 ITQ 4 Do the mass numbers and atomic numbers balance for beta decay?

Chapter 26 Radioactivity more damaging than the same quantity of energy from gamma rays.

This is because alpha particles are absorbed in a thinner layer of material, and so their energy is concentrated into fewer cells, causing greater damage to each one. 92 91 Mass number, A Sr 90 β Y 89 88 37 38 39 40 Atomic number, Z Figure 26.6 Beta decay of strontium-90 to yttrium-90. Gamma emission Gamma rays are electromagnetic radiations, not material particles. When a nucleus emits a gamma ray, the nucleus keeps the same atomic number, Z, and the same mass number, A. The gamma ray only carries away energy and the nucleus becomes more stable. Gamma emission may occur following an alpha or beta emission in which the product nuclide is left in an excited (raised energy) state. The excess energy is emitted as a photon of electromagnetic radiation. Cobalt-60 is a common gamma-emitting nuclide. Co* → 60 27Co + γ 60 27 The SI unit of absorbed radiation dose is the gray (Gy); 1 Gy = 1 J kg−1 (1 joule per kilogram). To account for the difference in the genetic damage caused by different types of radiation the dose in grays is multiplied by a number called the quality factor (QF) of the radiation, to give the equivalent dose. The QF of X-rays, gamma rays and beta particles is 1, for alpha particles it is 20. The product of the quality factor and the dose in grays is the equivalent dose in sieverts (Sv). Tables 26.2 and 26.3 gives some sources and effects of acute (one-off) and long-term exposure to various radiation doses. Examples of single exposure would be as a result of a medical procedure, nuclear accident or explosion. Table 26.2 Radiation doses and their effects; single exposure Dose / mSv Source / effect 0.01 dental X-ray 0.02 chest X-ray 0.05 long haul airline flight (cosmic rays) 1.5 CT head scan 10 Threshold above which radiation sickness appears. The symptoms of radiation sickness include nausea, vomiting, skin burns and reduced white blood cell count. 5000 50% of those exposed will die within a month (* = in excited, unstable state) Radiation hazards Biological effects of nuclear radiation The energy of nuclear radiations damages any biological cells in which it is absorbed. High radiation doses kill cells; in humans this leads to radiation burns, acute radiation sickness and, if the dose is sufficient, death. Lower doses can cause genetic damage (damage to the cell DNA), which may lead to radiation-induced cancers and birth defects, sometimes years after exposure. The severity of genetic damage depends on both how much energy is absorbed per unit mass of the organism, and how localized the absorbed energy is; for a example, energy deposited by alpha particles is about 20 times ITQ 5 Write balanced equations for the following nuclear decays: (a) the alpha decay of polonium, 210 84Po to lead, Pb. (b) the beta decay of potassium, 40 19K to argon, Ar. whole body CT scan 1000 10 000+ always fatal within weeks or less Table 26.3 Radiation doses and their effects; long-term exposure Dose / mSv Source / effect per year 2 2–3 3 typical government exposure limit for members of the public to radiation from artificial sources typical dose from natural background radiation average exposure to radiation from medical imaging in the USA 9 average exposure of long haul airline crew 20 typical exposure limit for radiation workers 100 1000 lowest dose rate at which a long-term increase in cancer rates can be detected accumulated dose which causes a 5% increase in long-term cancer rate ITQ 6 Suggest why alpha emitters are regarded as less hazardous than beta and gamma emitters when they are outside the body, but much more hazardous if they are ingested or inhaled. 421 422 Unit 2 Module 3 Atomic and nuclear physics Background radiation We are continually exposed to naturally occurring radiations from outer space as well as from rocks, the air, building materials, our food and trace elements in our own bodies. The radiations coming from outer space are called cosmic rays. Cosmic rays – which are mainly high energy protons and, to a lesser extent, nuclei of atoms other than hydrogen that have been stripped of their electrons – account for about 16% of our annual natural radiation dose. In many places, the air contains radioactive radon gas. Radioactive radon is an alpha emitter, which is produced naturally from the decay of the element uranium present in rocks. Radon is the single biggest contributor to our natural radiation exposure. This gas cannot be seen, tasted or smelt. The annual intake of radon is increased in modern, draught-proof houses. The United States Environmental Protection Agency estimates that inhaled radon may contribute to 20 000 lung cancer deaths per year in America and kills 1 in 10 000 people per year worldwide. In some parts of the UK the government gives money for testing for, and making houses proof against, radon gas. Our food contains small amounts of natural radioactive materials. For example, many foods contain tiny amounts of radioactive potassium-40, lead-210 and polonium-210. These radioactive substances become concentrated in fish and other animals and are passed to us when we eat these animals as foods. the 1000 tonne concrete lid of the reactor was ripped off. Massive amounts of radionuclides from the reactor core were thrown into the air and, in time, were detected all over the northern hemisphere. This was not a nuclear explosion, but the polluting effects on the environment were severe. More recently, in 2011, the Fukushima Daiichi Nuclear Power Plant suffered major damage when a magnitude 9.0 earthquake and tsunami hit Japan. This resulted in the release of radionuclides, contaminating the surrounding area. Iodine-131 was released into the air. This is a beta and gamma emitter with a half-life of 8 days. It is anticipated that, in the long term, the exposure to this iodine isotope will have increased the incidence of thyroid cancer in the local population by 70%. Caesium137 was released into the ocean. This is a beta and gamma emitter with a half-life of 30 years. Caesium forms soluble salts in water and so can be absorbed by marine plants and animals. Some fish caught near the power plant a year after the disaster had caesium levels many times government safety levels. The fishing industry in the region has been seriously affected. Radiation biohazards in the Caribbean Human beings have evolved with the natural radiation background and scientists believe that our cells are adapted to repair the damage that low-level radiation exposure causes.

However, the extent to which the repair mechanisms are always successful and/or can be overwhelmed by increased radiation levels is not yet clear. The Fukushima disaster highlights the risks of building a nuclear power plant in a region of high seismic activity such as Japan or the Caribbean, where earthquakes and tsunamis are to be anticipated. The North East trade winds and strong ocean currents, characteristic of the Caribbean region, would transport radionuclides accidentally released into the environment over a wide area.

The release of a radionuclide such as caesium-137 into the Caribbean Sea could have a serious impact on the sensitive ecosystems of the region and on the fishing industry. Tourism would also suffer if tourists stayed away through fear of exposure to radiation. Man-made radiation hazards Handling and disposal of radioactive material Man-made radiation hazards include those we expose ourselves to voluntarily, such as medical scans and X-rays, and involuntary exposures such as the accidental contamination of the environment by radiation leaks from nuclear power plants and waste processing plants, or the deliberate testing of nuclear weapons. Radioactive materials are increasingly widely used in medicine, industry, education and even in our homes. Smoke alarms, for example, generally contain a small radioactive source. The disposal of radioactive waste materials from medical procedures and redundant equipment is a growing issue. There are detailed regulations for the safe use, handling, transportation, storage and disposal of radioactive sources.

Properly trained personnel must implement these regulations. Unless you are trained, you must not handle radioactive Well-documented incidents of nuclear accidents include the disaster in 1986 at Chernobyl, in what was then the USSR. During a test, the reactor cooling water boiled and the cooling system exploded. The force was such that Chapter 26 Radioactivity sources without direction from an expert. Some general principles, which must always be followed when handling small radioactive sources in the laboratory, are as follows: t=0s N = 200 a b decayed nucleus t=1s N = 180 ■ never eat or drink in the laboratory ■ minimize the time spent near any artificial radiation source ■ keep an adequate distance from such sources (the radiation dose received from a source diminishes with your distance from it) ■ always use suitable shielding as instructed by trained personnel ■ do not handle sources with your fingers – use tongs ■ remove sources from their shielding only when required, and replace them immediately in their shielded box when an experiment is completed t=2s N = 162 c d t=3s N = 146 ■ always wash your hand after working in the laboratory. The handling and storage of radioactive waste from nuclear power plants is a highly technical process. Spent nuclear fuel rods can be placed into a pond for storage underwater before reprocessing. After plutonium and other valuable radionuclides are separated from the used fuel, the remaining waste is still highly radioactive.

It is formed into a stable glass and sealed in drums, which are then stored underground, potentially for thousands of years. Radioactive decay laws Figure 26.7 A sample of 200 radioactive nuclei. Modelling radioactive decay Figure 26.7 shows a model representing 200 radioactive nuclei at time t = 0 s. Suppose the probability of a decay for this nuclide is 0.1 per second. This means that, on average one tenth of the nuclei will decay in the next second. We cannot say which ones these will be, but we can predict that since that are 200 nuclei in the sample, 0.1 × 200 = 20 will decay in the following second. t / s N (number remaining) 0 200 1 180 2 162 ITQ 7 Copy and complete Table 26.4, rounding N to the nearest whole number at each stage. (Note: in practice, since radioactive decay is a random process, the actual number that decay may be more or less than 20 for by chance; but on average the sample will follow the decay law that we predict.) 3 Plot a graph of N against t.

From your graph determine the time taken for the number of nuclides to decrease: 4 (a) from 200 to 100 5 (b) from 100 to 50 6 (c) from 50 to 25. Figure 26.7b shows the situation at time t = 1 s. There are now 180 radioactive atoms remaining. In the following second, one tenth of these (0.1 × 180 = 18) will decay on average, leaving 162 at t = 2 s (Figure 26.7c). At t = 3 s there will be 162 − 0.1 × 162 = 146 remaining (Figure 26.7d). 8 9 10 11 12 13 14 Table 26.4 Radioactive decay with decay probability 0.1 per second 423 Unit 2 Module 3 Atomic and nuclear physics Table 26.5 Half-lives of some nuclides 1200 Nuclide 1000 Number of nuclei, N 424 800 dN at t = 5 s dt 600 400 dN at t = 25 s dt 200 0 0 10 20 30 40 Half-life, T 12 Notes H-3 12.26 year hydrogen isotope called tritium C-14 5730 year hard to study K-40 1.26 × 10 year found in the human body Co-60 5.24 year a common source Sr-90 27.7 year a common laboratory β source I-131 8.05 day used in medicine Cs-137 30.0 year spread by the Chernobyl explosion Rn-220 54 s causes lung cancer, especially in the UK and USA U-238 4.5 × 10 year the common isotope of uranium Pu-240 6580 year used in nuclear weapons 9 9 50 Time, t / s Figure 26.8 The rate of radioactive decay. The shape of the graph of N against t is an example of exponential decay.

The magnitude of its gradient, dN/dt, at any instant is the rate of decay (the number of decays per second) at that instant. This decreases as the number of radioactive nuclides decreases. ITQ 7 demonstrates that the rate of radioactive decay (the number of decays per second) is proportional to the number of nuclei present (Figure 26.8). If the number of nuclei present is N then we can write, rate of decay ∝ N and so, rate of decay = λN where λ is the probability per second of a single nuclide undergoing decay; λ is called the decay constant. Each type of radioactive nuclide has a different value for λ.

A special feature of exponential decay is that the time taken for the number of nuclides to decrease by half (from 200 to 100, or 100 to 50 for example) is a constant. This is called the half-life for the decay and is given the symbol T 1 . 2 The half-life of a nuclide is characteristic of that nuclide. It is governed by the structure of the nucleus. It is not affected by temperature, pressure, or any physical condition. Half-lives vary from many years to a fraction of a second. Some half-lives are given in Table 26.5. Activity The rate of radioactive decay (the number of decays per second) is called the activity A of the source. For a source ITQ 8 What is the half-life of the sample in ITQ 7? Worked example 26.1 Q A sample contains 1 000 000 atoms of iodine-131.

How many atoms of iodine-131 remain after 8.05 days, 16.1 days and 40.25 days? A 8.05 days is one half-life for this nuclide, 16.1 days is two half-lives, 40.25 days is five half-lives. In general the average number remaining after n half-lives will be 1n of the initial 2 number. Thus, ■ after 1 half-life, average number remaining = 1 000 000 × 11 = 500 000 2 ■ after 2 half-lives, average number remaining = 1 000 000 × 12 = 250 000 2 ■ after 5 half-lives, average number remaining = 1 000 000 × 15 = 31 250 2 with N active nuclei the rate of decay is equal to λN, where λ is the decay constant. Thus, A = λN ....(26.1) The SI unit of activity is the becquerel (Bq).

A source has an activity of 1 Bq if its average rate of decay is one disintegration per second. Another unit of activity is in common use. This is the curie (Ci). A source has an activity of 1 Ci if its rate of decay is 3.7 × 1010 disintegrations per second (this is approximately the activity of 1 g of radium-226). A typical radioactive source used in a school laboratory might have an activity of 5 μCi. However, a hospital radiotherapy machine might contain 1000 Ci of caesium137 or cobalt-60. Unshielded exposure of just a few minutes to radiation from such a powerful source will produce potentially fatal radiation sickness. Chapter 26 Radioactivity Rearranging we have, Worked example 26.2 Q ■ decay constant of U-238 = 4.9 × 10−18 s−1 ■ relative isotopic mass of U-238 = 238.05 ■ Avogadro constant = 6.022 × 1023 Calculate: (a) the activity of a sample containing 1021 U-238 nuclei (b) the activity of 238.05 g of U-238 (c) the activity of 2 g of U-238 A (a) A = λN = 4.9 × 10 × 10 = 4.9 × 10 Bq (b) From the definition of the Avogadro constant, 238.05 g of U-238 contains 6.022 × 1023 atoms. Therefore A = λN = 4.9 × 10−18 × 6.022 × 1023 = 2.95 × 106 Bq (c) activity of 238.05 g = 2.95 × 106 Bq, therefore activity of 2 g is given by, A = 2 × 2.95 × 106 = 2.48 × 104 Bq 238.05 –18 21 3 δN = −λN δt or in the limit (using calculus) as δt tends to zero, dN = −λN dt We can use integration to solve this equation to find an equation which gives N as a function of time. Rearranging the equation above we have, dN = −λdt N Therefore, N dN ∫N 0 N =− ∫ 0 λ dt t where N0 is the number of active nuclei at time t = 0 and N is the number remaining after time t. Performing the integrations gives, ln N − ln N0 = −λt Getting it right! The Avogadro constant and the mole The mole is defined as the amount of substance that contains the same number of particles as there are atoms in 12 grams of carbon-12; this number is the Avogadro constant, 6.022 141 29 × 1023. The mass of 1 mole of a nuclide is equal to its relative isotopic mass expressed in grams. The relative isotopic mass of U-238 is 238.05, and therefore 238.05g of U-238 contains the 1 × Avogadro number nuclei; 1 g will contain 238.05 23 6.022 × 10 nuclei. The relative isotopic mass (238.05) differs very slightly from the mass number (238) due to nuclear binding energy (this is discussed in the next chapter); assuming that the relative isotopic mass and the mass number are equal will give sufficient accuracy for most calculations of radionuclide activity. Decay law The number of radioactive nuclei that decay per second is equal to λN.

The number of decays in a given time must be the same as the number by which N decreases in that time. Let δN be the change in N in the time interval δt, then, δN = −λNδt δN is negative because N is decreasing. where ln N is the natural logarithm (log to base e) of N. Therefore, N = exp(−λt) N0 And so we have the exponential decay law, N = N0 exp(−λt) ....(26.2) Worked example 26.3 contains a calculation using this equation. Decay constant and half-life As we have seen, the time T for the activity of a source to 2 reduce by half is described as the 'half-life'. When N = T 1 , N = 0 and so, 2 2 N0 = N0 exp(−λT 1) 2 2 exp(−λT 1) = 2 Taking reciprocals, exp(λT 1) = 2 2 Taking natural logarithms of both sides, λT 1 = ln 2 2 Therefore, T1 = 2 ITQ 9 What are the activities of a 5 μCi source and a 1000 Ci source expressed in becquerels? 1 2 ln 2 0.693 = λ λ ....(26.3) Worked example 26.4 contains a calculation using this equation. 425 426 Unit 2 Module 3 Atomic and nuclear physics Worked example 26.3 Worked example 26.4 Q The decay constant of caesium-137 is 7.3 × 10−10 s−1. In an accident at a nuclear power plant a quantity of this radionuclide is released into the environment. Calculate the time in years for the activity of this material to decrease to 1% of its initial value. Q A Activity, A = λN, therefore (since λ is constant), A ∝ N. Let t be the time for the activity to decay to 1% of the initial activity, then, N = exp(−λt) = 0.01 N0 Taking natural logarithms of both sides, ln 0.01 = −λt − In 0.01 4.61 t= = = 6.3 × 109 s = 200 years 7.3 × 10−10 λ The half-life of cobalt-60 is 5.24 years. (a) Calculate the initial activity in curies of a 1 g source of this radionuclide. (b) The source in part (a) has been installed in a radiotherapy machine for 10 years; by what percentage has its activity decreased? relative isotopic mass of cobalt-60 = 59.93 Avogadro's constant = 6.022 × 1023 A (a) T 1 = 5.24 years = 5.24 × 365 × 24 × 3600 s = 1.65 × 108 s 2 0.693 0.693 = = 4.19 × 10−9 s−1 λ= T1 1.65 × 108 Applications of radioisotopes The radiations emitted from radionuclides have many applications. Such uses are usually based on the penetrating power of the radiation, its ability to kill living cells, and the energy carried by the radiations. Medical uses In medicine, radioisotopes are used both as diagnostic tools to identify the causes of disease and as therapeutic tools to treat disease. Technetium-99 is the most widely used nuclide for diagnosis. It has a short half-life (6 hours) and emits relatively low-energy gamma rays (140 keV). It can be injected into the body as part of various chemical compounds designed to concentrate in specific organs or potential tumours, for example in the lungs or the brain. The gamma rays emitted by the technetium are detected outside the body with a gamma ray camera, which produces an image of the area under investigation. 2 Since 59.93 g of Co-60 contain 6.022 × 1023 atoms, the number in 1 g is given by, 1 N= × 6.022 × 1023 = 1.005 × 1022 59.93 A = λN = 4.19 × 10−9 × 1.005 × 1022 = 4.21 × 1013 Bq Now 1 Ci = 3.7 × 1010 Bq and therefore the activity of the source in Ci is given by, 4.21 × 1013 A= = 1137 Ci 3.7 × 1010 (b) Activity is proportional to N. Using, N = exp(−λt) N0 N = exp(−4.19 × 10−9 × 10 × 365 × 24 × 3600) = exp(−1.32) = 0.27 N0 The activity is therefore 27% of its initial value and has decreased by 73%. Alternative solution to part (b): 1 In n half-lives the activity reduces to a fraction 2n of the initial activity.

10 10 years is equal to 5.24 half-lives = 1.91 half lives. 1 = 0.27 = 27% of the Therefore the activity is reduced to 1.91 2 initial activity. The activity has decreased by 73%. Radionuclide therapy (RNT) involves embedding a small source – usually a beta or gamma emitter – in the target area. For example iodine-131 is embedded to treat thyroid cancer by irradiating the cancerous cells; this is one of the most successful of all cancer treatments.

This type of treatment gives a smaller overall dose to the body than external radiation and therefore has fewer side-effects. An ideal therapeutic isotope is a strong beta emitter, but which also emits gamma rays so enabling imaging of the treatment site.

Lutetium-177 and yttrium-90 are the most significant RNT nuclides.
ITQ 10 How does the activity of a radioactive source change over time?
ITQ 11 What characteristics of technetium make it suitable for use in medical diagnosis? For radiotherapy, a radionuclide may either be inserted into the body to localize its effect on a tumour, or radiation from a radionuclide located externally may be directed onto a tumour. Chapter 26 Radioactivity External irradiation of tumours may carried out with gamma rays from a cobalt-60 or caesium-137 source; however, high-energy X-ray sources called LINACs (see Chapter 25) are now more commonly used to produce the required radiation. Gamma sterilization The preparation of medical supplies has made increasing use of radiation for sterilization. Packages, like those containing syringes and needles, are irradiated by gamma rays. All the germs present are killed. This eliminates the need for heavy and cumbersome sterilization apparatus. Radiation is also used in food preservation. For example, freshly caught shrimps have an ice-storage life of about 15 days. After being exposed to gamma rays, they are still in good condition after 7 weeks in ice. Carbon dating Because radionuclides decay at known rates, and some have long half-lives, they can be used to pinpoint the age of ancient materials. The Earth's atmosphere is 0.04% carbon dioxide. About 1 in 109 of the carbon atoms in atmospheric CO2 is the radioactive isotope C-14. This fraction remains constant over time as a result of production of C-14 from N-14 through nuclear reactions produced by cosmic rays. The half-life of carbon-14 is 5730 ± 40 years.
Carbon-14 decays into nitrogen-14 through beta decay. A steady state exists in which the rate of C-14 production is exactly balanced by the rate of decay. Plants take carbon dioxide from the atmosphere through photosynthesis. This process starts the food chains that provide the carbon used by all living things to build their bodies. When a plant or animal dies, it stops taking in carbon-14. Its existing carbon-14 decays with a half-life of 5730 years. Analysis of the carbon-14 activity of a small quantity of material from a specimen that was once part of a living thing (a bone or a wooden handle, for example) reveals the age of the material, provided it is less than about 60 000 years old. For example, a human body was found buried in ice in Northern Europe in 1991. At first it was thought that the body was only 200 years old, but radiocarbon dating showed that the man died 5300 years ago. Tracers in agriculture In agriculture, radionuclides are used as tracers to study plants, insects and other animals.
For example, phosphorus-32, a beta emitter, can be added to plant fertilizer.
The phosphorus is taken up by the plant and its distribution can be measured by recording the beta radiation from the various parts of the plant.
Radiation has been used to control the screw-worm fly pest in South America. A large number of the male of the species were exposed to gamma radiation. When the males were released back into the wild and mated with wild females, sterile eggs resulted in no new flies being born. Detectors of radiation Photographic plates Scientists working with radiation are required to wear a sealed badge containing a photographic film as a personal radiation monitor. The film is replaced and developed at regular intervals of time and the amount of blackening of the film indicates the level of exposure.
If the level is too high, the worker must be checked by a doctor and the area they worked in checked for leakage of radiation.
There is a limit as to how much radiation a worker can safely receive during a year. Ionization detectors The three radiations we have studied (alpha, beta and gamma) all deposit energy in matter as they pass through it. This energy may eject electrons from atoms, producing ionization. The ionization may in turn be detected by the resulting changes in electrical conductivity, or the flashes of light emitted when electrons and ions recombine. The detectors described below use ionization in various ways to detect and measure the radiations. Spark counter When nuclear radiations pass through the gas between a positively charged wire grid and a negatively charged metal plate, ionization occurs. The ions and electrons metal plate spark gap fine wire very high voltage insulator ITQ 12 Suggest why radionuclides are being replaced by X-ray sources to provide radiation for cancer treatment. Figure 26.9 A simple spark counter. 427 428 Unit 2 Module 3 Atomic and nuclear physics produced enable a large current to suddenly pass through the air under the high voltage between the grid and the plate (Figure 26.9). A spark can be seen and heard or it may be registered by an electronic device (such as a scaler). Cloud chamber ON OFF 2 1 3 0 9 START STOP counter RESET TUBE VOLTAGE A diffusion cloud chamber reveals the tracks of charged particles (α and β radiations). The base of the chamber is cooled by dry ice to about −80 °C.
A felt ring inside the top of the chamber is moistened with alcohol (Figure 26.10). The alcohol vapour diffuses downwards, becoming cooled and ready to condense. Each time a particle is emitted from a radioactive source, it produces ions along its path and the alcohol vapour then condenses around these ions. The condensed alcohol droplets reflect light and so can be seen as narrow white lines against a black background. G-M tube thin mica window argon gas at low pressure incoming particle ionizes gas The cloud chamber provides evidence that something is being emitted from radioactive materials. Cloud chamber photographs do not show the actual radiation, but only the alcohol droplets, which form on the ions produced by the radiation. radioactive source felt ring soaked in alcohol + insulator – high-voltage supply unit to counter Figure 26.11 A Geiger–Müller (GM) counter. The Geiger–Müller (GM) tube (Figure 26.11) is one of the most important instruments for detecting radiations. The metal tube walls are earthed. A thin mica window at one end allows radiations to enter the tube. The tube is filled with argon gas at low pressure. A high-voltage power supply maintains a large potential difference between a thin wire that runs along the axis of the tube and the tube walls. When nuclear radiation enters the GM tube, argon atoms are ionized. The ions and electrons produced are attracted to the tube walls and central wire respectively. This results in a sudden large current pulse through the tube and the external circuit. This pulse can be detected by a scaler (counter) or ratemeter.
A scaler records the total number of counts, whereas a ratemeter records the number of pulses or counts per second. transparent plastic lid vapour felt ring soaked with alcohol radium source wire black floor light 'dry ice' foam cushion The combination of a GM tube and a scaler or ratemeter is usually referred to as a Geiger–Müller counter. GM counters are simple and, with care, long lasting. They have the disadvantage that a given radiation either triggers the tube or it does not. The response does not depend on the energy of the radiation. However, a GM tube is not very efficient. It records about 10% of beta radiation falling on it and less than 1% of gamma radiation. Scintillation counter One of the early methods of detecting radioactivity was with the use of a fluorescent screen in conjunction with a microscope.
When nuclear radiation strikes the screen, light spots or scintillations are produced. The modern scintillation counter consists of a fluorescent crystal placed in contact with a photomultiplier. When radiation strikes the crystal, light is emitted. The light is detected by the photomultiplier and amplified by its electronic circuit. Such a counter is much more sensitive than a simple fluorescent screen. It has the advantage that the intensity of the light pulse that it produces depends on the energy deposited in the crystal. insulation Semiconductor detectors removable base Figure 26.10 A diffusion cloud chamber. levelling wedges These small devices are basically reversed biased p-n junctions (diodes). The current through a reverse biased diode is normally small, but when ionizing radiation passes through it, a large number of electron-hole pairs are created in the vicinity of the junction.
This produces a pulse of current, which can be counted electronically. The size of the pulse is proportional to the energy deposited in the detector by the radiation.
One use of semiconductor detectors is in alpha particle spectroscopy – the measurement of the spectrum (range) of energies of the alpha particles from a source.
The detector's small size and thin construction enable it to be placed close to the source, ensuring that all the alpha particle energy is deposited in its active layer. Getting it right! Deducting the background A key process when using any radiation detector is to record the background radiation. Background radiation will be present throughout any experiment and will result in a systematic error if it is not accounted for. The background count must be deducted from each reading. Summary ✓ Unstable nuclei decay to more stable states by emitting alpha, beta and/or gamma radiation. ✓ Alpha particles are the nuclei of helium atoms 4 2 He, beta particles are electrons, gamma rays are short-wavelength photons of electromagnetic radiation. ✓ Deflection by electric and magnetic fields demonstrates that alpha particles carry a positive charge, beta particles are negatively charged, gamma rays carry no charge. ✓ Alpha particles are absorbed by thick paper, beta particles are absorbed by a few mm of aluminium, a beam of gamma rays is halved in intensity by ~1 cm of lead (depending on the gamma ray energy). ✓ Radioactive decay is a random process; it is not possible to say when a particular nucleus will decay. ✓ The decay constant λ is the probability per second that a given nucleus will decay; this is characteristic of each type of radionuclide, and it not affected by temperature or pressure.
✓ The activity A is the number of decays per second of a sample; for a sample of N nuclei it is given by A = dN = λN. dt ✓ A rate of decay of 1 Bq = 1 decay per second; 1 Ci = 3.7 × 1010 decays per second. ✓ The half-life is the time for the number of radioactive nuclei, and therefore the activity, to decrease by half; T 1 = ln 2 = 0.693 λ 2 λ ✓ In general, the decay law is given by N = N0 exp(−λt) ✓ Nuclear radiation is hazardous; exposure to excessive doses of radiation can cause cancer, birth defects, radiation sickness and death. ✓ 1 gray (1 Gy) is a radiation dose of 1 J kg−1. The dose in sieverts (Sv) is equal to the dose in grays multiplied by the quality factor (QF) of the radiation. ✓ A typical annual dose from background radiation is 2–3 mSv. ✓ A typical government limit for annual exposure of the general population to radiation from artificial sources is 2 mSv. ✓ Nuclear radiation is used in medicine for diagnosis and radiotherapy. ✓ An isotope used internally for diagnosis should be a low-energy gamma emitter with a short half-life. ✓ An isotope used for internal radiotherapy should be a short half-life beta emitter and ideally also a gamma emitter; the beta radiation is absorbed at the treatment site and the gamma radiation allows external imaging. ✓ Radiation detectors include spark counters, Geiger–Müller tubes, cloud chambers, scintillation counters and solid-state detectors. ✓ When recording count rates from radiation sources with a detector, the background count must always be recorded and subtracted from the total count. 429 430 Unit 2 Module 3 Atomic and nuclear physics Review questions (c) What are the values of p and q in the nuclear equation given below? Avogadro constant = 6.022 × 1023 X → 230 90Y + 2α p q 1 Ci = 3.7 × 1010 Bq Nuclear instability, Types of radiations and their properties 1 How is it possible for an atomic nucleus to contain only positively charged protons and neutrons with no charge? Why don't the protons fly apart as a result of the repulsion between their like charges? 2 Your friend says that he heard a super heavy new element had been discovered with atomic number 1000. Explain briefly why this report is almost certainly untrue. 3 (a) Copy and complete Figure 26.12a to show how the particles and rays are deflected and at which material each of them is practically stopped. (b) Figure 26.12b shows the deflection of radiations from a radioactive source by a uniform magnetic field. Explain two factors that bring about the differences in the deflections. a electrically charged plates thin sheet of aluminium + – source of α, β and γ rays thin paper thick lead sheet b Radiation hazards, Radioactive decay laws 6 (a) Discuss briefly the biological effects of nuclear radiation. (b) What are the natural and artificial sources of nuclear radiation to which human beings are exposed? (c) Why are exposure limits for the general public set so far below the single radiation dose that causes radiation sickness? 7 Use the internet to research proposals to build nuclear power stations in the Caribbean region. Write a report comparing the potential environmental advantages and disadvantages of nuclear power for the Caribbean. 8 (a) The gaseous radionuclide radon-222 is produced by the decay of uranium and thorium in rocks. Radon-222 is an alpha emitter with a half-life of 3.8 days. Discuss why, as a gas, radon is more hazardous to health than the uranium and thorium, which are also alpha emitters. (b) (i) Calculate the decay constant of radon-222. (ii) From your answer to part (i), find the number of radon-222 nuclei present in a freshly prepared 2 mg sample that decay per second. (iii) How many days would it take for the number of radon-222 nuclei in the sample to decline to 0.1% of the initial number? 9 The table below gives the activity of a radionuclide measured over a period of 60 minutes with a radiation detector connected to a ratemeter. The background count rate was recorded as 37 counts per minute.
Figure 26.12 Equations of nuclear decay 4 What type of radiation is emitted when polonium-218 decays to astatine? Complete the decay equation. Po → 218 84 218 85 At + X 5 (a) An isotope of nitrogen can be represented as 147N. What is the significance of each of the numbers 14 and 7? (b) Find a and b in the following radioactive decay equation: X → abY + −10e + γ 214 82 Time / min Count rate / min−1 Corrected count rate / min−1 0 9970 5 881 10 772 15 670 20 592 25 523 30 450 35 404 40 350 45 319 50 279 55 243 60 220 Chapter 26 Radioactivity (a) Correct the data for the background, then plot a suitable graph from which you can determine the half-life of the nuclide. (b) Use your value for the half-life to calculate the decay constant. (c) Assuming that detector records 2% of the decays undergone by the sample, calculate the number of active nuclei present in the sample at the start of the experiment. 10 (a) Define the decay constant, activity and the half-life of a sample of a radionuclide. A nuclide, F, has a half-life of 2.5 hours. (b) What percentage of the original number of atoms of the isotope would be left after 10.0 hours? (c) How have the decay constant, activity and half-life of the sample changed in that time? 11 (a) Explain what is meant by the spontaneous nature of radioactive decay. (b) The radioactive isotope 232 90Th decays... radium nucleus of radioactive decay. (c) A sample of a certain nuclide, which has a half-life of 1500 years, has an activity of 32 000 counts per hour at the present time. (i) Sketch a graph of the activity of this sample over the period in which it will reduce to one-sixteenth of its present value. (ii) If the sample of the nuclide could be left for 2000 years, what would be the activity then? 12 The activity of a Cs-137 source at a particular time is 1 Ci. Cs-137 has a half-life of 30 years. (a) Express the activity of the source in becquerel (Bq). (b) Calculate the number of Cs-137 atoms in the source. (c) If the energy released by each Cs-137 decay is 1.2 MeV, calculate the rate of energy release by the source in watts. Applications of radioisotopes, Detectors of radiation 13 A 2 g sample of carbon taken from a leather shoe found in an ancient burial site has a C-14 activity of 21 disintegrations per minute. Living plants have a C-14 activity of 15.3 disintegrations per minute per 1.00 g of carbon they contain. The half-life of C-14 is 5730 years. Estimate the age of the leather from which the shoe is made. 14 The nuclide U-238 decays through a series of disintegrations to eventually produce the stable nuclide Pb-206. The effective half-life for this process is 1.4 × 1017 s. Chemical analysis of a rock sample reveals that the ratio of the number of Pb-206 atoms to U-238 atoms is 1 : 6. Assuming that the rock initially contained no Pb-206, estimate the age of the rock in years. 15 A small quantity of a solution containing a radioactive tracer is injected into the bloodstream of a patient. The tracer has a half-life of 15 hours. The initial activity of the solution was 300 Bq. After 30 hours the activity of 1 cm3 of blood taken from the patient was found to be 0.016 Bq. Find the patient's blood volume. 16 At time t = 0 the count rate recorded from a sample of a radionuclide is 786 counts per minute. After 48 minutes the count rate had dropped to 136 counts per minute. When the sample was completely shielded from the detector, the count rate fell to 46 counts per minute. What is the half-life of the sample? Practical exercises and challenges 1 Radioactive decay simulation You will need: ■ 500 small dice or small wood/plastic cubes with one face marked ■ a large tray 1 Throw the set of dice/cubes into the tray and spread them into a layer one cube deep. 2 Dice showing 6 on the top face (or cubes with the marked face up) have 'decayed'. Remove them from the set, counting them as you do so. Record the number removed and the number remaining, N, after the first trial. 3 Repeat the trial with the remaining dice. Record the number of decays, the number remaining and the trial number (n = 2). 4 Continue this process – recording the number of decays, the number remaining and the trial number each time – until less than 10 dice remain. 5 Plot a graph of the number of dice remaining, N, against the trial number, n. (a) What would you predict for the shape of this graph? Use your graph, determine a value for the half-life of the dice in the simulation (in this case the units of half-life will be 'trials', not seconds). 431 432 Unit 2 Module 3 Atomic and nuclear physics (c) The equation of your graph is given by the decay equation N = N0 exp(−λn), where N is the number remaining after n trials. Taking natural logarithms of both sides we can obtain a linear relationship between ln N and n, ln N = lnN0 − λn. From the gradient calculate λ. From the intercept with the y axis find N0. Discuss how the values you obtain from the graph compare to the values you would predict for this simulation. (d) Use the relationship T 1 = 1 ln 2 to calculate the half-life from the gradient determined λ 2 in part (c). Which half-life would you expect to be more accurate: the one read from the graph in part (b) or the one determined from the gradient measured in part (c)? Explain your conclusion. 2 Measurement of the half-life of a radioisotope with a short half-life You will need: ■ the apparatus shown in Figure 26.13 ■ a stopwatch Follow safety procedures as directed by your teacher. clip valves ratemeter G-M tube squeezy polythene bottle radon gas and air thorium hydroxide powder ON OFF Figure 26.13 Apparatus for the determination of the half-life of radon-220. Radon-220 is a radioactive gas with a short half-life. It is constantly replenished in the source bottle by the decay of thorium.
The detector is connected to a ratemeter, which records the number of counts per second. The background count rate is recorded. 1 A sample of radon-220 is puffed into the sealed flask. 2 Count rate and time are recorded every 20–30 seconds until the count rate returns to the background level. 3 The corrected count rate, R, is proportional to N, the number of radon-220 atoms in the sample at time t. R = kN = kN0 exp(−λt) where k is a constant. Taking natural logarithms, ln R = ln (k N0) − λt 4 A graph of ln R against t is thus a straight line with gradient − λ. Since T 1 = 1 ln 2, the half-life can be calculated from the gradient of the graph. 2 Chapter 26 Radioactivity Answers to ITQs Answers to Review questions 1 The strong nuclear force holds the nucleons together, overcoming the electrostatic repulsion between the protons. 1 2 The attractive strong nuclear force between nucleons is about 100 times stronger than the electromagnetic repulsion between protons. The attraction is short range, but the repulsion is long range, and a proton is affected by the repulsion of all the other protons in the nucleus.
When the (atomic (proton) number) reaches about 100 the repulsive electromagnetic force on a proton is sufficient to overcome the short-range nuclear force. 3 Beta particles. 4 Yes: A = A + 0; Z = Z + 1 − 1 210 206 4 5 (a) 84Po → 82Pb + 2α 40 0 (b) 40 19K → 20Ar + −1e 6 When the source is outside the body alpha particles are significantly absorbed by the air, clothing and skin before damaging internal body cells. Inside the body, tissues that are particular sensitive to damage such as the lining of the lungs are exposed directly to the alpha radiation. 7 (a) 6.9 s (b) 6.9 s (c) 6.9 s 6.8 6.9 s 95 μCi = 5 × 10−6 Ci = 5 × 10−6 × 3.7 × 1010 Bq = 1.9 × 105 Bq 1000 Ci = 1000 × 3.7 × 1010 Bq = 3.7 × 1013 Bq 10 Since A = λN, the activity A is proportional to N.
Therefore the activity is in proportion to N. The activity is halved in each successive half-life. 11 It is a low-energy gamma emitter with a relatively short half-life. Since gamma rays are very penetrating they can be detected outside the body to produce an image of the feature under investigation. The short half-life means that it will not be active in the body for too long a period. 12 X-ray sources are more easily controlled than radionuclide sources and do not require shielding when not in use. 0 − 1 e 2 (b) a = 214, b = 83 (c) p = 238, q = 86 8 (b) (i) 2.1 × 10−6 s (ii) 5.4 × 1018 s−1 (iii) 38 days 9 (a) 25 min = 1.5 × 103 s (b) 0.0276 min−1 = 4.6 × 10−4 s−1 (c) 1.7 × 106 10 6.25% 11 (c) (ii) 12700 min−1 12 (a) 3.7 × 1010 Bq (b) 5.1 × 1019 (c) 7.1 mW 13 3100 years 14 9.9 × 108 years 15 4.7 litres 16 16 minutes 433 434 Chapter 27 Energy from the nucleus Learning objectives ■ Define and calculate mass defect and binding energy. ■ Use the relationship between energy and mass in nuclear reactions; E = mc2.
■ Calculate the energy released in nuclear fission, nuclear fusion and nuclear decay. ■ Use the atomic mass unit (u) as a unit of energy. ■ Represent graphically the relationship between binding energy per nucleon and nucleon number. ■ Explain the relationship between binding energy per nucleon and nuclear fission and fusion. ■ Demonstrate that nucleon number, energy (mass) and charge are all conserved in nuclear processes. 1 2 3 ■ Represent and interpret nuclear reactions in the form 1H + 1H = 2He. Introduction to nuclear energy In the 19th century the source of the energy that powers the stars, including our Sun, was a mystery.
Calculations by Hermann von Helmholtz (a German physician and physicist) and by Lord Kelvin showed that known energy sources could not have powered the Sun long enough for continents to form and life to evolve on Earth. If the Sun was a ball of burning coal or oil, for example, its energy would have been exhausted in a few thousand years; gravitational energy, released as the matter that formed the solar system collapsed towards a central point, could have sustained the Sun's output for up to 60 million years, but this was still much less than the age of the Earth as indicated by its geology and biology.
The discovery of radioactivity gave the first clue to the nature of the Sun's energy source. Rutherford suggested that nuclear processes, such as the decay of radium and other radioisotopes, were capable of providing sufficient energy. We now know that radioactive decay inside the Earth contributes to the heat that keeps parts of the mantle and core in a fluid state, powering movements of the Earth's plates and volcanism; but radioactivity is not the energy source of the stars. Stars are powered not by nuclear decay, but by nuclear fusion – a process in which light nuclei combine to form heavier ones. Full understanding of the physics that makes the stars shine had to await the work of Albert Einstein and his famous equation: E = mc2. Figure 27.1 A nuclear explosion. The forces that hold the nucleus together are orders of magnitude greater than those that hold atoms together in chemical compounds.
The energy released in a nuclear reaction is correspondingly greater than the energy released in a chemical reaction. The explosion of a nuclear weapon containing just 6 kg of plutonium releases as much energy as the explosion of 20 000 tonnes of TNT (a chemical explosive); it causes terrible devastation. With his theory of relativity, Einstein showed that mass and energy are equivalent. In any process we should not talk separately about the conservation of energy and the conservation of mass; rather the appropriate conservation principle is the conservation of mass–energy.
When energy is transferred from one object to another, the object that gains energy gains an equivalent amount of mass; the object that loses energy loses an equivalent amount of mass. The total mass–energy of an isolated system is conserved. Chapter 27 Energy from the nucleus Worked example 27.1 Worked example 27.2 Q A 1 kg mass is raised through a height of 5 m near the Earth's surface. Calculate: (a) the energy transferred to the mass (b) the change in the mass of the object. Q A (a) Increase in gravitational potential energy, ΔE, is given by ΔE = mgh = 1 kg × 9.8 m s−2 × 5 m = 49 J (b) Let Δm be the corresponding increase in mass. ΔE = Δmc 2 ΔE Δm = 2 c 49 = (3 × 108)2 Again consider a stationary 1 kg mass. Suppose that nuclear processes transfer 0.1% of the mass–energy of the sample to (a) the kinetic energy of the resulting particles (perhaps in an explosion). Assume that initially there is no energy loss to the surroundings. Calculate: (a) the kinetic energy released (b) the total rest mass of all the particles that remain after the explosion. A (a) Δm = 1 g = 0.001 kg ΔE = Δmc2 = 0.001 × (3 × 108)2 = 9 × 1014 J (This is as much energy as there is chemical energy in 20 million litres of gasoline!) (b) mass equivalent of kinetic energy released = 0.001 kg Therefore new rest mass of sample = 0.999 kg = 5.4 × 10−15 kg The concept of mass–energy equivalence allows us to convert between mass units and energy units.
The conversion factor, c2, is the speed of light squared. In a process that involves an energy change (ΔE) as a result of a transfer of energy into or out of a system, there will be an equivalent mass change (Δm), which may be calculated from the equation ΔE = Δmc2 ....(27.1) In Worked example 27.1 the change of mass is extremely small. In general, for processes involving objects at the laboratory scale and the forces of gravity and/or electromagnetism only, the mass changes predicted by E = mc2 are so small as to be undetectable. Nuclear reactions are different: because the energy changes in nuclear processes are so much greater (millions of times those of chemical reactions), the mass changes can be detected and measured, and may be used to calculate the energies involved, as in Worked example 27.2. Getting it right! Rest mass Einstein's equation tells us that mass and energy are equivalent. Initially the sample in Worked example 27.2 is stationary – we say that its rest mass is 1 kg. One gram of this rest mass is then transferred to kinetic energy.
The total mass–energy of the sample is conserved and so is still 1 kg, but now there is 0.999 kg of rest mass and 0.001 kg of kinetic energy: if all the particles were brought to rest (by transferring their kinetic energy to the surroundings) their rest mass would be found to be 0.999 kg. Mass defect and binding energy Binding energy You are 'bound' to the surface of the Earth by the gravitational force. In order to escape from the Earth's surface you must be supplied with energy – from the burning fuel in a rocket engine for example. The energy required to completely separate two objects held together by an attractive force is called the binding energy. It is equal to the work that must be done to separate the objects against the attractive force that acts between them. The binding energy of a nucleus is the energy required to separate it into individual nucleons (protons and neutrons) at rest. It is therefore also equal to the energy released when initially stationary nucleons are brought together from infinity to form the nucleus. The greater the average binding energy per nucleon, the more stable the nucleus. Mass defect Atomic masses can be measured very accurately with a mass spectrometer. The nuclear mass is found by subtracting the masses of the orbital electrons. These measurements show that nuclear masses are not exactly equal to the sum of the rest masses of the constituent protons and neutrons. They are (with the exception of ITQ 1 Is the total energy of a nucleus more or less than the energy of the same number of neutrons and protons when they are widely separated? 435 436 Unit 2 Module 3 Atomic and nuclear physics Worked example 27.3 Q Use the data in Tables 27.1 and 27.2 to calculate binding energies and binding energies per nucleon of the following nuclei: 238 (a) 42He (b) 56 26Fe (c) 92U Express your results in MeV. A (a) The 42He nucleus contains 2 neutrons and 2 protons nuclear rest mass = 4.001 506 u rest mass of constituent particles = 2mp + 2mn = 2 × 1.007 276 + 2 × 1.008 665 = 4.031 882 u mass defect = 4.031 882 – 4.001 506 = 0.030 376 u binding energy = mass defect = 0.030 376 u = 0.030 376 × 931.5 MeV = 28.3 MeV binding energy per nucleon = 28.3 ÷ 4 = 7.08 MeV Note: in this solution we started with the nuclear mass and so did not add the mass of the orbital electrons to find the total rest mass of the constituent particles.
However, if we are given the atomic mass to work from, then we must include the mass of the electrons as follows (remember a neutral atom will have Z orbital electrons where Z is the atomic number). atomic rest mass = 4.002 603 u rest mass of constituent particles = 2mp + 2mn + 2me = 2 × 1.007 276 + 2 × 1.008 665 + 2 × 0.000 548 6 = 4.032 979 u mass defect = 4.032 979 − 4.002 603 = 0.030 376 u As we would expect, this method gives the same result. (b) The 56 26Fe nucleus contains 30 neutrons and 26 protons.
nuclear rest mass = 55.920 67 u rest mass of constituent particles = 26mp + 30mn = 26 × 1.007 276 + 30 × 1.008 665 = 56.449 13 u mass defect = 56.449 13 − 55.920 67 = 0.528 45 u binding energy = mass defect = 0.528 45 u = 0.528 45 × 931.5 MeV = 492 MeV binding energy per nucleon = 492 ÷ 56 = 8.79 MeV (c) The 238 92U nucleus contains 146 neutrons and 92 protons. nuclear rest mass = 238.000 32 u rest mass of constituent particles = 92mp + 146mn = 92 × 1.007 276 + 146 × 1.008 665 = 239.934 48 u mass defect = 239.934 48 − 238.000 32 = 1.934 16 u binding energy = mass defect = 1.934 16 u = 1.934 16 × 931.5 MeV = 1802 MeV binding energy per nucleon = 1802 ÷ 238 = 7.57 MeV Table 27.1 Rest masses in atomic mass units (u) Particle and mass symbol Rest mass / u electron, me 0.000 548 6 proton, mp 1.007 276 neutron, mn 1.008 665 Table 27.2 Atomic and nuclear rest masses ITQ 2 Is the rest mass of a carbon-12 nucleus greater than, the same as, or less than the sum of the rest masses of six protons and six neutrons? ITQ 3 Use Einstein's equation E = mc2 to show that the energy equivalent of the atomic mass unit is given by: Atomic rest mass / u Nuclear rest mass / u 1 u = 931.5 MeV 1 1 1.007 825 1.007 276 4 2 4.002 603 4.001 506 12 6 12 exactly 11.996 71 Note: in order to obtain the energy equivalent of the atomic mass unit to four significant figures, you must express the mass (1.6605 × 10−27 kg) and the velocity of light (2.9979 × 108 m s−1) to at least this number of significant figures in the calculation. 56 26 55.934 937 55.920 67 234 90 234.043 60 233.994 23 238 92 238.050 79 238.000 32 Isotope H He C Fe Th U ITQ 4 Explain why the mass of Z electrons must be subtracted from the atomic mass to find the nuclear mass. Chapter 27 Energy from the nucleus Worked example 27.4 Q A A The equation for this nuclear decay is: 238 234 4 92U → 90Th + 2α + energy Applying the principle of the conservation of mass–energy we have 234 4 mass of 238 92U = mass of 90Th + mass of 2α + mass equivalent of energy released Therefore, 234 4 mass equivalent of energy released = 238 92U – mass of 90Th – mass of 2α = 238.000 32 − 233.994 23 − 4.001 506 = 0.0046 u Now 1 u = 931.5 MeV therefore energy released = 0.0046 × 931.5 MeV = 4.28 MeV In this problem we could have worked directly with atomic rest masses instead of nuclear rest masses; this is because the orbital electrons do not take part in the reaction and therefore their masses will cancel on either side of the equation. 238 92 U nucleus decays spontaneously by emitting an alpha particle to produce Th. Write an equation for this process and calculate the energy released. 234 90 hydrogen-1) always less than the combined rest masses of the separated nucleons. The difference between the rest mass of its constituents and the rest mass of a nucleus is called the mass defect. mass defect, Δm = total rest mass of nucleons − rest mass of nucleus Einstein's equation E = mc2 explains the mass defect. The small mass difference is the mass equivalent of the binding energy released when the nucleus is formed. In terms of mass–energy, binding energy and mass defect are equivalent. binding energy = Δmc2 ....(27.2) Nuclear energy: units and calculations We noted in Chapter 26 that the joule is an inconveniently large energy unit for atomic and nuclear physics. Similarly, the kilogram is also more suited as a unit of mass for objects at the 'human' scale, rather than for subatomic particles.
The following units for energy and mass are generally used for nuclear energy calculations: ■ energy: the mega electron volt (MeV) 1 MeV = 106 × 1.6022 × 10−19 J = 1.6022 × 10−13 J ■ mass the atomic mass unit (u) 1 u = 12 mass of one carbon-12 atom = 1.6605 × 10−27 kg 1 Tables 27.1 and 27.2 give a selection of particle, atomic and nuclear rest masses expressed in atomic mass units.
The values given are used in Worked examples 27.3 and 27.4. The nuclear rest mass is obtained by subtracting Z × me from the atomic rest mass, where Z is the atomic number and me is the rest mass of the electron. Getting it right! Conserved quantities in nuclear decays and reactions In any nuclear process the mass number A, the charge and the mass–energy are conserved.
The particles that take part in the process are represented by symbols of the form AZX where A is the mass number (number of protons plus neutrons), and Z is the charge. For a nucleus the charge Z is equal to the proton number. Z is conserved if it is regarded as the charge, but the number of protons may change. For example, in a beta decay a neutron is transformed into a proton and an electron, so the number of protons increases by one: X→ A Z Y + −10e A Z+1 Binding energy per nucleon Figure 27.2 is a plot of binding energy per nucleon against mass number for the elements. There are number of important features of this plot that allow us to predict the possibility of various nuclear reactions and processes, including nuclear fission and fusion. You should note the following key features, along with some of their consequences. ■ The binding energy of a single proton and a single neutron in 21H (deuterium) is close to 1 MeV per nucleon. As more nucleons are added to the nucleus the binding energy per nucleon increases, up to a maximum of a little less than 9 MeV per nucleon for 56 26Fe. So, in general, nuclear stability increases with increasing A up to A = 56. ITQ 5 Confirm that you obtain the same result for Worked example 27.4 if you use atomic rest masses instead of nuclear rest masses. ITQ 6 Compared to the graph in Figure 27.2, how would a graph of the mass defect per nucleon against mass number appear? 437 Unit 2 Module 3 Atomic and nuclear physics 10 16 8O 12 6C 56 26 Fe 120 50 Sn 235 92 U 238 92U 8 4 2 He fission 6 235 92 U fusion Binding energy per nucleon / MeV 438 4 235 92 U 235 92U 3 2 He 2 1 0n 2 1H 0 0 50 100 150 200 235 92 U 250 Number of nucleons, A (mass number) 235 92 U Figure 27.2 Variation of binding energy per nucleon with atomic mass. ■ The increase in binding energy per nucleon with increasing A for the light elements means that energy will be released if light nuclei can be made to combine (fuse) to produce heavier ones. This demonstrates the possibility of nuclear fusion of light nuclei as an energy source.
■ For A greater than about 80 the binding energy per nucleon decreases with increasing A. ■ The decrease in binding energy with A for heavy nuclei means that energy will be released if heavier nuclei undergo splitting (fission) into smaller fragments. This demonstrates the possibility of nuclear fission of heavy nuclei as an energy source. 4 ■ The 2He nucleus has an exceptionally high binding energy per nucleon for a light nucleus, and is therefore particularly stable. It therefore requires less energy (on average) for alpha particles to escape from a nucleus than individual nucleons or other small groups of nucleons. This helps to explain why alpha emission is a common form of nuclear decay. smaller nuclei heavy nucleus neutron Figure 27.3 Nuclear fission. energy 235 92 U Figure 27.4 An idealized explosive chain reaction. Fission and fusion Nuclear fission In nuclear fission a heavy nucleus splits into nuclei with smaller masses. One or more neutrons may also be released (Figure 27.3). The total rest mass of the fragments is slightly less than that of the original nucleus. The mass difference is accounted for, through ΔE = Δmc2, by the kinetic energy of the fragments, which is ultimately converted into heat energy by collisions with other atoms, and by gamma rays. One way in which a 235 92U nucleus undergoes fission is summarized by the equation below. 136 1 U + 10n → 98 42Mo + 54Xe + 2 0n + energy 235 92 This is an example of an induced fission reaction: the uranium nucleus is induced to split when a slow moving neutron collides with it. The final products of the reaction are stable isotopes of molybdenum and xenon and two further neutrons (see Worked example 27.5 for the energy calculation). The process has several intermediate steps. The initial products of the fission are highly radioactive nuclides. These undergo a sequence of gamma and beta decays before finally reaching the stable end state. The overall equation above is not balanced (Z on the right is greater than that on the left), but when the intermediate beta decays are taken into account (remember a beta decay converts a neutron into a proton), Z is conserved correctly. Chapter 27 Energy from the nucleus Chain reaction The neutrons released by the fission of a uranium-235 nucleus can trigger off the fission of other nuclei, which in turn release more neutrons. Not all the neutrons released will cause further fissions – some will escape from the surface of the uranium sample without interacting with a nucleus. However, if on average the fission of one nucleus releases one neutron that goes on to induce another fission reaction, then a self-sustaining chain reaction will result. When the number of neutrons causing fission is less than this, the reaction will die out, but when it is greater the chain reaction will grow explosively (Figure 27.4). Critical mass Worked example 27.5 Q (a) Given the following atomic rest masses, calculate the energy released by the induced fission of one U-235 nucleus in the uranium, molybdenum and xenon reaction.
Express your answer in atomic mass units, MeV and joules.
State any assumptions made. Atomic rest mass / u 235 92 U 235.043 93 98 42 Mo 97.905 41 136 54 Xe 135.907 22 (b) One mole (235.04 g) of uranium-235 contains 6.0221 × 1023 atoms. If all these atoms undergo fission in the way described, how much energy is released? How much water could be boiled by this amount of energy? A (a) Mass–energy is conserved, therefore, 98 136 mass of 238 92U + mn = mass of 42Mo + mass of 54Xe + 2mn + mass equivalent of energy released Assume the initial kinetic energies of the uranium nucleus and the neutron are negligible. Rearranging, 98 136 mass equivalent of energy released = mass of 238 92U − mass of 42Mo − mass of 54Xe − mn = 235.043 93 − 97.905 41 − 135.907 22 − 1.008 665 = 0.223 u = 0.223 × 931.5 = 208 MeV = 208 × 1.602 × 10−13 = 3.33 × 10−11 J (b) Fission of one uranium-235 nucleus releases 3.33 × 10−11 J. fission of 1 mole of atoms releases 3.33 × 10−11 × 6.0221 × 1023 J = 2.01 × 1013 J To raise 1 g water through 1 °C requires 4.18 J. If room temperature is 20 °C then to boil 1 g of water needs (100 − 20) × 4.18 J = 334.4 J so 2.01 × 1013 J would boil 2.01 × 1013 ÷ 334.4 g = 6.0 × 1010 g or 60 000 tonnes of water More neutrons escape from a small sample than produce fission, and so a chain reaction is not produced. The number of uranium-235 nuclei that a neutron might hit before escaping from the sample surface increases with the volume and therefore with the mass of the sample. The sample must thus exceed a critical mass before a chain reaction occurs. The chance of a neutron escaping without causing fission increases with the surface area of the sample. A sphere has the smallest surface area for a given volume and therefore a sample of uranium-235 in the form of a sphere will have the smallest critical mass. ■ The critical mass of a sphere of pure U-235 is 52 kg.
This corresponds to a sphere 17 cm in diameter. ■ The critical mass of Pu-239 is 9.9 kg, which corresponds to 9.9 cm sphere. The neutrons emitted during fission are usually moving too fast to be absorbed by other fissionable nuclei. These neutrons must be slowed down by a suitable material called a moderator, for example graphite or heavy ITQ 7 How many intermediate beta decays are required to conserve Z in the uranium to molybdenum and xenon reaction? water (made with the isotope of hydrogen containing a proton and a neutron – deuterium, 21H). Slowing down the neutrons ensures that more collisions occur with the uranium nuclei. The first successful fission reaction was that of German chemists Otto Hahn and Fritz Strassman in 1939. It was later discovered that the element plutonium also undergoes fission and produces more neutrons when bombarded with slow neutrons. Almost pure uranium-235, or plutonium-239, is used in an atomic bomb. The destructive power of the atomic bomb comes from the energy released in a chain reaction. The first atomic bomb was exploded in 1945 at a test site in New Mexico, USA. Shortly afterwards, atomic bombs were dropped on Hiroshima and Nagasaki in Japan, killing up to 200 000 people and destroying the two cities. The radiations and radioactive fallout from the bomb harmed many living things for years afterwards. The dropping of the bombs ended World War II. 439 440 Unit 2 Module 3 Atomic and nuclear physics electricity transmission lines water as moderator and coolant pressuriser concrete steam steel vessel steam-driven turbines control rods generator fuel rods heat exchanger pump pump cooling water from river, lake or sea warmed water to river, lake or sea coolant flow pump water from condensed steam Figure 27.5 A simplified pressurized water reactor. Nuclear energy The same chain reaction system can be used in a controlled way to liberate useful energy from fission. The world's first commercial nuclear power station was opened in 1956 at Calder Hall, England. In a nuclear power station, the fuel is arranged so that it can never explode as an atomic bomb does. The supply of neutrons is controlled so that the reaction is self-sustaining but no more. Fission in a nuclear reactor provides energy to turn water into steam. The steam is used to turn a turbine, which is connected to a generator. The generator produces electricity. Careful control of the chain reaction ensures energy is released at a steady rate. There are many different types of nuclear reactor. The pressurized water reactor (PWR) is the most common reactor type used today (Figure 27.5). The fuel (uranium oxide enriched with uranium-235) is contained in a steel pressure vessel. Energy released during the fission process heats the water to around 325 °C. The heated water is carried in pipes at high pressure (155 atmospheres) to the steam generator. The pressurized water also acts as a moderator to slow down the fission neutrons, thereby maintaining the chain reaction.
Boron-steel control rods inserted in the fuel elements are used to control the reaction. The boron absorbs the neutrons, slowing down the fission process. The chain reaction and hence the temperature can be controlled by raising or lowering the control rods. In an emergency all the control rods are lowered to shut down the reactor. The chain reaction is stopped within minutes, but the activity of the fission products continues to generate heat for days after shutdown, and the reactor must continue to be cooled. Failure of the cooling water pumps at the Fukushima Daiichi Nuclear Power Plant, following an earthquake and tsunami in Japan in 2011, led to conventional explosions that released radionuclides into the environment. The whole reactor is enclosed in a concrete shield to prevent any nuclear radiation from escaping and to protect workers.
Used fuel elements are stored underwater before being sent to a reprocessing plant when their activity has dropped below a set level. At the reprocessing plant, unused uranium and other useful nuclides are separated from the radioactive waste. Nuclear fusion In fission, a heavy nucleus is split into two to release energy. The reverse process can also produce large amounts of energy.
The reaction of two light nuclei to form a heavier nucleus is called nuclear fusion. For example, two deuterium nuclei can be smashed together to form a nucleus of helium-3 and a neutron (Figure 27.6). The reaction can be written as H + 21H → 32He + 10n + energy 2 1 energy He-3 Figure 27.6 Nuclear fusion. deuterium nucleus neutron Chapter 27 Energy from the nucleus Worked example 27.6 Q Given the following atomic rest masses, calculate the energy released when two deuterium nuclei undergo fusion according the equation 2 2 3 1 1H + 1H → 2He + 0n + energy. Express your answer in mass units, MeV and joules.
Atomic rest mass / u 2 1 H 2.014 102 3 2 He 3.016 029 A Mass–energy is conserved, therefore 2 × mass of 21H = mass of 31He + mn + mass equivalent of energy released Rearranging: mass equivalent of energy released = 2 × mass of 21H − mass of 31He − mn = 2 × 2.014 102 − 3.016 029 − 1.008 665 = 0.003 51 u = 0.003 51 × 931.5 = 3.27 MeV = 3.27 × 1.602 × 10−13 = 5.24 × 10−13 J The total energy released in the fusion reaction of two deuterium nuclei is less than that of uranium fission, but the energy per nucleon is much greater. In other words, less material is required to produce the energy. To start the fusion reaction the two nuclei must be brought sufficiently close to each other. It is not easy to do this as they repel each other with a very large electrical force due to their positive charges. One way to bring the nuclei together is to heat them up to an extremely high temperature (>108 K), so that they gain enough kinetic energy to overcome the repulsion. At these extreme temperatures all the atoms are completely ionized (their electrons are stripped) and the material is in the form of a gas of charged particles called plasma. Containment of plasma is difficult because any solid surface it makes contact with would be vaporized. One solution is to contain the charged particles in a 'magnetic bottle' created by appropriately designed magnetic fields as described in Chapter 20. The 'Tokamak' shown in Figure 27.7 is a containment device based on this principle. See for more information on Tokamak design and fusion science. At present fusion is not a net source of energy, as more energy has to be supplied to trigger and contain the reaction, than is produced. But it is hoped that later in this century, self-sustaining reactions will be created and nuclear fusion power may become a reality. There will be no shortage of fuel: hydrogen is found in water and two-thirds of the surface of the Earth is covered by the oceans. Nuclear fusion is the energy source for the Sun and stars. The solar energy received by us on Earth is due to the fusion of hydrogen nuclei in the Sun. Uncontrolled fusion here on Earth can be seen in the hydrogen bomb. The initial high temperature required is obtained by using an atomic bomb to trigger off the fusion. A hydrogen bomb releases much more energy than an atomic bomb. confining magnets containment vessel coils producing magnetic field plasma central magnets Figure 27.7 Nuclear fusion reactor design. One approach to developing a fusion power generator is to use magnetic fields to contain the plasma in a doughnut-shaped ring inside a device called a Tokamak.
A temperature of over 400 million K must be achieved to trigger the fusion of deuterium nuclei. 441 442 Unit 2 Module 3 Atomic and nuclear physics Review questions Summary ✓ Mass and energy are equivalent; they are related through Einstein's equation E = mc2. ✓ The energies of nuclear processes are sufficiently great for the corresponding mass changes to be measured. ✓ The atomic mass unit (u) is defined by 1 u = 1 12 mass of one carbon-12 atom. ✓ 1 u = 931.5 MeV ✓ The binding energy of a nucleus is the energy required to completely separate its component nucleons; this is equal to the energy released when the nucleons are brought together from infinity. ✓ The mass of a nucleus is always less than the rest mass of its constituent nucleons. ✓ Mass defect, Δm = total rest mass of nucleons − rest mass of nucleus ✓ Δmc2 = binding energy ✓ The binding energy per nucleon increases with mass number for light nuclei, indicating the possibility of nuclear fusion as an energy source.
✓ The binding energy per nucleon decreases with increasing mass number for heavy nuclei, indicating the possibility of nuclear fission as an energy source.
✓ In any nuclear process the mass number, charge and mass–energy are conserved. ✓ The energy released by the fission of one 23592U nucleus is of the order of 200 MeV.
✓ The energy released by the fusion of two 21H nuclei is 3.27 MeV. ✓ Nuclear fission is used to produce energy through a controlled chain reaction in a nuclear reactor. ✓ Nuclear fusion is the energy source that powers the stars, including our Sun. ✓ A practical fusion reactor for energy production has yet to be developed. c = 3.0 × 108 m s−1; 1 u = 931.5 MeV; mproton = 1.007 276 u; mneutron = 1.008 665 u; E = mc2 Mass defect and binding energy 1 Explain the meaning of each term in Einstein's mass– energy relation E = mc2. 2 Calculate the energy equivalent of a 1 kg mass.
3 A mobile phone battery stores 2 × 104 J of energy when fully charged. Explain why the mass of a fully charged battery is, in theory, greater than the mass of the battery when it is discharged. Calculate the mass difference and discuss whether or not it is measureable. 4 The Sun's power output (the solar luminosity) is 3.839 × 1026 W. Estimate the mass transformed into radiant energy by nuclear processes in the Sun each year. Express this mass as a percentage of the Sun's total mass (1.99 × 1030 kg). 5 Define the mass defect and the binding energy of a nucleus. Explain how these two quantities are related. 6 Is the energy of a nucleus greater or less than the energy of its constituent protons and neutrons when they are separated and at rest?
Discuss how the binding energy of a nucleus can be determined through measurements of mass. 7 Research the principle of the mass spectrometer as used to determine atomic masses. Write a brief description of its operation. Nuclear energy: units and calculations; Binding energy per nucleon 8 Define the atomic mass unit, u. 9 Show that 1 u = 931.5 MeV. 10 Calculate the mass defect, the binding energy and the binding energy per nucleon of the nuclides in the table below. Nuclide Nuclear mass / u 3 1 H 3.015 50 12 6 C 11.996 71 239 94 Pu 239.000 59 32 11 The radionuclide 32 15P undergoes beta decay to 16S according to the equation 0 P → 32 16S + −1e 32 15 Chapter 27 Energy from the nucleus Given the following values for the nuclear masses and the mass of the emitted electron, show that the total energy released in this process is 1.7 MeV. Particle or nucleus Mass / u 0 −1 32 15 31.965 68 32 16 31.963 29 P S 12 (a) Sketch a graph to show how the binding energy per nucleon varies with the nucleon (mass) number A. Your graph should indicate the value of A at which the binding energy per nucleon is greatest. (b) To which element does this peak value belong? (c) What is the approximate value of the maximum binding energy per nucleon in MeV? (d) Explain the features of the graph that indicate the possibility of nuclear fusion and nuclear fission as energy sources. Fission and fusion 13 Define nuclear fission and nuclear fusion. 14 Explain what is meant by a chain reaction. 15 (a) Use the atomic masses below to calculate the energy released in this fission reaction: U+ n→ 235 92 1 0 141 56 92 36 1 0 Ba + Kr + 3 n + energy (b) Explain briefly why, in this calculation, it is possible to work directly with either atomic masses or nuclear masses. (c) Explain the feature of this process that leads to the possibility of a chain reaction. (d) Given that, on average, the total energy release resulting from a 235 92U nucleus undergoing fission is 200 MeV, calculate the energy generated per gram of 235 92U that undergoes fission in a reactor. (e) Estimate the mass of 235 92U undergoing fission per year in a 500 MW nuclear reactor. State any assumptions you make. Particle or atom Mass / u 1 0 1.008 665 92 36 91.926 16 n Kr 141 56 140.914 41 235 92 235.043 93 Ba U 6 21H → 2 42He + 2 11H + 2 10n + energy 0.000 548 6 e 16 The overall reaction between the deuterium nuclei in a fusion reactor is summarized by the equation: In practice this reaction takes place in a series of steps in which deuterium nuclei first fuse with each other, and then the reaction products fuse with further deuterium nuclei. (a) Demonstrate that charge and mass number are conserved in the overall fusion equation. (b) Given the atomic masses below, calculate the energy released in the reaction. Explain briefly why in this calculation it is possible to work directly with either atomic masses or nuclear masses. (c) Calculate the energy generated per gram of deuterium fuel undergoing fusion in a fusion reactor. (d) Estimate the mass of deuterium fuel undergoing fusion per year in a 500 MW fusion generator. State any assumptions you make. (e) Briefly describe the difficulties of constructing a fusion reactor in practice. Particle or atom Mass / u 1 0 n 1.008 665 1 1 H 1.007 825 2 1 H 2.014 102 4 2 He 4.002 603 Practical exercises and challenges 1 Research one or more of the following topics: (a) Progress in the development of power generation by nuclear fusion. (b) The fusion processes that power the stars. (c) The technology of nuclear powered ships and submarines. Write a report on your findings. 2 It is possible (but not easy) to produce nuclear fusion in a well-equipped school laboratory by building a device called a 'fusor'. Deuterium is injected into a small evacuated globe. A high-voltage electric field causes ionization, and then accelerates the deuterium ions produced towards the centre of the globe, where they collide. The energy released is tiny in comparison to the energy supplied, and so fusors are not a source of power, but neutrons detected outside the globe indicate when fusion has occurred.
Find out more about fusor design and amateur fusion research on the internet. 443 444 Unit 2 Module 3 Atomic and nuclear physics 3 You can set up a model of a chain reaction with mousetraps and table tennis balls. For an example go to click on 'videos', select 'quantum physics and relativity' and play 'mousetrap fission'. You may like to try this for yourself! Answers to ITQs 1 Less – since energy must be supplied to separate the nucleons. 2 Less – since the nucleus has less energy, and therefore less mass, than the separated nucleons. 3 E = mc2 = 1.6605 × 10−27 × (2.9979 × 108)2 = 1.4924 × 10−10 J = 1.4924 × 10−10 = 931.5 MeV 1.6022 × 10−13 4 A neutral atom with atomic number Z has Z orbital electrons. The mass of these electrons must be subtracted from the measured atomic mass to find the nuclear mass. 5 Repeating the calculation with atomic masses we have: 234 4 mass equivalent of energy released = 238 92U − mass of 90Th − mass of 2α = 238.050 79 − 234.043 60 − 4.002 603 = 0.0046 u 6 The mass defect is directly proportional to the binding energy (Δmc2 = binding energy). The graph would therefore be the same apart from the units on the vertical axis, which would be atomic mass units instead of MeV. 7 The charge Z on the left of the equation is +92. The charge Z on the right is 42 + 54 = +96. Each beta decay carries away a charge of −1, leaving an extra positively charged proton in the nucleus. Therefore there must have been four intermediate beta decays if charge is conserved (92 = 96 − 4). Answers to Review questions 2 9.0 × 1016 J 3 2 × 10−13 kg 4 1.35 × 1015 kg; 6.8 × 10−14 % 10 Nuclide 3 1 H 12 6 C 239 94 Pu Nuclear mass / u Mass defect / u Binding energy / MeV Binding energy per nucleon / MeV 3.015 50 0.009 106 8.482 2.827 11.996 71 0.098 936 92.159 7.680 239.000 59 1.939 78 1806.9 7.560 15 (a) 173.3 MeV (d) 5.1 × 1023 MeV = 8.2 × 1010 J (e) 190 kg, assuming 500 MW is the thermal power (heat output) of the reactor. 16 (b) 43.25 MeV (c) 2.2 × 1024 MeV = 3.6 × 1011 J (d) 45 kg, assuming 500 MW is the thermal power (heat output) of the reactor. 445 Chapter 28 The photoelectric effect and wave–particle duality Learning objectives ■ Use the relationship E = hf to solve problems. ■ Describe the phenomenon of photoelectric emission. ■ Discuss the inability of classical physics to explain aspects of the photoelectric effect. ■ Use the photon model as the basis for explaining the classical paradoxes associated with photoelectric emission. ■ Discuss the photoelectric effect as evidence for the particulate nature of electromagnetic radiation. ■ Define 'work function' (Φ), 'threshold frequency' (f0), 'cut-off wavelength' (λ0) and 'stopping potential' (Vs). 1 ■ Use the relationship hf = Φ + 2 mv2 or hf = Φ + eVs to solve problems. ■ Explain the wave–particle nature of matter. ■ Describe and interpret the evidence provided by electron diffraction for the wave nature of particles. ■ Discuss interference and diffraction as evidence of the wave nature of electromagnetic radiation. ■ Use the relation for the de Broglie wavelength λ = h p. a b Figure 28.2 (a) A continuous wave. (b) Photons (wave packets). Figure 28.1 A solar panel generates power directly from sunlight. Photocells convert the solar energy into electrical energy. The transfer of energy from light photons to electrons is called the photoelectric effect.
Photons: E = hf Does light consist of particles or waves? Before the 20th century it was believed this question had been answered.
Observations of interference and diffraction had demonstrated that light behaves as a wave; however, this was not the end of the story. Experiments on the photoelectric effect (described below) show that light waves are not continuous – they are composed of discrete units or 'quanta'. A quantum of electromagnetic radiation is called a photon. Water waves and sound waves are continuous waves, like the waves shown in Figure 28.2a. A continuous wave may have any amount of energy, depending on its amplitude and frequency. There is no minimum value for the energy of the wave; the energy can be reduced to an arbitrarily small value simply by decreasing the amplitude of the wave. The photoelectric effect, however, shows that electromagnetic waves travel as 'wave packets', as illustrated in Figure 28.2b. Each wave packet is a photon with a definite amount of energy, which is proportional to the frequency f of the radiation. The constant of proportionality is the Planck constant h (see Chapter 25) and so the energy of a photon is given by E = hf ....(28.1) It is not possible to have a fraction of a photon and therefore hf is the minimum possible energy for light with frequency f. 446 Unit 2 Module 3 Atomic and nuclear physics The Planck constant UV lamp The Planck constant has the value h = 6.63 × 10−34 J s. As you will see in the following section, this value can be determined by measurement of the photoelectric effect. Light thus has a dual nature: it exhibits both wave characteristics (frequency, wavelength, diffraction, interference – see Chapter 9) and particle characteristics (it is divided into individual units). We say that light is quantized, and that it exhibits 'wave–particle duality'. zinc – – – – – Worked example 28.1 – – – – – Q (a) What is the energy of a photon of visible light with frequency 5.0 × 1014 Hz? (b) What is the energy of an X-ray photon with frequency 1018 Hz? A (a) E = hf = 6.63 × 10 × 5.0 × 10 J = 3.3 × 10−19 J (b) E = hf = 6.63 × 10−34 × 1018 J = 6.63 × 10−16 J −34 – – 14 electroscope Figure 28.3 A demonstration of the photoelectric effect. The photoelectric effect was not fully understood until Albert Einstein used the photon model of electromagnetic radiation to explain the observations. The success of his explanation helped to demonstrate that the photon theory was correct. The key to understanding the nature of the effect was the observation of the threshold frequency. The photoelectric effect Threshold frequency Figure 28.3 shows a simple demonstration of the photoelectric effect.
Electromagnetic radiation from a visible lamp and an ultraviolet lamp is directed, in turn, onto a freshly cleaned zinc plate, which forms part of a charged electroscope.
The following observations are made: Radiation below a certain frequency does not eject electrons from a metal surface, no matter how intense the light is made.
This frequency is called the threshold frequency, f0. Visible light does not eject electrons from zinc, but ultraviolet light (higher frequency radiation) does. This is illustrated in Figure 28.4. ■ a negatively charged plate is discharged by ultraviolet If visible light is a continuous wave, the existence of the threshold frequency cannot be understood – if the visible light is made bright enough it should eventually give the electrons enough energy to escape, but this does not happen. Einstein was able to use the photon picture of light to explain the threshold frequency. He reasoned that a certain minimum amount of energy was needed to eject an electron from the metal surface. If a single electron can gain energy only by absorbing the energy of a single photon, then the electron will escape only if the energy light (the electroscope leaf falls), but is not discharged by visible light ■ neither visible light nor ultraviolet light discharge a positively charged zinc plate ■ inserting glass between the ultraviolet lamp and the plate halts the discharge (the electroscope leaf ceases to fall), but removing the glass plate allows the discharge to proceed. Understanding the photoelectric effect In the photoelectric effect it is apparent that the energy of UV light is transferred to the electrons that give the electroscope a net negative charge. The electrons are given sufficient energy to escape from the plate, thus discharging the electroscope so the leaf falls.
What is not so obvious is why the energy of visible light fails to discharge the electroscope. ITQ 1 Can you suggest reasons for the following?
(a) A negatively charged zinc plate is discharged by ultraviolet light but a positively charged plate is not? (b) Visible light does not discharge the plate, but ultraviolet light does?

(c) A sheet of glass between an ultraviolet lamp and the zinc stops this discharging process? (d) The surface of the zinc must be cleaned at the start of the experiment? Chapter 28 The photoelectric effect and wave–particle duality visible photons UV photons f < f0 f > f0 no emission of electrons e– e– Worked example 28.2 e– e– e– e– metal surface photoelectrons e– e– metal surface Figure 28.4 The threshold frequency. The threshold frequency is the minimum frequency of the electromagnetic radiation that will cause the photoelectric effect for a particular metal.

The corresponding wavelength (λ0 = c/f0) is called the cut-off wavelength; λ0 is the maximum wavelength of the radiation that will eject photoelectrons from the metal. of the photon exceeds the minimum required.

This minimum energy is hf0 – the energy of a photon with the threshold frequency. For visible light hf < hf0 and so none of the photons have sufficient energy to eject electrons and therefore photoemission does not occur. For ultraviolet light hf > hf0 and so the photons can eject electrons. As the frequency of the incident light is increased photoelectric emission takes place when the threshold frequency is exceeded. Work function The work function, Φ, is the minimum amount of energy needed to eject an electron from a metal surface. It is the work needed to remove the negatively charged electron from the metal, overcoming the force of attraction to the positively charged atomic nuclei of the metal atoms. The value of the work function differs from one metal to another. Table 28.1 gives work functions of various metals Metal Work function, Φ / eV chromium 4.44 iron 4.60 lead 4.00 nickel 5.15 platinum 5.63 silver 4.44 tin 4.28 zinc 4.33 Q (a) Express the work function of zinc in joules. (b) Calculate the threshold frequency and cut-off wavelength for photoelectric emission from zinc. h = 6.63 × 10−34 J s; e = 1.60 × 10−19 C; c = 3.0 × 108 m s−1 A (a) From Table 28.1 the work function of zinc is given by Φ = 4.33 eV Now 1 eV = 1.60 × 10−19 J, therefore Φ = 4.33 × 1.60 × 10−19 J = 6.93 × 10−19 J (b) The work function is the minimum energy needed to remove an electron therefore, hf0 = Φ where f0 is the threshold frequency for photons that will just cause photoelectric emission. Thus for zinc, Φ 6.93 × 10−19 f0 = = = 1.05 × 1015 Hz h 6.63 × 10−34 Also, c λ0 = 108 = 2.9 × 10−7 m λ0 = = f0 1.05 × 1015 Investigating the photoelectric effect Figure 28.5 shows the apparatus used to investigate in detail the factors affecting photoelectric emission. The metal to be investigated is made one of the electrodes in an evacuated glass bulb. The bulb, together with the two electrodes it encloses, makes a simple photocell. Light of different frequencies and intensities is shone through the glass onto the metal surface. The electrons emitted from the surface travel through the vacuum to the second electrode, creating a current, which is indicated on the micrometer. The current reading is thus a measure of the number of electrons being emitted.

Note: to investigate the photoelectric effect produced by frequencies in the UV region of the spectrum, the incident light f electrode 1 electrode 2 e– e– evacuated glass bulb V I μA ITQ 2 How does the typical work function for a metal compare to the ionization energy of the hydrogen atom (see Chapter 25)?

Discuss the difference. variable voltage Figure 28.5 Apparatus used to investigate the photoelectric effect. 447 448 Unit 2 Module 3 Atomic and nuclear physics I intensity 2 intensity 1 – Vs V + for negative voltages the current falls as fewer electrons reach the electrode. As the negative voltage is increased, the current eventually drops to zero. The voltage at which the current becomes zero is the stopping potential Vs. This is equal to the maximum kinetic energy of the photoelectrons. If Vs is expressed in volts. To convert from eV to joules we multiply by the basic unit of charge 1.6 × 10−19 C, therefore: 1 Emax = 2 mv2 = eVs b l metal 2 metal 1 - Vs2 Vs1 V + Figure 28.6 (a) Variation of photocell current with applied voltage for two different values of the incident light intensity. (b) Variation of photocell current with applied voltage for two different metals exposed to light of the same frequency and intensity. photocell bulb must be made from fused quartz glass or fitted with a quartz window. Standard laboratory glass absorbs UV; quartz glass is relatively transparent to UV radiation. The external circuit shown is used to apply a variable potential difference between the electrodes. Effect of light intensity Figure 28.6a is a plot of the current against the potential applied, for two different incident intensities, I1 and I2, of light of a given frequency. For positive voltages the electrons emitted from plate A are all attracted to plate B. For intensity I1 the current for positive voltage values is constant – all the electrons emitted travel between the plates. If the light intensity is increased to I2 then more electrons are emitted and so a new constant value. This equation can be understood from the principle of the conservation of energy: if the electrons have maximum 1 energy 2 mv2 = eVs as they leave the metal surface, then they will lose this energy in moving through a potential difference of –Vs to reach the second electrode. If the voltage of the second electrode is decreased any further, none of the electrons will have sufficient energy to reach it. Note that in Figure 28.6a the stopping voltage is the same for both values of the light intensity.

This shows that the maximum energy does not depend on the intensity of the light. Figure 28.6b shows the variation of photocell current with voltage for two different metals exposed to light of the same frequency and intensity. Now the stopping voltages are different, indicating that the maximum kinetic energy does depend on the properties of the metal. Effect of light frequency Figure 28.7 shows how the stopping potential (the maximum kinetic energy) of the emitted electrons varies with light frequency for two different metals. The straightline graphs indicate that the electron energy increases linearly with incident light frequency above the threshold 5 4 3 2 The energy of the photoelectrons can be found by reversing the potential between the plates.

When the second plate is made negative with respect to the first, electrons are repelled from it, and only those that start with sufficient energy will reach the second electrode; 5V metal 2 1 Vs / V 0 –1 Measurement of electron energy: stopping potential metal 1 2 4 6 8 10 12 14 16 18 20 22 f / 1014 Hz 24 26 –2 –3 –4 12.2 x 1014 Hz –5 –6 Figure 28.7 Plots of stopping potential Vs against incident light frequency, f, for two different metals.

Chapter 28 The photoelectric effect and wave–particle duality frequency. When the lines are extrapolated backwards, their intercepts with the y-axis are equal to the work functions of the metals (expressed in electron volts); this result is explained by Einstein's equation in the next section. The key results of experiments with the photoelectric apparatus may be summarized as follows: ■ electrons are emitted only when the frequency of the incident light exceeds a certain threshold frequency, f0 ■ the threshold frequency is different for different metals ■ below the threshold frequency there is no photoemission no matter how bright the light ■ above the threshold frequency photoemission takes place no matter how dim the light ■ above the threshold frequency the energy of the emitted electrons increases with increasing frequency of light ■ the energy of the electrons does not depend on the brightness (intensity) of the light; increasing the intensity increases the number of electrons emitted, not their energy. Energy of photons The observations of the dependence of the photoelectric effect on light frequency and intensity can be explained if light (and other forms of electromagnetic radiation) is made up of photons that carry energy in fixed amounts determined by the frequency of the radiation.

The energy of a single photon is given by the expression E = hf. In the photoelectric effect, if the threshold frequency is exceeded, then even a single photon has sufficient energy to eject an electron from the metal surface.

This explains why the photoelectric effect occurs even in very dim light when the frequency is greater than the threshold frequency. However, if the frequency is below the threshold frequency then, no matter how bright the light, none of the photons has sufficient energy to eject an electron from the metal surface, and the photoelectric effect does not take place.

The Einstein equation Einstein's equation gives the energy of the electrons emitted during the photoelectric effect. It is derived as follows. Suppose the work function of the metal surface is Φ. The threshold frequency, f0, for light that will just eject an electron from the surface is given by: hf0 = Φ The energy of electrons ejected by light of frequency f greater than the threshold frequency will thus be: E = hf − Φ = hf − hf0 This energy is the kinetic energy of the emitted electrons, so: 1 2 2 mv = hf − hf0 Therefore: 1 hf = hf0 + 2 mv2.............(28.2) This is Einstein's equation for the photoelectric effect. Einstein's equation may be understood in terms of the conservation of energy. The energy of the incident photon, hf, is transferred to the work done in removing the electron from the metal surface (hf0) plus the kinetic energy of the electron. 1 Now, since 2 mv2 = eVs and hf0 = Φ, Einstein's equation can also be written as: hf = Φ + eVs................(28.3) Rearranging this equation we can write the equation of the straight lines in Figure 28.7 as: Vs = h Φ f − e e The gradient of both lines is thus the Planck constant h divided by e. The intercept with the y-axis (when the lines are extrapolated backwards) is −Φ/e. Worked examples 28.3 and 28.4 show how the Einstein equation can be applied. Applications of the photoelectric effect The photoelectric effect has practical application in two key areas: ■ the detection of light and the measurement of its ITQ 3 How do the following changes correspond to changes to the photons in a beam of light: (a) an increase in the light intensity (b) an increase in the light frequency? intensity ■ the generation of electrical energy from light energy.

Various devices based on the photoelectric effect have been developed for these purposes.

449 450 Unit 2 Module 3 Atomic and nuclear physics Worked example 28.3 Q A (a) Calculate the Planck constant from Figure 28.7. (b) Determine the work functions of the two metals. Hence, by reference to Table 28.1, identify the metals. rise h 5.0 V = = 4.1 × 10−15 V s run e 12.2 × 1014 Hz Now, e = 1.6 × 10−19 C, therefore, h = 1.6 × 10−19 × 4.1 × 10−15 J s = 6.6 × 10−34 J s Thus, h = 6.6 × 10−34 J s (b) y-axis intercept = − Φ e = − work function expressed in eV Therefore, reading the intercepts, Φ for metal 1, Φ = 4.0 eV ■ metal 2, Φ = 5.6 eV With reference to Table 28.1 we can identify metal 1 as lead and metal 2 as platinum. A exciter lamp to amplifier and loudspeaker photocell UV radiation with frequency 1.2 × 1015 Hz is used in an investigation of the photoelectric effect with the metals listed in Table 28.1. (a) Calculate the energy of the UV photons in joules and in electron volts. (b) Which of the metals will not exhibit photoelectric emission with radiation of this frequency? (c) Calculate the maximum kinetic energy and maximum speed of the photoelectrons emitted when the radiation is incident on lead. h = 6.63 × 10−34 J s; e = 1.602 × 10−19 C; me = 9.11 × 10−31 kg; c = 3.00 × 108 m s−1 A (a) E = hf = 6.63 × 10−34 × 1.2 × 1015 J = 7.96 × 10−19 J = 7.96 × 10−19 1.602 × 10−19 eV = 4.97 eV (b) Nickel and platinum will not emit photoelectrons since their work functions are greater than the energy of the sound track. (c) From the Einstein equation: hf = hf0 + 12 mv2 KEmax = 12 mv2 = hf − hf0 But, hf0 = Φ, so KEmax = hf − Φ = 4.99 eV − 4.01 eV = 0.98 eV = 0.98 × 1.602 × 10−19 J = 1.57 × 10−19 J 2KE 2 × 1.57 × 10−19 v = m s−1 = 9.11 × 10−31 = 5.87 × 105 m s−1 film picture slit R picture (a) gradient = Worked example 28.4 Q film sound track Figure 28.8 Reading the sound track from a movie film. photocell lens lens lamp conveyor belt to counter Figure 28.9 Electronic counting using a photocell. Photoemissive devices Photoemissive devices use the original photocells, based on the same principle as the apparatus used to investigate the photoelectric effect, shown in Figure 28.5. Figure 28.8 shows a photocell used to generate a signal from the sound track recorded on a movie film. The signal is recorded as variations in the transparency of the track. The changes in the intensity of the light transmitted by the track are converted to an electric signal by the photocell as the film passes. The signal is amplified and used to drive loudspeakers. Figure 28.9 shows how a photocell could be used to count boxes on a conveyor belt. As a box passes between the light source and the cell the output of the cell decreases, then increases again. An external circuit counts the pulses electronically. Photoconductive devices In modern applications photoconductive devices, which change their electrical resistance when light falls on them, have largely replaced photoemissive cells. Photoconductive devices are based on semiconducting materials such as cadmium sulfide. Photons of visible light carry sufficient energy to excite electrons from the valence band to the conduction band in these materials.

These electrons can then move freely through the material (like the free electrons in a metal), decreasing the electrical resistance. Figure 28.10a shows a light-dependent resistor (LDR) based on this principle. In the dark the resistance of the Chapter 28 The photoelectric effect and wave–particle duality A Matter waves b symbol We have seen that electromagnetic (EM) radiation exhibits wave–particle duality.

EM radiation has frequency and wavelength (wave behaviour), but it is quantized as discrete photons, each with energy E = hf. Photons behave as massless particles. metal electrodes on surface of cadmium sulfide Figure 28.10 (a) A light-dependent resistor (LDR). (b) A lightening circuit. device is 10 MΩ. In daylight its resistance falls to 1 kΩ. Figure 28.10b shows how an LDR may be used with a transistor switch to turn a light bulb on when it is dark and off when it is light. Photovoltaic devices generate an e.m.f. when light falls on them. When light is incident on the junction region of a semiconductor diode, electron–hole pairs are generated as electrons are excited from the valence band to the conduction band. Because a barrier potential exists across the p–n junction region, the electrons can flow only in one direction across the junction, while holes can flow only in the opposite direction. A current is thus generated in an external circuit connected across the junction, and energy is transferred from the light photons to the current. The selenium photocell shown in Figure 28.11 is used as the light sensor in a photographic exposure meter. You will find more details about the construction and applications of semiconductor photoconductive and photovoltaic devices in Chapter 22. + polarity of photo-emf metal collector strip transparent conducting film + n-type cadmium oxide – + p-n junction – p-type selenium metal back plate + – In his theory of relativity Einstein had shown that, even though it is massless, a photon has momentum, p, given by E p = c ....(28.4) where c is the speed of light and E is the photon energy. Photovoltaic devices light In 1924 Louis de Broglie proposed that particles with mass, electrons for example, may also exhibit wave– particle duality, and so have wave characteristics as well as their particle properties. Now for a photon of frequency f and wavelength λ, E = hf and c = fλ Therefore the photon momentum, p = hf h = fλ λ By analogy, de Broglie argued that a particle with mass m and velocity v will have a wavelength given by the same equation, p = h where the momentum is now given by λ p = mv. Thus for all particles, with or without mass, λ = h p ....(28.5) This equation, known as the de Broglie relation, links momentum to wavelength. Shortly after de Broglie's proposal, American physicists Clinton Davisson and Lester Germer and independently George Paget Thomson (J. J. Thomson's son), showed that a beam of cathode rays is diffracted when it passes through a crystal lattice – in a similar way to light passing through, or reflected from, a diffraction grating.

The wavelength of the electrons, calculated from the spacing of the atoms in the crystal and the diffraction angles, agreed precisely with de Broglie's predictions. holes electrons Figure 28.11 A selenium photocell. – ITQ 4 Why is a p-n junction required to produce a steady current from a semiconductor-based photocell? 451 452 Unit 2 Module 3 Atomic and nuclear physics A Worked example 28.5 Q Calculate the wavelength associated with electrons that have been accelerated through 500 V in a cathode ray tube. h = 6.63 × 10−34 J s; me = 9.11 × 10−31 kg; e = 1.602 × 10−19 C A The energy of the electrons, 1 Ek = eV From the relationship 1 Ek = 2 mv2 we have 2 mv2 = eV A So, p = √2meV b c Hence from the de Broglie relation, h h λ 6.63 × 10−34 = = p √2meV √ 2 × 9.11 × 10−31 × 1.602 × 10−19 × 500 Worked example 28.6 Q A Q A Calculate the wavelength of an electron accelerated through V volts is eV joules, where e is the charge of the electron. Thus 12 mv 2 = eV Also, p = mv p 2 Therefore, 12 mv 2 = eV Or, p = √2meV h of λ = p Hence, from the de Broglie relation, h h 6.63 × 10−34 λ = √2meV = √2 × 9.1 × 10−31 × 1.602 × 10−19 × 500 Calculate the wavelength of a cricket ball of mass 0.16 kg travelling at 40 m s−1 (90 mph). Compare this wavelength to the size of an atomic nucleus (~ 10−14 m). h h 6.63 × 10−34 = = 1.04 × 10−34 m p mv 0.16 × 40 This wavelength is 10−20 the size of a nucleus. This is so tiny in comparison to the size of the cricket ball that quantum effects such as probability wave diffraction will never be observed for objects at this scale. λ= The diffraction pattern produced when a narrow beam of electrons passes through a thin sheet of graphite can be observed on a fluorescent screen or recorded with a photographic film. The significance of matter waves De Broglie's relation applies to all particles, from photons, electrons, nucleons and atoms, to macroscopic objects, such as cricket balls (see Worked example 28.6), people and even planets and stars. Everything, at least in principle, has wave–particle duality.

The theory of quantum mechanics has established that the wave associated with a particle is a probability wave. The intensity of this wave at a particular location (its amplitude squared) is proportional to the probability of finding the particle at that point. The meaning of the probability wave is illustrated by electron diffraction. When a beam of electrons passes ITQ 5 Is the wavelength of a cricket ball greater or less than that of an electron travelling at the same speed? e Figure 28.12 How a diffraction pattern builds up as individual electrons pass through a pair of slits. through a crystal the beam is spread into a diffraction pattern with maxima and minima. However, an individual electron is not 'smeared' across the pattern, it arrives at a definite point on the screen or film, where it is observed. We cannot follow the path of an individual electron through the crystal, nor predict exactly where it will end up in advance of observing it. However, we can calculate that it has a high probability of arriving at a point in the maximum of the diffraction pattern and a low probability of arriving at a minimum. The overall pattern is the result of many individual electrons arriving at different points on the screen, with a distribution determined by the diffraction pattern of the probability wave (Figure 28.12). So why if wave–particle duality and quantum theory apply to all matter are people not diffracted as they walk through doorways, or cricket balls as they pass by the stumps? You will recall from Chapter 9 that interference and diffraction become significant only when the width of slits or the size of objects is similar to the wavelength of the wave that passes through or by them. Chapter 28 The photoelectric effect and wave–particle duality Worked example 28.6 demonstrates that the wavelength of a cricket ball is so tiny (10−20 the diameter of a nucleus) that diffraction effects are completely negligible. This is why cricket balls behave in a predictable way. The same is not true of electrons and other subatomic particles. Their wavelengths are of the same order as the size of atoms – this is why wave–particle duality and quantum theory are needed to understand how matter behaves at the atomic and subatomic scales.

Getting it right! Understanding quantum theory Quantum theory is one of the most successful and accurate scientific theories ever developed. Without it we could not understand atomic energy levels, the properties of semiconductors, or the physics of the stars; computers, mobile phones and other modern technologies would not have been invented. Yet the ideas underlying quantum theory can seem peculiar. Its principles – particles governed by probability waves, for example – are unfamiliar and contrary to everyday experience, so don't worry if you find these ideas strange the first time you meet them. Niels Bohr famously said: Those who are not shocked when they first come across quantum theory cannot possibly have understood it. Another Nobel Prize winner, Richard Feynman, observed: Electrons, when they were first discovered, behaved exactly like particles or bullets, very simply. Further research showed, from electron diffraction experiments for example, that they behaved like waves. As time went on there was a growing confusion about how these things really behaved – waves or particles, particles or waves? Everything looked like both ... There is one simplification at least.

Electrons behave in this respect in exactly the same way as photons; they are both screwy, but in exactly the same way. Summary ✓ In the photoelectric effect, incident light energy causes electrons to escape from a metal surface. ✓ Experiments with a photocell show that the photoelectric effect occurs only when the frequency of the incident light exceeds a threshold frequency, f0. ✓ The threshold frequency is not consistent with the classical picture of light as a continuous wave, but can be explained if light consists of photons, each of which carries energy E = hf.

✓ The work function, Φ, is the energy required to just eject an electron from the surface of a particular metal: hf0 = Φ, where h is the Planck constant. ✓ The maximum kinetic energy of photoelectrons can be found by measuring the stopping voltage in a photocell: 12 mv2 = eVs ✓ The Einstein equation (hf = hf0 + 12 mv2 or hf = Φ + eVs) relates the frequency of the incident light to the threshold frequency and the kinetic energy of the photoelectrons. ✓ A graph of stopping voltage Vs against incident light frequency f for a photocell is a straight line −Φ with gradient h e and intercept e . ✓ Both electromagnetic radiation and particles with mass exhibit wave–particle duality. ✓ The wavelength of a photon or particle with mass is related to its momentum by the de Broglie relation: λ = h p ✓ The wave nature of electrons is confirmed by observations of electron diffraction: diffraction and interference are wave characteristics. ✓ The wave associated with a particle is a probability wave; the intensity (amplitude squared) of the wave at a point is proportional to the probability of observing the particle at that point. 453 454 Unit 2 Module 3 Atomic and nuclear physics Review questions h = 6.63 × 10−34 J s; me = 9.11 × 10−31 kg; e = 1.6 × 10−19 C 1 (a) Explain what is meant by the 'photoelectric effect'. (b) State the factors that affect photoelectric emission. (c) Explain the terms: (i) photon (ii) threshold frequency (iii) cut-off wavelength (iv) work function (v) electron-volt. 2 (a) Calculate the energies in eV of photons of the following wavelengths: 5.0 × 10−7 m; 1.0 × 10−8 m. (b) The work function for a given surface is 2.9 eV. Find: (i) the cut-off wavelength (ii) the threshold frequency. 3 Refer to Table 28.1. Calculate the threshold frequencies and cut-off wavelengths for photoelectric emission by tin and platinum. Investigating the photoelectric effect. The Einstein equation 4 (a) Draw a diagram of the apparatus used to investigate the energy of the electrons emitted from a metal surface as a result of the photoelectric effect. (b) Explain how the maximum energy of the electrons is determined. 5 (a) Sketch a graph showing the variation of current (y-axis) with applied voltage (x-axis) for a photocell exposed to light of a given frequency and intensity. The x-axis of your graph should cover both positive and negative voltage values. Label the point corresponding to the stopping voltage, Vs. (b) Account for the main features of the graph you sketched in part (a) in terms of the energy and number of the photoelectrons emitted in the photocell and the effect of the applied voltage on these electrons. (c) On the same axes sketch a second curve to show the effect of increasing the intensity of the incident light without changing the frequency. Account for any differences/similarities between the two curves. 6 The Einstein equation for the photoelectric effect may be written as hf = hf0 + 2 mv2 or hf = Φ + eVs. Explain the significance of each of the terms in these expressions. 7 The cut-off wavelength for a certain metal is 3.6 × 10−7 m. Calculate: (a) the threshold frequency (b) the work function for the metal (c) the voltage needed to stop the emission stimulated by light of wavelength 2.2 × 10−7 m. 8 The threshold frequency for photoelectric emission from sodium is 4.41 × 1014 Hz. Calculate: (a) the work function in electron volts (b) the stopping potential for photoelectrons released by ultraviolet light of wavelength 3.00 × 10−7 m.

9 The table below shows the stopping potential, Vs, and the corresponding frequencies for a certain photocell. Stopping potential, Vs / V 1.25 2.50 3.75 5.00 6.25 Frequency / × 1015 Hz 1.38 1.68 1.98 2.30 2.60 (a) Plot a graph of stopping potential against frequency. (b) From your graph determine: (i) the threshold frequency (ii) the cut-off wavelength (iii) the work function of the metal (iv) a value for the Planck constant. 10 The minimum frequency of light that will cause photoelectric emission from a certain metal surface is 5.25 × 1014 Hz. When the surface is illuminated by another radiation, electrons are emitted with a maximum speed of 7.25 × 105 m s−1. Calculate: (a) the work function of the metal surface (b) the maximum kinetic energy of the emitted electrons (c) the frequency of the second source. 11 In an experiment to investigate the photoelectric effect, the stopping voltage, Vs, required to just prevent photoelectric emission is measured as a function of the frequency, f, of the incident light for lead, zinc and platinum. (a) Referring to Table 28.1, sketch graphs (on the same set of axes) of Vs against f for these metals. (b) Explain why there is a threshold frequency, below which no photoelectric emission occurs. Is this frequency the same for all three metals?

Explain. (c) Explain how the Planck constant, h, could be determined from the data.

Chapter 28 The photoelectric effect and wave–particle duality Matter waves 12 Describe briefly the evidence that electrons behave as: (a) particles (b) waves. Assuming the LED has negligible resistance then the electrical energy supplied by the battery must be equal to the energy emitted as light. Now if V is the voltage drop across the diode then, 13 Both radiation and matter are said to exhibit 'wave– particle' duality. Explain the meaning of this statement. What is the nature of the wave associated with a particle such as the electron? energy change per electron = eV c energy of photons = hf = h λ c eV/λ therefore, eV = h λ ∝ Procedure 1 Set up the circuit as shown. 2 Start with the variable resistor set at its maximum value. 3 Slowly reduce the resistance until the LED just switches on. 4 Note the voltage drop across the LED when it just starts to operate. 5 By adjusting the variable resistor, record data to plot a graph of current against voltage for the LED. 6 Use a spectrometer to measure the wavelength of the light emitted by the LED (alternatively look up this wavelength on the datasheet for the particular LED you are using). 14 Write down the de Broglie relation that links the wave and particle characteristics of a particle. Explain the meaning of the terms in this equation. Does the de Broglie relation apply to quanta of electromagnetic radiation? 15 Calculate the wavelengths of the following particles: (a) an electron that has been accelerated through a p.d. of 1000 V (b) a neutron with energy kT, where k is the Boltzmann constant and T is room temperature (293 K) (c) a soccer ball (m = 0.43 kg) travelling at 20 m s−1. 16 Looking back at question 15, discuss the circumstances in which it would be possible to observe the diffraction of electrons and neutrons. Explain why the wave characteristics of soccer balls cannot be detected. mn = 1.67 × 10−27 kg; k = 1.38 × 10−23 J K−1 Practical exercises and challenges 1 Use a photocell to design a burglar alarm. Your design could be based on the circuit shown in Figure 28.10b. 2 Figure 28.13 shows a circuit you can use to find the Planck constant by measuring the voltage drop across an LED (light-emitting diode) and the wavelength of the radiation it emits. spectrometer V Figure A Figure 28.13 Measurement of h with an LED. Theory When an LED conducts, electrons lose energy by combining with holes in the junction region. The energy lost is emitted as photons of visible light. Analysis The voltage value you require is the voltage drop when the diode just operates with minimum current. You can assume that all the energy supplied by the battery is transferred to photons, and energy lost through resistive heating is zero. Obtain the best value for V by extrapolating your graph of V against I back to I = 0. Calculate h from your measured value for V and λ using the equation above.

Answers to ITQs 1 (a) This suggests that ultraviolet light can discharge a negatively charged plate by giving electrons enough energy to escape from its surface. Ejecting electrons from a positively charged plate would require more energy (since the electrons are attracted back to the plate) and would increase the net positive charge of the plate. (b) This suggests that ultraviolet light can transfer sufficient energy to an electron for it to escape from the plate; but visible light cannot. (c) This suggests that the glass absorbs ultraviolet light.

(d) This suggests that other electrons cannot travel through the layer of oxide that forms on zinc exposed to the atmosphere; or the layer absorbs UV radiation. 455 456 Unit 2 Module 3 Atomic and nuclear physics 2 The ionization energy of hydrogen is 13.6 eV. This is the energy required to completely remove the electron from the atom. The work function of a metal is typically about one third of this energy. There are two reasons why this energy is less than the ionization energy of hydrogen: firstly the outer electron of a metal atom is further from the nucleus and less tightly bound than the electron in hydrogen (the first ionization energy of a zinc atom is 9.4 eV); secondly the free electrons in a metal have already partly escaped from the influence of individual atoms, and so less energy is required to remove a free electron from the metal surface than to ionize an isolated atom; for zinc this energy is 4.33 eV (the work function).

3 (a) An increase in light intensity increases the number of photons. (b) An increase in light frequency increases the energy of each photon (E = hf). 4 The photovoltaic effect takes place in a pure semiconductor will generate electron–hole pairs that can flow in any direction through the component; this will produce a change in resistance but not current without an externally applied e.m.f. A p-n junction creates a potential barrier that permits charge of a given sign to flow in one direction only, and so the photoelectric effect produces a net current in that direction. 5 Wavelength is inversely proportional to momentum (λ = p). The wavelength of a cricket ball is therefore very much less than that of an electron travelling at the same speed. Answers to Review questions 2 (a) 2.5 eV, 120 eV (b) (i) 4.3 × 10−7 m (ii) 7.0 × 1014 Hz 3 tin 1.0 × 1015 Hz, 2.9 × 10−7 m; platinum 1.4 × 1015 Hz, 2.2 × 10−7 m 7 (a) 8.3 × 1014 Hz (b) 3.4 eV (c) 2.2 V 8 (a) 1.8 eV (b) 2.3 V 9 (b) (i) 0.97 × 1015 Hz (ii) 2.8 × 10−7 m (iii) 4.0 eV (iv) 6.4 × 10−34 J s 10 (a) 2.25 eV (b) 2.39 × 10−19 J = 1.50 eV (c) 8.81 × 1014 Hz 15 (a) 3.9 × 10−11 m (b) 1.8 × 10−10 m (c) 7.7 × 10−35 m 457 CAPE SBA The Caribbean Advanced Level Proficiency Examination (CAPE) Physics School Based Assessment (SBA) is the internal assessment portion of the CAPE physics syllabus. All practical work, including the SBA, is worth THREE main reasons: 1 The SBA component is worth 20% of the marks for each unit.

If you do not fulfil the requirements for the SBA component, the results of the written examination are of no consequence, Therefore the SBA cannot be neglected. 2 While the theory component is important in the understanding of science subjects (chemistry, physics and biology), remember that discoveries in these subjects are made by experimentation. Therefore, the relationship between the experimental work and the theory behind it is invaluable.

3 To go on to higher education where advanced experimental skills will be required. You should therefore consider this a way to develop the basics of proper laboratory thinking and practices. Skills involved There are four skills that will be used for the SBA assessment.

You can find more details about these skills on pages 2, 3 and 4 of the syllabus. 1 Analysis and Interpretation (AI) ■ Understand cause and effect relationships. ■ Make deductions from experimental results.

■ Predict outcomes of unknown scenarios. ■ Do calculations correctly. ■ Give assumptions and limitations in experiments. 2 Manipulation and Measurements (MM) 3 Observation, Recording and Reporting (ORR) ■ Observation is the ability to perceive objects using your senses. There can be changes in length, colour, phases, size, temperature and readings on an instrument that must be noted. ■ Recording involves the accurate recording of measurements and the use of appropriate presentation formats for their display. In calculations you should pay special attention to significant figures and the maximum error associated with the various instruments in use. ■ Reporting involves the ability to organize and present the completed experiment in a logical and concise manner. Punctuation, grammar and spelling are important in this category.

It also involves introspection, where the student examines how things could be done differently if the experiment was repeated. 4 Plan and Design Experiments (PD) In this category you may employ your creativity.

You will be presented with a scenario and asked to plan and design an experiment to address the scenario. You will be marked on: ■ problem recognition within the scenario ■ development of a testable hypothesis ■ design an experiment to test the hypothesis ■ variable recognition A hypothesis is simply a statement (which may be valid or invalid) that is made about a parameter within the scenario. The statement should be testable by the experiment and the resulting data would allow you to confirm its truth or untruth. This involves: Here is an example. ■ successfully following detailed instructions ■ assembling apparatus Scenario: The scenario ■ assembling apparatus Observation and Mr Jones has just died and his family is dividing his estate. His five-cent Canadian coin collection is reputed to be worth 10 million dollars! However, with the proliferation of dated counterfeit coins, Mr Jones' relatives are arguing about the true worth of the coins in the collection.

They have hired you to appraise the collection. investigation ■ using apparatus appropriately and taking relevant measurements. 458 CAPE SBA It is well known that coins minted between 1908 and 1919 had a composition of over 90% silver by mass while those minted after 1919 have very little silver but are composed of over 90% nickel. Coins made before 1919 may be worth approximately $200 000 each, while those minted after 1919 are worth nothing. What features must you include in an experiment to determine the true worth of the coin collection? Procedure The way the procedure is expressed may vary depending on the type of experiment. The general criteria for expressing a procedure are as follows: ■ it must be in the past tense ■ it must be in the passive voice ■ it should be concise ■ it should follow the logical sequence of events Laboratory notebook ■ It should not include any observations or inferences. Here are some rules of thumb for presenting an excellent laboratory notebook. Results and calculations The notebook 1 The pages of the laboratory notebook should be numbered. 2 There should be a table of contents. 3 All experiments should be dated. The experiments Here is the general format for laboratory experiments: ■ Experiment title ■ Date ■ Topic ■ Aim ■ Apparatus and materials ■ Procedure ■ Results and calculations ■ Theory ■ Discussion ■ Conclusions Guidelines Topic and aim Your topic and your aim should be consistent with the actual experiment being performed. Apparatus and materials Separate lists of apparatus and materials should be included. Making separate lists allows you to see if any key lists of apparatus or any key materials have been missed. This is especially useful for 'plan and design' experiments. As far as possible, results should be tabulated. This reduces the space utilized and correlates the results obtained to the factors performed. Calculations should be concise and to the point. All associated formulae should be stated. Take care that all fundamental and derived units are correct. Plot all graphs that are necessary. In some experiments, data will be extrapolated from the graph to be used in other calculations (for example, the gradient); pay keen attention to the values and units. Theory The theory section should give a brief background of the law, principle or equation under investigation or used in the analysis of data. Discussion Consider precautions, limitations and (with hindsight) possible improvements. Answer all other questions associated with the experiment. Conclusions This should be a clear, concise statement of your result, whether numerical or factual, related to the problem posed in the aim. Good writing and not-so-good examples This section will be a comparison of the work from two different students. The reports will be analysed and the good features and bad features of each report will be highlighted. In these examples, symbols (e.g. l for current) will not be printed in italics, because you will not be able to do this in a handwritten report. CAPE SBA MM and AI: Student A Calculations Sample calculation from Table SBA.1. Title January 23, 2012 t20 = 8.91 8.91 T= 20 T = 0.45 Topic T2 = (0.45)2 Vertical oscillation of a spring–mass system. = 0.20 An investigation of simple harmonic motion. Date Aim 0.8 2 0.7 Period squared, T / s To investigate simple harmonic motion in the 'small amplitude' oscillations of a spring–mass system. 2 Apparatus and materials Apparatus 0.6 Retort stand Stop watch Materials Two identical spiral springs Standard masses 0.5 0.4 0.3 0.2 0.1 0 0 0.05 0.10 0.15 Procedure 0.20 0.25 0.30 0.35 0.40 0.45 0.50 Mass, m / kg 1 One spring was suspended from the retort stand and a mass m1 was hung from it. A small vertical displacement was given to it and the time t20 was determined for 20 oscillations and hence the period T1 was obtained for m1. This was repeated for seven other masses. 2 The results were tabulated and a graph plotted to determine the value of the spring constant k. 3 The two springs were suspended in series from the retort stand and mass m1 was hung from the bottom.

The time t20 for 20 oscillations and the period T were determined. The experiment was repeated with the springs in parallel. Results and calculations Results Table SBA.1 Results of oscillations for one spring Mass m / kg Time for 20 oscillations t20 / s Period T/s Period squared T2 / s2 0.10 8.91 0.45 0.20 0.15 10.01 0.51 0.26 0.20 11.37 0.57 0.32 0.25 12.63 0.63 0.40 0.30 14.65 0.73 0.53 0.35 14.03 0.77 0.59 0.40 16.01 0.80 0.64 0.45 17.01 0.85 0.72 Figure SBA.1 Results shown as a graph of period squared (T2) against mass (m, in kilograms). The slope (a) of the graph is given by: y –a ∝ x2 – x1 2 0.72 – 0.18 0.45 – 0.10 = 1.5 s2 kg−1 = The value of the spring constant k in magnitude is given 4 π2 by k = a 4 × π2 spring constant, k = 1.5 = 26.2 The value of k in dimensions is 26.2 N m−1 Spring–mass system with two springs in series with 300 g mass attached The value of t20 = 19.78 s period T = 20 19.78 s = 20 = 0.99 s m ks 4π2m ks = T2 4π2 × 0.3 = 0.992 = 12.07 using T = 2π 459 460 CAPE SBA Spring–mass system with two springs in parallel with 300 g mass attached The value of t20 = 9.78 s period T = 20 9.78 s = 20 = 0.49 s m using T = 2π kp 4π2 × 0.3 kp = 0.492 = 49.3 k 26.2 s = ks 12.07 s = 2.2 k = k2 ks kp 49.3 = the ratio k 26.2 the ratio kp Discussion ■ To compensate for differences in reaction time in starting and stopping the clock and also to avoid measuring small intervals, twenty small oscillations were used. This helped to avoid errors.

■ It was assumed that if the experiment was not affected by wind. ■ The value of the spring constant k on the earth would be the same value on the moon, since the period of the spring–mass system does not depend on the value of gravitational field strength. k k ■ The ratios and ≈ were proved to be equal to two. ks k ■ If the experiment was repeated it could be made more accurate by taking the times at the mid-point of the motion since this is where the speed of the pendulum is greatest. Conclusion Simple harmonic motion was observed in the spring–mass system. = 1.9 ≈2 k ■ To calculate the effective spring constant when two springs with constants k1 and k2 are connected in series this formula is utilized: 1 1 1 ■ = keffective k1 k2 ■ To calculate the effective spring constant when two springs with constants k1 and k2 are connected in parallel this formula is utilized: keffective = k1 + k2 Theory ■ Simple harmonic motion can be defined as the motion of a body in which the acceleration is directly proportional to its displacement from a fixed point and is always directed to that point, m , which gives the period of ■ The equation T = 2π k the motion, was used to determine k, where T is the period of oscillation, k is the spring constant and m is the mass attached.

■ When springs are placed in parallel, the effective spring constant is greater than the spring constant of one spring. Alternatively when springs are placed in series the effective spring constant is less than the spring constant of one spring. Analysis Student A's report was excellent for a number of reasons. 1 The topic and aim were related to the purpose of the experiment. 2 Apparatus and materials were listed separately. This allowed the student to account for everything used. It also made it easier for the examiner to check that everything used was recorded. 3 The 'Procedure' was written in the past tense and the passive voice as required. It was also written in a logical sequence and it was also written in point form, which made it easier to follow. 3 The results were tabulated. The table had a title and each quantity had the correct unit. 5 The graph had the correct quantities and units and the scale was appropriate. 6 In the 'Calculations' there were several good points: ■ each solution had an associated formula ■ the units were correct for each calculation ■ the number of significant figures was kept constant ■ where repeat calculations were required, the first was used as a sample calculation and the result for the others were tabulated. 7 The discussion was well written in a logical order. CAPE SBA 7 In the 'Theory' section a definition was given for the principle being investigated and the associated equation was stated. 0.8 0.7 0.6 8 The account of the precautions taken to avoid error was relevant and specific – though no details were provided of what these precautions were. Period 0.5 T there is one gap in student A's report. 0.4 0.3 1 There is no explanation why a graph of T2 against m gives the value for k in the calculations that immediately follow Figure SBA.1. There is some explanation in the section that considers the two springs. 0.2 0.1 0 0.05 0.10 0.20 0.25 0.30 0.35 0.40 0.45 0.50 Mass, m / kg Figure SBA.2 Results shown as a graph of T2 against m. MM and AI: Student B Calculations Experiment 1 Date Topic Sample calculation: 8.25 T= 20 = 0.41 Simple Harmonic Motion. T2 = 0.17 Aim Springs in series January 20, 2012 To investigate Simple Harmonic Motion in the 'small amplitude' oscillations of a spring–mass system. Ks = 4π2 × 0.2 0.892 Apparatus and materials Retort stand, Stop watch, 2 identical spiral springs and standard masses Springs in parallel Procedure One spring was suspended from the retort stand and a mass m1 was hung from it. The time t20 was determined for 20 oscillations and hence the period T1 was obtained for m1.

This was repeated for seven other masses. The two springs were suspended in series from the retort stand and mass m1 was hung from it.

The time t20 for 20 oscillations and the period T were determined. The experiment was repeated with the springs in parallel. Results Table SBA.2 Mass M Time for 20 oscillations t20 Period T Period squared T2 0.1 8.25 0.41 0.17 0.15 10.68 0.53 0.28 0.2 11.45 0.57 0.32 0.25 12.63 0.63 0.40 0.3 13.90 0.69 0.48 0.35 15.1 0.76 0.58 0.4 16.25 0.81 0.66 0.45 17.57 0.88 0.77 0.15 = 9.97 kg s−2 Kp = 4π2 × 0.2 0.432 = 42.70 kg s−2 Effective spring constant in series = 1 = 1 + 1 kef k1 k2 Proving that k kp = k2 ks k k 2.2 ks 9.97 ks kp 42.70 = 1.99 k 21.47 kp 2 k ≈ –y The slope of the graph is: a = x2 – x1 2 0.669 – 0.154 a = 0.40 – 0.12 a = 1.839 s2 kg−1 461 462 CAPE SBA The value of the spring constant k in magnitude is given (4π2) by k = a (4 × π2) spring constant k = 1.839 K = 21.47 6 The graph actually shows period squared against mass, which is not what the axis labels says. Theory ■ The apparatus and materials were not separated. ■ Simple harmonic motion can be defined as the motion of a body in which the acceleration is directly proportional to its displacement from a fixed point and is always directed to that point, m , which gives the period of ■ The equation T = 2π k the motion, was used to determine k, where T is the period of oscillation, K is the spring constant and m is the mass attached. Discussion 7 Some of the solutions did not have the correct units. 8 The number of significant figures was not consistent. Conclusion The use of both k and K for the spring constant was inconsistent. ORR, MM and AI: Student A Title The specific heat capacity of several metals Date February 3, 2012 One of the sources of error was human reaction time. It was ensured that the amplitude of oscillation was small. Topic The spring mass system, as in this experiment, demonstrates simple harmonic motion.

When the equation F = −kx is satisfied the mass executes harmonic motion with the period.

Aim Conclusion In small amplitude oscillations, simple harmonic motion may be produced. Specific heat capacity To determine the specific heat capacity of three unknown metals x, y and z by the method of mixtures. Apparatus and materials Apparatus 3 Results were tabulated and sample calculation done. Tripod stand Gauze Lid Beakers Bunsen burner Beam balance Thermometer Materials Styrofoam cups String Metals x, y and z 4 The format of the report is as expected. Procedure However, student B's report had the following weaknesses. 1 The masses of the three metals x, y and z were determined using the beam balance. Strings were then attached to them. Analysis Student B's report was just about average. There were some good features in the report. 1 The procedure was in the past tense. 2 It was in a passive voice. 1 The topic and aim were related but topic was lacking in relation to the experiment. 2 The steps in the procedure were not separated into point form, but were instead one long paragraph. 3 The table did not have any title. 4 The quantities in the table had no units. 5 There was no explanation as to why the graph of T2 against m was plotted and what was worked out from the slope of the graph. 2 Water was poured into three beakers labelled x, y and z. The metals were then lowered into there respective beakers, allowing the strings to hang over. The beakers were then placed on the tripod stand allowing the water to boil. 3 Three Styrofoam cups were labelled x, y and z, there masses were determined and water poured into them. The mass of each cup with water was determined. The initial temperature was also measured using a thermometer. CAPE SBA 4 The initial temperature of the boiling water was measured and this was also recorded as the initial temperature of the metals. The hot metals were quickly transferred to their respective cups. Each was gentle stirred and the temperature taken when it became steady. Results Table SBA.3 Showing results obtained for three metals x, y and z Metal y EH = mcΔT heat lost by metal = heat gained by water mmcmΔTm = mwcwΔTw 50.8 × 10−3 × cm × 60 °C = 99.4 × 10−3 × 4.2 × 103 × 4 3.40 × c = 1669.92 1669.92 c = 3.40 = 491.2 J kg−1 °C−1 Metal x Metal y Metal z Mass of metal / kg 270 × 10−3 50.8 × 10−3 369 × 10−3 Mass of empty cup / kg 2.5 × 10−3 2.2 × 10−3 2.2 × 10−3 Mass of cup with water / kg 116.7 × 10−3 101.9 × 10−3 98.6 × 10−3 Mass of water / kg 114.2 × 10−3 99.4 × 10−3 96.4 × 10−3 Metal z EH = mcΔT Initial temperature of water / °C 27 27 26.5 Initial temperature of metal / °C 99 98 99 Final temperature of water / °C 37 31 34 mmcmΔTm = mwcwΔTw Change in temperature of water / °C 10 4 7.5 369 × 10−3 × cm × 64 = 96.4 × 10−3 × 4.2 × 103 × 7.5 Change in temperature of metal /°C 60 67 64 Calculations 1 Sample calculation mass of water = mass of cup and water − mass of empty cup = 116.7 × 10−3 kg − 2.5 × 10−3 kg = 114.2 × 10−3 kg change in temp. of water = final temp. of water − initial temp.

of water = 39 °C − 27 °C = 12 °C change in temp of metal = initial temp of metal − final temp of metal and water = 99 °C − 39 °C = 60 °C 2 Specific heat capacity of water = 4200 kg−1 °C−1 Metal x EH = mcΔT heat lost by metal = heat gained by water mmcmΔTm = mwcwΔTw 270 × 10−3 × cm × 60 = 114.2 × 10−3 × 4.2 × 103 × 12 16.2 × cm = 5755.68 5755.68 cm = 16.2 = 355.3 J kg−1 °C−1 heat lost by metal = heat gained by water 23.62 × c = 3027.15 3027.15 c= 23.62 = 128.16 J kg−1 °C−1 Theory ■ The specific heat capacity of a substance or an object is the heat energy required to change the temperature of 1 kg of the metal by 1 K (1 °C). The specific heat capacity depends on the material of the object or substance. ■ To determine specific heat capacity, the following equation is utilized: EH = mcΔT where EH is the heat energy used up or given off when the temperature of the object changes, c is the specific heat capacity of the object and ΔT is the change in temperature of the object. Discussion Styrofoam cups were used in this experiment because they are good insulators of heat and therefore reduce heat loss to the environment, hence improving the accuracy of the results. A lid was also used and this helped to reduce heat loss to the environment. It was ensured that the metals were transferred quickly to reduce heat loss. Care was taken to prevent the splashing of any of the water from the containers. Care was taken to prevent zero error on the beam balance and parallax error on the thermometer. If the experiment should be repeated it would be best to use a digital balance instead of the beam balance, as this would give more accurate readings. 463 464 CAPE SBA ORR, MM and AI: Student B Conclusion The specific heat capacity of metals x, y and z are as follows: Metal x: 355.3 J kg−1 °C−1 Metal y: 491.2 J kg−1 °C−1 Metal z: 128.16 J kg−1 °C−1 Title Date Metal z: 491.2 J kg−1 Experiment 3 February 3, 2012 −1 −1 −1 Topic Analysis Specific heat capacity Aim Student A presented an excellent report and would score highly in all areas. To determine the specific heat capacity of three unknown metals x, y and z by the method of mixtures. The good points are as follows: Apparatus and materials The aim was related with the overall purpose of Tripod stand, Gauze, Lid, Bunsen burner, Beam balance, Thermometer, beakers, string, metals x, y and z. the experiment. ■ The apparatus list and materials list were comprehensive with apparatus separated from materials. Procedure ■ The procedure was concise, in a logical order, in the passive voice and in the past tense. ■ Importantly the procedure was in point form that made it easy to follow. ■ The use of the table made the results clear and easily understood. ■ The table had a title, columns and rows were labelled correctly and included quantity and units. ■ Sample calculations were used to show how results were derived in the table. ■ Each calculation was preceded by the associated formula. ■ The final answers had the correct units.

■ The discussion explained the topic, gave precautions taken, and sources of error. ■ The conclusion gave the solution sought by the aim. ■ Number of significant figures was consistent; this was facilitated by the use of standard form. Although the report was excellent there were a few weaknesses. ■ The equations in the equations were without units, hence it was not shown how the unit for the final answer was derived. Procedure The masses of the three metals x, y and z were determined using the beam balance. Strings were then attached to them. Water was poured into three beakers labelled x, y and z. The metals were then lowered into their respective beakers, allowing the strings to hang over. The beakers were then placed on the tripod stand allowing the water to boil. The initial temperature of the boiling water was measured and this was also recorded as the initial temperature of the metals. The mass of each empty Styrofoam cup was determined, they were then labelled x, y and z, their masses were determined and water poured into them. The mass of each cup with water was determined. The initial temperature was also measured using a thermometer. The initial temperature of the boiling water was measured and this was also recorded as the initial temperature of the metals. The hot metals were quickly transferred to their respective cups.

Each was gently stirred and the temperature taken when it became steady. Results Table SBA.4 Metal x Metal y Metal z Mass of metal / kg 0.05 0.085 0.0057 Mass of empty beaker / kg 2.5 × 10−3 2.6 × 10−3 2.5 × 10−3 Masses of water and beaker / kg 0.0849 0.0969 0.0883 Mass of water / kg 0.0824 0.0843 0.0868 Initial temperature of water / °C 24 24 25 Initial temperature of metal / °C 99 98 100 Final temperature of metal and water / °C 30 37 31 ■ The report did not tell the importance of making sure Change in temperature of water / °C 6 13 6 that the metals were totally immersed in the water. Change in temperature of metal /°C 69 61 69 ■ There are minor inconsistencies with the symbols used (sometimes c, sometimes cm). ■ Spelling is poor in some places, in particular the difference between 'there' and 'their'. CAPE SBA Calculations Specific heat capacity of water = 4200 J kg−1 °C−1 A lid was used and this helped to reduce heat loss to the environment. Metal x EH = mcΔT heat lost by metal = heat gained by water mmcmΔTm = mwcwΔTw 0.05 kg × cm × 69 °C = 0.0824 kg × 4200 J kg−1 °C−1 × 6 °C 3.45 kg °C × cm = 2076.48 J It was ensured that the metals were transferred quickly to reduce heat loss. Care was taken to prevent the splashing of any of the water from the containers. Care was taken to prevent zero error on the beam balance and parallax error on the thermometer. Conclusion The specific heat capacities are: 2076.48 J cm = 3.45 kg °C−1 cm = 601.88 J kg−1 °C−1 metal x: 601.88 J/kg/°C metal y: 992.1 J/kg/°C Metal y metal z 482.86 J/kg/°C EH = mcΔT heat lost by metal = heat gained by water Analysis mmcmΔTm = mwcwΔTw Student B's report was fairly good. However, there were some flaws. 5.19 kg °C × cm = 5148.78 J The good points were: 5148.78 J 5.19 kg °C−1 cm = 992.1 J kg−1 °C−1 ■ The report was well organized and contained all of the required features. −1 −1 cm = Metal z EH = mcΔT heat lost by metal = heat gained by water mmcmΔTm = mwcwΔTw 0.0657 kg × cm × 69 °C = 0.0868 kg × 4200 J kg−1 °C−1 × 6 °C 4.53 kg °C × cm = 2187.36 J 2187.36 J 4.53 kg °C cm = 482.86 J kg−1 °C−1 cm = Theory The specific heat capacity of a metal is the heat energy required to change the temperature of 1 kg of the metal by 1 K (1 °C). The specific heat capacity depends on the type of metal and its mass of it. To determine specific heat capacity, the following equation is used: EH = mcΔT where EH is the heat energy used up or given off when the temperature of the object changes, c is the specific heat capacity of the object and ΔT is the change in temperature of the object. Discussion Styrofoam cups were used in this experiment because they are good insulators of heat and therefore reduce heat loss to the environment, hence improving the accuracy of the results. ■ The results were placed in a table that had a title, quantities and correct units. 3 The procedure was written in a logical order, in the passive voice and the past tense. 4 The calculations were very good; the formulas were given and the quantities were written with the correct units. The units for the answers were also correct. 5 The conclusion was to the point and offer the solution the aim seeks. The bad points are: 1 There is no comment about the wide differences between the metals. 2 The number of significant figures used in the table was not constant. Standard form could have been used to make it easier.

3 The procedure should have been written in point form instead of one paragraph.

This would make it easier to follow. 4 Student B did not appreciate the use of a less heat conductive material (for example, Styrofoam) to put water and metal in. 5 Specific heat capacity doesn't depend on the mass of the material (as stated in the Theory section). 6 At this level it is more usual to use '−1' rather than '/' in 465 466 CAPE SBA units such as J kg−1 °C−1 (not J/kg/°C). Both forms are used by Student B.

P&D: Student A In these examples no results are included, because it is the structure of the work which is being assessed. Title [No title was provided] Date March 3, 2012 Topic Specific latent heat of vaporization and fusion, for 'Planning and Design' Problem Theory suggests that more energy is required to convert one kilogram of water to steam than to convert the same quantity of ice to water. Plan and design an experiment, using an electrical method, to investigate this statement. Hypothesis More energy is required to convert 1 kilogram of water to steam than to convert the same quantity of ice to water. 4 While evaporation taking place, take the reading on the voltmeter and ammeter. 5 The equation: energy = voltage × current × time E = VIt can then be used to calculate the energy. B Specific latent heat of fusion of ice 1 Set up a series circuit with the apparatus to supply energy to the heater. 2 Place the heater in a beaker with ice (same mass as the water used previously). 3 Close the heater switch and simultaneously start the clock. 4 While melting is taking place, take the voltmeter and ammeter readings. 5 Take the time when all the ice is melted. The equation energy = voltage × current × time E = VIt can then be used to calculate the energy. Aim To determine if more energy is required to convert one kilogram of water to steam than to convert one kilogram of ice to water. Apparatus and materials Electrical heater, digital scale, 2 beakers, with water, beakers, thermometer, a rheostat, an electrical heater, a voltmeter across the heater and an ammeter in series. Place the heater in the beaker with water. Turn on the heater and check the temperature. When it reaches 100 °C, start the clock and note the time it took for all the water to vaporize. Take the readings from voltmeter and ammeter. Use the formula: energy = voltage × current × time to find energy. Apparatus: Stop watch Thermometer Rheostat Ammeter Voltmeter Ammeter Materials Water Ice Procedure A Specific latent heat of vaporization 1 Set up a series circuit with the apparatus to supply energy to the heater. Formula 2 Pour some water (note the mass) in the beaker, then with the heater immersed in the beaker close the switch. 2 EH = m lf 3 When the temperature is steady at 100 °C, start the stop watch. Record the time it takes for all the water to change to steam. 1 EH = m lv 3 EH = IVt ∝ m lf 5 IVt ∝ m lf 5 IVt = m lv Current Time Energy CAPE SBA Theory The bad points were the following. The specific latent heat of fusion (lf) of a substance is the energy required to change unit mass of the substance from solid to liquid without a change in temperature. The unit is J/kg. 1 There is no title. Equation: 2 The aim was changed to one which is not the same as the problem. 3 The headings in the procedure were not written in point form. 4 The units are missing from the table. EH = m lf P&D: Student B Theory The specific latent heat of vaporization (lv) of a substance is the energy required to change unit mass of the substance from liquid to vapour without a change in temperature. The unit is J/kg.

The difference between the specific latent heats of fusion and vaporization. Date March 3, 2012 Topic Equation: Specific latent heat of vaporization and fusion. EH = m lv Problem where lv is the latent heat of vaporization, EH is the energy used up and m is the mass of the substance. Variables Controlled: mass of water and ice Manipulated: current Theory suggests that more energy is required to convert one kilogram of water to steam than to convert the same quantity of ice to water.

Plan and design an experiment, using an electrical method, to investigate this statement. Hypothesis Responding: time for evaporation and melting Precautions 1 Ensure none of the water spills. 2 Take care to stop the clock when all the ice is melted. Sources of error It takes more energy to convert 1 kilogram of water to steam than to convert the same quantity of ice to water. Precautions Aim 1 Ensure that the rheostat is adjusted to maintain the current necessary. 2 Ensure stop watch is started when the switch is closed. Limitations 1 Faults occurs in the circuit. 2 Error in reading the instruments. Analysis The report is a good one for many reasons. 1 The aim is related to the problem and relevant to the experiment. 2 The hypothesis was well stated.

3 The procedure was written as instructions, which were clear and precise. 4 The procedure was written in a logical order. 5 The table of expected results was drawn. 6 Variables were identified. 7 The formulae to be used were given. 8 A diagram was given. To determine if more energy is required to convert one kilogram of water to steam than to convert one kilogram of ice to water. Apparatus and material Electrical heater, digital scale, 2 beakers, with water, beakers, thermometer. Place one beaker containing water on a balance. Set up a series circuit with the source, a rheostat, an electrical heater, a voltmeter across the heater and an ammeter in series. Place the heater in the beaker with water. Turn on the heater and check the temperature. When it reaches 100 °C, start the clock and note the time it took for all the water to vaporize. Take the readings from voltmeter and ammeter. Use the formula: energy = voltage × current × time to find energy. Repeat the experiment with ice. Place a container of ice in a container, immerse the heater in the ice. Turn on the heater while simultaneously start the clock. Check also the voltage across and current through the heater. Use the formula: energy = voltage × current × time to calculate the energy used. 467 468 CAPE SBA Analysis Expected results Table SBA.5 The report was an average one. The good points were: Voltage / V Current / A Time / s Energy / J Evaporation Melting 1 The topic and aim were appropriate for the experiment. 2 The hypothesis was good. 3 The procedure was written in a logical order. 4 The table was complete with both quantity and units. EH = m lv Formulae were given that would be used. EH = m lf The bad points were: EH = VIt 1 The procedure was written in point form; which would make it easier to follow. VIt = m lf VIt = m lv Theory The specific latent heat of fusion (lf) of a substance is the energy required to change unit mass of the substance from solid to liquid without a change in temperature. The unit is J/Kg. Equation: EH = m lf where lf is the latent heat of fusion of the substance, EH is the energy used up to change the state and m is the mass of the substance. The specific latent heat of vaporization (lv) of a substance is the energy required to change unit mass of the substance from liquid to vapour without a change in temperature. The unit is J/kg. Equation: EH = mlv where lv is the latent heat of vaporization of the substance, EH is the energy used up to change the state and m is the mass of the substance. Variables Controlled: mass of water and ice Manipulated: substance used Responding: time taken for evaporation and melting Precautions 1 Ensure none of the water spills. 2 Take care to stop the clock when all the ice is melted. Sources of error Parallax error when reading thermometer. Human reaction time. 2 A diagram should be clear. 3 Some of the apparatus and materials were not mentioned. 4 There is no comment about the truth or otherwise of the hypothesis. 469 Appendix 1: Base unit definitions in the S.I. The metre (symbol m) is the length of

the path travelled by light in vacuum during a time interval of 1/299 792 458 of a second. The kilogram (symbol kg) is equal to the mass of the international prototype of the kilogram (a platinum- iridium mass kept at the International Bureau of Weights and Measures in France). The second (symbol s) is the duration of 9 192 631 770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the caesium-133 atom. The ampere (symbol A) is that constant current which, if maintained in two straight parallel conductors of infinite length, of negligible circular cross-section, and placed 1 metre apart in vacuum, would produce between these conductors a force equal to 2 × 10−7 newton per meter of length. The kelvin (symbol K), the unit of thermodynamic temperature, is the fraction 1/273.16 of the thermodynamic temperature of the triple point of water. The mole (symbol mol) is the amount of substance of a system which contains as many elementary entities as there are atoms in 0.012 kilogram of carbon-12. [This corresponds to the Avogadro constant, 6.022 141 5 × 1023, but this number isn't part of the 'official definition'.] More information can be found at Appendix 2: Prefixes and conversions Some prefixes used in the S.I. …

= a2 − b2 ■ This formula was used in Chapter 2, page 22, in deriving the equation of motion v2 = v02 + 2a(x − x0) Help in obtaining various types of regression graphs and equations using paired data can be found in Excel®, many graphing calculators or on the internet. Regression may include a choice among linear, polynomial, logarithmic or exponential best fits to the data. Dimensional consistency of equations One way of checking the consistency of units in an equation is through dimensional analysis. This topic is discussed on page 5. 4: Trigonometric relationships Pythagoras' theorem Definitions of sin, cos and tan For the right-angled triangle ABC shown in Figure A-2, where angle B is a right angle (i.e. b is a hypotenuse), Pythagoras' theorem states that The following are definitions of sine (sin), cosine (cos) and tangent (tan) of an angle, using the right-angled triangle in Figure A-2: b2 = a2 + c2 sin A = opposite side a = hypotenuse b cos A = adjacent side c = hypotenuse b tan A = opposite side a = adjacent side c C b A a c B Figure A-2 Note: A, B and C represent angles BAC, CBA and ACB, respectively. 471 472 Appendix Two angles for each of sin, cos and tan The general triangle It can be shown that there are two angles that have each the same sin, cos and tan. Try various angles for yourself using your calculator and see that this is indeed so … For any general triangle, ABC (Figure A-3) a2 = b2 + c2 − 2bc cos A and a = b = c sin A sin B sin C C sin A = sin (180° − A) cos A = −cos (180° − A) b a tan A = tan (180° + A) A ■ Worked example 3.1 illustrates why one needs to be aware of two solutions when working out angles associated with sin, cos or tan. Some special right-angled triangles ■ Triangles whose sides are in the ratio 3 : 4 : 5 or 5 : 12 : 13 are right-angled triangles. You can confirm this using Pythagoras' theorem. ■ Referring back to Figure A-2, if angle A = 30°, 1 and since sin 30° = 2 , then it can be shown, using Pythagoras' theorem, that 1 2 = cos 60° 1 = sin 60° cos 30° = √3 ■ If a = c, then angle A = 45° and tan 45° = 1 sin 30° = Other trigonometric relationships sin 2θ = 2 sin θ cos θ c B Figure A-3 Angles measured in radians The radian measure of an angle, θ, is given by θ= arc length, s, of a circle, subtended by angle radius, r, of the circle θ= s r For small angles in radians, the following hold true: sin θ ≈ θ cos θ ≈ 1 tan θ ≈ θ This small angle approximation was used in Chapter 7, page 112 (derivation of period, T, for a simple pendulum), and in Chapter 9, page 150 (Worked example 9.6, Diffraction). sin2 θ + cos2 θ = 1 ■ The equation for the range of a projectile (case 2, page 29) was derived using the relationship sin 2θ = 2 sin θ cos θ 5: Logarithms ■ Logarithms were introduced in Chapter 10, pages 165 to 167, in the section 'Response of the human ear to sound levels'. Logarithms to base 10 (common logarithms or 'logs') The logarithm of a number is the power to which a base must be raised in order to obtain the number. Thus, if a base of 10 is used, we say that the 'log' of 1000 is 3, because 103 = 1000. Similarly, the log of 1 is 0 since 100 = 1. Thus, Logarithms to other bases Other bases are in use. We use the term 'log' when the base is 10. The term 'natural log' (abbreviated ln) applies when the base is the irrational number e = 2.718… Rules of logarithms If we let A = 103 and B = 102, then log (AB) = log 105 = 5 But, log A + log B = log 103 + log 102 = 3 + 2 = 5 Hence, we see that log 10x = x(i) log (AB) = log A + log B(ii) ■ We can use a calculator to find the log of a number.

Other rules include: A log = log A − log B (iii) B log Ax = x log A(iv) ■ If we are given the log of a number we can use the 'log−1' (log inverse) function on a calculator to find the number.

( ) Appendix Equations (i) to (iv) hold true for any other base. For base e, for example, ln ex = x(i)(e) ln (AB) = ln A + ln B(ii)(e) A ln = ln A − ln B (iii)(e) B ( ) ln A = x ln A(iv)(e) x The decibel The ratio of comfortable hearing of sound of greatest intensity (Im) to the threshold intensity of hearing (I0) is approximately 1 W m−2 to 10−12 W m−2, or 1012 (which is 1 trillion!). Sound level, β, measured in bels (B), corresponding to a sound of intensity I is therefore defined instead by the logarithmic ratio I (10.4) β (B) = log10 I0 ( ) For a range of intensities from maximum intensity of 1 W m−2 to the threshold intensity of 10−12 W m−2, the logarithmic ratio log10 II ranges from 0 to 12, which is a 0 more easily comprehensible range than from 1 to 1 trillion. ( ) 1 The decibel is 10 of a bel, and therefore, I β (dB) = 10 log10 I0 ( ) giving the usual range of values for sound levels in decibels from 0 to 120. Logarithmic scales and graphs The equal intervals on a logarithmic scale do not represent equal amounts of a physical quantity. Logarithmic scales are, however, useful when the range of values to be plotted is very large, as in the case of frequencies that can be heard by humans, ranging from 20 Hz to 20 kHz. Figure 10.11 (page 167) shows a good example of a logarithmic scale of frequencies. The dB scale shown in Figure 10.11 is also logarithmic. Note that it is not true to say that 2 dB + 3 dB = 5 dB, since the decibel is a logarithmic ratio. Worked example 10.6 (page 166) illustrates this. In Figure 23.18b (page 375), both the dB and the frequency scales are logarithmic since numerical ranges in op-amp measurements are huge. 6: Binary numbers The base of 2 is used to express numbers in binary form, especially in digital electronics. Binary and decimal 7: Graphical analysis numbers are discussed on pages 390 and 391, especially as binary operations apply to logic gates. V1 v1 = mv2 + c Linear graphs from linear equations Linear graphs are obtained from linear equations. A linear equation will be of the form v1 = mv2 + c slope of graph = m c or y = mx + c where v1, v2, y and x are variables, and m and c are constants. When v1 is plotted on the vertical axis against v2 on the horizontal axis (or y is plotted on the vertical axis against x on the horizontal axis), a linear (straight line) graph is always obtained. The slope or gradient of the straight line is m and the v1- (or y-) intercept is always c (see Figure A-4). Conversely, if a linear graph is obtained experimentally when a variable v1 is plotted against another variable v2, V2 Figure A-4 the equation describing the relationship between v1 and v2 is given by v1 = mv2 + c 473 474 Appendix Obtaining linear graphs from non-linear equations The simple pendulum equation t ln V = ln [V0 e−(RC)] l T = 2π g where T and l are variables and g is a constant, does not yield a linear graph if T is plotted against l. To determine the constant, g, from a linear graph based on this equation, we square the equation and get T2 = Taking the natural logarithm of both sides of the equation, we get 4π2 l g ( ) A graph of T2 against l will yield a straight line, since these 2 is a are the two variables, v1 and v2, respectively, and 4π g 2 4π constant.

The slope of the line will be g , from which g can be determined. ( ) ( ) In Chapter 2, we see two examples of linear graphs, one from a linear equation and the other from a non-linear equation: ■ Figure 2.6: x = x0 + vt; with a plot of x versus t 1 ■ Figure 2.7b: x = 2 at2; with a plot of x versus t2 In both these examples, x and t (and hence t2) are variables and x0 and a are constants. Applying logarithms to non-linear equations t = ln V0 + ln e−(RC) −t = ln V0 + RC −t + ln V0 = RC This last equation is of the form y = mx + c −1 . Hence, if ln V is plotted against t, the slope would be RC If R is known, the slope can therefore be used to determine C. Such a graph is shown in Figure 18.11. Graphs of best fit Graphs plotted using data obtained experimentally are not usually 'smooth' when the points are simply joined. This is because of experimental error (usually random error) in the measurements.

A smooth curve or straight line is usually drawn by 'eye-balling' what seems to be the most reasonable best fit pattern. A linear graph of best fit is shown in Figure A-5.

A non-linear graph of best fit is shown in Figure A-6. y Logarithms can be used to investigate the value of the exponent in a non-linear relationship. For example, if H = kLb, where H and L are variables and k and b are constants, we take the log of both sides of the equation and get log H = log kLb = log k + log (Lb) = log (Lb) + log k = b log L + log k If log H is plotted against log L, a linear graph is obtained. The slope of this graph is b. x Figure A-5 y The constant k can also be determined from the graph since the log H intercept will be log k, from which k can be calculated. Chapter 18, page 286, discusses an example of the application of natural logarithms (ln) to the equation curve of best fit t V = V0 e−(RC) x Figure A-6 Appendix 8: Vectors D Vector addition and subtraction by drawing A vector quantity can be represented in magnitude and direction by a directed straight line drawn to scale. Addition of vectors can be done: (a) by placing the vectors tip-to-tail, in which case, the resultant vector is one drawn from the tail of the first vector to the tip of the last vector; (b) by placing the vectors tail-to-tail, in which case, the resultant vector is obtained by first completing the parallelogram whose two adjacent sides are represented by the two vectors. The diagonal drawn from the two tails to the opposite corner of the parallelogram represents the resultant vector. For the special case in which two vectors, AB and BC are at right angles to each other, the resultant, R, is the hypotenuse of a right-angled triangle drawn from B, the point of intersection of the two tails (Figure A-7). Do not make the mistake of treating either AC or CA, rather than BD, as the resultant vector.

A D R B C Figure A-7 ■ See pages 11 and 12 for a treatment of tip-to-tail vector addition and subtraction. ■ See pages 11 and 12 for a treatment of tail-to-tail vector addition and subtraction (parallelogram method). Orthogonal components of a vector Any vector, V, can be replaced by two vectors, V x and V y, drawn at right angles to each other, as shown in Figure A-8. V x and V y are adjacent sides to a rectangle. V x and V y are called orthogonal components of the vector V. Vy V α θ Vx Figure A-8 The geometry of Figure A-8 shows that: V = √Vx2 + Vy2(i) (Note that the V's that are not bold represent magnitudes of the respective vectors.) Vx = V cos θ(ii) Vy = V sin θ(iii) Vy (iv) tan θ = Vx For these relationships to hold, it is important that the angle θ be always drawn between V x and V, and is designated positive when drawn anticlockwise from V x. A treatment of orthogonal components of vectors can be found on page 12. (Note that the orthogonal component of a vector in any direction is the projection of the vector along that direction.

As such, the orthogonal component is always the magnitude of the vector times the cosine of the angle between the vector and its component. Thus, Vx = V cos θ Vy = V cos α = V sin θ, since, by trigonometry, cos α = sin θ (see Figure A-8).

[illegible]

been authorized, sponsored or otherwise approved by CXC. • Case studies provide examples for reference and help students to develop critical thinking skills • Multiple choice and extended essay-type questions help to assess progress through the syllabus and prepare students for their exams • Definitions, examples from various Caribbean territories and a range of exercises enable students to engage actively and relate to the content Helen Jacobs Novelette Sadler-McKnight Stewart McLean Patrice Piggot-Cumberbatch Graeme Corbin Mike Taylor CHEMISTRY CHEMISTRY FOR CAPE EXAMINATIONS FOR CAPE EXAMINATIONS ® CAPE® is a registered trade mark of the Caribbean Examinations Council (CXC). Chemistry for CAPE® Examinations is an independent publication and has not been authorized, sponsored or otherwise approved by CXC. Other titles for CAPE® Examinations: CONNECT WITH US: Find us on Facebook www.facebook.com/macmillancaribbean ® Jacobs Sadler-McKnight McLean Piggot-Cumberbatch Corbin Taylor Dr. Novelette Sadler-McKnight is a lecturer and researcher of Chemistry at the University of the West Indies, Mona. She has been the coordinator for workshops on Spectroscopy and Chromatography for students and teachers for CAPE Chemistry since 2003. Dr. Stewart McLean graduated in Chemistry from Glasgow University and earned his Ph.D. in Organic Chemistry from Cornell University. As Assistant Professor at the University of Toronto he established a long lasting collaboration with chemists at the University of the West Indies and the University of Guyana. Dr. Patrice Piggot-Cumberbatch is an Assistant Professor at the University of Trinidad and Tobago. She obtained her B.Sc. and Ph.D. degrees from the University of the West Indies and has been actively involved in academia ever since, having authored a book and five research papers for international journals to-date. Dr. Graeme Corbin is an educator at Harrison College, Barbados.

He holds a B.Sc. (Hon) in Chemistry & Biology and Ph.D. in Chemistry, both from the University of the West Indies Cave Hill, Barbados. Dr. Mike Taylor has taught Chemistry at CAPE level for many years in schools, colleges and at university. He is regarded as an expert in matching explanations of the concepts of modern chemistry to the needs of particular student groups. www.macmillan-caribbean.com FOR CAPE EXAMINATIONS Chemistry for CAPE® Examinations provides comprehensive coverage of Units 1 and 2 of the CAPE Chemistry syllabus. It has been written with the expectations of prospective employers and university teachers in mind. This text develops an enhanced understanding of fundamental chemical principles, specific knowledge of chemical processes, materials, and applications, and an awareness of important chemical industries and their effects on our environment. It encourages the ability to apply scientific methods to a range of academic and everyday problems. Key features: • Worked examples and in-text questions help students build their understanding and provoke analytical thought • Review questions (including answers) at the end of each chapter test knowledge and how to apply it • The School-Based Assessment section gives step-by-step tips to maximise CAPE coursework success Authors Dr. Helen Jacobs is currently Professor of Organic Chemistry at the University of the West Indies, Mona. She has been involved for over 15 years in the delivery of the first year undergraduate chemistry course at UWI. CHEMISTRY ® • The School-Based Assessment section gives in-depth tips to maximise CAPE coursework success. Other titles available for CAPE® Examinations: I S B N 978-0-230-03782-3 Find us on Twitter @MacCaribbean 9 780230 037823 9780230037823_Cover.indd 1 07/05/2014 14:02 JEROME PITTERSON MANAGEMENT OF BUSINESS FOR CAPE® EXAMINATIONS JEROME PITTERSON • A variety of illustrations, photographs, diagrams, tables and worked examples support challenging areas of the text • Keywords are highlighted to develop vocabulary MANAGEMENT OF BUSINESS FOR CAPE® EXAMINATIONS Dr. Jeniffer Mohammed is an experienced educator and author and is currently Head of the School of Education at the University of the West Indies, St. Augustine. Her expertise lies in the sociology of education, the sociology of health, the teaching of social studies and qualitative research. JENIFFER MOHAMMED • A research component at the end of each chapter provides guidance on how to complete a study based on the themes covered I S B N 978-0-230-40043-6 www.macmillan-caribbean.com Find us on Facebook Follow us on Twitter www.facebook.com/macmillancaribbean @MacCaribbean 9 780230 400436 CAPE MoB Cover.indd 1 Find us on Facebook /macmillancaribbean Find us on Twitter @MacCaribbean www.macmillan-caribbean.com 25/11/2014 14:33 I S B N 978-0-230-43157-7 9 780230 431577 Caribbean Studies for CAPE_Final.indd 11 Find us on Facebook /macmillancaribbean Find us on Twitter @MacCaribbean www.macmillan-caribbean.com 02/02/2015 14:23 I S B N 978-0-230-03785-4 9 780230 037854 ALEC FARLEY AND DAVID GLOVER •The School-Based Assessment section gives step-by-step tips to maximise CAPE coursework success PHYSICS CAPE® is a registered trade mark of the Caribbean Examinations Council (CXC). Physics for CAPE® Examinations is an independent publication and has not been authorized, sponsored or otherwise approved by CXC. FOR CAPE® EXAMINATIONS Physics for CAPE® Examinations provides comprehensive coverage of Units 1 and 2 of the CAPE Physics syllabus. A non-calculus treatment of Physics, it has been written to facilitate a smooth transition from high school to tertiary level Physics at regional and international institutions.