

PART TWO

VIBRATIONS AND WAVES

Having explored the behaviour of matter and the principles of heat in Part One, we now turn to motions that repeat and disturbances that travel through matter and space. Vibrations and Waves examines oscillatory motion such as simple harmonic motion, the propagation of mechanical waves, the nature of sound, the behavior of light in physical optics, electromagnetic waves, and the Doppler effect. These ideas explain how energy is transferred without the permanent movement of matter and help us understand systems ranging from vibrating strings and sound waves to optical instruments, radio transmission, and modern communication technologies. Mastering this part will deepen your understanding of periodic motion and wave behaviour, and prepare you for more advanced studies in modern physics and applied science.

Chapter 6

SIMPLE HARMONIC MOTION

INTRODUCTION

The trouble began, as it often does at Miono Secondary, on a Tuesday afternoon at the school sports field. Kipanga was attempting, with limited success, to push Kipute on the metal swing near the boundary wall. Each time she reached the lowest point and started moving forward, his push arrived a fraction too late, by which point she was already pulling away from him, and his hand bumped gently against the back of an empty seat. On the return swing he pushed too early. By the seventh attempt the swing was making the kind of half-hearted excursion a swing makes only when it suspects it is being mocked.

“Kipanga,” Kipute said over her shoulder, in the tone she reserved for moments when his physics required correction, “you are fighting the swing.”

“I am pushing the swing.”

“You are pushing it at the wrong time. Wait until I reach the bottom and am moving away from you. Push then. Then let me come back. Then push at exactly the same moment of the next cycle. Match the swing’s own rhythm. Stop arguing with it.”

Kipanga waited. He pushed once, gently, at exactly the moment Kipute crossed the lowest point. He waited again. He pushed at the same moment of the next cycle. By the fifth push the swing was rising satisfyingly. By the tenth, Kipute was higher than the boundary wall and Kipanga was looking unreasonably proud of himself.

“Why,” he asked the air, “does it matter when I push?”

Mr. Akilikubwa, who had a habit of materialising at boundary walls when good questions were imminent, materialised at the boundary wall. He had the contented look of a man who had been waiting twenty-three years for exactly this question.

“Because the swing has its own natural frequency,” he said. “It wants to oscillate at one particular rate, determined entirely by its length and the local strength of gravity. If you push at that rate, in resonance with its rhythm, every push adds a little energy and the amplitude grows. If you push at any other rate, you fight the swing and your energy is wasted, mostly into the muscles of your arm. Kipute discovered this by instinct. Newton discovered it by mathematics. The mathematics begins in this chapter.”

He walked over, caught the seat, and brought the swing gently to rest. *“This swing is the simplest oscillator in nature. One restoring force, one natural frequency, one rhythm. But the same mathematics governs a car suspension, a guitar string, a quartz crystal in your phone, the columns of air in your lungs when you speak, and the basilar membrane in your inner ear that lets you tell those columns apart. All of them are, in their hearts, the same thing. They obey one equation, and that equation rests on one observation: when you pull a thing away from where it wants to be, the harder you pull, the harder it pushes back. Force proportional to displacement, always pointing back home. That is **simple harmonic motion**. That is what this chapter is about.”*

Welcome to Chapter 6. The previous five chapters spent themselves on matter that pushes back when you squeeze it (Chapter 1), matter that warms up when you heat it (Chapter 2), matter that conducts that warmth to its neighbours (Chapter 3), matter that keeps its temperature by the busy random motion of its molecules (Chapter 4), and matter that converts heat into work in heat engines and into nothing useful in everyday life (Chapter 5). All of that was about state and slow change. This chapter is the first to be about repetition. It is about systems that do the same thing, over and over, at the same rate, until they get tired and stop, and then start again the moment something gives them a nudge in time.

If you remember the molecular potential well from Chapter 1 (the picture in which two bonded molecules sit at the bottom of a curve that looks like a spring near its minimum), you already know why simple harmonic motion is universal. Anything that lives near the bottom of a smooth potential well, when displaced a little, behaves like a tiny mass on a tiny spring. The world is full of such wells, and this chapter is about the motion that happens inside them.

We also have an old promise to keep. Back in Chapter 1, when we discussed the shear modulus, an experimental method for measuring it was deferred: the one in which a wire is suspended from a rigid support, a horizontal disc is fixed at its lower end, and the disc is set into torsional oscillations. We could

not write down the period of those oscillations at the time, because the period belongs to the language of simple harmonic motion. We will keep that promise in this chapter, in the section on torsional oscillations, where the answer comes out so cleanly it almost feels unfair.

And we will need a third tool from earlier in the series. The chapter on the rotation of rigid bodies introduced the moment of inertia, the parallel axis theorem, and the radius of gyration. Those ideas were treated as endpoints in their own right. Here, they become starting points. Three sections of this chapter, on the compound pendulum, the torsion pendulum, and the rolling oscillation in a bowl, could not be written without them. If those concepts are not fresh in your memory, take a moment to revisit the rotational dynamics chapter before you reach the second half of this one. The calculation style here is direct, but only if the moment of inertia is not a mystery.

THE OSCILLATION VOCABULARY

Before we write down a single equation, we need to agree on a small vocabulary. Almost every word in this chapter will refer to some feature of a body that is moving back and forth. If the words are slippery, the equations will feel slippery; if the words are sharp, the equations will feel inevitable. So we begin slowly.

Periodic motion and oscillation

A motion is **periodic** if *it repeats itself, exactly, after a fixed interval of time*. The motion of the hands of a clock is periodic. The orbit of the moon around the Earth is periodic. The bouncing of a ball on a hard floor, if no kinetic energy were lost, would be periodic. Periodic motion need **not** pass through the same position twice in opposite directions; the moon, for example, simply returns to where it began without ever reversing its direction of travel. Repetition is enough.

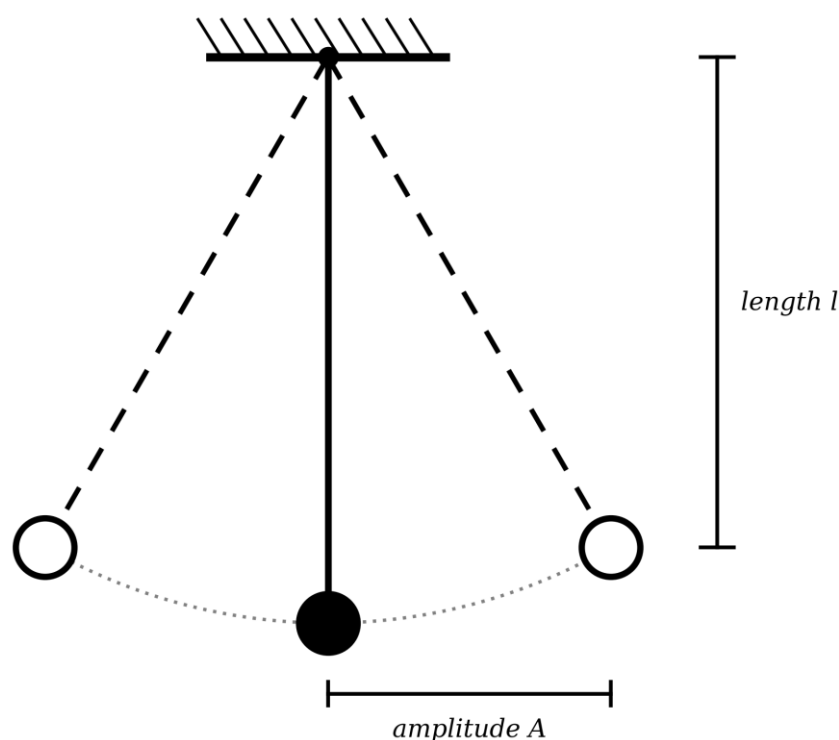
An **oscillation** is the special kind of periodic motion that involves *a back-and-forth excursion about a fixed central position*. A swing oscillates. A clock pendulum oscillates. A guitar string oscillates. A child on a trampoline, neglecting air resistance, oscillates. The moon does not oscillate. Every oscillation is periodic, but not every periodic motion is an oscillation. In this chapter the word *oscillation* will always mean a back-and-forth motion of this sort.

Equilibrium position, displacement, and amplitude

The fixed central position about which the body oscillates is called the **equilibrium position**. It is *the position the body would settle at if it were left alone with no nudge*: the bottom of the swing, the natural length of the spring, the still surface of a pond. In every oscillation we will study, the equilibrium position is the position at which the net force on the body is zero. Pull the body away from this position and a force appears pulling it back.

The **displacement**, written x , is *the signed distance of the body from its equilibrium position at any instant*. We choose one direction along the line of motion to be the positive direction; then x is positive on that side of equilibrium and negative on the other. Displacement is therefore a number that may be positive, negative, or zero, and which changes continuously as the oscillation proceeds. At the equilibrium position itself, $x = 0$.

The **amplitude**, written A , is *the largest value the displacement ever takes during an oscillation*. It is always a positive number. The body therefore moves between $x = +A$ and $x = -A$ on either side of equilibrium, and the **total length** of its excursion from one extreme to the other is therefore $2A$. For a pendulum we may use the angular amplitude (the largest angle from vertical) instead of a linear amplitude, depending on which is more convenient; the following figure shows both senses on the same diagram.



$$x = -A$$

$$\begin{array}{l} PE \text{ max} \\ KE = 0 \end{array}$$

$$x = 0$$

$$\begin{array}{l} PE = 0 \\ KE \text{ max} \end{array}$$

$$x = +A$$

$$\begin{array}{l} PE \text{ max} \\ KE = 0 \end{array}$$

Figure: A simple pendulum at three positions of its oscillation. At the extreme left ($x = -A$) and the extreme right ($x = +A$) the bob is momentarily at rest, so its kinetic energy is zero and its potential energy is maximum. At the equilibrium position ($x = 0$) the bob is moving fastest, so its kinetic energy is maximum and its potential energy is zero. The dotted curve marks the bob's path; the bracket on the right shows the pendulum length l ; the bracket below shows the amplitude A .

Period, frequency, and angular frequency

The **period**, written **T**, is the time taken to complete one full oscillation: that is, one full back-and-forth round trip that returns the body to the same position and the same direction of motion. Its SI unit is the second. A clock pendulum has a period of about 2s; a guitar string playing the note A above middle C has a period of about 2.27ms; the basilar membrane in the inner ear, when responding to a high-pitched whistle, may have a period as short as 0.05ms.

The **frequency**, written **f**, is the number of complete oscillations per unit time. It is the reciprocal of the period:

$$f = \frac{1}{T}$$

Its SI unit is the **hertz** (1Hz = one oscillation per second), named after Heinrich Hertz, the physicist who first generated and detected radio waves in the laboratory about 140 years ago. A clock pendulum has a frequency of about 0.5Hz; a guitar string at the same A note has a frequency of 440Hz; the human ear can hear frequencies between roughly 20Hz and 20kHz, with the upper limit shrinking gently as we get older.

The **angular frequency**, written ω (the Greek letter omega), is the frequency expressed in radians per second rather than oscillations per second. Since one full oscillation corresponds to 2π radians, the conversions are:

$$\omega = 2\pi f \quad \text{or} \quad \omega = \frac{2\pi}{T}$$

Angular frequency may look like an unnecessary luxury (*why not stick with the friendly f in hertz?*), but every equation in this chapter will contain an ω rather than an f . The reason is geometric. We will derive the equations of oscillation by comparing them with uniform motion in a circle, and a particle moving in a circle naturally wants its speed described in radians per second. Once you accept that one round trip is 2π radians of phase, the formulas come out clean and stay clean. We will let the swing of the pendulum itself decide which symbol to use, and ω will win nine times out of ten.

The energy picture

An oscillation is, at heart, a continual exchange between two forms of energy: **kinetic energy (KE)**, *the energy of motion*, and **potential energy (PE)**, *the energy stored in the displacement away from equilibrium*. At the extreme positions, the body is momentarily at rest, so its KE is zero and all the energy of the system is locked up in PE. At the equilibrium position, the body is moving fastest, so its KE is maximum and its PE is zero. In between, the energy is shared in changing proportions, but if no energy is lost to friction or air resistance, the **total** energy KE + PE remains exactly **constant** throughout the motion.

This is the deepest pattern in the whole chapter, and it is worth pausing on. An oscillation is, in effect, a system that has too much of one kind of energy and is constantly trying to convert it into the other kind, and then back again. The conversion never stops. The total never changes. The figure earlier in this section showed the two extreme positions and the equilibrium position with the energy state of each marked beneath the bob. Look at it again now, and notice that the labels at the two extreme positions are the same: the system does not care which side of equilibrium it is on; it only cares how **far** from equilibrium it is. That symmetry is the oscillation speaking, and we will hear it speak many times before the chapter ends.

Free, forced, and damped oscillations

One last triple of vocabulary, then on to the equations. An oscillation is called **free** if *the only force acting on the body is the restoring force toward equilibrium*, with no external driver and no friction or air resistance to drain the energy away. A free oscillation has constant amplitude forever, and it occurs at one particular frequency, called the **natural frequency**, that is determined entirely by the properties of the system itself (the length of the pendulum, the stiffness of the spring, the density of the liquid). Free oscillations are an idealisation; in real life, every oscillation eventually loses energy to its surroundings.

An oscillation is called **damped** when *an external force, usually friction or air resistance, removes energy from the system on every cycle*. The amplitude of a damped oscillation decreases steadily over time, and the motion eventually dies out. Kipute's swing, left to itself after Kipanga walks away, will undergo damped oscillations: each cycle is a little smaller than the last, and after a few minutes the swing is back at rest. We will treat damping in detail later in the chapter, in a section devoted to it.

An oscillation is called **forced** when *an external periodic force is applied to keep it going*, as Kipanga's pushes did. A forced oscillation eventually settles into a steady state at the frequency of the driving force, not at the natural frequency of the system. When the driving frequency **matches** the natural frequency, however, something dramatic happens: the amplitude builds up to very large values, and the system is said to be in **resonance**. Resonance is the reason Kipanga's timed pushes worked when his random pushes did not. It is also the reason a tuned radio receives one station and not another, the reason a wineglass shatters at the right soprano note, and the reason an opera-singing physicist must be careful with the chandelier. The full treatment of resonance closes the chapter, and that is also where Kipanga returns to the swing for the last time.

These three words, free, damped, and forced, will return repeatedly in the chapters on waves, sound, and alternating current. They are the working vocabulary of every oscillating system in the syllabus, and the more comfortable you are with them now, the easier the rest of the book becomes.

THE DEFINING CONDITION OF SIMPLE HARMONIC MOTION

We have the vocabulary; we may now ask the proper question. *What must a back-and-forth motion obey, exactly, to qualify as **Simple Harmonic Motion (SHM)**?* The answer is one of the cleanest in physics. It is also one of the most powerful, because it turns out to apply to so many systems that you start to suspect nature is showing off.

The two conditions

Imagine that you take any oscillating system you like (a swing, a spring with a mass on it, a guitar string, a child on a mattress, a charged particle at the centre of a crystal, the bob of a pendulum) and you displace it a small amount from its equilibrium position. Two things must be true about the resulting motion if it is going to be simple harmonic.

Condition 1: *The acceleration of the body must be **proportional to its displacement** from equilibrium.* If you displace the body twice as far, it must accelerate twice as fast. If you displace it three times as far, it must accelerate three times as fast. The relationship must be exactly linear.

Condition 2: *The acceleration must always be directed **toward the equilibrium position**.* Whenever the body is to the right of equilibrium, the acceleration must point left; whenever it is to the left, the acceleration must point right. The *acceleration must oppose the displacement*, every time, without exception.

Both conditions can be wrapped into one elegant equation. Let x be the displacement of the body from equilibrium and let a be the acceleration. Then SHM occurs if, and only if,

$$a = -\omega^2 x$$

Where ω is some positive constant whose value depends on the particular system (the mass and the stiffness of the spring, the length of the pendulum and the strength of gravity, the depth and density of the liquid in the U-tube, and so on). The proportionality between a and x is captured by the factor ω^2 . The fact that the acceleration points back toward equilibrium is captured by the minus sign. The constant ω is the angular frequency of the oscillation that the equation produces, the same ω we met in the vocabulary section.

Why the minus sign matters

It would be tempting, on first sight, to read that minus sign as a small algebraic detail and to forget it after one or two pages. Resist that temptation. The minus sign in $a = -\omega^2 x$ is the entire reason the motion oscillates rather than explodes.

To see why, suppose for a moment that someone wrote the equation without the minus sign:

$$a = +\omega^2 x$$

What kind of motion would this describe? Begin the body at a small positive displacement, $x > 0$. The equation tells us its acceleration is also positive, $a > 0$. So the body accelerates further in the positive direction. Its displacement grows. But as the displacement grows, the acceleration grows along with it (since a is proportional to x). So the body accelerates more and more rapidly, and its displacement runs away from equilibrium without bound. The “equilibrium” is no longer an equilibrium at all; it is a point that the body flees from at increasing speed. There is no oscillation, no rhythm, no return: just an exponential explosion.

The minus sign reverses all of this. With $a = -\omega^2 x$, a positive displacement produces a *negative* acceleration: the body is pushed back toward equilibrium. As it accelerates back, it overshoots equilibrium (because by the time it arrives there it has built up a velocity), and finds itself on the other side at a negative displacement, where the equation now says a is positive. It gets pushed back the other way. The result is a graceful, indefinite oscillation about $x = 0$, with the body returning, again and again, to the equilibrium it cannot help loving.

Therefore, the minus sign is the difference between a stable equilibrium (a body sitting in a valley, which oscillates) and an unstable one (a body balanced on a hilltop, which falls off). All simple harmonic motion lives in valleys. The minus sign is the valley. The figure below shows the picture in its simplest form: a parabolic potential well, with the body at three positions and the restoring force at each one pointing back toward $x = 0$.

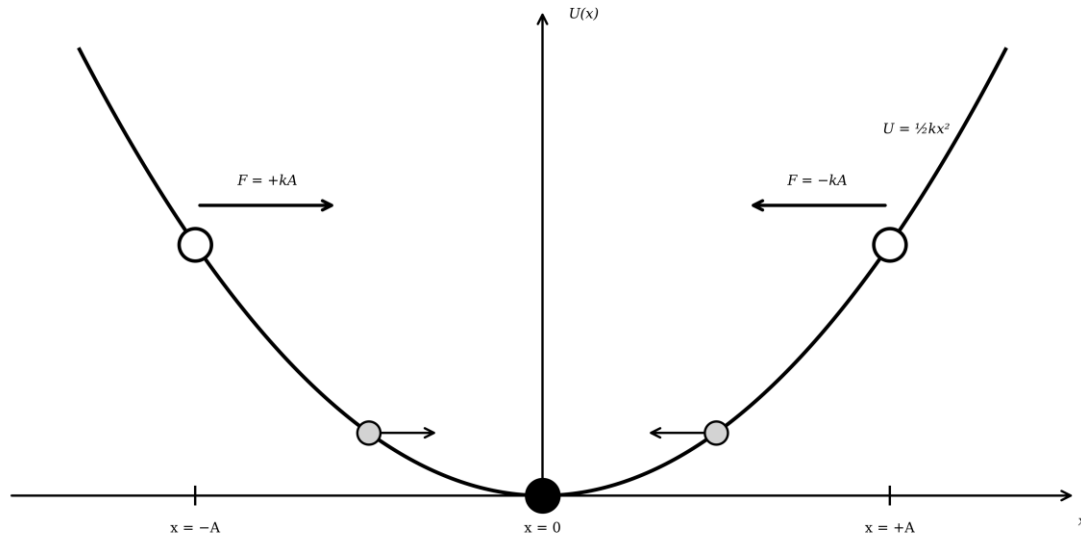


Figure: The parabolic potential well $U(x) = \frac{1}{2}kx^2$ in which simple harmonic motion lives. At every position other than equilibrium, the restoring force $F = -kx$ points back toward the bottom of the well. At the extremes ($x = \pm A$) the force has its maximum magnitude kA ; at intermediate positions it is smaller; at the equilibrium $x = 0$ it is zero. This is the geometric meaning of the minus sign in $a = -\omega^2 x$: every disturbance is met with a push back home.

The condition as a differential equation

Mathematically, the condition $a = -\omega^2 x$ is a **differential equation**. Acceleration is the second derivative of displacement with respect to time, so the same condition can be written in calculus form as:

$$\frac{d^2x}{dt^2} = -\omega^2 x$$

This is a request to find a function $x(t)$ that, when differentiated twice with respect to time, returns a constant ($-\omega^2$) times itself. To search for such a function, we need an angular variable that grows linearly with time, so that the function repeats. Let us define this variable explicitly. As always, we use the Greek letter θ for an angle in radians. Here we set:

$$\theta = \omega t$$

The product ωt has units of radians, since ω is in radians per second and t is in seconds. So θ is a perfectly good angle, and it is the angle that grows linearly with time at angular speed ω . (In the next section we will see that θ has a beautiful geometric meaning: it is the angular position of a particle moving uniformly around a reference circle. For the moment we keep it as a definition.)

There are exactly two basic functions that satisfy the SHM equation: the sine and the cosine of θ . We check this by direct differentiation. Take $x = \sin \theta = \sin(\omega t)$. Differentiating once with respect to t (using the chain rule, since the inner function is ωt):

$$\frac{dx}{dt} = \frac{d}{dt}(\sin \omega t) = \omega \cos \omega t$$

Differentiating once more:

$$\frac{d^2x}{dt^2} = \frac{d}{dt}(\omega \cos \omega t) = -\omega^2 \sin \omega t = -\omega^2 x$$

So $x = \sin(\omega t)$ satisfies $d^2x/dt^2 = -\omega^2 x$. The same calculation for $x = \cos \theta = \cos(\omega t)$ gives:

$$\frac{dx}{dt} = -\omega \sin \omega t$$

$$\frac{d^2x}{dt^2} = -\omega^2 \cos \omega t = -\omega^2 x$$

So $x = \cos(\omega t)$ satisfies the SHM equation as well. Any linear combination of the two also passes the test (we shall demonstrate this in the second worked example below). The full family of solutions to the SHM differential equation is therefore

$$x(t) = a \sin \omega t + b \cos \omega t \quad (\text{or equivalently } x(\theta) = a \sin \theta + b \cos \theta)$$

Where a and b are constants determined by the initial position and initial velocity of the body. In the next section we will derive these solutions properly using a beautiful geometric trick (the reference circle), but for now we note only that they are sinusoidal: every simple harmonic motion is, when plotted against time, a sine curve. The two basic conditions on a and x have produced a single basic shape for the motion $x(t)$.

We have the conditions, the equation, and the family of solutions. Now we put them to work. Three worked examples follow.

BINDER Example 1

Each of the following expressions describes a force F acting on a body of mass m moving along a coordinate axis (x , y , or angle θ). For each expression, decide whether the resulting motion would be simple harmonic, and justify your answer using the two conditions for SHM.

- (a) $F = -2x^2$ (b) $F = -3y$ (c) $F = 10x$ (d) $F = -7\theta$

Solution

Apply Newton's second law in the form $F = ma$, so that

$$a = \frac{F}{m}$$

We then check the two conditions on the resulting acceleration: (1) is it proportional to the displacement? (2) is it directed back toward equilibrium (opposite in sign to the displacement)?

- (a) $F = -2x^2$. Substituting into $a = F/m$:

$$a = \frac{-2x^2}{m}$$

The sign is negative when x is positive (so the acceleration points back toward equilibrium), but the dependence is on x^2 , not on x . Condition 1 fails. Hence it is **not SHM**.

- (b) $F = -3y$. Substituting:

$$a = \frac{-3y}{m} = -\left(\frac{3}{m}\right)y$$

Linear in the displacement and directed back toward equilibrium. Both conditions are met. Hence it is **SHM**, with $\omega^2 = 3/m$.

- (c) $F = 10x$. Substituting:

$$a = \frac{10x}{m} = +\left(\frac{10}{m}\right)x$$

Linear in the displacement, but the sign is positive: the acceleration points in the **same** direction as the displacement, so the body is pushed **away** from equilibrium. Condition 2 fails. Hence it is **not SHM**; this is the exponential-explosion case discussed earlier.

- (d) $F = -7\theta$. Substituting:

$$a = \frac{-7\theta}{m} = -\left(\frac{7}{m}\right)\theta$$

Linear in the angular displacement and directed back toward $\theta = 0$. Both conditions are met. Hence it is **Rotational SHM**, with $\omega^2 = 7/m$.

So only (b) and (d) describe simple harmonic motion.

Making Sense of the Answer: The four expressions are designed to fail in different ways. (a) gets the direction right but the dependence wrong. (c) gets the dependence right but the direction wrong. Failing either condition kills SHM. The exercise is good practice in checking **both** conditions, not just the one that catches your eye first.

Think Like a Physicist: A common mistake is to look only at the minus sign and conclude that any force with a minus sign produces SHM. That is wrong: $F = -2x^2$ has a minus sign and is not SHM. The sign tells you the direction; the linearity in x tells you the proportionality. You need both. A second useful habit: once you have decided a force is SHM, read off ω^2 as the coefficient of the displacement (after dividing by m). The angular frequency is then immediate, and so is the period $T = 2\pi/\omega$.

BINDER Example 2

A particle moves so that its displacement from equilibrium at time t is:

$$x(t) = a \sin \omega t + b \cos \omega t$$

Where a , b , and ω are constants. (i) Show that the motion is simple harmonic. (ii) Find the amplitude of the oscillation in terms of a and b . (iii) If $a = 3\text{cm}$ and $b = 4\text{cm}$, find the amplitude.

Solution

(i) Differentiate x once with respect to t :

$$\frac{dx}{dt} = \frac{d}{dt}(a \sin \omega t + b \cos \omega t) = a\omega \cos \omega t - b\omega \sin \omega t$$

Differentiate once more:

$$\frac{d^2x}{dt^2} = \frac{d}{dt}(a\omega \cos \omega t - b\omega \sin \omega t) = -a\omega^2 \sin \omega t - b\omega^2 \cos \omega t$$

Factor $-\omega^2$ out of the right-hand side:

$$\frac{d^2x}{dt^2} = -\omega^2(a \sin \omega t + b \cos \omega t)$$

The bracket on the right is exactly x . Therefore:

$$\frac{d^2x}{dt^2} = -\omega^2 x,$$

which is the SHM condition. The motion is therefore simple harmonic with angular frequency ω .

(ii) The trick now is to combine the sine and cosine into a single sinusoid. Suppose we can write x as:

$$x = A \sin(\omega t + \phi)$$

For some amplitude A and phase ϕ . Apply the sine-of-a-sum identity $\sin(\alpha + \beta) = \sin \alpha \cos \beta + \cos \alpha \sin \beta$ with $\alpha = \omega t$ and $\beta = \phi$:

$$x = A(\sin \omega t \cos \phi + \cos \omega t \sin \phi) = (A \cos \phi) \sin \omega t + (A \sin \phi) \cos \omega t$$

Comparing this with the original $x = a \sin \omega t + b \cos \omega t$ term by term gives two equations:

$$A \cos \phi = a \quad \text{and} \quad A \sin \phi = b$$

Square each equation:

$$A^2 \cos^2 \phi = a^2 \quad \text{and} \quad A^2 \sin^2 \phi = b^2$$

Add the two squared equations:

$$A^2 \cos^2 \phi + A^2 \sin^2 \phi = a^2 + b^2$$

$$A^2 (\cos^2 \phi + \sin^2 \phi) = a^2 + b^2$$

Use the Pythagorean identity $\cos^2 \phi + \sin^2 \phi = 1$ to simplify:

$$A^2 = a^2 + b^2$$

$$A = \sqrt{a^2 + b^2}$$

The amplitude is therefore the square root of the sum of the squares of the two coefficients. (The phase angle is then found from $\tan \phi = b/a$, although we did not need it here.)

(iii) Substitute $a = 3\text{cm}$ and $b = 4\text{cm}$ into $A = \sqrt{a^2 + b^2}$:

$$A = \sqrt{(3\text{cm})^2 + (4\text{cm})^2} = 5\text{cm}$$

The amplitude is 5cm.

Making Sense of the Answer: We have shown two things at once. First, that any sum of a sine and a cosine of the same frequency is itself simple harmonic motion. Second, that the sum can always be rewritten as a single sinusoid with a single amplitude and a single phase. The amplitude of the combined motion is not $(a + b)$, as a casual look might suggest, but $\sqrt{a^2 + b^2}$, the diagonal of a right-angled triangle whose two perpendicular legs are a and b . With $a = 3\text{cm}$ and $b = 4\text{cm}$, that triangle is the famous 3-4-5 triangle from Form Two geometry, and the amplitude is a clean 5cm.

Think Like a Physicist: The recipe in part (i) is one to remember: differentiate twice, factor out $-\omega^2$, notice that what remains in the bracket is x , then read off the SHM condition. The amplitude formula $\sqrt{a^2 + b^2}$ is the Pythagorean theorem in disguise, and once you have written it down once, you will recognise it everywhere: in the addition of two perpendicular SHM motions, in the addition of two waves of the same frequency, in the magnitude of an alternating-current phasor, in the resultant of two perpendicular electric fields. The pattern repeats. Almost every chapter from now until the end of the book will contain at least one moment where it appears.

REAL Example 3

Read the following exchange. Then use the two conditions for simple harmonic motion to settle the argument once and for all.

Kipanga: Sir, I have been thinking. Everything in the world is simple harmonic motion. The swing is SHM. The pendulum is SHM. The car suspension is SHM. Even the football that bounces on the playground when it has been kicked too softly to leave the ground, that is SHM, surely? It bounces up and down at a regular rate.

Kipute: Kipanga, no. A bouncing ball is not SHM.

Kipanga: But it is back-and-forth! It returns to the floor again and again! It has a period! What more do you want?

Solution

Applying both conditions to the bouncing football.

Condition 1. Is the acceleration proportional to displacement from some equilibrium position? While the ball is in the air (which is where it spends nearly all of its motion), the only force on it is gravity, of magnitude mg directed downward. Its acceleration is therefore g downward, the same value at every height, regardless of how far above the ground it happens to be. Acceleration is constant, not proportional to displacement. The first condition for SHM fails immediately.

Condition 2. Is the acceleration always directed toward an equilibrium position? There is no equilibrium position in the air. The ball is in free fall, accelerating downward whether it is rising, falling, near the ground, or far from it. The only restoring force the ball ever feels is the brief contact force from the ground at the bottom of each bounce, lasting a small fraction of a millisecond. For almost the entire motion, the ball is in free fall with no restoring force at all. The second condition for SHM also fails.

The bouncing ball is therefore **nearly periodic, but not simple harmonic**. Kipanga has confused a much weaker condition (repetition at regular intervals) with the much stronger condition (a linear restoring acceleration toward equilibrium throughout the motion).

Making Sense of the Answer: The temptation Kipanga fell into is a common one: to see back-and-forth motion at a regular rate and conclude it must be SHM. The conditions are stricter than that. SHM requires a restoring force that grows linearly with displacement, throughout the motion, with no jumps and no breaks. The bouncing ball spends most of its life in free fall, with no restoring force at all, so it is disqualified.

Think Like a Physicist: *Most periodic motions in nature are not simple harmonic. The rotation of the Earth, the orbit of the moon, the bouncing of a ball, the beating of a heart, the firing of a neuron: all of these are periodic, none of them are SHM. SHM is the special case in which Hooke's law (or its analogue for angular motion, fluids, electric forces, or torsion) applies cleanly to the entire motion, not just part of it. That special case is, however, common enough (at the bottom of every smooth potential well in nature) to dominate physics in the same way that uniform circular motion dominates astronomy. Recognising what counts as SHM, and what does not, is the first true skill of this chapter.*

Three characteristics of simple harmonic motion

The equation $a = -\omega^2 x$ has three immediate consequences that we will use repeatedly throughout the chapter. We collect them here so you have them at your fingertips.

Characteristic 1: the period is independent of amplitude (the isochronous property)

The angular frequency ω depends only on the constants of the system (its mass, its stiffness, the local strength of gravity, the density of the liquid, and so on), not on how far you displaced the body to start the motion. The period $T = 2\pi/\omega$ is therefore also independent of amplitude. A pendulum displaced through 5° and the same pendulum displaced through 1° (assuming both angles are small) will swing at exactly the same period. This property is the reason the pendulum was a useful timekeeper for three centuries: as it loses energy to friction and its amplitude shrinks, the rate of swinging does not change, so the clock continues to tell time accurately even as it winds down.

Characteristic 2: the acceleration is maximum at the extremes and zero at equilibrium

From $a = -\omega^2 x$: when the displacement has its largest magnitude, $|x| = A$, the acceleration also has its largest magnitude, $|a| = \omega^2 A$. When the displacement is zero (the body is passing through equilibrium), the acceleration is zero. So the body is decelerating most strongly at the very moment it momentarily comes to rest at the extremes, and feels no net force at all as it speeds through the equilibrium position. This pattern of acceleration is the opposite of what intuition often suggests, and it is worth pausing on.

Characteristic 3: the total energy is conserved

The continual exchange of kinetic and potential energy that we described in the vocabulary section keeps $KE + PE$ exactly constant throughout the oscillation, provided no energy is lost to friction or air resistance. We will derive the explicit expression for the total energy a few sections from now, but the principle is already in your hands.

These three characteristics, together with the equation $a = -\omega^2 x$ and the family of solutions $x(t) = a \sin \omega t + b \cos \omega t$, are the entire conceptual apparatus of simple harmonic motion. Everything from this point on in the chapter is, in one way or another, an application of these tools to a particular physical system. The next section turns the abstract solution into a concrete set of equations for displacement, velocity, and acceleration as functions of time, by means of the geometric trick mentioned earlier (the reference circle), which is, in our experience, the single most useful idea in the whole chapter.

THE EQUATIONS OF SIMPLE HARMONIC MOTION

We have shown that any simple harmonic motion can be written in the form $x(t) = a \sin \omega t + b \cos \omega t$, and that this expression can always be re-arranged as a single sinusoid $x(t) = A \sin(\omega t + \phi)$. It is now time to give the individual equations of SHM in usable form. We will need three: one for the displacement x , one for the velocity v , and one for the acceleration a , each as a function of time. These equations, together with their phase relationships, are the calculation engine of the chapter. Almost every numerical question on SHM, from the simplest mass-spring exercise to the most ambitious damped-oscillator problem, can be solved by some combination of them.

Rather than solving the SHM differential equation by formal calculus, we will derive the equations using a beautiful geometric trick known as the reference circle. The trick goes back to Galileo, and it has the great merit of making the sinusoidal solution look obvious instead of mysterious. Once you have seen the trick once, you will not forget it.

The reference circle

Imagine a small particle P moving uniformly around a circle of radius A , in the counterclockwise sense, with constant angular speed ω . Place the centre of the circle at the origin O of an x - y coordinate plane. At time

$t = 0$ let P be on the positive x -axis. After a time t has elapsed, P has swept out an angle $\theta = \omega t$ from the positive x -axis. Its coordinates at time t are therefore:

$$x_P = A \cos \theta = A \cos \omega t$$

$$y_P = A \sin \theta = A \sin \omega t$$

Now imagine that we drop a perpendicular from P down to the x -axis. The foot of that perpendicular, call it M , lies on the x -axis at the position $x = A \cos \omega t$. As P moves around the circle, M moves back and forth along the x -axis, between $x = +A$ (when P is on the right) and $x = -A$ (when P is on the left). The point M is the *projection* of P on the x -axis, and it executes simple harmonic motion. We will now demonstrate this directly as follows:

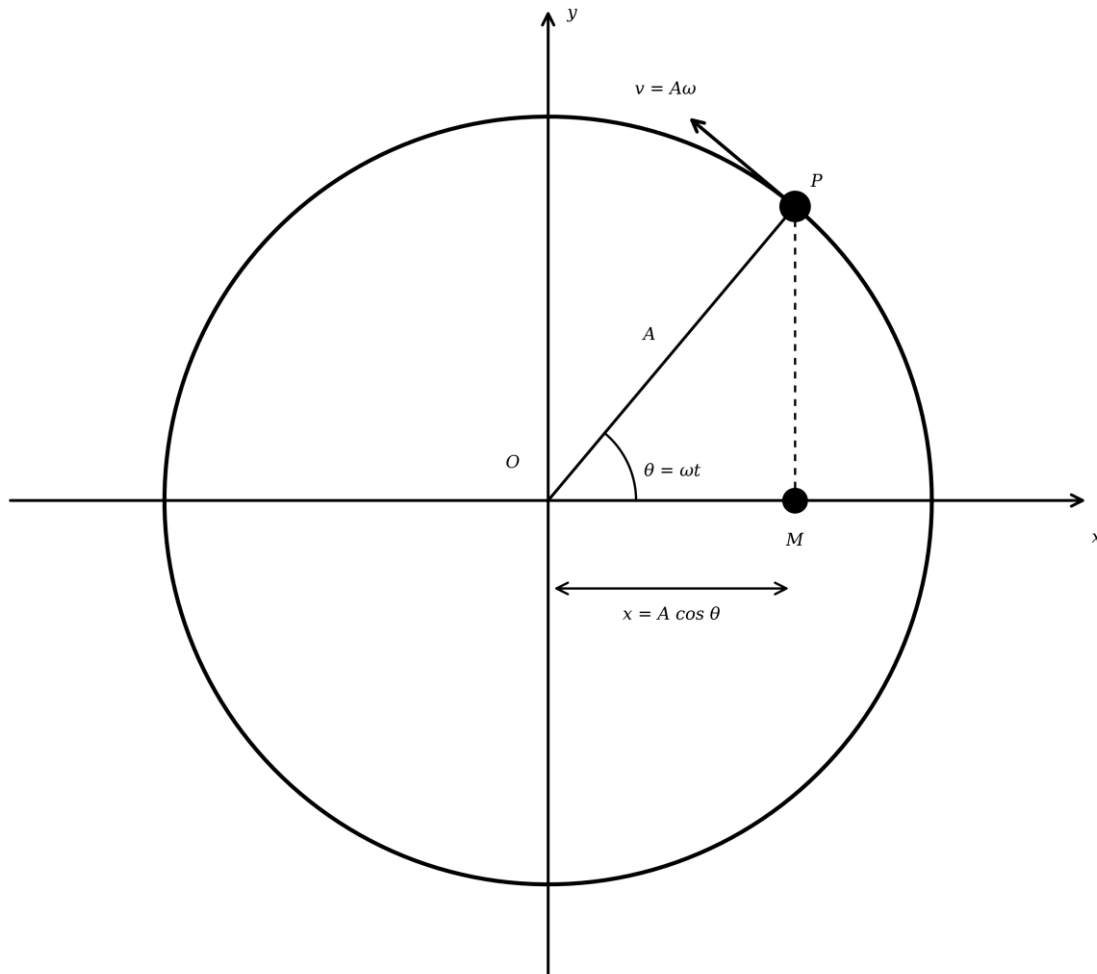


Figure: The reference circle for SHM. A particle P moves uniformly around a circle of radius A at angular speed ω . Its angular position at time t is $\theta = \omega t$. The projection M of P onto the x -axis lies at $x = A \cos \theta$, and as P moves around the circle, M moves back and forth along the x -axis between $x = +A$ and $x = -A$. The motion of M is simple harmonic.

To prove that the motion of M is SHM, we need to show that its acceleration satisfies $a = -\omega^2 x$.

Differentiate the x -coordinate of M with respect to time once to obtain its velocity:

$$v = \frac{dx}{dt} = \frac{d}{dt}(A \cos \omega t) = -A \omega \sin \omega t$$

Differentiate once more to obtain the acceleration:

$$a = \frac{dv}{dt} = \frac{d}{dt}(-A \omega \sin \omega t) = -A \omega^2 \cos \omega t$$

Recognising that $A \cos \omega t = x$, the right-hand side becomes:

$$a = -\omega^2 x,$$

which is exactly the SHM condition. The projection M therefore obeys SHM with angular frequency ω , and the constants of the circle (its radius A and the angular speed ω) become directly the amplitude and the angular frequency of the SHM. The reference circle has done its work in two lines of calculus.

From this point on we have a complete physical picture: every simple harmonic motion can be thought of as the shadow of uniform circular motion on a diameter. The *radius* of the circle is the **amplitude**, the *angular speed around* the circle is the **angular frequency**, and one full *revolution* corresponds to **one full oscillation**. The **speed** of the circulating particle, which is constant and equal to $A\omega$, becomes the **maximum speed** of the oscillation, attained as the projection M crosses the centre of the circle. We will use these relationships in the worked examples shortly.

Displacement, velocity, and acceleration

Two equivalent forms are commonly used for the displacement, depending on the initial condition. If the body starts at the equilibrium position ($x = 0$ at $t = 0$), then we choose the sine form:

$$x = A \sin \omega t$$

If the body starts at the extreme position ($x = A$ at $t = 0$), we choose the cosine form:

$$x = A \cos \omega t$$

For the most general initial condition (**the body starts somewhere in between**), we use the form with an initial phase ϕ :

$$x = A \sin(\omega t + \phi)$$

From here on we will work with the sine form $x = A \sin \omega t$ unless a problem requires otherwise; the algebra is the same up to a re-labelling of $t = 0$.

Differentiate the displacement once with respect to time to obtain the **velocity**:

$$v = \frac{dx}{dt} = \frac{d}{dt}(A \sin \omega t) = A\omega \cos \omega t$$

The velocity is largest when $\cos \omega t = \pm 1$, which happens at $t = 0, T/2, T, \dots$ (whenever the body is passing through equilibrium). The maximum speed is therefore

$$v_{\max} = A\omega$$

The velocity is zero when $\cos \omega t = 0$, which happens at $t = T/4, 3T/4, \dots$ (whenever the body is at the extreme positions $x = \pm A$). This matches the physical picture: the body must momentarily come to rest at the extremes before reversing direction.

It is often more useful to express the velocity in terms of the displacement x rather than the time t . We do this by eliminating t between the equations for x and v . Square both:

$$x^2 = A^2 \sin^2 \omega t$$

$$v^2 = A^2 \omega^2 \cos^2 \omega t$$

Divide the second by ω^2 , then add the two:

$$x^2 + \frac{v^2}{\omega^2} = A^2 \sin^2 \omega t + \frac{A^2 \omega^2 \cos^2 \omega t}{\omega^2} = A^2 \sin^2 \omega t + A^2 \cos^2 \omega t$$

$$x^2 + \frac{v^2}{\omega^2} = A^2 (\sin^2 \omega t + \cos^2 \omega t) = A^2 \text{ (using the identity } \sin^2 \omega t + \cos^2 \omega t = 1).$$

Solving for v gives the very useful form:

$$v = \pm \omega \sqrt{A^2 - x^2}$$

This is the equation we will reach for whenever a question gives us the displacement and asks for the speed (or vice versa) without telling us the time. The $\pm$ sign reminds us that, at any given displacement other than

the extremes, the body could be moving in either direction along the line: through equilibrium going right, or through equilibrium going left.

Differentiate the velocity once more to obtain the **acceleration**:

$$a = \frac{dv}{dt} = \frac{d}{dt}(A\omega \cos \omega t) = -A\omega^2 \sin \omega t$$

Recognising that $A \sin \omega t = x$, we recover the SHM condition:

$$a = -\omega^2 x,$$

and we read off the maximum acceleration immediately. It occurs at the extremes (where $|x| = A$):

$$a_{\max} = \omega^2 A$$

The acceleration is zero at the equilibrium position, in agreement with the second of the three characteristics we listed in the previous section. The full set of equations is summarised below.

The equations of SHM (sine form):

- $x = A \sin \omega t$
- $v = A\omega \cos \omega t \quad v = \pm \omega \sqrt{A^2 - x^2}$
- $a = -A\omega^2 \sin \omega t \quad a = -\omega^2 x$
- $v_{\max} = A\omega \quad a_{\max} = \omega^2 A$

Phase relationships between x, v, and a

The three quantities x , v , and a all oscillate sinusoidally at the same angular frequency ω , but they reach their peaks at different times. We can read the phase shifts directly from the equations.

Compare $v = A\omega \cos \omega t$ with $x = A \sin \omega t$. Using the identity $\cos \theta = \sin(\theta + \pi/2)$, we can write the velocity as

$$v = A\omega \sin(\omega t + \pi/2)$$

So v has the same form as x but with the argument advanced by $\pi/2$ radians. We say that the velocity **leads** the displacement by a quarter cycle ($\pi/2$, or $T/4$). Whatever value of x occurs at time t , the same fractional value of v has already occurred a quarter period earlier.

The acceleration, on the other hand, is given by $a = -A\omega^2 \sin \omega t$. Using $-\sin \theta = \sin(\theta + \pi)$, we may write

$$a = A\omega^2 \sin(\omega t + \pi)$$

So a has the same form as x but with the argument advanced by a half cycle (π , or $T/2$). The acceleration is **anti-phase** with the displacement: when x is at $+A$, a is at its most negative; when x is at $-A$, a is at its most positive. This is exactly what the equation $a = -\omega^2 x$ already tells us in compact form.

Putting the two phase shifts together: v leads x by $T/4$, and a leads v by $T/4$ (which makes a lead x by $T/2$, as the equation says). The following figure shows the three curves on the same time axis. The pattern repeats every period: x rises from zero, v reaches its peak as x passes equilibrium, x reaches its peak as v passes zero, and a reaches its most negative value at the same instant.

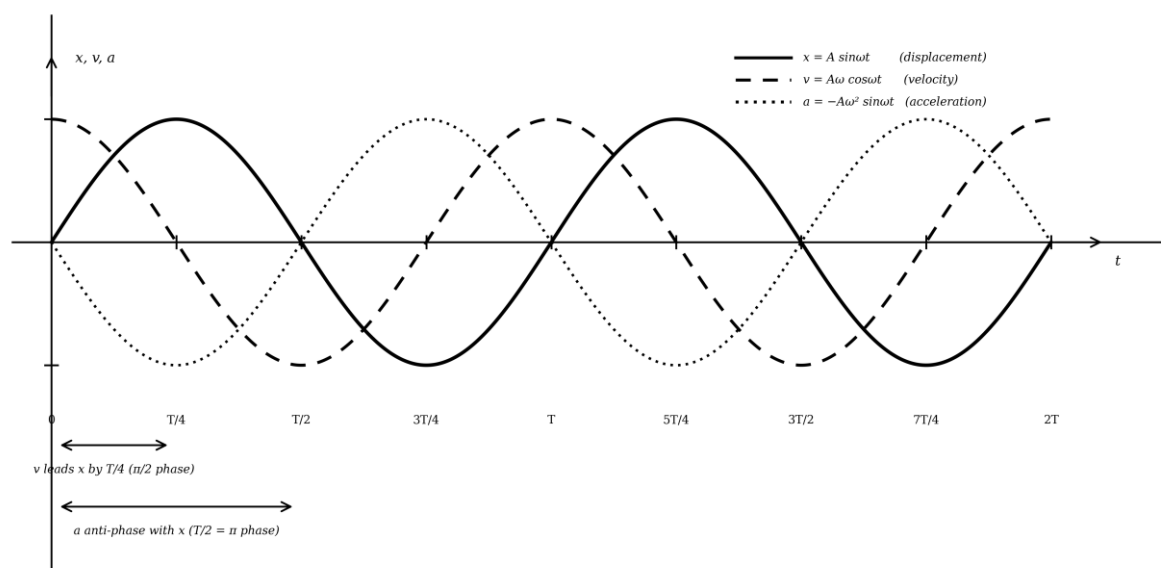


Figure: Displacement $x = A \sin \omega t$ (solid), velocity $v = A \omega \cos \omega t$ (dashed), and acceleration $a = -A \omega^2 \sin \omega t$ (dotted) plotted against time. The three quantities oscillate at the same angular frequency ω but at different phases. The velocity leads the displacement by a quarter period ($\pi/2$), and the acceleration is anti-phase with the displacement (π , or half a period). The amplitudes have been normalised in the figure so that the phase relationships are clearly visible; in a real oscillation the actual peak values are A , $A\omega$, and $A\omega^2$ respectively.

Five worked examples follow. They will work the equations of motion in every direction we are likely to be asked about in practice.

BINDER Example 4

A particle executes simple harmonic motion with a period $T = 0.50\text{s}$ and amplitude $A = 6.0\text{cm}$. Calculate (a) the maximum speed of the particle, (b) the speed at the instant the particle is at half the amplitude.

Solution

First find the angular frequency ω from the period using $T = 2\pi/\omega$:

$$\omega = \frac{2\pi}{T} = \frac{2\pi}{0.50\text{s}} = 4\pi \text{ rad/s}$$

(a) Maximum speed occurs at $x = 0$ (the equilibrium position), and is given by $v_{\text{max}} = A\omega$:

$$v_{\text{max}} = (0.06\text{m}) \times (4\pi \text{ rad/s}) = 0.24\pi \text{ m/s} = 0.754 \text{ m/s}$$

The maximum speed is approximately 0.75m/s .

(b) Use the displacement-form $v = \pm \omega \sqrt{A^2 - x^2}$ with $x = A/2 = 0.030\text{m}$ and $A = 0.060\text{m}$.

Substituting numerical values:

$$v = (4\pi \text{ rad/s}) \times \sqrt{(0.060\text{m})^2 - (0.030\text{m})^2} = 0.654\text{m/s}$$

The speed at $x = A/2$ is approximately 0.65m/s , taking the magnitude (the sign indicates direction along the line of motion).

Making Sense of the Answer: At $x = A/2$, the speed is about 87% of the maximum speed. This is consistent with the fact that the particle loses very little speed as it moves from equilibrium out to halfway, and then rapidly decelerates over the second half of its journey to the extreme. The result also illustrates an important pattern: the speed of an SHM oscillator is not a linear function of position; it follows the elliptical relationship $v^2 = \omega^2(A^2 - x^2)$.

Think Like a Physicist: Always work in SI units (metres and seconds) before plugging numbers in, even when the question quotes centimetres and milliseconds. The angular frequency ω is naturally in rad/s, and the speed $A\omega$ comes out in m/s only if A is in metres. A common slip is to use A in cm and ω in rad/s, which

gives a speed in cm/s; that is correct, but the unit must be reported faithfully. The shortest way to avoid confusion is to convert to SI at the start and to think only in SI thereafter.

BINDER Example 5

A particle executing simple harmonic motion has an acceleration of 8.0 m/s^2 at the instant when its displacement from the equilibrium position is 0.50 m . Calculate the period of the motion.

Solution

Use the SHM relation $a = -\omega^2 x$, which gives the magnitude relation $|a| = \omega^2 |x|$. Solve for ω^2 :

$$\omega^2 = \frac{|a|}{|x|} = \frac{8.0 \text{ m/s}^2}{0.50 \text{ m}} = 16 \text{ rad}^2/\text{s}^2$$

Take the positive square root for ω :

$$\omega = \sqrt{16 \text{ rad}^2/\text{s}^2} = 4.0 \text{ rad/s}$$

The period follows from $T = 2\pi/\omega$:

$$T = \frac{2\pi}{\omega} = \frac{2\pi}{4.0 \text{ rad/s}} = \frac{\pi}{2} \text{ s} = 1.57\text{ s}$$

The period of the motion is approximately 1.57 s .

Making Sense of the Answer: An angular frequency of 4 rad/s is roughly that of a slowly-swinging child on a long swing, and a period of about 1.6 s matches that physical picture nicely. The key step is the very compact use of $|a| = \omega^2 |x|$, which gives ω directly from any single (a, x) measurement, without needing to know the amplitude or the time. This is one of the most useful relations in the chapter: every oscillator carries its angular frequency stamped on every (a, x) pair it visits.

Think Like a Physicist: Notice that the question gave us only one piece of data about each of two quantities (a and x at the same instant), and the SHM relation extracted ω from the pair. The amplitude A never entered the calculation, which makes physical sense, because ω depends only on the system, not on how violently it has been disturbed. If the same particle were released from a much larger amplitude, the relation between a and x at any instant would still give the same ω . The isochronous property is at work in the algebra.

BINDER Example 6

The velocity of a particle executing simple harmonic motion is 16 cm/s at a distance of 8 cm from the mean position, and 8 cm/s at a distance of 12 cm from the mean position. Determine the amplitude of the motion.

Solution

Apply the displacement-form of the velocity relation, $v^2 = \omega^2(A^2 - x^2)$, at each of the two given instants. At the first instant, $v = 16\text{ cm/s}$ and $x = 8\text{ cm}$:

$$\begin{aligned}(16 \text{ cm/s})^2 &= \omega^2(A^2 - (8 \text{ cm})^2) \\ 256 \text{ cm}^2/\text{s}^2 &= \omega^2(A^2 - 64 \text{ cm}^2) \quad \dots\dots(i)\end{aligned}$$

At the second instant, $v = 8\text{ cm/s}$ and $x = 12\text{ cm}$:

$$\begin{aligned}(8 \text{ cm/s})^2 &= \omega^2(A^2 - (12 \text{ cm})^2) \\ 64 \text{ cm}^2/\text{s}^2 &= \omega^2(A^2 - 144 \text{ cm}^2) \quad \dots\dots(ii)\end{aligned}$$

Divide equation (i) by equation (ii) to eliminate ω^2 :

$$\begin{aligned}\frac{256 \text{ cm}^2/\text{s}^2}{64 \text{ cm}^2/\text{s}^2} &= \frac{A^2 - 64 \text{ cm}^2}{A^2 - 144 \text{ cm}^2} \\ A &= 13.06 \text{ cm}\end{aligned}$$

The amplitude of the motion is 13.06 cm .

Making Sense of the Answer: The two velocity readings, taken at two different displacements of the same oscillation, are enough to pin down the amplitude exactly, without needing to know either the period or the angular frequency. The trick is the $v^2 = \omega^2(A^2 - x^2)$ relation written twice and divided to cancel ω^2 . The

resulting amplitude (13cm) is larger than either of the two displacements (8cm and 12cm), as it must be, since both displacements are less than the amplitude.

Think Like a Physicist: This question pattern (two velocities at two displacements, find A) appears frequently in examination papers, and the recipe should become reflexive: write $v^2 = \omega^2(A^2 - x^2)$ twice, divide one equation by the other to kill ω^2 , solve for A. Once you have A, you can also recover ω by substituting back into either equation, and from there get the period and the maximum speed if the question asks. The same trick is used elsewhere in physics whenever two readings of the same kind have been taken under different conditions, and you need to extract the system parameters.

REAL Example 7

Read the dialogue below. Then use the equations for SHM to settle the question Kipanga raises.

Kipanga: Kipute, this chapter is breaking my brain. I have just read that the acceleration of an SHM oscillator is **biggest** exactly when the oscillator is **not moving**. At the extreme positions, $x = \pm A$, the body is momentarily at rest, and yet that is where the acceleration is largest. I have been taught since Form Two that an object that is not moving has zero acceleration.

Kipute: No. Form Two never said that. Form Two said that an object in **equilibrium** has zero acceleration. An object that is **momentarily at rest** but **not in equilibrium** can have any acceleration at all.

Kipanga: I do not see the difference. The body is at rest. $v = 0$. So $a = 0$.

Use the SHM equations to explain to Kipanga why his reasoning is wrong, and why the acceleration is in fact **maximum** at the extreme positions where the body is momentarily at rest.

Solution

Kipanga has confused two different physical situations. Velocity zero does not imply acceleration zero. Acceleration measures the rate at which velocity changes, not the value of velocity itself. We can have velocity zero at one instant and acceleration enormous at the same instant; the body is simply about to start moving rapidly in the opposite direction.

The SHM equations make this very precise. The acceleration is

$$a = -\omega^2 x$$

and the velocity is

$$v = \pm \omega \sqrt{A^2 - x^2}$$

These two equations say that a depends on x , while v depends on the difference between A^2 and x^2 . At the extreme positions $x = \pm A$, the bracket $A^2 - x^2$ is zero, so v is zero. But $x = \pm A$ is also where $|x|$ is largest, so $|a|$ is at its maximum value, equal to $\omega^2 A$. Both the velocity-zero and the acceleration-maximum statements refer to the same instant in the same motion. There is no contradiction.

Making Sense of the Answer: The key idea, both in this dialogue and in the SHM equations, is that velocity and acceleration are independent in the following sense: the value of one tells you nothing directly about the value of the other. They are related by the equation $a = dv/dt$, which is a **rate of change**, not a value. A high rate of change of velocity is compatible with the velocity itself being zero at that instant, just as a sharp turn in a road requires a large sideways acceleration even if the car is moving slowly.

Think Like a Physicist: This counter-intuition is one of the deepest sources of confusion in introductory mechanics, and it returns in many guises throughout physics: in projectile motion at the apex of the trajectory, in the spring-mass system at the turning points, in any oscillator at the extremes, and in alternating current circuits where the current can be zero at the very moment the rate of change of current is at its highest. The general lesson is that the value of a quantity and the value of its rate of change are different physical things; they do not have to vanish together, and they do not have to peak together. The pendulum swings most rapidly through the bottom and accelerates most strongly at the top. Once you have absorbed that pattern, the rest of oscillation physics becomes much easier.

HOT Example 8

A particle oscillating with simple harmonic motion has, at the instant when it is 0.30m from its equilibrium position, a speed of 1.6m/s and an acceleration of magnitude 4.8m/s². Find (a) the angular frequency ω of the motion, (b) the amplitude A, (c) the period T, and (d) the maximum acceleration.

Solution

(a) From the SHM relation $|a| = \omega^2|x|$, we solve directly for ω :

$$\omega^2 = \frac{|a|}{|x|} = \frac{4.8\text{m/s}^2}{0.30\text{m}} = 16 \text{ rad}^2/\text{s}^2$$

$$\omega = \sqrt{16\text{rad}^2/\text{s}^2} = 4.0 \text{ rad/s}$$

(b) From the displacement-form of the velocity relation $v^2 = \omega^2(A^2 - x^2)$, solve for A^2 :

$$A^2 = x^2 + \frac{v^2}{\omega^2}$$

$$A^2 = (0.30\text{m})^2 + \frac{(1.6\text{m/s})^2}{16 \text{ rad}^2/\text{s}^2}$$

$$A = \sqrt{0.250\text{m}^2} = 0.50\text{m}$$

(c) The period follows from $T = 2\pi/\omega$:

$$T = \frac{2\pi}{\omega} = \frac{2\pi}{4.0\text{rad/s}} = \frac{\pi}{2} \text{ s} = 1.57\text{s}$$

(d) The maximum acceleration occurs at the extremes ($|x| = A$) and equals $a_{\text{max}} = \omega^2 A$:

$$a_{\text{max}} = (16 \text{ rad}^2/\text{s}^2) \times (0.50\text{m}) = 8.0\text{m/s}^2$$

Making Sense of the Answer: All four quantities have come out to clean values. The amplitude $A = 0.50\text{m}$ is comfortably larger than the given displacement $x = 0.30\text{m}$ (as it must be, since the particle is on the way to or from one extreme but has not yet reached it). The maximum acceleration 8.0m/s^2 is larger than the given acceleration 4.8m/s^2 (as it must be, since the body was not at the extreme when measured), and indeed the ratio $(8.0)/(4.8) = 5/3$ equals exactly $A/x = 0.50/0.30$, which is the consistency check guaranteed by the SHM relation $a = -\omega^2 x$.

Think Like a Physicist: This is the most common multi-step SHM problem in the syllabus: given (v, a, x) at one instant, find $(\omega, A, T, a_{\text{max}})$. The recipe is the same every time: (i) get ω from a/x , (ii) get A from $v^2 = \omega^2(A^2 - x^2)$, (iii) get T from $2\pi/\omega$, (iv) get a_{max} from $\omega^2 A$. Four steps, four well-defined quantities, and one consistency check (the ratio a_{max}/a should equal A/x). Once the recipe is in your fingertips you can solve any version of this problem in under five minutes.

With these equations in hand, every numerical question that the syllabus is likely to ask about a simple harmonic oscillator becomes a matter of picking the right formula and substituting carefully. The equations are the calculation engine of the rest of the chapter, and we will reach for them often. But the equations are silent about a feature of the motion that we promised, in the vocabulary section, to explore: the way the oscillator passes its energy back and forth between the moving body and the spring (or pendulum, or column of liquid, or whatever else is providing the restoring force). The energy is doing something specific at every position and at every instant, and the next section makes that something explicit.

ENERGY IN SIMPLE HARMONIC MOTION

An SHM oscillator carries energy in two forms. There is the kinetic energy of the moving body itself, which is largest when the body sweeps through the equilibrium position at maximum speed. And there is the potential energy stored in the restoring agent (the spring, the bent guitar string, the displaced liquid in the U-tube, the lifted bob of the pendulum), which is largest when the body is at the extreme of its excursion and the restoring agent is most strongly deformed. As the body oscillates back and forth, energy flows continually between these two forms, but the **total** energy stored in the system, undamped, never changes. We make this statement quantitative in this section.

Potential energy

The potential energy of an SHM oscillator is the work that must be done against the restoring force to move the body from the equilibrium position out to a displacement x . The restoring force on the body at position x is $\mathbf{F} = -\mathbf{k}x$ (where $\mathbf{k}$ is the *stiffness of the system*, with $\mathbf{k} = m\omega^2$ in the SHM language). The agent doing the work against this force must apply a force $+\mathbf{k}x$ in the direction of the displacement. The work done by this external agent over a small further displacement $d\mathbf{x}$ is:

$$dW = kx \, dx$$

The total work done in moving the body from equilibrium ($x = 0$) out to displacement x is found by integrating:

$$U(x) = \int_0^x kx' \, dx' = \frac{1}{2} kx^2$$

Where x' is a dummy integration variable. So the potential energy of an SHM oscillator at displacement x is

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

Using $k = m\omega^2$ (which we will derive properly when we treat the mass-spring system in the next section), this may be written equivalently as:

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

The potential energy is therefore proportional to the **square** of the displacement. It is zero at the equilibrium position ($x = 0$) and reaches its maximum value at the extremes ($x = \pm A$), where its value is:

$$U_{\max} = \frac{1}{2} kA^2 = \frac{1}{2} m\omega^2 A^2$$

Plotted against x , the potential energy is a parabola opening upward, exactly the shape of the potential well we drew earlier (page 338). SHM lives in a parabolic well, and the potential energy **is** the well.

Kinetic energy

The kinetic energy of the body at displacement x is $K = \frac{1}{2}mv^2$, with v given by the displacement-form of the SHM velocity equation:

$$v = \pm \omega \sqrt{A^2 - x^2}$$

Squaring,

$$v^2 = \omega^2 (A^2 - x^2)$$

And substituting into $K = \frac{1}{2}mv^2$,

$$K = \frac{1}{2} m\omega^2 (A^2 - x^2)$$

Or equivalently $K = \frac{1}{2}k(A^2 - x^2)$, since $k = m\omega^2$. The kinetic energy is therefore zero at the extremes (where $x = \pm A$ and the body is momentarily at rest) and reaches its maximum value at the equilibrium position ($x = 0$):

$$K_{\max} = \frac{1}{2} m\omega^2 A^2 = \frac{1}{2} kA^2$$

Plotted against x , the kinetic energy is a parabola opening downward, reaching $K_{\max}$ at $x = 0$ and falling to zero at $x = \pm A$. The two parabolas (the upward U -curve and the downward K -curve) are mirror images of each other across a horizontal line whose value we are about to determine.

Total energy is conserved

Add U and K at any displacement x :

$$E = U + K = \frac{1}{2} m\omega^2 x^2 + \frac{1}{2} m\omega^2 (A^2 - x^2)$$

The terms in x^2 cancel cleanly,

$$E = \frac{1}{2} m\omega^2 A^2$$

leaving the total energy independent of x and independent of t . The total energy of an undamped SHM oscillator is therefore **constant**, equal to $\frac{1}{2}m\omega^2 A^2 = \frac{1}{2}kA^2$ at every instant. This is the third characteristic of SHM that we listed earlier, now derived rather than asserted.

Notice that the total energy is proportional to the **square** of the amplitude. Doubling the amplitude does not double the energy of the oscillator; it increases the energy by a factor of four. Trebling the amplitude multiplies the energy by nine. This square-law dependence is universal across all kinds of oscillators (mass-spring, pendulum, U-tube, guitar string, alternating-current circuit, sound wave) and it is one of the most important relationships in this chapter. We will return to it when we discuss damping, where the rate at which energy is lost from an oscillator is shown to be proportional to A^2 , and the amplitude therefore decays exponentially with time.

Energy graphs

Two graphs of the SHM energies are useful enough that examiners ask for them by name. The first plots PE, KE, and total E against the displacement x ; the second plots them against time t .

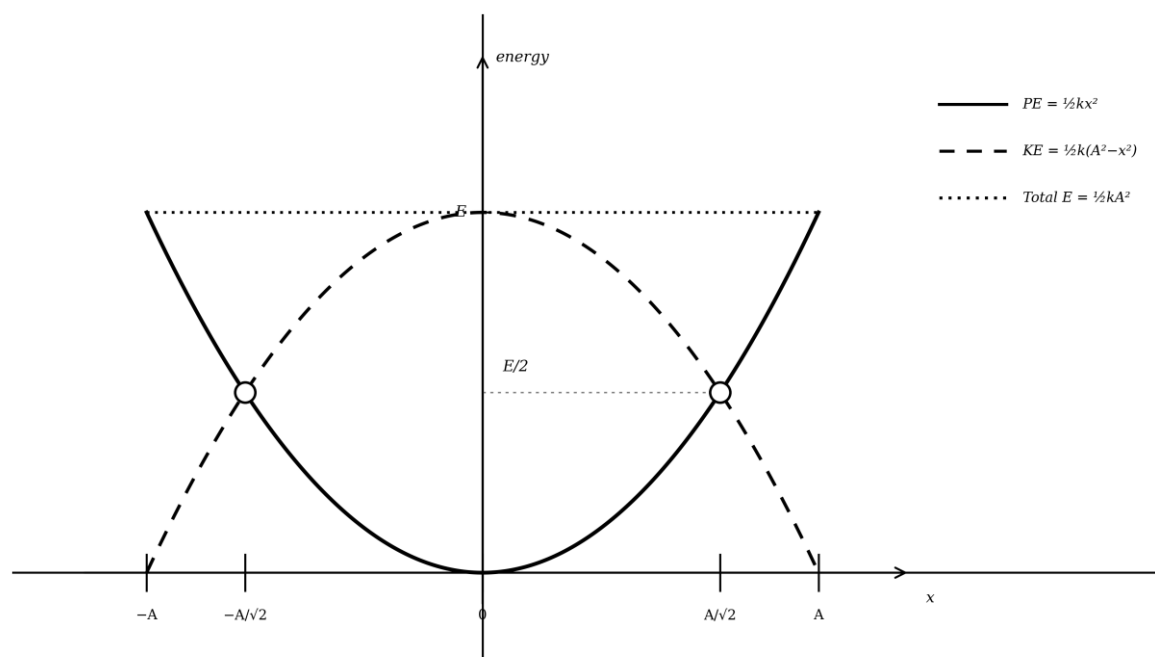


Figure: Energy of an SHM oscillator as a function of displacement x . The potential energy $U = \frac{1}{2}kx^2$ (solid) is a parabola opening upward; the kinetic energy $K = \frac{1}{2}k(A^2 - x^2)$ (dashed) is a parabola opening downward; the total energy $E = \frac{1}{2}kA^2$ (dotted) is a horizontal line independent of x . The two curves cross at $x = \pm A/\sqrt{2}$, where $U = K = E/2$.

From the energy-displacement graph, we can read directly an important position. The two parabolas cross when $U = K = \frac{1}{2}E$. Setting $U = \frac{1}{2}E$:

$$\begin{aligned}\frac{1}{2}kx^2 &= \frac{1}{2}\left(\frac{1}{2}kA^2\right) = \frac{1}{4}kA^2 \\ x^2 &= \frac{A^2}{2} \\ x &= \pm \frac{A}{\sqrt{2}} \approx \pm 0.707A\end{aligned}$$

So the kinetic and potential energies are equal when the displacement is $(1/\sqrt{2}) \approx 0.707$ of the amplitude, or about 70.7% of the way out to the extreme. This is one of the most asked-about positions in the chapter, and it is worth memorising the result.

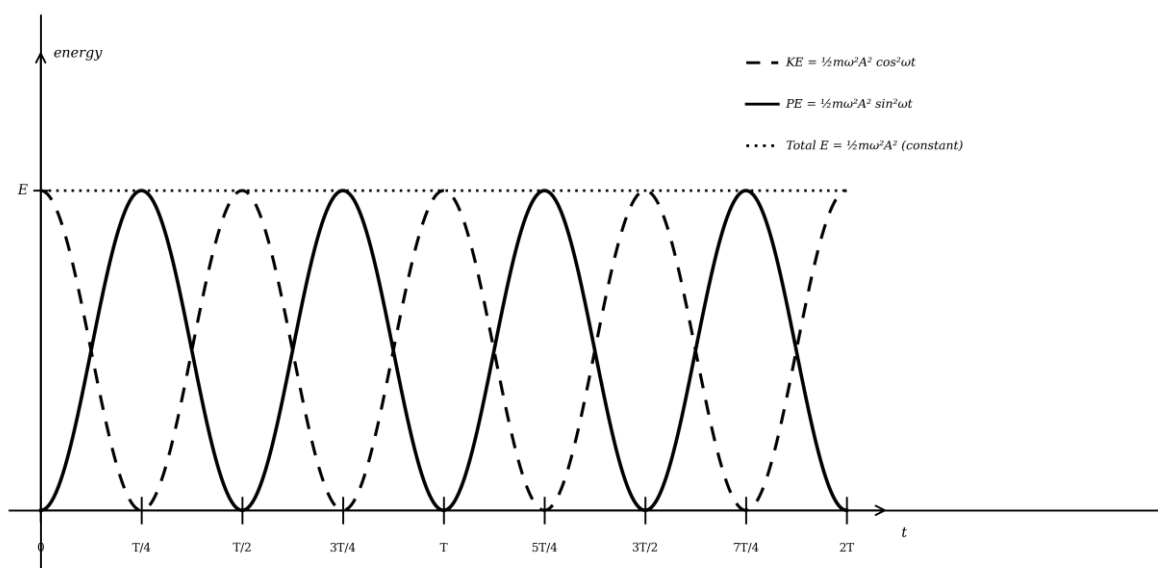


Figure: Energy of an SHM oscillator as a function of time t . The kinetic energy $K = \frac{1}{2}m\omega^2 A^2 \cos^2 \omega t$ (dashed) and potential energy $U = \frac{1}{2}m\omega^2 A^2 \sin^2 \omega t$ (solid) both oscillate at twice the angular frequency of the displacement (period $T/2$ instead of T), and they are perfectly out of phase with each other: when one is maximum, the other is zero. The total energy $E = \frac{1}{2}m\omega^2 A^2$ (dotted) is constant in time.

The energy-time graph repays a moment of attention. Both KE and PE oscillate, but they oscillate at twice the frequency of the displacement itself (or equivalently with half the period). The reason is that energy depends on the **square** of position or velocity, and the square of a sinusoid completes a full oscillation in half the period of the original. So a body oscillating at frequency f exchanges its energy back and forth at frequency $2f$. The total of course remains a flat line, since the sum of the two squared sinusoids is always equal to the constant E .

Four worked examples follow. They will look familiar in the same way that all the worked examples in this chapter look familiar; the trick is to know which formula they are quietly testing.

BINDER Example 9

A particle of mass 0.50kg vibrates with simple harmonic motion of period 1.0s. If its amplitude is 20cm, find its maximum kinetic energy.

Solution

First find the angular frequency ω from the period:

$$\omega = \frac{2\pi}{T} = \frac{2\pi}{1.0\text{s}} = 6.283 \text{ rad/s}$$

The maximum kinetic energy occurs at the equilibrium position ($x = 0$) and is given by $K_{\max} = \frac{1}{2}m\omega^2 A^2$. With $m = 0.50\text{kg}$, $\omega = 6.283\text{rad/s}$, and $A = 0.20\text{m}$:

$$K_{\max} = \frac{1}{2}(0.50\text{kg})(6.283\text{rad/s})^2(0.20\text{m})^2 = 0.395\text{J} \approx 0.40\text{J}$$

The maximum kinetic energy of the particle is approximately 0.40J.

Making Sense of the Answer: An energy of 0.40 joules is what a half-kilogram brick gains by being lifted about 8cm against gravity, so the result is modest but tangible. Notice that the maximum kinetic energy depends on m , ω , and A together, and is sensitive to all three. If we doubled the amplitude we would quadruple $K_{\max}$; if we halved the period (doubling ω), we would also quadruple it; doubling the mass would only double it.

Think Like a Physicist: The maximum kinetic energy equals the total energy of the oscillator. So the same formula $\frac{1}{2}m\omega^2 A^2$ answers any question phrased as “maximum KE,” “total energy,” “maximum PE,” or “energy stored in the oscillator.” This is one of those formulas that quietly does the work of four when you know how to read it.

BINDER Example 10

A particle is executing SHM of amplitude 5cm. At what distance from the mean position is the kinetic energy of the particle equal to its potential energy?

Solution

The kinetic and potential energies are equal when each equals half the total energy. Set the potential energy equal to half the total energy:

$$\begin{aligned}\frac{1}{2}kx^2 &= \frac{1}{2}\left(\frac{1}{2}kA^2\right) \\ x^2 &= \frac{A^2}{2} \\ x &= \frac{A}{\sqrt{2}} = \frac{5\text{cm}}{\sqrt{2}} = 3.54\text{cm}\end{aligned}$$

So the kinetic energy equals the potential energy at a displacement of about 3.54cm from the mean position, on either side.

Making Sense of the Answer: The result is the same fraction of the amplitude ($1/\sqrt{2} \approx 0.707$) regardless of what the amplitude actually is. An oscillator with amplitude 5cm crosses its $KE = PE$ point at 3.54cm; an oscillator with amplitude 1m crosses it at 70.7cm. The position is a property of the SHM, not of the particular oscillator. The point also marks the place where the body is moving at exactly $(1/\sqrt{2})$ of its maximum speed.

Think Like a Physicist: The $KE = PE$ position is a hidden symmetry of the energy graph. Look at the energy-displacement figure: the two parabolas cross at exactly two points ($x = \pm A/\sqrt{2}$) and at no others. There is no oscillator anywhere in nature, of any size, made of any material, in which the kinetic and potential energies are equal at any displacement other than $\pm A/\sqrt{2}$. The result belongs to the geometry of the parabolas, not to the physics of any specific spring.

REAL Example 11

Read the dialogue below. Then explain, using the energy formula for SHM, why Kipanga's prediction is wrong and what the correct answer is.

Kipanga: Sir, last week you said that the total energy of an SHM oscillator is constant. So for the same spring with the same mass on it, if I displace it twice as far before letting go, the oscillator will have twice the energy when it starts swinging. Right?

Kipute: No. Four times the energy.

Kipanga: Four? But I only doubled the displacement. How does the energy quadruple from one extra factor of two?

Solution

Kipanga has assumed that energy is proportional to amplitude. The SHM energy formula says otherwise. The total energy of an SHM oscillator is

$$E = \frac{1}{2}m\omega^2 A^2,$$

and the amplitude appears as **a squared**, not as A. If we double the amplitude ($A \rightarrow 2A$), the new energy is

$$E_{\text{new}} = \frac{1}{2}m\omega^2 (2A)^2 = 4 \times \left(\frac{1}{2}m\omega^2 A^2\right) = 4E$$

So doubling the amplitude multiplies the energy by 4. If we trebled the amplitude, the energy would multiply by 9. If we increased it tenfold, the energy would multiply by 100. Energy depends on the **square** of amplitude.

Making Sense of the Answer: The square-law dependence of energy on amplitude is responsible for many physical facts that, at first sight, look unrelated. A guitar string vibrating with twice the amplitude is four times as loud (in energy terms), not twice. A pendulum clock with twice the swing carries four times the energy through the air on each cycle, and loses energy correspondingly faster. A vibrating molecule with

twice the amplitude stores four times the energy in the bond. The pattern is not a coincidence; it follows from the fact that potential energy in any Hooke-law system is quadratic in displacement.

Think Like a Physicist: Whenever you read a physical statement about an oscillator and find yourself thinking “linear in amplitude,” stop and check. Almost no quantity associated with an SHM oscillator is linear in A . The maximum speed ($A\omega$) is linear in A ; the maximum acceleration ($\omega^2 A$) is linear in A ; almost everything else (energy, average power delivered, intensity, loudness, brightness, signal strength) is quadratic. The square is the rule, the linear the exception. Make a habit of asking which one applies.

HOT Example 12

An object of mass 2kg executes SHM with a frequency of 2Hz and an amplitude of 2cm. Calculate (a) its maximum velocity and (b) its maximum potential energy. (c) At what position is the potential energy equal to three-quarters of the total energy?

Solution

First find ω from f :

$$\omega = 2\pi f = 2\pi \times 2.0\text{Hz} = 12.566\text{rad/s}$$

(a) Maximum velocity occurs at $x = 0$ and equals $A\omega$. With $A = 0.020\text{m}$ and $\omega = 12.566\text{rad/s}$:

$$v_{\max} = A\omega = (0.020\text{m}) \times (12.566\text{rad/s}) = 0.251\text{m/s}$$

The maximum velocity is approximately 0.25m/s.

(b) Maximum potential energy equals the total energy of the oscillator, $\frac{1}{2}m\omega^2 A^2$. With $m = 2\text{kg}$:

$$U_{\max} = \frac{1}{2}m\omega^2 A^2$$

$$U_{\max} = \frac{1}{2}(2\text{kg})(12.566\text{rad/s})^2(0.020\text{m})^2 = 0.0632\text{J} \approx 63.2\text{mJ}$$

The maximum potential energy is approximately 63.2mJ.

(c) Use the ratio U/E . From $U = \frac{1}{2}kx^2$ and $E = \frac{1}{2}kA^2$,

$$\frac{U}{E} = \frac{x^2}{A^2}$$

Setting $U/E = 3/4$:

$$\frac{x^2}{A^2} = \frac{3}{4}$$

$$x = \frac{\sqrt{3}}{2}A = 0.866 \times 2.0\text{cm} = 1.73\text{cm}$$

The potential energy reaches three-quarters of the total energy when the displacement is approximately 1.73cm from the mean position, on either side.

Making Sense of the Answer: The three results ($v_{\max} = 0.25\text{m/s}$, $U_{\max} = 63\text{mJ}$, x at 1.73cm) describe the same oscillator from three complementary angles. The maximum velocity is fixed by ω and A ; the maximum potential energy is fixed by m , ω , and A ; and the position where $U = \frac{3}{4}E$ depends on A alone. None of the three depends on what kind of restoring agent is doing the work, and the same answers would apply to a 2kg mass on any spring tuned to $\omega = 12.6\text{rad/s}$.

Think Like a Physicist: Part (c) used the very compact relation $U/E = (x/A)^2$. It is worth memorising in this form, because it lets you read off the displacement at which the energy is any fraction of the total. $PE = \frac{1}{2}E$ at $x = A/\sqrt{2}$; $PE = \frac{3}{4}E$ at $x = (\sqrt{3}/2)A$; $PE = \frac{1}{4}E$ at $x = A/2$; $PE = E$ at $x = A$. Symmetric pairs of these positions divide every oscillation into energy-equal slices, and they are useful in problems on oscillator average behaviour and on the time spent in different parts of the cycle.

We now have the equations of SHM, the energies of SHM, and the geometric picture of the reference circle. What we do not yet have is a single physical system whose oscillation we have analysed in full from first principles. We have spoken of springs and pendulums and U-tubes in passing, as examples of systems that

obey $a = -\omega^2 x$, but we have not derived the angular frequency of any of them from the system's own physics. That is the work of the next several sections.

We start with the simplest and most commercially useful of all oscillators, the mass on a helical spring, and we will see that the angular frequency falls out of Newton's second law and Hooke's law in two lines. The ω we have been carrying through every formula in this chapter will, for the first time, become a number we can compute from the mass of the body and the stiffness of the spring. From there the rest of the chapter follows: every other physical system we look at will be analysed by the same recipe, and each will give us its own ω .

THE MASS-SPRING SYSTEM

We have promised, since the very beginning of this chapter, that the angular frequency ω that appears in every SHM equation is determined entirely by the properties of the system itself, and is not a free parameter to be fitted from one problem to the next. We have also promised that for the simplest oscillating system in nature, a mass on a helical spring, the relationship comes out cleanly and quickly. The time has come to deliver on both promises. The mass-spring system is the workhorse of the chapter, the most exam-friendly oscillator in the syllabus, and the system most directly connected to everyday objects: car suspensions, pogo sticks, vibrating weighing scales, the up-and-down jiggle of a hanging chandelier when somebody slams a door. We start with the horizontal case, where the algebra is cleanest, then move to the vertical case, where gravity makes a cameo appearance and is then politely shown to make no difference, then finally to two springs combined either in series or in parallel.

Horizontal mass-spring system

Consider a mass m attached to one end of a horizontal helical spring of force constant k , with the other end of the spring fixed to a rigid wall. The mass rests on a frictionless surface, so the only horizontal force on it is the force from the spring. When the mass is displaced from its equilibrium position by a distance x (stretching or compressing the spring), Hooke's law gives the restoring force as:

$$F = -kx$$

Where the minus sign reflects that the force always points back toward equilibrium. Apply Newton's second law:

$$ma = F = -kx$$

Divide both sides by m :

$$a = -\frac{k}{m}x$$

Compare this with the SHM condition $a = -\omega^2 x$ term by term. The acceleration is proportional to the displacement and points back toward equilibrium, so the motion is simple harmonic. Reading off the angular frequency:

$$\omega^2 = \frac{k}{m}, \quad \text{so} \quad \omega = \sqrt{\frac{k}{m}}$$

and the period follows from $T = 2\pi/\omega$:

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

This is the headline formula of the section, and it deserves a moment of contemplation. The period depends on the mass and on the stiffness of the spring, and on absolutely nothing else. It does not depend on the amplitude (because SHM is isochronous), it does not depend on the initial velocity, it does not depend on which direction you displaced the mass first, and it does not depend on gravity (because we are on a horizontal surface and gravity is balanced by the normal reaction of the table). A heavier mass means a longer period; a stiffer spring means a shorter period; both dependencies are square-root, so quadrupling either changes the period by a factor of two.

Vertical mass-spring system

Now hang the same spring vertically from the ceiling and attach the mass at the bottom. Gravity now pulls the mass downward with a steady force mg , in addition to the spring tension. At first sight you might worry that this changes the period: the mass has more force on it, after all. The first sight is wrong. The figure below shows what happens.

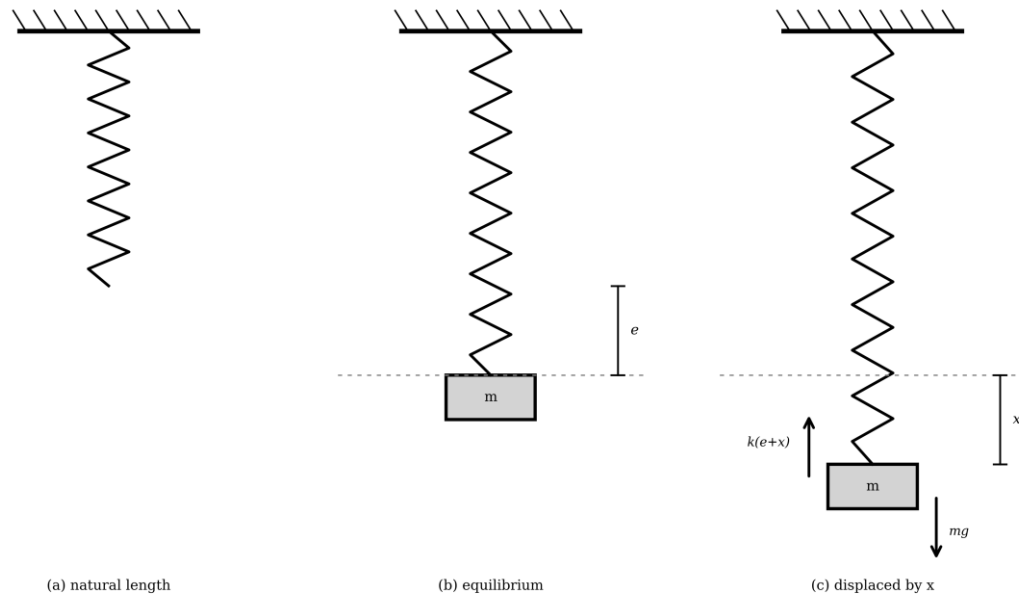


Figure: The vertical mass-spring system in three states. (a) The spring at its natural length, no mass attached. (b) The mass attached and the system in equilibrium: the spring has stretched by a static extension e , and the upward spring tension ke balances the downward weight mg . (c) The mass pulled further down by an additional displacement x and released: the spring tension is now $k(e + x)$, and the net downward force on the mass is $mg - k(e + x) = -kx$. The motion that follows is simple harmonic with the same period as the horizontal case.

When the mass is in equilibrium under gravity alone (figure (b)), the spring is stretched by some static extension e . At equilibrium the upward spring tension ke balances the downward weight mg :

$$ke = mg$$

Now displace the mass by an additional distance x below this equilibrium position (figure (c)). The spring is now stretched by a total amount $(e + x)$, so the upward spring tension is $k(e + x)$. The downward weight is still mg . The net downward force on the mass is

$$F_{\text{net}} = mg - k(e + x) = (mg - ke) - kx = 0 - kx = -kx$$

since $mg = ke$ at equilibrium. So the net force on the mass, measured from the equilibrium position, is just $-kx$, exactly the same as the horizontal case. Apply Newton's second law:

$$ma = -kx, \quad \text{so} \quad a = -\frac{k}{m}x$$

and the period is again:

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

Gravity has not changed the period at all. What it changed was the equilibrium position: the mass now oscillates about the position it would settle at when left alone, not about the natural length of the spring. Once we measure displacement from the new equilibrium, the equation of motion is identical to the horizontal case. This is one of the most pleasant cancellations in physics, and it is worth pausing on. Hang

any mass from any spring, in a uniform gravitational field of any strength, and the period of small oscillations is the same as it would be if you took the whole apparatus out into deep space and oscillated it horizontally. Gravity shifts the equilibrium; it does not change the rhythm.

There is one further useful relation that comes free with the vertical case. From $mg = ke$ we have $k = mg/e$, so the period can be written entirely in terms of the static extension e and the local strength of gravity g :

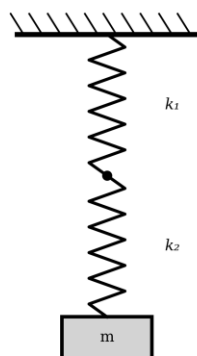
$$T = 2\pi \sqrt{\frac{e}{g}}$$

This form is particularly useful in the laboratory, because the static extension e is much easier to measure than the spring constant k itself. Hang the mass on the spring, measure how far it descends, plug into the formula. A spring extended by 25mm in static equilibrium will give an oscillation period of about $2\pi\sqrt{(0.025/9.81)} \approx 0.32\text{s}$, regardless of what the spring is made of and regardless of what the mass actually is.

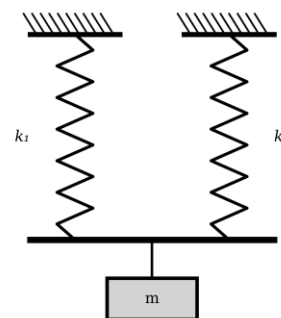
Recall also the chapter on elasticity, where we treated a 2kg mass hanging from a steel wire and asked for the period of small vertical oscillations. The answer came out at about 34ms. The same machinery is at work there: the steel wire behaves as a spring with $k = YA/L$ (where Y is Young's modulus, A is the cross-sectional area, and L is the original length of the wire), and $T = 2\pi\sqrt{(m/k)}$ gives the very short period that surprised us at the time. The physical-world distinction between a spring and a wire is largely cosmetic; both store potential energy as $\frac{1}{2}kx^2$ and both produce SHM by the same mechanism.

Two springs combined: series and parallel

Many real systems use more than one spring. A car body sits on four springs (one above each wheel) and a diving board is essentially two long flexing springs working together. The two ways to combine springs, series and parallel, give different effective spring constants, and the period of oscillation depends on which is which. The figure below shows the two arrangements.



(a) Two springs in series



(b) Two springs in parallel

Figure: Two ways to combine springs. (a) *Series:* the two springs lie end to end, each carrying the full weight of the load. The total extension is the sum of the individual extensions. (b) *Parallel:* the two springs lie side by side, sharing the weight of the load. Both springs are stretched by the same amount.

Series: Two springs of constants k_1 and k_2 are connected end to end, with a single mass hanging from the lower end. Both springs carry the same downward force F (the weight of the load is transmitted through the lower spring up into the upper spring without loss). The extension of spring 1 is $e_1 = F/k_1$ and the extension of spring 2 is $e_2 = F/k_2$. The total extension is

$$e = e_1 + e_2 = \frac{F}{k_1} + \frac{F}{k_2}$$

Defining the effective spring constant by $F = k_{\text{eff}} \times e$:

$$e = \frac{F}{k_{\text{eff}}} = \frac{F}{k_1} + \frac{F}{k_2}$$

$$\frac{1}{k_{\text{eff}}} = \frac{1}{k_1} + \frac{1}{k_2}$$

For n identical springs of constant k in series:

$$\frac{1}{k_{\text{eff}}} = n \times \frac{1}{k}$$

$$k_{\text{eff}} = \frac{k}{n}$$

The combined system is therefore **softer** than a single spring (smaller effective k), and the period of oscillation is **longer**, by a factor of $\sqrt{n}$.

Parallel: Two springs of constants k_1 and k_2 are arranged side by side, both attached to a rigid crossbar from which the load hangs. Both springs are stretched by the same amount e (because they are connected at both ends to rigid supports). The total upward force from the two springs is

$$F = F_1 + F_2 = k_1 e + k_2 e = (k_1 + k_2)e$$

So the effective spring constant is:

$$k_{\text{eff}} = k_1 + k_2$$

For n identical springs of constant k in parallel, $k_{\text{eff}} = nk$. The combined system is **stiffer** than a single spring (larger effective k), and the period of oscillation is **shorter**, by a factor of $\sqrt{n}$. A car on four springs in parallel oscillates twice as fast as it would on one spring of the same individual stiffness.

Five worked examples follow. The mass on the spring is the most heavily-tested oscillator in this chapter, so it is worth doing them slowly and not because they are exciting.

BINDER Example 13

A small mass of 0.2kg is attached to one end of a helical spring and produces an extension of 15mm. The mass is now pulled down 10mm and set into vertical oscillation of amplitude 10mm. Find: (a) the period of oscillation, (b) the maximum kinetic energy of the mass, and (c) the potential energy of the spring when the mass is 5mm below the centre of the oscillation.

Solution

(a) Use the static-extension form $T = 2\pi \sqrt{\frac{e}{g}}$ with $e = 0.015\text{m}$ and $g = 9.8\text{m/s}^2$:

$$T = 2\pi \sqrt{\frac{0.015\text{m}}{9.8\text{m/s}^2}} = 0.246\text{s}$$

The period of oscillation is approximately 0.25s.

(b) First find ω from the period:

$$\omega = \frac{2\pi}{T} = \frac{2\pi}{0.246\text{s}} = 25.57\text{rad/s}$$

The maximum kinetic energy occurs at the equilibrium position and equals $\frac{1}{2}m\omega^2 A^2$.

With $m = 0.2\text{kg}$, $\omega = 25.57\text{rad/s}$, and $A = 0.010\text{m}$:

$$K_{\text{max}} = \frac{1}{2} (0.2\text{kg}) (25.57\text{rad/s})^2 (0.010\text{m})^2$$

$$K_{\text{max}} = 6.54 \times 10^{-3}\text{J} \approx 6.5\text{mJ}$$

(c) First find the spring constant:

$$k = \frac{mg}{e} = \frac{(0.2\text{kg})(9.81\text{m/s}^2)}{0.015\text{m}} = 130.8\text{N/m}$$

Then $U = \frac{1}{2}kx^2$ at $x = 0.005\text{m}$:

$$U = \frac{1}{2}(130.8\text{N/m})(0.005\text{m})^2$$

$$U = 1.64 \times 10^{-3}\text{J} \approx 1.6\text{mJ}$$

Making Sense of the Answer: All three answers are small in joules but quite manageable in millijoules, which is normal for laboratory-scale oscillators. Notice the relationship between the energies: the maximum kinetic energy (6.5mJ) is the total energy of the oscillation. The potential energy at $x = A/2 = 5\text{mm}$ is 1.6mJ, which is exactly one-quarter of the total energy. That ratio ($1/4$) arises because $U/E = (x/A)^2$, and $(5/10)^2 = 1/4$.

Think Like a Physicist: The static-extension form $T = 2\pi\sqrt{e/g}$ is the most useful single formula in the section. It lets you compute the period of any vertical mass-spring system from a single ruler measurement of the static extension, without ever needing to know the spring constant or the mass separately.

BINDER Example 14

When a metal cylinder of mass 0.2kg is attached to the lower end of a light helical spring, the upper end of which is fixed, the spring extends by 0.16m. The metal cylinder is then pulled down a further 0.08m and left free to oscillate vertically in simple harmonic motion. Calculate the period of vertical oscillations.

Solution

Use $T = 2\pi\sqrt{e/g}$ directly. The 0.08m “pulled down” is the amplitude of the oscillation, which (by the isochronous property) does not affect the period. Substituting $e = 0.16\text{m}$ and $g = 9.8\text{m/s}^2$:

$$T = 2\pi\sqrt{\frac{0.16\text{m}}{9.8\text{m/s}^2}} = 0.803\text{s}$$

The period of vertical oscillation is approximately 0.80s.

Making Sense of the Answer: A period of about 0.8s for a 0.2kg cylinder on a spring that stretches 16cm under its weight is comfortable to imagine: roughly the rhythm of a slow, deliberate human heartbeat. The 0.08m amplitude information is deliberately included by the question setter as a distractor; the period of an undamped SHM oscillator is always independent of amplitude, and a careful reader will notice that no amplitude information appears in the formula $T = 2\pi\sqrt{e/g}$.

Think Like a Physicist: When you see numerical information in a question that appears to play no role in the formula, do not assume you have made an error. The information may be there to test whether you know it is irrelevant. The isochronous property is the most heavily tested fact about SHM, precisely because students who do not understand it are tempted to substitute the amplitude into formulas where it does not belong.

BINDER Example 15

A 2.0kg ball is attached to a horizontal spring of negligible mass and force constant $k = 50\text{N/m}$ on a frictionless table. The ball is displaced 0.20m from equilibrium and released. Find (a) the angular frequency, (b) the period, (c) the maximum speed, and (d) the maximum acceleration.

Solution

(a) Angular frequency from the formula $\omega = \sqrt{k/m}$:

$$\omega = \sqrt{\frac{k}{m}} = \sqrt{\frac{50\text{N/m}}{2.0\text{kg}}} = 5.0\text{rad/s}$$

(b) Period from $T = 2\pi/\omega$:

$$T = \frac{2\pi}{\omega} = \frac{2\pi}{5.0\text{rad/s}} = 1.26\text{s}$$

(c) Maximum speed at equilibrium, $v_{\text{max}} = A\omega$:

$$v_{\text{max}} = A\omega = (0.20\text{m}) \times (5.0\text{rad/s}) = 1.0\text{m/s}$$

(d) Maximum acceleration at the extremes, $a_{\text{max}} = \omega^2 A$:

$$a_{\max} = \omega^2 A = (5.0 \text{ rad/s})^2 \times (0.20 \text{ m}) = 5.0 \text{ m/s}^2$$

Making Sense of the Answer: A spring of $k = 50 \text{ N/m}$ is fairly soft (about the stiffness of a typical retractable-pen spring), and 2 kg is a moderately heavy mass for it. The resulting period of about 1.3 s is leisurely, the maximum speed of 1 m/s is gentle, and the maximum acceleration of 5 m/s^2 is about half of g . All four quantities are in the comfortable middle range of laboratory oscillators, which is what a spring constant of 50 N/m and a 2 kg mass naturally produce.

Think Like a Physicist: Once you have ω from the system constants k and m , the rest of the standard SHM quantities follow from one formula each: period from $T = 2\pi/\omega$, maximum speed from $v_{\max} = A\omega$, maximum acceleration from $a_{\max} = \omega^2 A$, maximum KE from $K_{\max} = \frac{1}{2}m\omega^2 A^2$. The whole calculation collapses into four quick lines, and a question that asks for any of these is essentially testing whether you can navigate the formula tree from k and m at the root to the quantity asked for at the leaf.

REAL Example 16

A car with a mass of 1300 kg is constructed so that its frame is supported by four springs. Each spring has a force constant of 20000 N/m . If two people riding in the car have a combined mass of 160 kg , find the frequency of vibration of the car after it is driven over a pothole in the road.

Solution

The car body and its passengers oscillate together as a single mass M supported by four springs in parallel. Total oscillating mass:

$$M = 1300 \text{ kg} + 160 \text{ kg} = 1460 \text{ kg}$$

Springs in parallel combine by addition of spring constants. With four identical springs of $k = 20000 \text{ N/m}$:

$$k_{\text{eff}} = 4k = 4 \times 20000 \text{ N/m} = 80000 \text{ N/m}$$

Angular frequency:

$$\omega = \sqrt{\frac{k_{\text{eff}}}{M}} = \sqrt{\frac{80000 \text{ N/m}}{1460 \text{ kg}}} = 7.40 \text{ rad/s}$$

Frequency:

$$f = \frac{\omega}{2\pi} = \frac{7.40 \text{ rad/s}}{2\pi} = 1.18 \text{ Hz}$$

The car body oscillates at a frequency of about 1.2 Hz after a pothole (which is roughly one full bounce per second).

Making Sense of the Answer: A frequency of about 1 Hz is the comfortable sweet spot for vehicle suspensions worldwide. Frequencies much higher than this feel jittery and make passengers uncomfortable on long journeys; much lower and the car rocks queasily after every bump. Engineers tune the spring constants and the damping (which we will treat later in this chapter) so that loaded family cars settle around this $1\text{--}1.5 \text{ Hz}$ range. The Tanzanian road context is real: anyone who has driven the Dar–Mbeya highway will testify that pothole-induced car oscillations are not just an academic example.

Think Like a Physicist: Notice how naturally the parallel-springs rule simplifies a real engineering question. Without it, you would have to track four separate forces from four separate springs, then add them. With it, the four springs behave as a single equivalent spring of constant $4k$, and the period formula reads off in one step. The modular approach (replace any combination of springs by a single equivalent spring, then use $T = 2\pi\sqrt{M/k_{\text{eff}}}$) scales to any number of springs in any combination of series and parallel, and is the practical way to design real oscillating systems.

REAL Example 17

A delivery lorry plying the Dar es Salaam to Iringa highway carries sacks of maize. The driver notices that when the lorry is fully loaded with sacks, it bounces noticeably less violently after passing over a pothole than when it is running empty on the return trip. Using the mass-spring model of vehicle suspension, explain this observation. Address in particular: (i) how the period of bounce changes when the lorry is loaded, (ii) how the maximum speed of the bounce changes, and (iii) why the loaded ride feels gentler to the driver and to the cargo, even though more mass is being moved.

Solution

The body of the lorry rests on its leaf springs, and when the lorry hits a pothole the spring-mass system is set into vertical SHM about its equilibrium position. The springs together provide the same effective spring constant k_{eff} regardless of the load; what changes between loaded and empty is the total mass M of the body plus its cargo. Three SHM relations describe the bounce: the period $T = 2\pi\sqrt{\frac{M}{k_{\text{eff}}}}$, the maximum bounce speed $v_{\text{max}} = A\omega = \frac{2\pi A}{T}$, and the maximum bounce acceleration $a_{\text{max}} = \omega^2 A = \left(\frac{2\pi}{T}\right)^2 A$, where A is the amplitude (set by the depth of the pothole, which is the same on both trips).

(i) The period of bounce: Under the same set of springs, $T = 2\pi\sqrt{\frac{M}{k_{\text{eff}}}}$ varies as the square root of the total mass M . When the lorry is fully loaded, M is much larger than when it is empty. The period therefore increases in proportion to the square root of the increase in mass. As a result, the loaded lorry oscillates more slowly than the empty one.

(ii) The maximum speed of the bounce: With the same amplitude A but a larger period T , the maximum bounce speed $v_{\text{max}} = \frac{2\pi A}{T}$ is smaller when the lorry is loaded. Each bounce, although it moves the body through the same vertical distance, takes longer and therefore proceeds at a lower peak velocity. The cargo and the driver experience a gentler maximum speed of motion.

(iii) The acceleration the driver feels: The human body, and most cargo, senses jolts not as positions or speeds but as accelerations. The maximum bounce acceleration is $a_{\text{max}} = \left(\frac{2\pi}{T}\right)^2 A$, which falls as the inverse square of the period. So the loaded lorry has a longer period than the empty one, and thus its maximum acceleration is significantly reduced. This means the suspension responds more gently, so the driver experiences smaller jolts on the spine, the maize sacks bounce less violently, and the dashboard rattles much less.

This is why the loaded lorry feels more comfortable, even though it is moving more total mass.

The same principle is the reason that empty pickup trucks famously have a harsh ride while heavily-loaded lorries feel comparatively smooth to their drivers. The spring stiffness is the same; the mass is doing all the work.

Making Sense of the Answer: The result is, at first sight, counter-intuitive. One might expect that a heavier load would strain the suspension more and produce a rougher ride. The physics says the opposite. More mass slows the oscillation and reduces both the bounce speed and (more strongly) the bounce acceleration. Comfort comes from gentle acceleration, and the simple inverse-square dependence of acceleration on period is what links a heavier load to a more comfortable ride.

Think Like a Physicist: Real engineers exploit this systematically. Heavy goods vehicles are designed with springs stiff enough to support a fully-loaded mass, which means they ride harshly when empty (because then the effective mass is too small for the spring constant). Air suspensions on luxury buses go a step further: they vary the effective spring constant in real time as passengers board and disembark, keeping the period in the comfortable one-to-one-and-a-half hertz range regardless of how full the bus is. Behind every comfortable ride is a piece of SHM physics being applied deliberately, not accidentally.

HOT Example 18

A tray of mass 4.0kg is supported by two identical springs. When the tray is displaced slightly and released, it executes simple harmonic motion with a period of 1.0s. (a) Find the spring constant of each spring. (b) A block of mass m is now placed on the tray; the period of oscillation increases to 2.0s. Find m .

Solution

(a) The two springs are in parallel (they support the tray side by side). The effective spring constant is $k_{\text{eff}} = 2k$. From $T = 2\pi\sqrt{\frac{M}{k_{\text{eff}}}}$:

$$T_1 = 2\pi\sqrt{\frac{M}{2k}}$$

Square both sides and solve for k:

$$T_1^2 = 4\pi^2 \frac{M}{2k} = \frac{2\pi^2 M}{k}$$

$$k = \frac{2\pi^2 M}{T_1^2}$$

Substitute $M = 4.0\text{kg}$ and $T_1 = 1.0\text{s}$:

$$k = \frac{2\pi^2 \times 4.0\text{kg}}{(1.0\text{s})^2} = 79.0\text{N/m}$$

Each spring has a constant of approximately 79N/m .

(b) With the block of mass m on the tray, the total oscillating mass is $(M + m)$ and the new period is

$$T_2 = 2\pi \sqrt{\frac{M+m}{2k}}. \text{ Squaring:}$$

$$T_2^2 = \frac{4\pi^2(M+m)}{2k} = \frac{2\pi^2(M+m)}{k}$$

Divide this by the equation for T_1^2 :

$$\frac{T_2^2}{T_1^2} = \frac{M+m}{M}$$

Solving for m :

$$m = M \left(\frac{T_2^2}{T_1^2} - 1 \right)$$

$$m = (4.0\text{kg}) \left(\frac{(2.0\text{s})^2}{(1.0\text{s})^2} - 1 \right) = (4.0\text{kg})(4 - 1) = 12\text{kg}$$

The block has a mass of 12kg .

Making Sense of the Answer: When the period doubled, the mass quadrupled. This is the inverse of the more familiar relation that the period varies as the square root of the mass. The recipe of dividing T_2^2 by T_1^2 to cancel the spring constant is one of the cleanest tricks in the chapter.

Think Like a Physicist: Whenever a problem gives you the same oscillator before and after some change (a mass added, a spring removed, the system taken from horizontal to vertical), look for a quantity that is unchanged by the change and use it to set up a ratio. Here the unchanged quantity is the spring constant $2k$, and the ratio T_2^2/T_1^2 eliminated it automatically. Ratio methods of this kind almost always shorten a problem and reduce the number of arithmetic operations required.

We have analysed the mass-spring system inside out: horizontal, vertical, in series, in parallel, with passengers, with extra blocks. Every variation reduced to one formula, $T = 2\pi\sqrt{m/k}$, and one geometric picture, the body sitting in a parabolic well of stiffness k . This is the recipe we will now apply to every other oscillating system in the chapter: identify the restoring force as a function of displacement, read off the effective stiffness, divide by the effective mass, take the square root and 2π , done. The next system on the list is the one Kipanga encountered first: the simple pendulum.

THE SIMPLE PENDULUM

The mass-spring system was kind to us because the spring did the heavy mathematical lifting. Hooke's law, $F = -kx$, is already in the SHM form before we even start. We did very little work to extract the period; we identified k and m , took a square root, and multiplied by 2π . The simple pendulum, is more interesting precisely because it has no spring. Its restoring force comes entirely from gravity, with no Hooke's law to lean on, and yet, after a careful look at the geometry, the SHM equation $a = -\omega^2 x$ reappears, with a new and beautiful meaning for ω .

The simple pendulum was Kipanga's very first encounter in this chapter. He tried, and failed, to push Kipute on the school swing. We now have the mathematical equipment to explain his failure and to dignify his

eventual success. Behind every working clock pendulum, every grandfather clock that has measured weddings and burials for a hundred years, every metronome on every piano, the same equation lives. We are about to derive it.

What a simple pendulum is, and is not

A simple pendulum is an idealisation. It consists of a small, heavy bob, treated as a point mass, suspended from a fixed pivot by a long, light, inextensible string. The word “light” means we ignore the mass of the string entirely. The word “inextensible” means the string never stretches; its length is constant and equal to the distance from the pivot to the centre of the bob. The phrase “point mass” means the bob is small enough that we may treat it as concentrated at a single point.

No real pendulum is simple in this strict sense. The bob has finite size; the string has finite mass; and if the string is replaced by a metal rod, as in a clock pendulum, the rod is heavy enough that its own moment of inertia matters. Such a rod-and-disc arrangement is called a compound pendulum, and we shall treat it later in the chapter. The simple pendulum is the cleanest version of the device, the one that lets us isolate the gravity-and-length physics from any complication of mass distribution. Once we have the simple pendulum well in hand, the compound pendulum will be a small extension.

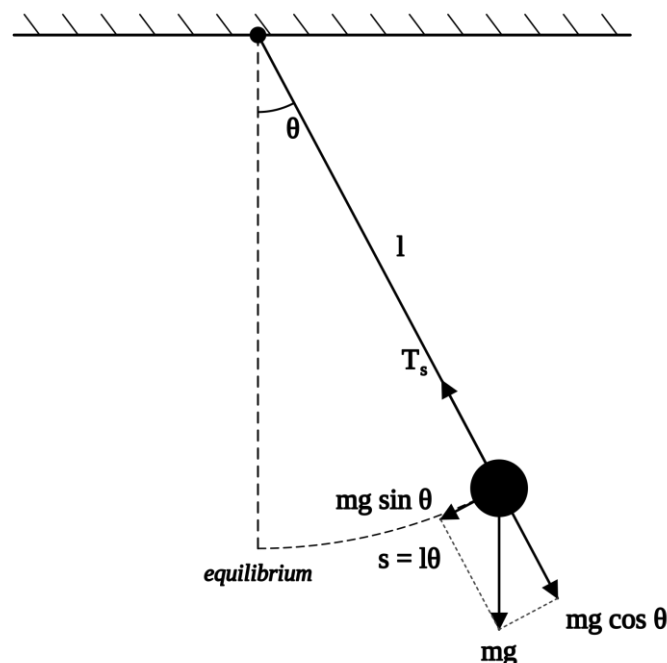


Figure: *A simple pendulum displaced through a small angle θ from the vertical. The bob hangs from a fixed pivot by a light inextensible string of length l . The arc length from the equilibrium position is $s = l\theta$. Two forces act on the bob, the weight mg straight down and the tension T_s along the string. The weight resolves into a component $mg\cos\theta$ along the string (balanced by the tension) and a component $mg\sin\theta$ tangent to the arc, which is the restoring force always directed back toward equilibrium.*

From the swinging bob to the period formula

Place the pendulum at rest. The string hangs vertically, and the bob hangs at its lowest point. This is the equilibrium position, the only position at which the net force on the bob is zero. The two forces present, the weight mg straight down and the tension T_s straight up along the string, exactly cancel.

Now displace the bob through a small angle θ from vertical and release it. The bob is no longer at its lowest point; it stands at a **height $l(1 - \cos \theta)$** above equilibrium, swung along an arc of length **$s = l\theta$** (with θ in radians). Two forces still act on the bob: the tension T_s along the string, now pointing up and back toward the pivot rather than vertically; and the weight mg , still pointing straight down.

To analyse the motion along the arc, we resolve the weight into two components. One component is along the string, of magnitude $mg\cos\theta$; the other is tangent to the arc, of magnitude $mg\sin\theta$. The component along the string is balanced by the tension. The component tangent to the arc is the restoring force. It always points back toward equilibrium, since when the bob is displaced to the right of vertical (positive θ), the tangential component points to the left, and the reverse holds on the other side. Calling the tangential force F_t :

$$F_t = -mg\sin\theta$$

The negative sign captures the restoring nature of the force, in the same way it did for the mass on a spring. Whenever θ is positive, F_t is negative; whenever θ is negative, F_t is positive. The force is always pulling the bob back home.

Newton's second law along the arc says that this tangential force equals mass times tangential acceleration.

The tangential acceleration is the second derivative of arc length with respect to time, $a = \frac{d^2s}{dt^2}$, where $s = l\theta$ is the arc length. Therefore:

$$m \frac{d^2s}{dt^2} = -mg\sin\theta$$

We have one equation. What we need now is to express the right-hand side in terms of the displacement s , so that we can compare with the SHM condition $a = -\omega^2s$. Two substitutions do this. First, since $s = l\theta$, we have $\theta = s/l$. Second, for small angles measured in radians, $\sin\theta \approx \theta$. Putting both substitutions together:

$$\sin\theta \approx \theta = \frac{s}{l}$$

So the equation of motion becomes:

$$m \frac{d^2s}{dt^2} \approx -mg \frac{s}{l}$$

The mass m cancels from both sides, a fact whose physical meaning we shall pause on shortly:

$$\frac{d^2s}{dt^2} = a \approx -\frac{g}{l}s$$

This has exactly the form of the SHM condition $a = -\omega^2s$, with:

$$\omega^2 = \frac{g}{l}$$

And therefore the period of the simple pendulum is:

$$T = \frac{2\pi}{\omega} = 2\pi \sqrt{\frac{l}{g}}$$

This is the formula every Tanzanian advanced-level student is expected to write from memory. Let us look at it slowly. The period of a simple pendulum depends on only two things, its length l and the local strength of gravity g . It does not depend on the mass of the bob, nor on the amplitude of the swing (provided the small-angle approximation holds), nor on the colour, density, shape, or chemical composition of the bob. A pendulum of bronze, a pendulum of iron, and a pendulum of green mango, all swinging from strings of the same length on the same playground, have exactly the same period.

The cancellation of the mass is not a coincidence. It is the same cancellation that makes all bodies fall to the ground at the same rate: heavy bodies feel a larger gravitational pull, but they also have a larger inertia to be moved, and the two effects cancel exactly. The pendulum is therefore the cleanest possible test of the equality of gravitational and inertial mass.

The dependence on length is more useful in everyday practice. A pendulum twice as long as another has a period larger by a factor of $\sqrt{2} \approx 1.414$. To halve the period, you must shorten the length by a factor of four. To double the period, you must lengthen by a factor of four. This is why grandfather clocks have their pendulums about a metre long: that gives a period of just over two seconds, which the clock's escapement mechanism turns conveniently into the familiar tick-tock of one second per half-swing.

The dependence on g is much less noticeable in everyday life, since gravitational acceleration does not vary significantly from one place to another on Earth, but it becomes enormously important in scientific applications.

A pendulum is one of the most sensitive instruments ever invented for measuring g ; a careful experiment can determine g to four significant figures with apparatus that costs less than a stove. We shall set up that experiment shortly.

When the motion is truly simple harmonic

The derivation above contained one approximation: $\sin \theta \approx \theta$ for small θ in radians. For angles below about 10° , the small-angle approximation is correct, and the pendulum may be treated as performing simple harmonic motion. For 20° and above, the approximation begins to fail noticeably, and the pendulum's period starts to depend on its amplitude; large swings take measurably longer than small ones, and the motion is no longer strictly SHM.

This is the precise answer to the recurring examination question, “*Why is the motion of a simple pendulum not strictly simple harmonic?*” The restoring force is $mg\sin\theta$, not $mg\theta$. The two are equal only in the limit of zero amplitude. Every real pendulum oscillation involves a small but non-zero amplitude, and is therefore only approximately simple harmonic. The approximation is excellent for the small angles that dominate practical use, but it must be acknowledged as an approximation, not as an exact identity.

We may now state the conditions on the simple pendulum precisely. The motion is approximately simple harmonic with period $T = 2\pi\sqrt{l/g}$ provided that:

1. the bob is treated as a point mass, with finite mass and zero size;
2. the string is treated as light (zero mass) and inextensible (constant length);
3. the angle of swing is small enough that $\sin \theta \approx \theta$; in practice, $\theta < 10^\circ$ gives accuracy better than 0.5%;
4. the only forces on the bob are gravity and the tension in the string, with no air resistance and no friction at the pivot.

Whenever any of these conditions fails, the motion ceases to be strictly simple harmonic, although it may still be approximately periodic. Air resistance, in particular, causes the amplitude of the swings to decay over time, a topic we shall treat in detail in the section on damped oscillations later in this chapter.

The pendulum at zero gravity, and other places

Three thought experiments will fix the dependence of T on g in the imagination.

First, take a simple pendulum to the surface of the moon. The moon's surface gravity is about $g_{\text{moon}} \approx 1.62\text{m/s}^2$, about one-sixth of Earth's 9.81m/s^2 . The same pendulum (same l) on the moon has a period:

$$\frac{T_{\text{moon}}}{T_{\text{earth}}} = \sqrt{\frac{g_{\text{earth}}}{g_{\text{moon}}}} = \sqrt{\frac{9.81\text{m/s}^2}{1.62\text{m/s}^2}} = 2.46$$

So a one-second-period pendulum on Earth becomes a 2.46-second-period pendulum on the moon. A clock built for Earth gravity would run grossly slow on the moon, and one would need to shorten its pendulum by a factor of about six to recover the same period.

Second, take the same pendulum to the top of Mount Kilimanjaro, 5895 metres above sea level. Local gravity decreases with altitude, since g varies as $1/r^2$, where r is the distance from the centre of the Earth. At the summit, g is reduced by about 0.18%. The fractional change in period is half of this, since $T \propto 1/\sqrt{g}$, so about 0.09%. A clock taken from Dar es Salaam to Uhuru Peak (highest point on Mount Kilimanjaro) loses roughly 80 seconds in a day. The effect is tiny but real, and it is routinely measured by precise gravity surveys.

Third, take the pendulum aboard the International Space Station, in orbit at 400km altitude. The Station and everything in it are in free fall around the Earth. The effective gravity inside is zero, which is what “weightlessness” in orbit actually means: not that gravity has vanished, but that everything is falling together. With $g_{\text{eff}} = 0$:

$$T = 2\pi\sqrt{\frac{l}{0}} = \infty$$

The pendulum simply does not oscillate. Pulled aside and released, it stays where it is, because in a frame falling freely with the bob there is no restoring force at all. Examiners enjoy this question, and the answer is unambiguous: a simple pendulum experiment cannot be done inside a satellite. *The pendulum needs a non-zero gravitational acceleration to provide its restoring force.* Take that away, and the device becomes inert.

Astronauts on the Space Station can, however, conduct mass-spring experiments perfectly well. The spring does not need gravity; it provides its own restoring force. This is one of the conceptual reasons mass-spring oscillators are studied alongside pendulums in the syllabus: they teach the same SHM equation but they fail in different environments, and a student who understands both can predict which experiment will work in which place.

Using the pendulum to measure g

Of all the experiments performed in school physics laboratories worldwide, the simple pendulum experiment for the determination of g is perhaps the most common, the most reliable, and the most rewarding. Its appeal lies in three properties at once: its apparatus is cheap (a length of string, a heavy bob, a stopwatch, a metre rule); its formula is clean ($T = 2\pi\sqrt{l/g}$ inverts to $g = 4\pi^2 l/T^2$); and the measurement uncertainty can be reduced indefinitely by patient repetition.

The basic procedure runs as follows:

1. Suspend a heavy bob (a brass cylinder, a steel ball with a hook, or a smooth stone with a string firmly tied) from a fixed pivot. A clamp on a retort stand is ideal. Measure the length l of the pendulum from the pivot to the centre of mass of the bob, using a metre rule.
2. Pull the bob aside through a small angle (less than 10° from vertical, to keep the small-angle approximation valid), and release it without giving it any initial sideways velocity.
3. Time the duration of n complete oscillations using a stopwatch. Choosing n large (say $n = 20$ or $n = 50$) reduces the influence of the human reaction time on each measurement, since the reaction time is fixed in absolute terms but the time for n oscillations grows linearly with n . The period is then:

$$T = \frac{\text{total time}}{n}$$

4. Compute g from the formula:

$$g = \frac{4\pi^2 l}{T^2}$$

A more refined version of the experiment varies the length l over five or six values, measures T at each, and plots T^2 (vertical axis) against l (horizontal axis). The graph should be a straight line through the origin, with a gradient equal to $4\pi^2/g$. The slope of this best-fit line gives g without ever needing to compute T at any single value; it averages over all the data and minimises systematic errors.

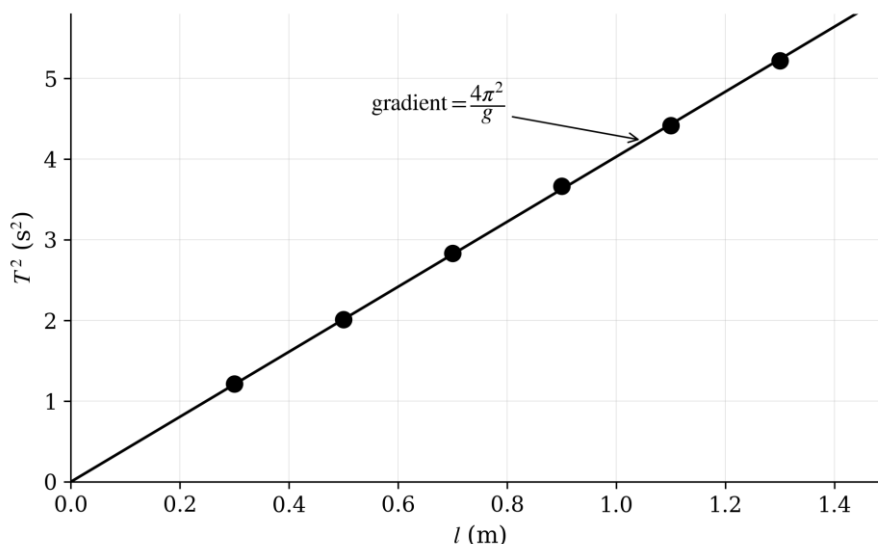


Figure: An idealised T^2 against l graph for a simple pendulum experiment. Six measured points lie on a straight line through the origin. The gradient of the best-fit line equals $4\pi^2/g$, from which g may be computed without using any single (l, T) pair.

The principal sources of uncertainty in this experiment are:

- **the measurement of l** (limited by the precision of the metre rule, typically $\pm 0.001\text{m}$, and by the difficulty of locating the centre of mass of the bob exactly), and
- **the timing of the oscillations** (limited by the human reaction time at the moments the stopwatch is started and stopped, typically $\pm 0.1\text{s}$ on each press).

Propagating these uncertainties through $g = 4\pi^2 l/T^2$ gives:

$$\frac{\Delta g}{g} = \frac{\Delta l}{l} + 2 \frac{\Delta T}{T}$$

The factor of 2 in front of $\Delta T/T$ comes from the fact that T appears squared in the denominator. Notice that the timing uncertainty contributes twice as heavily as the length uncertainty, fractionally, which is why students are urged to time many oscillations rather than few. Timing 100 oscillations instead of 10 reduces the fractional timing error by a factor of 10, and brings the dominant uncertainty down to the level of the length measurement.

Four worked examples now follow.

BINDER Example 19

In a school physics laboratory, a student timed 20 complete oscillations of a simple pendulum of length $l = 100.0\text{cm}$. The total time, measured by a stopwatch reading to the nearest 0.1s , was 40.0s . The metre rule used to measure the length reads to the nearest 0.1cm . (a) Calculate the value of g obtained in this experiment. (b) Estimate the maximum percentage error in the measured value of g .

Solution

(a) The period is the total time divided by the number of oscillations:

$$T = \frac{40.0\text{s}}{20} = 2.00\text{s}$$

The length is $l = 100.0\text{cm} = 1.000\text{m}$. Substituting into $g = 4\pi^2 l/T^2$:

$$g = \frac{4\pi^2 (1.000\text{ m})}{(2.00\text{ s})^2}$$

$$g = 4\pi^2 \times 0.250\text{ m/s}^2 = 9.87\text{ m/s}^2$$

The obtained value of g is 9.87m/s^2 .

(b) Propagating the errors:

Length uncertainty: $\Delta l = 0.1\text{cm} = 0.001\text{m}$, so $\Delta l/l = (0.001\text{ m})/(1.000\text{ m}) = 0.10\%$.

Timing uncertainty: the stopwatch reads to $\pm 0.1\text{s}$ on each press, so the worst-case error in the total time is $\pm 0.2\text{s}$ (one press at start and one at stop, each contributing $\pm 0.1\text{s}$). The error in $T = (\text{total time})/20$ is therefore $\pm 0.2\text{s}/20 = \pm 0.01\text{ s}$, so $\Delta T/T = (0.01\text{ s})/(2.00\text{ s}) = 0.50\%$.

Combined fractional error in g :

$$\frac{\Delta g}{g} = \frac{\Delta l}{l} + 2 \frac{\Delta T}{T} = 0.10\% + 2 \times 0.50\% = 1.10\%$$

The maximum percentage error is 1.10% .

So $\Delta g \approx 1.10\% \times 9.87\text{ m/s}^2 \approx 0.11\text{ m/s}^2$. The result of the experiment, including its uncertainty, is therefore $g = (9.87 \pm 0.11)\text{ m/s}^2$.

Making Sense of the Answer: The accepted value of g at sea level in Tanzania is approximately 9.81m/s^2 . The experimentally determined 9.87m/s^2 lies within the calculated uncertainty of $\pm 0.11\text{m/s}^2$ of this accepted value, so the experiment is consistent with the textbook number, which is reassuring. Notice that the timing error dominates the uncertainty, contributing 1.0% out of the 1.1% total, while the length error contributes only 0.1% . This is the pattern in nearly every real pendulum experiment.

Think Like a Physicist: If you wanted to reduce the percentage error to below 0.5% , the cheapest improvement would not be a more precise metre rule, nor a more expensive stopwatch. It would simply be

to time 100 oscillations instead of 20. The total reaction-time error remains $\pm 0.2s$, but it is now spread over a total time of about 200s instead of 40s, so $\Delta T/T$ falls by a factor of five, from 0.5% to 0.1%. The length error stays at 0.1%, and the combined fractional error becomes $0.1\% + 2 \times 0.1\% = 0.3\%$, comfortably under the half-percent target. This is the practical wisdom of the pendulum experiment: be patient with the timing, generous with the count, and the formula will do the rest.

BINDER Example 20

Two simple pendulums of lengths $l_1 = 1.00m$ and $l_2 = 0.81m$ hang side by side from the same horizontal support and are set swinging at the same instant, in phase. After how many complete oscillations of the longer pendulum do they next swing exactly in phase again? Take $g = 9.8m/s^2$.

Solution

Let the period of pendulum 1 (longer) be T_1 and the period of pendulum 2 (shorter) be T_2 .

From the simple-pendulum formula:

$$T_1 = 2\pi \sqrt{\frac{l_1}{g}}, \quad T_2 = 2\pi \sqrt{\frac{l_2}{g}}$$

Each pendulum returns to the position and direction of motion it started with only after an integer number of complete oscillations. If pendulum 1 has performed n_1 complete oscillations, the time elapsed is $n_1 T_1$; if pendulum 2 has performed n_2 complete oscillations, the time elapsed is $n_2 T_2$. The two pendulums are observed at the same physical instant, so the two elapsed times must be the same number. That single requirement gives us the in-phase condition:

$$t = n_1 T_1 = n_2 T_2$$

Each side of this equation is the total elapsed time, written in two different ways. Both pendulums are simultaneously back to their starting phase only at instants when this equation is satisfied with positive integers n_1 and n_2 . The next in-phase moment after release is the **smallest** such pair of integers.

Rearrange to isolate the ratio of integers:

$$\frac{n_1}{n_2} = \frac{T_2}{T_1} = \sqrt{\frac{l_2}{l_1}}$$

The square root in the last expression came from the period formulas: $T_2/T_1 = \sqrt{l_2/l_1}$, with the common factor $2\pi/\sqrt{g}$ cancelling.

We now substitute numbers. With $l_1 = 1.00m$ and $l_2 = 0.81m$:

$$\frac{n_1}{n_2} = \sqrt{\frac{0.81m}{1.00m}} = 0.9 = \frac{9}{10}$$

The smallest positive integers satisfying $n_1/n_2 = 9/10$ are $n_1 = 9$ and $n_2 = 10$. (The next in phase after this will occur at $n_1 = 2 \times 9 = 18$, $n_2 = 2 \times 10 = 20$.)

So to be in phase again the longer pendulum completes 9 oscillations (while the shorter completes 10).

Checking the answer:

The total elapsed time, in either form, is:

$$t = n_1 T_1 = 2\pi n_1 \sqrt{\frac{l_1}{g}}$$

Substituting:

$$t = 9 \times 2\pi \sqrt{\frac{1.00m}{9.8m/s^2}} = 18.06s$$

Or equivalently, $t = 10T_2 = 10 \times 1.806s = 18.06s$, as it must be.

Making Sense of the Answer: The clean 9:10 ratio is no accident. The two lengths, 1.00m and 0.81m, were chosen deliberately so that $l_2/l_1 = 0.81 = (9/10)^2$. Taking the square root gives $T_2/T_1 = 9/10$, so the simplest integer match is $n_1 : n_2 = 9 : 10$. If the lengths had been chosen randomly, the integers could have been very large indeed; in the worst case, the lengths are incommensurable and the pendulums never return exactly to phase. The choice of lengths in this question is therefore a small kindness to the student, made obvious by the fact that $0.81 = 0.9^2$ is a perfect square in disguise. Notice also a methodological point: by working symbolically first and substituting numbers only at the end, we obtained the answer with one cleaner formula, $t = 2\pi n_1 \sqrt{l_1/g}$, rather than two separate period calculations followed by a ratio. Wherever the algebra allows, this is the better order of operations.

Think Like a Physicist: The principle at work here is the simplest possible case of resonance, although that word will not officially appear in this chapter until the section on forced oscillations. Two oscillators with a simple integer ratio of their periods come into phase often, exchange energy efficiently if coupled, and produce all the phenomena of musical harmony. Two oscillators whose period ratio is irrational, the golden ratio for example, almost never come into phase, never exchange energy efficiently, and produce the perpetual unfinished sound of an out-of-tune instrument. The integer ratios of the pendulum lengths in this question are the same integer ratios that govern the consonant intervals of music. We shall meet them again, in different clothes, in the chapter on stationary waves.

REAL Example 21

In the staffroom at Miono Secondary, there is a wall clock whose pendulum is a brass rod with a heavy disc at the end. The clock has hung on that wall for years, and the headmaster's secretary, who has the unenviable job of calling the start and end of every lesson, has noticed something peculiar. In the cool months between June and August, when the school sits at twenty degrees Celsius from morning to night, the clock keeps perfect time. In the hot months between February and April, however, when the temperature inside the building reaches thirty-five degrees Celsius and the brass pendulum visibly glistens, the clock loses about twenty seconds a day. By the end of a hot week she has had to march into the staffroom and reset it.

Kipute, who has been reading ahead, raises this with Mr. Akilikubwa one Friday afternoon.

Kipute: "Sir, the wall clock is losing time again. The secretary is unhappy. Can you explain to her, in physics, why a hot brass pendulum is slower than a cool one?"

Mr. Akilikubwa, who has been waiting since the chapter on thermal expansion to be asked exactly this question, smiles a small triumphant smile.

Mr. Akilikubwa: "The pendulum is a brass rod. Brass, like all metals, expands when it is heated. From Chapter 2, the linear expansion of a rod under a temperature change ΔT is given by:"

$$l = l_0(1 + \alpha \Delta T)$$

Mr. Akilikubwa: "where α is the coefficient of linear expansion of brass, equal to about $19 \times 10^{-6}/K$, and ΔT is the temperature increase. Now, the period of a simple pendulum (and to a good approximation a rod-and-disc pendulum behaves like a simple pendulum of effective length close to the rod length) is:"

$$T = 2\pi \sqrt{\frac{l}{g}}$$

Mr. Akilikubwa: "Substitute the expanded length:"

$$T = 2\pi \sqrt{\frac{l_0(1 + \alpha \Delta T)}{g}} = T_0 \sqrt{1 + \alpha \Delta T}$$

Mr. Akilikubwa: "For the small fractional changes typical of thermal expansion, we may expand the square root to first order, using $(1+x)^{\frac{1}{2}} \approx 1 + \frac{x}{2}$ for small x :"

$$T \approx T_0(1 + \frac{1}{2} \alpha \Delta T)$$

Mr. Akilikubwa: "So the fractional change in period is half the fractional change in length. Plugging in numbers, a temperature increase of $\Delta T = 15K$ from $20^\circ C$ to $35^\circ C$ gives a fractional length change:"

$$\frac{\Delta l}{l_0} = \alpha \Delta T = (19 \times 10^{-6}/K)(15K) = 2.85 \times 10^{-4}$$

Mr. Akilikubwa: "and therefore a fractional period change of half this, about 1.43×10^{-4} . In a day of 86400s, the clock will be slow by:"

$$(86\,400\text{s})(1.43 \times 10^{-4}) \approx 12.4\text{s per day}$$

Mr. Akilikubwa: “Twelve seconds a day, in round figures, in agreement with the secretary’s complaint of about twenty seconds, a number subject to the rounding errors of human observation. The clock loses time in summer because its pendulum rod is thermally longer in summer.”

Kipute: “And what does the clockmaker do about it?”

Mr. Akilikubwa: “Two things, traditionally. The simple solution is to leave a small adjustment screw at the bottom of the rod that lets the position of the disc be moved up or down by a fraction of a millimetre. Moving the disc up shortens the effective pendulum length and speeds the clock; moving it down slows it. A patient secretary can compensate the seasonal change by hand.”

Mr. Akilikubwa: “The clever solution, invented by **John Harrison** in eighteenth-century England, uses two metals with different expansion coefficients arranged so that their expansions cancel. As the brass rod gets longer in summer, a steel rod, which expands less, on the other side of the system gets longer too, but by less, and the lever arrangement keeps the effective pendulum length constant. Such a pendulum is called a temperature-compensated pendulum, and a well-built one keeps time to within a second a week regardless of the season.”

Kipanga, who has just sauntered in for water and stayed for the conversation, looks up.

Kipanga: “So if I want the clock to be perfect, I should hire **Harrison**?”

Mr. Akilikubwa: “Harrison died in 1776. But you may hire any honest watchmaker on Samora Avenue, and they will do the same job today, for less.”

Making Sense of the Answer: The whole calculation rests on three ideas you already have: thermal linear expansion (Chapter 2), the pendulum period formula (this section), and the binomial approximation (BAM/Advanced mathematics). The numerical result, about twelve seconds a day for a fifteen-degree temperature change, matches what is observed in school clocks across Tanzania every dry season, and is one of the most direct everyday consequences of the temperature-dependent length of solids. The same physics is the reason a railway track buckles in the heat and a power line sags in summer; it just appears, in this case, in a slightly more elegant disguise.

Think Like a Physicist: This problem is a small example of a very large theme in physics: the physics of small effects. The fractional period change of 1.4×10^{-4} is impossible to notice in any single oscillation; you would not see it with a stopwatch. But twelve seconds compounded over a day is a noticeable amount. Many physical effects of historical importance, the precession of the equinoxes, the gravitational red shift of light, the slowing of clocks in motion, are of comparable smallness on a per-cycle basis, and become observable only when patiently accumulated. Physicists are people who believe that small effects, properly counted, can move mountains. Engineers are physicists who learned to add a temperature-compensated pendulum.

HOT Example 22

A simple pendulum of period $T = 2.00\text{s}$ swings through an angular amplitude of 2.0° on each side of the vertical. (a) Is the motion strictly simple harmonic? Justify your answer using a numerical comparison. (b) Calculate the maximum speed of the bob. (c) Calculate the height of the bob above its lowest position at the instant of maximum displacement. Take $g = 9.8\text{m/s}^2$.

Solution

The angular amplitude of the swing is $\theta_{\text{max}} = 2.0^\circ = 2.0 \times (\pi/180)\text{rad} = 0.034907\text{rad}$.

(a) The strict requirement for SHM is $\sin \theta_{\text{max}} \approx \theta_{\text{max}}$. Compute both:

$$\theta_{\text{max}} = 0.034907\text{rad}$$

$$\sin \theta_{\text{max}} = \sin (2.0^\circ) = 0.034899$$

The difference is 0.000008, a fractional error of about 2×10^{-4} , or 0.02%. The small-angle approximation is therefore correct to two parts in ten thousand at this amplitude, far better than any school stopwatch could detect, and far better than the timing accuracy of the experiment we conducted in Example 19. The motion is strictly simple harmonic to better than the limits of measurement, even though, in mathematical principle, the small-angle approximation introduces a tiny error.

(b) From $T = 2\pi\sqrt{l/g}$, solve for l :

$$l = \frac{gT^2}{4\pi^2} = \frac{(9.8\text{m/s}^2)(2.00\text{s})^2}{4\pi^2}$$

$$l = \frac{9.8 \times 4}{4\pi^2} \text{m} = \frac{9.8}{\pi^2} \text{m} = 0.9928\text{m}$$

The arc-length amplitude (the maximum displacement along the curve of motion) is:

$$A = l\theta_{\max} = (0.9928\text{m})(0.034907\text{rad}) = 0.0347\text{m} = 3.47\text{cm}$$

The angular frequency is:

$$\omega = \frac{2\pi}{T} = \frac{2\pi}{2.00\text{s}} = \pi \text{ (rad/s)} = 3.142\text{rad/s}$$

The maximum speed occurs at the lowest point (the equilibrium position) and equals $A\omega$:

$$v_{\max} = A\omega = (0.0347\text{m})(3.142\text{rad/s}) = 0.109\text{m/s}$$

So the bob's maximum speed at the bottom of its swing is approximately 0.109m/s, or about 11cm/s.

(c) The height above the lowest position at maximum displacement is:

$$h = l(1 - \cos \theta_{\max}) = 0.9928\text{m}(1 - \cos 2.0^\circ) = 6.05 \times 10^{-4}\text{m} \approx 0.6\text{mm}$$

We may **verify the answer** using energy conservation. At maximum displacement all the energy is gravitational potential energy; at the lowest point all the energy is kinetic. Setting them equal:

$$mgh = \frac{1}{2}mv_{\max}^2$$

$$h = \frac{v_{\max}^2}{2g} = \frac{(0.109\text{m/s})^2}{2 \times 9.8\text{m/s}^2} = 6.07 \times 10^{-4}\text{m}$$

Agreement to three significant figures. The two methods, SHM kinematics and energy conservation, give the same answer, as they must.

Making Sense of the Answer: A 1-metre pendulum swinging through 2° rises only 0.6mm above its lowest position at the extremes of its arc. This is much smaller than most students expect on first sight; the swing looks much larger to the eye than the height it actually achieves. The reason is geometric: at small angles the bob moves along the arc almost horizontally, with very little vertical component to its displacement. The vertical rise grows as the square of the angular amplitude, and at 2° the square is small indeed. By the time the swing reaches 30° , however, the bob is rising about 13cm, easily visible, and the small-angle approximation is no longer trustworthy.

Think Like a Physicist: The two methods agreed to three significant figures, which is the best confirmation we can ask for. SHM kinematics gave us $A = l\theta_{\max}$ for the arc amplitude, then ω , then $v_{\max} = A\omega$. Energy conservation gave us $h = v_{\max}^2/(2g)$ directly from the lowest-point speed. Both routes are valid. When two independent methods give the same number, you know your physics is right. This is one of the secret pleasures of doing physics, the relief of seeing the same answer arrive at the same destination from two different roads.

We have spent the last several pages on the simple pendulum. The same machine, a restoring force linear in displacement, an SHM equation, a period of the form 2π divided by the square root of a system constant, has appeared once again, but this time the system constant came not from a spring but from gravity acting along an arc. The pattern is now well-established: identify the restoring force, write Newton's second law, identify ω , take $2\pi/\omega$, and the period drops out.

The next three systems we shall study, a column of liquid in a U-tube, a floating cylinder bobbing in water, and a small ball sliding in the bottom of a curved bowl, will each have their own restoring force, their own derivation, and their own ω . None of them involves a spring. None of them involves gravity along an arc in the simple-pendulum sense. And yet, in each case, the same equation $a = -\omega^2x$ will reappear, and a corresponding period $2\pi/\omega$ will follow. The simple pendulum is the second of eight oscillating systems in this chapter, each different in physical detail and identical in mathematical bone structure.

OTHER STANDARD SHM SYSTEMS: U-TUBE, FLOATING BODY, AND BALL IN A BOWL

We promised at the close of the simple-pendulum section that the next three systems would have nothing in common physically and everything in common mathematically. Now is when we deliver on that promise. A column of liquid in a U-tube has no spring, no string, no gravity acting along an arc; it has only the weight of the unbalanced part of the liquid and a hollow tube to constrain it. A floating cylinder bobbing gently in calm water has no spring either; the restoring force comes from buoyancy alone. A small ball sliding in the bottom of a curved bowl has no string; gravity pulls it back to the bottom because the bowl is curved. And yet, in every one of these three systems, when we work through the geometry slowly and carefully, the same equation $a = -\omega^2 x$ emerges, and the same period $T = 2\pi/\omega$ follows.

Each of the three derivations follows the recipe established in earlier sections: write the picture before the formula, identify the restoring force as a function of small displacement, apply Newton's second law, read off ω^2 , and write down the period. The recipe is now well-rehearsed; we shall apply it three times in quick succession.

Liquid oscillations in a U-tube

Imagine a U-shaped glass tube of uniform cross-section A , standing vertically and containing a quantity of liquid of density ρ that fills both arms to the same equilibrium level. The total length of the liquid column, measured along the centre line of the tube from one free surface, around the bend, to the other free surface, is L . When the liquid is at rest, the two free surfaces are at the same height; the pressure at any point in the bend is the same from both sides; and there is no net force pushing the liquid one way or the other.

Now disturb the liquid: blow gently into one arm, or rock the tube to one side, so that the liquid surface in one arm is pushed down by a small distance y and the surface in the other arm rises by exactly the same distance y . The total mass of liquid is unchanged, of course; the liquid simply rearranges itself between the two arms while maintaining the same total volume. The two free surfaces are now $2y$ apart in height.

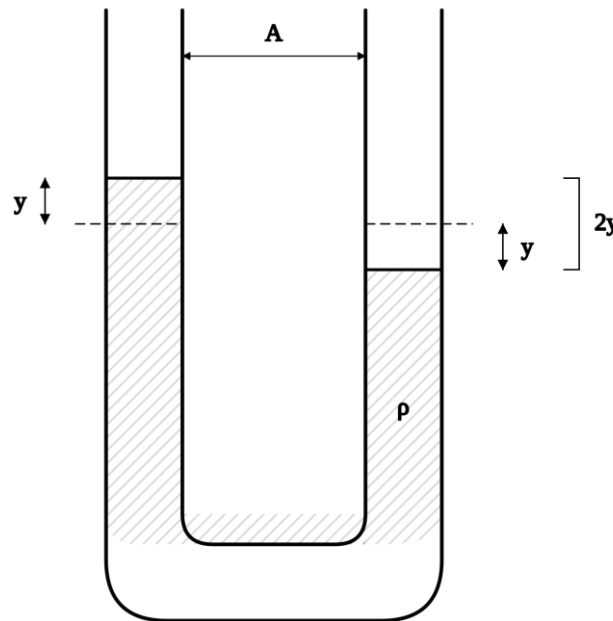


Figure: A U-tube of uniform cross-section A and density ρ , displaced from equilibrium so that one arm rises by y and the other falls by the same y . The two free surfaces are $2y$ apart, and the excess weight of the higher column provides the restoring force on the entire liquid mass.

What is the restoring force? Look only at the column of excess liquid: a column of height $2y$ in the arm that has been raised, of cross-section A , and density ρ . Its mass is $\rho A(2y)$, and its weight pulling downwards on the rest of the liquid is:

$$F = -2y\rho gA$$

The negative sign reflects that this weight always pulls the liquid back towards the equilibrium configuration: when y is positive (one arm raised), the force is downward on the raised side, which means it accelerates the liquid back to equilibrium; when y is negative (that arm lowered, the other raised), the force is in the opposite direction. It is a true restoring force, linear in the displacement, and proportional to y through the constant $2\rho gA$.

Newton's second law applies to the entire liquid column, of mass $m_{\text{liq}} = \rho AL$:

$$\rho AL \frac{d^2 y}{dt^2} = -2y\rho gA$$

The mass density and the cross-section both cancel, leaving:

$$\frac{d^2 y}{dt^2} = -\frac{2g}{L}y$$

This is the SHM equation, with $\omega^2 = 2g/L$. The period of oscillation is therefore:

$$T = \frac{2\pi}{\omega} = 2\pi \sqrt{\frac{L}{2g}}$$

It is sometimes more useful to write the period in terms of half the column length, $h = L/2$ or $L=2h$, which is the equilibrium height of liquid in each arm above the bend. Substituting:

$$T = 2\pi \sqrt{\frac{h}{g}}$$

Notice the form: it is identical to the period of a simple pendulum, with h in place of l . A liquid column of half-length 1.0m oscillates in a U-tube with the same period as a simple pendulum of length 1.0m. Neither the cross-section of the tube nor the density of the liquid matters; only the length of the liquid column does. Mercury and water in identical tubes oscillate at the same rate.

A floating cylinder bobbing in equilibrium

The next system is a hollow cylindrical tube of mass m and uniform cross-sectional area A , weighted at its lower end so that it floats vertically in a liquid of density ρ (think of a test tube with a small lead weight at its base, or a cork plumb bob, or a fisherman's float on a calm lake). At equilibrium, the cylinder sits with some depth h below the liquid surface; the upthrust on the submerged volume exactly balances the weight of the cylinder.

Setting weight equal to upthrust at equilibrium:

$$mg = \rho Ahg \Rightarrow h = \frac{m}{\rho A}$$

The equilibrium depth h is therefore the mass of the cylinder divided by the product of liquid density and cross-section. It is the depth the cylinder would naturally sink to if released gently from rest just at the surface.

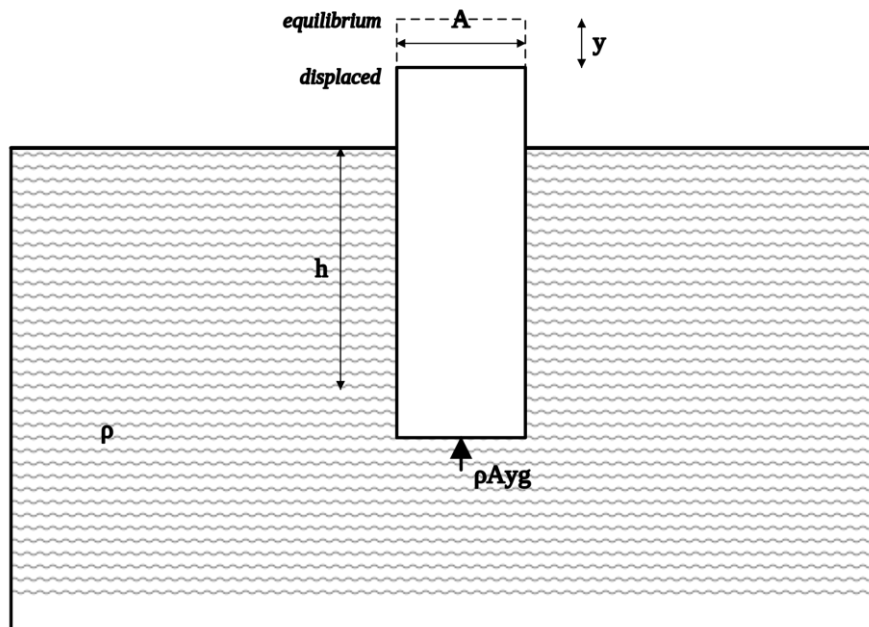


Figure: A floating cylinder of cross-section A and mass m at its equilibrium submerged depth h , then displaced downward by a further small distance y . The extra submerged volume Ay produces an excess upthrust $\rho A y g$ directed upward, which is the restoring force on the cylinder.

Now push the cylinder downward by a small additional distance y . The submerged volume increases by Ay , and so the upthrust increases by $\rho A y g$, while the weight is unchanged. The net force on the cylinder is now the difference between the new (larger) upthrust and the unchanged weight:

$$F_{\text{net}} = \rho A(h+y)g - mg = \rho A y g$$

Using $mg = \rho A h g$ from the equilibrium condition. If we measure y downward as positive, then the force is upward (negative y direction) when y is positive, and downward when y is negative. So:

$$F = -\rho A y g$$

Newton's second law:

$$m \frac{d^2 y}{dt^2} = -\rho A y g$$

Using $m = \rho A h$ from the equilibrium condition, and dividing both sides by m :

$$\frac{d^2 y}{dt^2} = -\frac{g}{h} y$$

This is the SHM equation again, with $\omega^2 = g/h$. The period is:

$$T = 2\pi \sqrt{\frac{h}{g}}$$

Once again the same form. $T = 2\pi\sqrt{h/g}$ for the U-tube, $T = 2\pi\sqrt{l/g}$ for the simple pendulum, and now $T = 2\pi\sqrt{h/g}$ for the floating cylinder. The same equation, three times, with three different physical meanings for the length scale h or l . The cross-sectional area, the density of the liquid, and the mass of the cylinder all play their part in setting the equilibrium depth, but once that depth is fixed the period depends only on it and on the local strength of gravity.

A small ball sliding in a curved bowl

The third system is a small smooth ball of radius r resting at the bottom of a polished concave bowl whose inner surface forms part of a sphere of radius R . The bowl is much larger than the ball ($R \gg r$), so when the ball is displaced slightly along the curved surface it slides without rolling, in the way a marble slides over an oiled glass surface. (The case in which the ball rolls without slipping, with friction supplying the rolling constraint, gives a slightly different period and is treated in a later section of this chapter.)

At equilibrium, the ball rests at the lowest point of the bowl, directly below the centre of curvature C . The line from C to the centre of the ball is vertical, of length $R - r$ (the centre of the ball lies on a smaller sphere of radius $R - r$ concentric with the bowl).

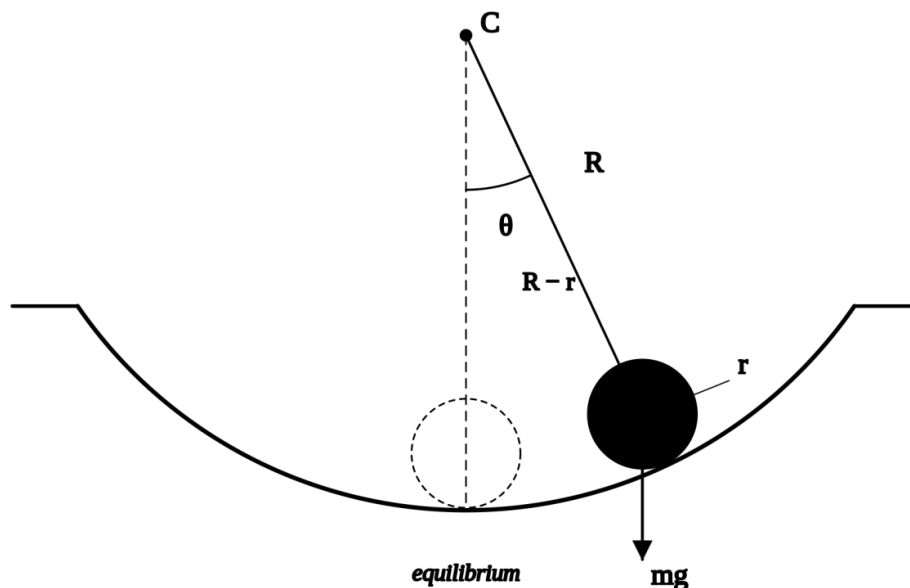


Figure: A small ball of radius r sliding in a concave bowl of radius R . The centre of curvature C is at the top, and the ball's centre travels on an arc of radius $R - r$ about C . When displaced through a small angle θ from the vertical at C , the tangential component of the weight along the arc is the restoring force.

Now displace the ball through a small angle θ from the vertical at C, measured at the centre of curvature. The ball is now resting on the sloping side of the bowl, and the weight mg resolves into two components, exactly as it did for the simple pendulum bob: a component $mg\cos\theta$ along the line from the ball back to C (balanced by the normal force from the bowl), and a component $mg\sin\theta$ tangent to the arc, directed back towards equilibrium. The tangential component is the restoring force.

The arc length of the ball's displacement, measured along the path of the ball's centre, is $s = (R - r)\theta$. For small angles, $\sin\theta \approx \theta$, so the restoring force is:

$$F_t = -mg\sin\theta \approx -mg\theta = -\frac{mg}{R - r}s$$

Apply Newton's second law along the arc:

$$m \frac{d^2s}{dt^2} = -\frac{mg}{R - r}s$$

The mass cancels:

$$\frac{d^2s}{dt^2} = -\frac{g}{R - r}s$$

SHM, with $\omega^2 = g/(R - r)$. The period is:

$$T = 2\pi \sqrt{\frac{R - r}{g}}$$

Once more the same shape. The role of the length scale is now played by $R - r$, the distance from the centre of curvature to the centre of the ball. If the ball is much smaller than the bowl, $r \ll R$, and the period reduces to $T \approx 2\pi\sqrt{R/g}$, the same as a simple pendulum of length equal to the bowl's radius.

Four worked examples now follow. Set the kettle on, prepare the calculator, and proceed. The first three are kind to the reader; the fourth is the kind of question that briefly makes a student question their life choices, then redeems itself in the final line.

BINDER Example 23

A wooden cylindrical bar of mass $m = 0.40\text{kg}$ and uniform cross-sectional area $A = 12.0\text{cm}^2$ floats vertically in fresh water, slightly weighted at one end so that its long axis stays vertical. The bar is pushed down a small distance and released. (a) Show that its motion is simple harmonic. (b) Calculate the period of oscillation. Take $\rho_{\text{water}} = 1000\text{kg/m}^3$ and $g = 9.8\text{m/s}^2$.

Solution

(a) Let the bar have mass m , cross-section A , and float in liquid of density ρ . When displaced downward by a small distance y , the extra submerged volume Ay produces an extra upthrust $\rho A y g$ directed upward, while the weight is unchanged. So the net force is restoring:

$$F = -\rho A g y$$

Newton's second law:

$$m \frac{d^2y}{dt^2} = -\rho A g y \quad \Rightarrow \quad a = \frac{d^2y}{dt^2} = -\frac{\rho A g}{m} y$$

The acceleration is proportional to the displacement (y) and directed back to equilibrium, so the motion is simple harmonic, with angular frequency:

$$\omega = \sqrt{\frac{\rho A g}{m}}$$

(b) The period follows from $T = 2\pi/\omega$:

$$T = 2\pi \sqrt{\frac{m}{\rho A g}}$$

Substitute now, with $A = 12.0\text{cm}^2 = 12.0 \times 10^{-4}\text{m}^2$:

$$T = 2\pi \sqrt{\frac{0.40\text{kg}}{(1000\text{kg/m}^3)(12.0 \times 10^{-4}\text{m}^2)(9.8\text{m/s}^2)}} = 1.16\text{s}$$

So the period is approximately 1.16s.

Making Sense of the Answer: A period of about 1.2s for a 0.40kg wooden bar of cross-section 12cm^2 in fresh water is comfortable to imagine: roughly one bob per second, of the kind one sees on any calm pond when something has just disturbed a floating reed. Notice that the formula $T = 2\pi\sqrt{m/(\rho A g)}$ can be rewritten using the equilibrium depth $h = m/(\rho A)$, giving the cleaner form $T = 2\pi\sqrt{h/g}$. This is exactly the form we derived in the prose, and it is identical to the simple pendulum formula with h in place of l .

Think Like a Physicist: If the bar were placed in salt water (density 1025kg/m^3 instead of 1000kg/m^3), the period would shorten very slightly, by about 1.2%. If instead the bar were placed in mercury, of density 13600kg/m^3 , it would barely sink at all (its equilibrium depth is now only about 2.5cm), and the period would shrink to about 0.31s, roughly three bobs per second. The bar would sit nearly on top of the mercury and twitch very rapidly. The same physics, the same formula, dramatically different numbers.

BINDER Example 24

A glass U-tube of uniform bore is partly filled with mercury. The total length of the mercury column, measured from one free surface around the bend to the other, is $L = 36.0\text{cm}$. The mercury is given a small displacement and allowed to oscillate freely. (a) Show that the oscillation is simple harmonic. (b) Calculate the period of oscillation. Take $g = 9.8\text{m/s}^2$.

Solution

(a) Let the cross-section be A , the density ρ , and the total liquid length L . When one arm is displaced downward by a small distance y , the other arm rises by the same y , and the excess column of height $2y$ has weight $2y\rho gA$, which acts as the restoring force on the entire mercury mass:

$$F = -2y\rho gA$$

Newton's second law applied to the whole mercury column, of mass ρAL :

$$\rho AL \frac{d^2y}{dt^2} = -2y\rho gA$$

$$a = \frac{d^2y}{dt^2} = -\frac{2g}{L}y$$

The acceleration is proportional to the displacement (y) and directed back to equilibrium. The motion is simple harmonic, with $\omega^2 = 2g/L$.

(b) The period is:

$$T = 2\pi \sqrt{\frac{L}{2g}}$$

Substitute $L = 0.360\text{m}$ and $g = 9.8\text{m/s}^2$:

$$T = 2\pi \sqrt{\frac{0.360\text{m}}{2 \times 9.8\text{m/s}^2}} = 0.852\text{s}$$

So the period of oscillation is approximately 0.85s.

Making Sense of the Answer: Notice what does and does not appear in the formula $T = 2\pi\sqrt{L/(2g)}$. The cross-section A does not appear; nor does the density ρ . A wide tube and a narrow tube, filled to the same

total liquid length, give exactly the same period; mercury and water filled to the same length oscillate at the same rate. The only system property that matters is the total length of the liquid column L . This cancellation of density and cross-section is one of the small surprises of the U-tube.

Think Like a Physicist: The form $T = 2\pi\sqrt{L/(2g)}$ can be rewritten as $T = 2\pi\sqrt{h/g}$ with $h = L/2$, the length of liquid in each arm above the bend. This is the same formula as for the simple pendulum and the floating cylinder; only the geometric meaning of the length scale changes from one system to the next. Whenever you see $T = 2\pi\sqrt{(\text{length})/g}$ written down in any future chapter, it is worth pausing to identify which physical length the formula is referring to.

REAL Example 25

On a calm dawn on Lake Tanganyika, a fisherman watches his small wooden float bob lazily on the slack of his line. Each time a small fish nibbles, the float dips downward by a centimetre or two and returns. The bobbing is steady, and to the fisherman's ear it has a definite rhythm, perhaps two complete cycles every second.

- (a) Explain why does the float bob at this particular rate, and discuss what would make it bob faster or slower?
- (b) For a wooden float of mass 30g and cross-sectional area 6cm^2 floating in fresh water, calculate:
 - (i) the equilibrium submerged depth h ,
 - (ii) the period of oscillation T , and show that the result agrees with the fisherman's observation of about two complete cycles per second.

Solution

The float behaves as a floating cylinder. Its period of small vertical oscillations is:

$$T = 2\pi\sqrt{\frac{h}{g}},$$

where $h = m/(\rho A)$ is the equilibrium submerged depth, the depth the float sinks to when it is left alone in still water under its own weight.

Three things determine h , and therefore T :

- 1) **The mass of the float, m .** A heavier float pushes more water aside before its weight is balanced, so its equilibrium depth is greater, and its period is longer. A wooden float and a brass cylinder of the same shape would bob at very different rates: the brass cylinder, being denser, would sink to a deeper equilibrium and bob more slowly.
- 2) **The cross-sectional area of the float at the waterline, A .** A wider float (larger A) needs less depth to displace the same weight of water, so its equilibrium depth is shallower and its period is shorter. A narrow needle-shaped float bobs more slowly than a fat barrel-shaped one of the same mass.
- 3) **The density of the liquid, ρ .** In denser water, less depth is needed to balance the float's weight, so h is smaller and the period is shorter. The fisherman would notice a tiny difference if he moved his float from the fresh water of Lake Tanganyika to the salt water of the Indian Ocean: the float would bob (about 1%) faster in salt water than in fresh.

- (b) (i) The equilibrium depth is:

$$h = \frac{m}{\rho A} = \frac{30 \times 10^{-3}\text{kg}}{(1000\text{kg/m}^3)(6 \times 10^{-4}\text{m}^2)} = 0.050\text{m} = 5.0\text{cm}$$

- (ii) And the period is:

$$T = 2\pi\sqrt{\frac{0.050\text{m}}{9.8\text{m/s}^2}} = 2\pi \times 0.0714\text{s} = 0.45\text{s}$$

$$f = \frac{1}{T} = \frac{1}{0.45\text{s}} = 2.22\text{s}^{-1}$$

The frequency is roughly two cycles every second, exactly as the fisherman's ear told him.

Making Sense of the Answer: The pleasure of this calculation is that it requires no measurement of the spring constant of anything, no estimation of the height the float was pushed down, and no assumption about how hard the fish tugged on the line. Once the fisherman picks his float (fixing its mass and width), he has fixed the period; he has essentially built a slow analogue clock that ticks at 0.45s whenever the line is twitched. Different floats give different periods, and an experienced fisherman recognises the 'rhythm' of his particular float so well that he can tell the difference between a fish nibbling and the line catching on a reed by the bobbing rate alone.

Think Like a Physicist: The cancellation in $T = 2\pi\sqrt{h/g}$ is the same one we saw in the simple pendulum: the period depends only on a length scale and on the local strength of gravity, not on the inertia or the restoring stiffness separately. The mass and the buoyancy stiffness conspire so that their effects cancel except through the equilibrium depth h . This is the third time we have met this kind of cancellation in this chapter, and it will not be the last. Wherever a system has a natural length scale that is set jointly by inertia and restoring force, the period takes the form $2\pi\sqrt{(\text{length})/g}$ or its non-gravitational analogue.

HOT Example 26

A hollow cylindrical test tube of uniform cross-sectional area $A = 1.50\text{cm}^2$ and total length $L = 20.0\text{cm}$ is loaded with lead shot at its sealed lower end so that it floats vertically in fresh water with the open end uppermost. The total mass of the tube and its load is M .

- Show that the equilibrium submerged depth is $h = M/(\rho A)$, and that the tube bobs in simple harmonic motion of period $T = 2\pi\sqrt{h/g}$.
- Find the maximum value M_{max} for which the tube can still float, that is, before it would be fully submerged at equilibrium.
- The tube is observed to bob with a period of $T = 0.70\text{s}$. Find M , and verify that $M < M_{\text{max}}$.
- An additional 5.0g of sand is now poured into the tube. Find the new period of oscillation.

Take $\rho_{\text{water}} = 1000\text{kg/m}^3$ and $g = 9.8\text{m/s}^2$.

Solution

- At equilibrium, the weight of the loaded tube is balanced by the upthrust on the submerged volume:

$$Mg = \rho Ahg \quad \Rightarrow \quad h = \frac{M}{\rho A}$$

When the tube is displaced downward by a small distance y , the extra submerged volume Ay produces an excess upthrust ρAyg upward, while the weight is unchanged. Newton's second law on the tube (whose total mass remains M) gives:

$$M \frac{d^2y}{dt^2} = -\rho Agy$$

$$a = \frac{d^2y}{dt^2} = -\frac{\rho Ag}{M}y = -\frac{\rho Ag}{\rho Ah}y = -\frac{g}{h}y$$

The motion is therefore simple harmonic, with $\omega^2 = g/h$ and period:

$$T = 2\pi\sqrt{\frac{h}{g}}$$

- The tube is fully submerged when h reaches the tube length L . Beyond that, no equilibrium configuration exists with the tube floating; any extra mass and the tube sinks. From $M_{\text{max}} = \rho AL$:

$$M_{\text{max}} = (1000\text{kg/m}^3)(1.50 \times 10^{-4}\text{m}^2)(0.200\text{m}) = 0.0300\text{kg} = 30.0\text{g}$$

- From $T = 2\pi\sqrt{h/g}$, solve symbolically for h then for M :

$$h = \frac{gT^2}{4\pi^2}, \quad M = \rho Ah = \frac{\rho Ag T^2}{4\pi^2}$$

Substituting numerical values:

$$M = \frac{(1000\text{kg/m}^3)(1.50 \times 10^{-4}\text{m}^2)(9.8\text{m/s}^2)(0.70\text{s})^2}{4\pi^2} = 0.01825\text{kg} \approx 18.2\text{g}$$

Comparing with the maximum: $M = 18.2\text{g}$ is well below $M_{\text{max}} = 30.0\text{g}$, so the tube does float, as required.

(d) After adding 5.0g of sand, the new total mass is $M' = 18.2\text{g} + 5.0\text{g} = 23.2\text{g}$. Rather than recomputing h' from scratch, exploit the proportionality $h \propto M$ (since $h = M/(\rho A)$ with ρ and A unchanged). The new period is:

$$\frac{T'}{T} = \sqrt{\frac{h'}{h}} = \sqrt{\frac{M'}{M}} = \sqrt{\frac{23.2}{18.2}} = 1.129$$

$$T' = 1.129 \times 0.70\text{s} = 0.79\text{s}$$

So the period lengthens from 0.70s to about 0.79s when the sand is added. (And $M' = 23.2\text{g}$ is still below $M_{\text{max}} = 30.0\text{g}$, so the tube continues to float.)

Making Sense of the Answer: The four parts of this question are linked: (a) sets up the formula, (b) finds the upper limit on M , (c) inverts the formula to recover the mass from a measured period, and (d) uses a ratio argument to handle a small change without redoing the full calculation. The ratio trick in (d), $T'/T = \sqrt{M'/M}$, is one of the most useful in oscillator physics; it lets you handle ‘what if I change the mass slightly’ questions in two lines instead of ten. The numerical answer 0.79s is comfortably close to the original 0.70s , as it should be when the mass changes by less than 30%.

Think Like a Physicist: This problem is a textbook example of how a single physical instrument (the loaded test tube) can be used as both a densitometer and a clock. Reading off the period gives you the equilibrium depth and therefore the total mass; reading off the equilibrium depth (with a ruler) gives you the period without ever needing a stopwatch. The instrument is self-calibrating in the sense that h and T are tied to one another by the same constant $4\pi^2/g$, which depends only on the local strength of gravity. Industrial densitometers, used to measure the density of fluids in chemical plants, are loaded test tubes of exactly this kind, instrumented so that the bobbing period can be read electronically to four significant figures. The physics is no more complicated than what you have just done.

We have analysed five different oscillating systems so far in this chapter. Each system gave us the equation $a = -\omega^2 x$ (or its angular twin), and each gave us a period of the form $2\pi/\omega$ in which the system properties combined to define a single length scale or stiffness scale. All of these systems share a deeper feature, however, that we have so far been able to ignore: the moving body in each case has been treated as a point mass. The ball in the bowl was a small ball; the cylinder bobbing in water was treated as a rigid block whose orientation never changed; the pendulum bob was a point. In every case the body moved without rotating about its own centre of mass.

The next three sections lift this restriction. We shall study three systems in which the oscillating body is a rigid extended object, and in which the rotation of the body about its own axis is an essential part of the motion. The compound pendulum (a real clock pendulum, with a rod and a disc), the torsional pendulum (a body suspended from a wire that twists), and the rolling oscillator (a ball that rolls inside a bowl rather than slides) all require the moment of inertia, the parallel axis theorem, and the radius of gyration that we developed in the *rotation of rigid bodies* chapter of **Foundations of Physics** book. If the moment of inertia and the parallel axis theorem are not fresh in your memory, take a moment to revisit the rotation of rigid bodies chapter before reading on.

THE COMPOUND PENDULUM (PHYSICAL PENDULUM)

We promised at the close of the previous section that the next three systems would lift one restriction we had been quietly keeping in the background. In every system so far, the moving body was treated as a point mass. Now we are going to allow the moving body to be a real, extended, rigid object that rotates as it oscillates. The system properties we shall need are:

- the moment of inertia I ,
- the parallel axis theorem $I_O = I_{\text{cm}} + Md^2$, and
- the radius of gyration k .

These were all developed in the **Foundations of Physics** book.

We begin with the most familiar of the three rigid-body oscillators: the **compound pendulum**, sometimes called the **physical pendulum**. The compound pendulum is the workhorse of every clockmaker's bench, every metronome, every grandfather clock face, and every laboratory experiment in which a metre rule with holes drilled along its length is hung from a pin and timed. It is not a simple pendulum, although its mathematics is built on top of the simple pendulum's, and its formula reduces to the simple pendulum's formula in a special limit we shall identify shortly.

What a compound pendulum is

A compound pendulum, sometimes called a physical pendulum, is any rigid body that is free to swing about a fixed horizontal axis that does not pass through its centre of mass. The axis is called the pivot, and the line from the pivot to the centre of mass is the body's effective swing arm. When the body is displaced through a small angle from its equilibrium position (in which the centre of mass hangs directly below the pivot), gravity supplies a restoring torque that pulls the body back, and the body oscillates.

The simple pendulum is the special case in which the swinging body is a point mass at the end of a light string. The string itself has no mass and contributes no moment of inertia; the entire mass of the system is concentrated at a single point at distance l from the pivot. The compound pendulum drops both of these idealisations: *the body has finite size, its mass is distributed throughout its volume, and the body's rotation about its own centre of mass during the swing must be accounted for.*

Examples of compound pendulums include: a metre ruler hung from a hole drilled near one end; the disc-and-rod assembly of a clock pendulum; a uniform bar pivoted at one end and allowed to swing; a wheel hung from a peg on its rim; even an athlete on a high bar at the moment they are fully extended below the bar before swinging up. All of these are real bodies of finite extent rotating about real axes that do not pass through their centres of mass, and all of them obey the same equation we are about to derive.

From restoring torque to the period formula

Consider a rigid body of mass M suspended from a fixed horizontal pivot at point O . The centre of mass of the body is at point C , at distance d from the pivot. At equilibrium, the line from O to C is vertical, and C hangs directly below the pivot. Now displace the body through a small angle θ from the vertical, as shown in the figure.

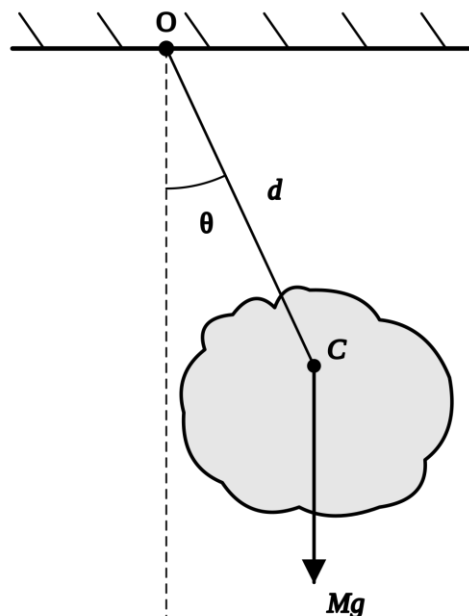


Figure: A rigid body of mass M suspended from a fixed pivot O . The centre of mass C is at distance d from O . Displaced by a small angle θ from the vertical, the body experiences a restoring torque produced by the weight Mg acting at C .

The only external forces on the body are its weight Mg (acting at C) and the reaction at the pivot (which exerts no torque about O , since its lever arm is zero). The torque about the pivot is therefore due entirely to gravity:

$$\tau = -Mg \times d \sin \theta$$

The negative sign captures the restoring nature of the torque: when θ is positive (body tilted to the right of vertical), the gravity torque rotates the body anticlockwise about O (back toward equilibrium), which is a negative torque in our sign convention. The factor d is the lever arm: the perpendicular distance from the pivot to the line of action of the weight (which is vertical at C, so the lever arm is $d \sin \theta$).

Newton's second law for rotation about the pivot states:

$$\tau = I_O \alpha = I_O \frac{d^2 \theta}{dt^2}$$

where I_O is the moment of inertia of the body about the pivot (not about the centre of mass), and $\alpha = d^2 \theta / dt^2$ is the angular acceleration. Equating the two expressions for the torque:

$$I_O \frac{d^2 \theta}{dt^2} = -Mgd \sin \theta$$

For small angles, $\sin \theta \approx \theta$, and the equation reduces to:

$$\alpha = \frac{d^2 \theta}{dt^2} = -\frac{Mgd}{I_O} \theta$$

This has the standard SHM form $d^2 \theta / dt^2 = -\omega^2 \theta$, with:

$$\omega^2 = \frac{Mgd}{I_O}$$

And the period of small oscillations is therefore:

$$T = \frac{2\pi}{\omega} = 2\pi \sqrt{\frac{I_O}{Mgd}}$$

This is the master formula for the compound pendulum. Three quantities determine the period: the moment of inertia about the pivot, I_O ; the mass of the body, M ; and the distance from the pivot to the centre of mass, d (along with the local strength of gravity g).

If the body were a point mass at distance d from the pivot, then $I_O = Md^2$, and the formula reduces to $T = 2\pi\sqrt{d/g}$, the simple pendulum formula. The compound pendulum is therefore a generalisation of the simple pendulum, and the simple pendulum is a special case of the compound.

The parallel-axis form, and the radius of gyration

It is often more convenient to express I_O in terms of the moment of inertia about the centre of mass, I_{cm} , which is usually a tabulated property of the body. The parallel axis theorem tells us how:

$$I_O = I_{cm} + Md^2$$

Substituting into the period formula:

$$T = 2\pi \sqrt{\frac{I_{cm} + Md^2}{Mgd}}$$

If we define the radius of gyration of the body about its centre of mass by $I_{cm} = Mk^2$. Then k has the dimensions of length, and the formula simplifies further:

$$T = 2\pi \sqrt{\frac{Mk^2 + Md^2}{Mgd}} = 2\pi \sqrt{\frac{k^2 + d^2}{gd}}$$

Notice that the mass M has dropped out completely. The period of a compound pendulum depends on the geometry of the mass distribution (through k and d) and on gravity, but not on the total mass at all. This is the same isolation of mass that we saw for the simple pendulum, and for exactly the same reason: *heavy bodies feel a larger gravitational torque, but they also have larger inertia to be turned, and the two effects*

cancel. Two compound pendulums of identical shape but different masses (one made of wood, the other of brass, with the same dimensions) oscillate at exactly the same period.

An equivalent form of this formula, useful for experimental work, is:

$$T^2 d = \frac{4\pi^2(d^2 + k^2)}{g}$$

This form is linear in d^2 when $T^2 d$ is plotted against d^2 , with slope $4\pi^2/g$ and intercept $4\pi^2 k^2/g$. By measuring T at several pivot positions d on the same body and plotting the graph, one extracts both g (from the slope) and k (from the intercept) from a single run of measurements. The perforated metre rule experiment in the laboratory works on exactly this principle.

The equivalent simple pendulum

Compare the compound pendulum period $T = 2\pi\sqrt{(k^2 + d^2)/(gd)}$ with the simple pendulum period $T = 2\pi\sqrt{l/g}$. For the two to give the same period, we must have $l = (k^2 + d^2)/d$. This length L_{eq} is called the **equivalent simple pendulum length**:

$$L_{eq} = \frac{k^2 + d^2}{d} = d + \frac{k^2}{d}$$

It is *the length of a simple pendulum that would oscillate at exactly the same period as the given compound pendulum*. Notice the structure: L_{eq} is the sum of two terms, d (the distance from pivot to centre of mass) and k^2/d (a correction term that captures the body's extended mass distribution). Since both terms are positive, $L_{eq} > d$ always; the equivalent simple pendulum is always longer than the lever arm of the gravity torque. A compound pendulum therefore swings more slowly than a simple pendulum of the same lever arm d ; the extra mass distributed throughout the body acts as additional rotational inertia and slows the swing. Only in the limit $k \rightarrow 0$ (a body whose mass is concentrated at a single point at distance d from the pivot, a simple pendulum) does L_{eq} reduce to d .

The minimum period

The expression $L_{eq} = d + k^2/d$ has an interesting feature: it depends on d , and not in a monotonic way. If the pivot is moved very close to the centre of mass ($d \rightarrow 0$), the term $k^2/d \rightarrow \infty$, and the period becomes very long; the body barely swings, since the gravity torque has a vanishing lever arm. If the pivot is moved very far from the centre of mass ($d \rightarrow \infty$), the term d itself dominates and L_{eq} grows without bound; the body swings slowly because the equivalent length is now very large. Somewhere between these two extremes there must be a value of d that minimises the period. To find it, differentiate L_{eq} with respect to d and set the result to zero:

$$\frac{dL_{eq}}{dd} = 1 - \frac{k^2}{d^2} = 0 \quad \Rightarrow \quad d = k$$

The minimum period therefore occurs when the pivot is placed at a distance equal to the radius of gyration from the centre of mass. At this special pivot position:

$$L_{eq, \min} = k + \frac{k^2}{k} = 2k$$

$$T_{\min} = 2\pi\sqrt{\frac{2k}{g}}$$

This is a universal result: it applies to any rigid body, of any shape. Watchmakers and pendulum-clock designers historically used this result to set the optimum balance between size and speed: a clock pendulum is positioned to give a period close to the minimum for its mass distribution, so that small adjustments to the pivot position produce only second-order changes in the period.

Five worked examples now follow. Put the kettle on, pick up your pen, and proceed. The fifth one is the kind of question that briefly makes a student wish they had not removed the disc from the wall clock at the beginning of the section, then rewards the persistence with a small physical insight.

BINDER Example 27

A uniform rod of mass $M = 1.20\text{kg}$ and length $L = 1.50\text{m}$ is suspended from one end and allowed to swing as a compound pendulum about that end. (a) Calculate the moment of inertia of the rod about the pivot. (b) Calculate the period of small oscillations. (c) Calculate the equivalent simple pendulum length and explain why it is not equal to the rod length, nor to the distance from the pivot to the centre of mass. Take $g = 9.8\text{m/s}^2$.

Solution

For a uniform rod pivoted at one end, the moment of inertia about the pivot is the standard tabulated result $I_O = ML^2/3$. The centre of mass of a uniform rod is at its midpoint, so $d = L/2$.

(a) Substitute $M = 1.20\text{kg}$ and $L = 1.50\text{m}$:

$$I_O = \frac{ML^2}{3} = \frac{(1.20\text{kg})(1.50\text{m})^2}{3} = \frac{2.70\text{kgm}^2}{3} = 0.900\text{kgm}^2$$

(b) Substitute into $T = 2\pi\sqrt{I_O/(Mgd)}$:

$$T = 2\pi\sqrt{\frac{0.900\text{kgm}^2}{(1.20\text{kg})(9.8\text{m/s}^2)(0.750\text{m})}} = 2.01\text{s}$$

Or, using the symbolic shortcut $T = 2\pi\sqrt{2L/(3g)}$ (which one obtains by substituting $I_O = ML^2/3$ and $d = L/2$ directly into the master formula):

$$T = 2\pi\sqrt{\frac{2 \times 1.50\text{m}}{3 \times 9.8\text{m/s}^2}} = 2.01\text{s}$$

(c) The equivalent simple pendulum length is:

$$L_{\text{eq}} = \frac{I_O}{Md} = \frac{0.900\text{kg} \cdot \text{m}^2}{(1.20\text{kg})(0.750\text{m})} = 1.00\text{m}$$

Or symbolically: $L_{\text{eq}} = (ML^2/3)/(M \cdot L/2) = 2L/3 = (2/3)(1.50\text{m}) = 1.00\text{m}$.

This is two-thirds of the rod length, longer than the lever arm $d = 0.75\text{m}$ (the distance from the pivot to the centre of mass) but shorter than the full rod length $L = 1.50\text{m}$.

The reason L_{eq} is not equal to d is that the rod is an extended body whose mass is distributed along its length, not concentrated at the centre of mass; the rod has its own moment of inertia about its centre of mass ($I_{\text{cm}} = ML^2/12$), which adds to the gravitational lever-arm contribution and slows the swing.

The reason L_{eq} is not equal to L is that the bottom end of the rod is moving faster (greater radius from pivot) than the top end during the swing, so the body's effective inertia is less than what one would calculate by treating the entire rod mass as concentrated at the bottom end.

Making Sense of the Answer: A 1.50m rod swings with the same period as a simple pendulum of length 1.00m , which has period $T = 2\pi\sqrt{1.00/9.8} \approx 2.01\text{s}$, exactly what we found above. The result $L_{\text{eq}} = 2L/3$ for a rod pivoted at one end is one of the most useful in compound-pendulum physics; it converts a question about a swinging bar into a question about a simple pendulum, which is much easier to think about. It is worth memorising.

Think Like a Physicist: The factor of $2/3$ comes from a small piece of integration: the moment of inertia of a rod about its end is $\int_0^L (m/L)x^2 dx = ML^2/3$, and the centre of mass is at $L/2$. The ratio $I_O/(Md) = 2L/3$ falls out of the integration alone. Whenever you see a simple, clean coefficient like $2/3$ in a compound-pendulum formula, suspect it of having come from an integration of x^2 over a uniform mass distribution. The whole machinery of moment-of-inertia integrals is always quietly behind the curtain.

BINDER Example 28

A uniform metre rule with holes drilled at intervals along its length is suspended on a horizontal pin so that it can swing as a compound pendulum. When the pin is placed in the hole at distance $d = 30.0\text{cm}$ from the centre of the rule, the period of small oscillations is observed to be $T = 1.53\text{s}$. The radius of gyration of a uniform metre rule about its centre is $k = 0.289\text{m}$. Use these data to determine the local value of g at the laboratory.

Solution

The compound-pendulum period in the radius-of-gyration form is $T = 2\pi\sqrt{(k^2 + d^2)/(gd)}$, which rearranges to:

$$g = \frac{4\pi^2(k^2 + d^2)}{T^2 d}$$

Substituting:

$$g = \frac{4\pi^2((0.289\text{m})^2 + (0.300\text{m})^2)}{(1.53\text{s})^2(0.300\text{m})} = 9.76\text{m/s}^2$$

So the local value of g is 9.76m/s^2 , which is consistent with the accepted value of 9.8m/s^2 at sea level.

Making Sense of the Answer: The compound-pendulum experiment is one of the most accurate gravimetric methods accessible in a school laboratory. Unlike the simple pendulum (which requires careful measurement of the string length and assumes a point mass at the end), the compound pendulum uses a single rigid body whose dimensions are easy to measure and whose moment of inertia is built in. A single experimental run at one pivot position, like the one above, gives g to within about a percent. A more refined experiment varies the pivot position over several values of d and plots $T^2 d$ against d^2 ; the resulting straight-line slope $4\pi^2/g$ can be measured to four significant figures, giving g correspondingly accurately.

Think Like a Physicist: The radius of gyration $k = 0.289\text{m}$ for a metre rule is itself a clean derived result: for a uniform rod of length L about its midpoint, $I_{\text{cm}} = ML^2/12$, so $k = \sqrt{I_{\text{cm}}/M} = L/(2\sqrt{3})$. For $L = 1.000\text{m}$, this gives $k = 1/(2\sqrt{3})\text{m} = 0.2887\text{m}$, which we have rounded to 0.289m in the calculation. The same formula structure $k = L/(2\sqrt{3})$ applies to every uniform rod regardless of mass; only the length matters. This independence of mass is one of the recurring small pleasures of moment-of-inertia calculations.

BINDER Example 29

A uniform circular disc of mass $M = 1.50\text{kg}$ and radius $R = 0.200\text{m}$ is suspended from a horizontal axis through a point on its rim, perpendicular to the plane of the disc. Calculate the period of small oscillations. Take $g = 9.8\text{m/s}^2$.

Solution

For a uniform disc, the moment of inertia about its centre of mass perpendicular to its plane is the standard result $I_{\text{cm}} = MR^2/2$. The pivot is at the rim, so $d = R$ (distance from rim to centre is one radius). By the parallel axis theorem:

$$I_O = I_{\text{cm}} + Md^2 = \frac{MR^2}{2} + MR^2 = \frac{3MR^2}{2}$$

Substitute into $T = 2\pi\sqrt{I_O/(Mgd)}$:

$$T = 2\pi\sqrt{\frac{\frac{3MR^2}{2}}{MgR}} = 2\pi\sqrt{\frac{3R}{2g}} = 2\pi\sqrt{\frac{3 \times 0.200\text{m}}{2 \times 9.8\text{m/s}^2}} = 1.10\text{s}$$

So the period of small oscillations is approximately 1.10s .

Making Sense of the Answer: A disc of 1.5kg and 20cm radius swung from its rim oscillates with about a one-second period. The factor $3R/(2g)$ inside the square root is general for a disc-at-rim, regardless of the mass; the period depends only on the radius of the disc and the local strength of gravity. The mass cancellation is by now a familiar feature of compound-pendulum problems.

Think Like a Physicist: Notice the role of the parallel axis theorem in this calculation. Without it, computing I_O by direct integration over the disc with the rim as the axis would have been a multi-step exercise in cylindrical coordinates. With it, the work collapses to one line: $I_O = I_{\text{cm}} + Md^2$. This is the recurring strategy in compound-pendulum physics: write down I_{cm} from a table, add Md^2 , and you have I_O for any pivot position you like. The same body, the same body, with seven different pivot positions, gives seven different periods, and the parallel axis theorem produces all seven from a single I_{cm} .

REAL Example 30

Kipanga and Kipute are watching the wall clock in Mr. Akilikubwa's laboratory. The pendulum, a thin brass rod ending in a heavy circular disc, swings back and forth with a steady rhythm. Kipanga turns to Kipute with a furrowed brow.

Kipanga: *"That clock pendulum is basically a heavy disc at the end of a thin rod, so it is almost a simple pendulum. Why does it not swing at the same rate as a simple pendulum of the same length?"*

Kipute thinks for a moment.

Kipute: *"Because a simple pendulum has all its mass at one point. The clock pendulum has mass spread out. The disc at the bottom has its own moment of inertia about the pivot, on top of the lever-arm moment of inertia Md^2 , and the rod has mass too. The total moment of inertia about the pivot is larger than Md^2 alone, so the period is longer."*

Kipanga: *"So the disc makes it slower?"*

Kipute: *"Not by much, because the disc's own moment of inertia about its centre of mass is $MR^2/2$, with the disc's radius R typically much smaller than its distance d from the pivot. The total moment of inertia is $Md^2 + MR^2/2$, and the second term is a small correction to the first. The period is correspondingly close to the simple-pendulum period of length d , but slightly larger."*

Kipanga: *"But Mr. Akilikubwa said the clock is designed so the period is exactly two seconds, so the seconds-hand ticks once per swing."*

Kipute: *"Yes! That is the whole point. The clockmaker adjusts the position of the disc on the rod to fine-tune the period. Moving the disc up shortens d and slightly changes the moment of inertia at the same time. There is a small adjustment screw at the bottom of every clock pendulum that does exactly this. The clockmaker tunes it once, and the clock keeps time forever, until the temperature changes or someone bumps it."*

Mr. Akilikubwa, who has been listening, walks over and adds:

Mr. Akilikubwa: *"And the beauty is that for small oscillations, the period is independent of amplitude. So whether the clock is wound up enough to swing through three degrees or one degree, the period is the same. That is the entire reason the pendulum clock was the most accurate timekeeping device in the world for nearly three centuries, from 1656 until quartz crystals took over in the 1930s. Galileo noticed the property; Huygens built the first working clock; Harrison perfected it for marine navigation; and the pendulum clock survived for three hundred years before electronics replaced it."*

Kipanga: *"What if the clock pendulum swings hard, with a large amplitude?"*

Mr. Akilikubwa: *"Then the small-angle approximation fails, $\sin \theta$ is no longer equal to θ , and the period does begin to depend on amplitude. The clock would run slow. That is why clock pendulums are designed to swing through very small arcs, typically less than two degrees on each side of vertical. A wall clock you can see from across the room has a noticeable swing, but if you measure it, it is almost always between one and three degrees. The visible motion is much smaller than it appears."*

Making Sense of the Answer: The dialogue captures three ideas that are easy to miss on first reading. **First:** the compound pendulum is slower than the simple pendulum of the same lever arm, because the body's own moment of inertia adds to the lever-arm moment of inertia. **Second:** clockmakers tune the period by adjusting the disc position, which changes both d and the moment of inertia at the same time. **Third:** the isochronous property of small oscillations is the reason pendulum clocks worked at all; a clock whose period depended on amplitude would lose time whenever the amplitude changed.

Think Like a Physicist: Pendulum clocks ruled timekeeping for three centuries, from Christiaan Huygens's first design in 1656 to the rise of quartz crystals in the 1930s. The key innovation was the escapement mechanism, which delivers a small impulse to the pendulum at each swing to compensate for air resistance and friction at the pivot, keeping the amplitude (and therefore the period) approximately constant over time. The compound pendulum at the heart of every grandfather clock is exactly the system we have just analysed; its period is set by the geometry of the rod and disc, and its accuracy is set by the quality of the escapement. When you next walk past a working pendulum clock, in a sitting room, in a railway station, in a museum, pause to admire it. It is three hundred years of compound-pendulum physics, distilled into a piece of furniture.

HOT Example 31

Consider again the uniform disc of Example 29: $M = 1.50\text{kg}$, $R = 0.200\text{m}$. The disc is now to be hung from a series of pivot positions along a radial line from the centre to the rim. (a) Find the pivot distance d (measured from the centre of the disc) that gives the minimum period of small oscillations. (b) Calculate this minimum period $T_{\min}$, and compare it with the period at the rim (Example 29). (c) Explain physically why the period at the rim is close to (but slightly larger than) the minimum, even though the rim is not the optimum pivot. Take $g = 9.8\text{m/s}^2$.

Solution

(a) The minimum period occurs at $d = k$.

The radius of gyration about the centre of mass is defined by $I_{\text{cm}} = Mk^2$. For a uniform disc, $I_{\text{cm}} = MR^2/2$, so:

$$k = \sqrt{\frac{I_{\text{cm}}}{M}} = \sqrt{\frac{R^2}{2}} = \frac{R}{\sqrt{2}}$$

Substitute $R = 0.200\text{m}$:

$$k = \frac{0.200\text{m}}{\sqrt{2}} = 0.1414\text{m} \approx 0.141\text{m}$$

So the optimum pivot is at distance $d_{\text{opt}} = k = 0.141\text{m}$ from the centre of the disc.

The pivot distance d that gives the minimum period is approximately 0.141m .

(b) The minimum period is $T_{\min} = 2\pi\sqrt{2k/g}$:

$$T_{\min} = 2\pi\sqrt{\frac{2 \times 0.1414\text{m}}{9.8\text{m/s}^2}} = 1.067\text{s}$$

Compare with the rim-pivot period from Example 29: $T_{\text{rim}} = 1.10\text{s}$. The ratio is:

$$\frac{T_{\text{rim}}}{T_{\min}} = \frac{1.10}{1.067} = 1.031$$

The rim-pivot period is therefore only about 3% longer than the minimum, even though the rim is not the optimum pivot.

(c) The function $L_{\text{eq}}(d) = d + k^2/d$ has a minimum at $d = k$ and is a relatively flat function of d in the neighbourhood of this minimum (because the derivative $1 - k^2/d^2$ is small near $d = k$). For the disc, the rim is at $d = R = 0.200\text{m}$ while the optimum is at $d = k = 0.141\text{m}$. These two pivot positions are close to each other (the ratio $d/k = R\sqrt{2}/R = \sqrt{2} \approx 1.41$, i.e., the rim is only about 40% farther from the centre than the optimum). The flatness of $L_{\text{eq}}(d)$ near its minimum then ensures that the period at the rim is only a few percent away from the minimum, even though the pivot positions differ by 40% in absolute terms.

Making Sense of the Answer: The condition $d = k$ for minimum period is **universal**: it applies to any rigid body, of any shape, as long as the body has a definite radius of gyration about its centre of mass. The same condition applies whether the body is a rod, a disc, a hollow sphere, an irregular blob; only the value of k changes.

Think Like a Physicist: The flatness of the period function near its minimum is what makes the compound pendulum a good gravimetric instrument. If the period varied sharply with pivot position, small experimental errors in setting up the pivot would translate into large errors in the period and therefore in the inferred g . Because the period varies slowly with d near the minimum, the experimentalist who happens to set the pivot a few millimetres off the optimum still measures a period within a small fraction of the true minimum, and can therefore extract g accurately.

We have spent the last several pages on the compound pendulum. The same recipe (find the restoring agent, write Newton's second law, identify ω^2 , take $2\pi/\omega$) has produced once again a clean SHM equation, this time for a rigid body rotating about a fixed axis. The restoring agent was gravity, acting on the centre of mass through the lever arm d . The inertia was I_O , computed via the parallel axis theorem from the body's own I_{cm} and the pivot displacement.

The next system, the torsional pendulum, replaces gravity as the restoring agent with a different mechanism entirely: the elastic resistance of a wire that has been twisted out of its natural shape.

TORSIONAL OSCILLATION OF A SOLID SPHERE AND CYLINDER

The compound pendulum used gravity as its restoring agent. The torsional pendulum, replaces gravity with a different mechanism altogether: the elastic resistance of a wire that has been twisted out of its natural shape. Hang a body from a vertical wire, twist the body about the wire's axis through a small angle, and the wire fights back with a restoring torque proportional to the twist. The mathematical structure will be familiar by now: identify the restoring agent, write Newton's second law for rotation, identify ω^2 , take $2\pi/\omega$, and the period drops out.

The torsional pendulum is, in a precise sense, the rotational twin of the mass-on-a-spring system. Where the mass-spring system has a **stiffness k** (force per unit displacement), the torsional pendulum has a **torsional constant c** (torque per unit angular displacement). Where the mass-spring system has **inertia m** (mass), the torsional pendulum has **inertia I** (moment of inertia about the wire axis). **Wherever the mass-spring formula has m or k , the torsional pendulum formula has I or c .** The translation is mechanical and the resulting period formula is the **rotational twin** of $T = 2\pi\sqrt{m/k}$.

The torsional pendulum

A torsional pendulum, in its simplest form, consists of a rigid body suspended from a fixed point by a thin vertical wire. The wire passes through the body's centre of mass (or close to it) and along the body's axis of symmetry. When the body hangs at rest, the wire is straight and untwisted; this is the equilibrium configuration. Now grasp the body and rotate it through a small angle θ about the wire's axis. The wire, attached to the body at the bottom and to a fixed support at the top, is now twisted along its length. Release the body, and the elastic resistance of the twisted wire rotates it back. The body overshoots the equilibrium, twists the wire the other way, and oscillates. This is the torsional pendulum, and we shall now derive its period.

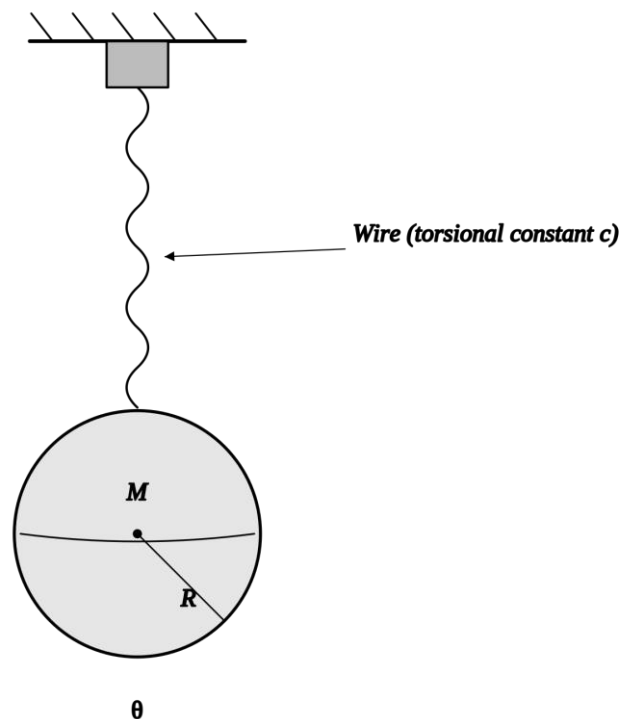


Figure: A solid body of mass M and moment of inertia I about the wire axis suspended from a vertical wire of torsional constant c . The wire is fixed at the upper end to a small support attached to the ceiling. When the body is twisted through a small angle θ about the wire axis, the wire provides a restoring torque proportional to θ , and the body oscillates.

From restoring torque to the period formula

When the wire is twisted through a small angle θ , it exerts on the body a restoring torque proportional to that angle. This is Hooke's law for torsion, and it holds as long as the wire remains within its elastic limit. Calling the proportionality constant c (the torsional constant of the wire, with units of newton-metres per radian):

$$\tau = -c\theta$$

The negative sign captures the restoring nature of the torque: when θ is positive, τ is negative, and the wire rotates the body back toward equilibrium.

Applying Newton's second law for rotation about the wire axis:

$$I \frac{d^2\theta}{dt^2} = -c\theta$$

Or, dividing both sides by I :

$$\alpha = \frac{d^2\theta}{dt^2} = -\frac{c}{I}\theta$$

This is the SHM equation in angular form, with:

$$\omega^2 = \frac{c}{I}, \quad T = 2\pi \sqrt{\frac{I}{c}}$$

Compare with the mass-spring formula $T = 2\pi\sqrt{m/k}$. The replacements are m by I and k by c . Every algebraic feature of the mass-spring system has its rotational twin in the torsional pendulum: the restoring force becomes a restoring torque, the displacement becomes an angular displacement, the spring constant becomes a torsional constant, the mass becomes a moment of inertia. The mathematical structure is identical, and the physical phenomena are correspondingly identical: oscillations at a single natural frequency, isochronous motion, conservation of energy in the undamped case, sensitivity to the system's inertial and elastic properties. All the intuition you developed for the mass-spring system applies here, with the dictionary translated.

Period for a solid sphere and a solid cylinder

The two most common bodies one suspends from a torsion wire in the laboratory are a uniform solid sphere and a uniform solid cylinder, each rotated about an axis through its centre that coincides with the wire. Their moments of inertia about this axis are standard tabulated results:

$$I_{\text{sphere}} = \frac{2}{5}MR^2, \quad I_{\text{cylinder}} = \frac{1}{2}MR^2$$

Substituting into $T = 2\pi\sqrt{I/c}$:

$$T_{\text{sphere}} = 2\pi \sqrt{\frac{2MR^2}{5c}}, \quad T_{\text{cylinder}} = 2\pi \sqrt{\frac{MR^2}{2c}}$$

The ratio of the two periods, for the same wire and equal mass and radius, is:

$$\frac{T_{\text{cylinder}}}{T_{\text{sphere}}} = \sqrt{\frac{I_{\text{cylinder}}}{I_{\text{sphere}}}} = \sqrt{\frac{1/2}{2/5}} = \frac{\sqrt{5}}{2} \approx 1.118$$

So the cylinder oscillates about 12% more slowly than the sphere of the same mass and radius. The reason is that for the same total mass, the cylinder has more of its mass at large radii from the axis (since a disc-cross-section is uniform-radius all the way out, while a sphere-cross-section is widest at the equator and narrows toward the poles). More mass at large radius means larger moment of inertia, means longer period.

Measuring the torsional constant

The torsional constant c of a wire is rarely tabulated; it depends on the wire's material, its length, its cross-sectional radius, and its surface treatment. It is therefore standard practice in the laboratory to measure c of a given wire by suspending a body of known moment of inertia from it, timing the oscillations, and rearranging $T = 2\pi\sqrt{I/c}$:

$$c = \frac{4\pi^2 I}{T^2}$$

A solid sphere of accurately known mass and radius is the body of choice, because its moment of inertia $(2/5)MR^2$ is one of the cleanest of all moments of inertia, and any small geometric imperfection in the sphere has only a quadratic effect on I (and therefore only a quadratic effect on the inferred c). Once c is known for a particular wire, the same wire can be used to measure the moment of inertia of any other body, just by hanging the body from the wire, timing the oscillations, and inverting:

$$I = \frac{cT^2}{4\pi^2}$$

The torsional pendulum is, in this sense, a rotational analogue of the spring balance: just as a spring of known stiffness lets you measure mass by reading the static extension, a wire of known torsional constant lets you measure moment of inertia by reading the period of small oscillations. The first instrument is gravimetric, the second is rotational; the underlying physics is the same.

Two worked examples now follow.

BINDER Example 32

A solid brass sphere of mass $M = 2.50\text{kg}$ and radius $R = 0.0500\text{m}$ is suspended from a vertical wire fixed to the ceiling. When the sphere is twisted through a small angle and released, it is observed to complete 8 full torsional oscillations in 20.0s. (a) Find the period of oscillation. (b) Calculate the moment of inertia of the sphere about the wire axis. (c) Determine the torsional constant of the wire.

Solution

(a) The period is the total time divided by the number of oscillations:

$$T = \frac{t}{n}$$

$$T = \frac{20.0\text{s}}{8} = 2.50\text{s}$$

(b) For a uniform solid sphere, the moment of inertia about an axis through its centre is:

$$I = \frac{2}{5}MR^2$$

$$I = \frac{2}{5}(2.50\text{kg})(0.0500\text{m})^2 = 2.50 \times 10^{-3}\text{kgm}^2$$

(c) The torsional constant follows from $T = 2\pi\sqrt{I/c}$:

$$c = \frac{4\pi^2 I}{T^2}$$

$$c = \frac{4\pi^2(2.50 \times 10^{-3}\text{kgm}^2)}{(2.50\text{s})^2} = 1.58 \times 10^{-2}\text{Nm/rad}$$

Making Sense of the Answer: A torsional constant of $\sim 16\text{mNm/rad}$ is what one expects from a thin steel wire about a metre long; it tells us that the wire applies a torque of about 16 millinewton-metres for each radian of twist. If we doubled the twist angle, we would feel double that resistance, in the wire's elastic regime. The brass sphere, with its small moment of inertia of $2.5 \times 10^{-3}\text{kgm}^2$, produces a period of 2.5s, which is comfortable for a stopwatch and gives clean timing data without requiring particularly fine instruments.

Think Like a Physicist: The clean numerical separation between the moment of inertia ($\sim 10^{-3}\text{kgm}^2$) and the torsional constant ($\sim 10^{-2}\text{Nm/rad}$) is a feature, not a bug. The two quantities differ by a factor of ten in their natural orders of magnitude, which is what makes the period (which depends on the ratio I/c) come out to a few seconds rather than a few microseconds or a few hours. In molecular vibrations or MEMS (Micro-Electro-Mechanical Systems) oscillators, the same formula gives periods of nanoseconds, simply because I in those systems is many orders of magnitude smaller than in our brass sphere.

BINDER Example 33

The brass sphere of Example 32 is now removed from the wire and replaced by a solid brass cylinder of the same mass $M = 2.50\text{kg}$ and the same radius $R = 0.0500\text{m}$, suspended so that the cylinder's axis of symmetry coincides with the wire. Without retiming, predict the new period and verify that it is about 12% longer than the period of the sphere.

Solution

The wire is unchanged, so its torsional constant c is the same. The mass and radius are unchanged. The only thing that changes is the moment of inertia:

$$I_{\text{sphere}} = \frac{2}{5}MR^2, \quad I_{\text{cylinder}} = \frac{1}{2}MR^2$$

The ratio of the new to the old period is:

$$\frac{T_{\text{cylinder}}}{T_{\text{sphere}}} = \sqrt{\frac{I_{\text{cylinder}}}{I_{\text{sphere}}}} = \sqrt{\frac{1/2}{2/5}} = \sqrt{\frac{5}{4}}$$

So the new period is:

$$T_{\text{cylinder}} = T_{\text{sphere}} \times \sqrt{5/4}$$

Substituting $T_{\text{sphere}} = 2.50\text{s}$:

$$T_{\text{cylinder}} = (2.50\text{s}) \times \sqrt{5/4} = (2.50\text{s})(1.118) = 2.795\text{s} \approx 2.80\text{s}$$

The new period is 2.80s.

The fractional increase is:

$$\frac{T_{\text{cylinder}} - T_{\text{sphere}}}{T_{\text{sphere}}} = \frac{2.8\text{s} - 2.5\text{s}}{2.5\text{s}} = 0.12 = 12\%$$

The percentage increase is 12% as predicted.

Making Sense of the Answer: Ratio reasoning is the cleanest tool for problems like this. We did not have to recompute the moment of inertia or re-derive the period. We used the fact that $T \propto \sqrt{I}$ and that the only quantity that changed between the two experiments was the moment of inertia. Two lines of algebra and one square root delivered the answer. The same trick works whenever a problem describes 'the same system but with one parameter changed': identify what is unchanged, write a ratio with the unchanged thing cancelling, and substitute.

Think Like a Physicist: The 12% difference between the sphere and cylinder periods, for the same mass and radius, is small but measurable. A torsional pendulum experiment that times 100 oscillations of each body can resolve this difference comfortably. Conversely, if you suspended an unknown solid body from the same wire and measured the period, comparing it to what you would expect for a sphere or a cylinder of the same mass and radius would reveal the body's shape category. If the unknown body's period were 2.5s, it would be sphere-like (mass concentrated near the centre); if it were 2.8s, cylinder-like (mass distributed at all radii); if it were larger still, hollow (mass concentrated near the rim). The torsional pendulum is, in this sense, a moment-of-inertia spectrometer.

We have seen the compound pendulum (gravity along a lever arm) and the torsional pendulum (elastic resistance of a twisted wire). The third and last of the rigid-body oscillators in this chapter is a body that rolls back and forth inside a curved bowl, without slipping.

ROLLING OSCILLATION ON A CURVED SURFACE

We have now seen two of the three rigid-body oscillators of this chapter. The compound pendulum used gravity acting along a lever arm; the torsional pendulum used the elastic resistance of a twisted wire. The third and last is a body that rolls back and forth, without slipping, in a curved bowl. Gravity is again the restoring agent, as in the compound pendulum, but the body's motion now combines two ingredients we have so far kept separate: translation of the centre of mass along an arc, and rotation of the body about its own axis. Rolling links these two motions, in a strict ratio set by the body's radius. The result is a period that

is longer than the simple ball-in-bowl period we derived earlier in this chapter, by a factor that depends only on the body's shape, not on its size or its mass.

There is a real reason to do this case carefully. The ball-in-bowl analysis we did earlier treated the bowl as frictionless, so that the ball merely slid up and down the curve without spinning. That idealisation is convenient, but it is rarely realistic. A marble in a steel mixing bowl rolls. A ball-bearing in a curved race rolls. A bottle rocking inside a half-pipe rolls. Rolling is the rule, not the exception, and it adds a measurable amount of inertia to the motion. The aim of this section is to make that addition precise.

The setup and the rolling constraint

A body of radius r (a sphere or a cylinder) is placed inside a smooth bowl whose inner surface is a portion of a sphere of radius R . The geometry is exactly that of the ball-in-bowl figure used earlier in this chapter. We reproduce it here for convenience, but the physics is now different: the body rolls without slipping, instead of sliding.

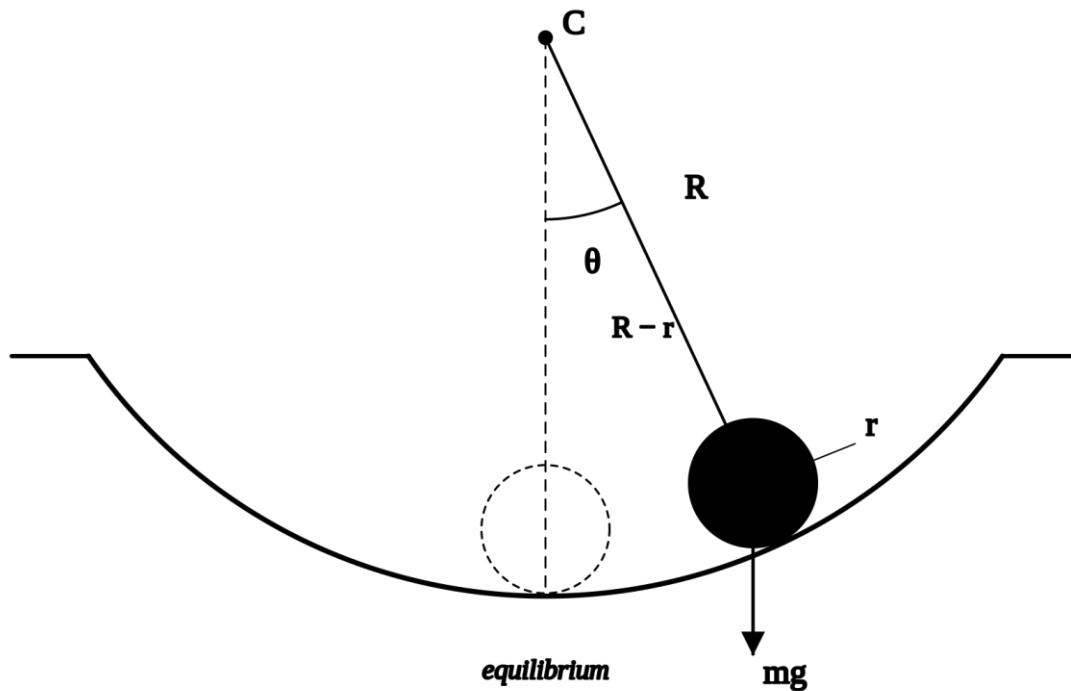


Figure: A body of radius r rolling without slipping inside a bowl of radius R . The centre of the body traces an arc of radius $R-r$ about the bowl's centre of curvature C . The angle θ measures angular position; gravity Mg supplies the restoring torque. Same geometry as the earlier ball-in-bowl figure, but the body now rolls.

Let θ be the angle, measured at the bowl's centre of curvature C , between the vertical and the line from C to the centre of the rolling body. When the body sits at the bottom of the bowl, $\theta = 0$; when it has rolled a small angle up either side, θ is positive. The centre of the body moves along a circular arc of radius $R-r$, so its arc length from the bottom is:

$$s = (R-r)\theta$$

Differentiating gives the linear speed of the centre of mass:

$$v = (R-r) \frac{d\theta}{dt}$$

Because the body rolls without slipping, its contact point with the bowl is instantaneously at rest. This forces a strict relation between the linear speed of the centre and the angular speed ω_s of the body about its own centre:

$$v = r\omega_s \quad \Rightarrow \quad \omega_s = \frac{v}{r} = \frac{R-r}{r} \frac{d\theta}{dt}$$

Notice the amplification: while the line from C sweeps through an angle $d\theta$, the body itself spins through $\frac{R-r}{r}d\theta$, which is much larger when r is small compared with R . A marble in a salad bowl spins many times faster about its own axis than about the bowl's centre. The bowl's curvature does small things; the marble's rotation does big things. Both contribute to the kinetic energy.

Translation plus rotation: the effective inertia

The body's total kinetic energy splits into two terms, one for translation of the centre of mass and one for rotation about it:

$$KE = \frac{1}{2}Mv^2 + \frac{1}{2}I\omega_s^2$$

Here v is the speed of the centre of mass of the rolling body, ω_s is its spin angular speed, and I is its moment of inertia about the centre.

Use the rolling constraint to replace ω_s with v/r :

$$KE = \frac{1}{2}Mv^2 + \frac{1}{2}I\frac{v^2}{r^2} = \frac{1}{2}\left(M + \frac{I}{r^2}\right)v^2$$

Writing the moment of inertia in terms of the radius of gyration, $I = Mk^2$:

$$KE = \frac{1}{2}M\left(1 + \frac{k^2}{r^2}\right)v^2$$

The factor in brackets is the effective inertia ratio. The rolling body behaves, for kinetic-energy purposes, like a sliding body of mass $M(1 + k^2/r^2)$. Rolling adds inertia. The hollower the body, the more inertia is added; the more concentrated its mass near its centre, the less. For the four standard cases:

$$\text{solid sphere: } \frac{k^2}{r^2} = \frac{2}{5}; \quad \text{solid cylinder: } \frac{1}{2}; \quad \text{thin spherical shell: } \frac{2}{3}; \quad \text{thin cylindrical shell: } 1$$

So the four effective-mass factors $1 + k^2/r^2$ are $7/5$, $3/2$, $5/3$, 2 , a hierarchy that mirrors the rotation-of-rigid-body table you have memorised. These four numbers are all you need for the rest of this section.

Period in the rolling bowl

We now stitch the energy method together, beginning with the potential energy. Looking back at the figure, when the body sits at the bottom of the bowl ($\theta = 0$), its centre is at the lowest point of its arc, a vertical distance $R-r$ below C. When the body is displaced by an angle θ , its centre is at vertical distance $(R-r)\cos\theta$ below C. The body has therefore risen above its equilibrium height by:

$$h = (R-r) - (R-r)\cos\theta = (R-r)(1 - \cos\theta)$$

Multiplying by Mg gives the gravitational potential energy:

$$PE = Mgh = Mg(R-r)(1 - \cos\theta)$$

For small θ , $1 - \cos\theta \approx \theta^2/2$ (the small-angle approximation, exactly as in the simple and compound pendulum sections). So:

$$PE = \frac{1}{2}Mg(R-r)\theta^2$$

Now the kinetic energy, with $v = (R-r)\frac{d\theta}{dt}$, from the equation we derived earlier:

$$KE = \frac{1}{2}M(R-r)^2\left(1 + \frac{k^2}{r^2}\right)\left(\frac{d\theta}{dt}\right)^2$$

Total energy is conserved (constant) during the motion (no friction does any work, since the rolling-contact friction acts at a point that is instantaneously at rest). So the time derivative of $E = KE + PE$ is zero. So:

Differentiating each term:

$$\frac{d}{dt}\left(\frac{d\theta}{dt}\right)^2 = 2\frac{d\theta}{dt}\frac{d^2\theta}{dt^2}, \quad \frac{d}{dt}\theta^2 = 2\theta\frac{d\theta}{dt}$$

Therefore:

$$\frac{dE}{dt} = M(R-r)^2\left(1 + \frac{k^2}{r^2}\right)\frac{d\theta}{dt}\frac{d^2\theta}{dt^2} + Mg(R-r)\theta\frac{d\theta}{dt} = 0$$

Every term on the left contains a common factor of $d\theta/dt$. Factoring it out:

$$\frac{d\theta}{dt} \left[M(R-r)^2 \left(1 + \frac{k^2}{r^2} \right) \frac{d^2\theta}{dt^2} + Mg(R-r)\theta \right] = 0$$

The body is generally moving, so $d\theta/dt$ is not always zero. The bracket must be zero on its own:

$$M(R-r)^2 \left(1 + \frac{k^2}{r^2} \right) \frac{d^2\theta}{dt^2} + Mg(R-r)\theta = 0$$

Now divide both sides by the coefficient of $d^2\theta/dt^2$ (that is, by $M(R-r)^2(1 + k^2/r^2)$). The factor M cancels and one factor of $(R-r)$ cancels between the two terms, leaving:

$$\frac{d^2\theta}{dt^2} = - \frac{g}{(R-r)(1 + k^2/r^2)} \theta$$

This is SHM with:

$$\omega^2 = \frac{g}{(R-r)(1 + k^2/r^2)}, \quad T = 2\pi \sqrt{\frac{(R-r)(1 + k^2/r^2)}{g}}$$

Compare with the sliding result obtained earlier in this chapter, $T_{\text{slide}} = 2\pi\sqrt{(R-r)/g}$. Rolling lengthens the period by exactly the square root of the effective-mass factor:

$$T_{\text{roll}} = T_{\text{slide}} \sqrt{1 + \frac{k^2}{r^2}}$$

A clean, parameter-free amplification. The bowl, the gravity, the body's mass, and even the body's radius all drop out of the ratio. Only the body's shape (through k^2/r^2) survives. This is the rolling fingerprint of the body.

With this understanding, we are now ready to see these ideas come alive through carefully chosen worked examples.

BINDER Example 34

A solid steel ball of radius 2.50cm is placed inside a hemispherical bowl whose inner radius is 30.0cm. The ball is given a small initial displacement and released, and rolls back and forth without slipping. Calculate the period of small oscillations. Take $g = 9.8\text{m/s}^2$.

Solution

The ball is a solid sphere, for which $k^2/r^2 = 2/5$, so the effective-mass factor is $1 + 2/5 = 7/5$. The period from the rolling-bowl formula is therefore:

$$T = 2\pi \sqrt{\frac{(R-r)(1 + k^2/r^2)}{g}} = 2\pi \sqrt{\frac{7(R-r)}{5g}}$$

This is a clean, small-letter formula worth memorising.

Substituting:

$$T = 2\pi \sqrt{\frac{7(0.300\text{m} - 0.0250\text{m})}{5 \times 9.8\text{m/s}^2}} = 1.245\text{s} \approx 1.25\text{s}$$

Making Sense of the Answer: About 1.25s per full oscillation. If we had ignored rolling and used the sliding formula $T_{\text{slide}} = 2\pi\sqrt{(R-r)/g}$, we would have got $2\pi\sqrt{0.275/9.8} = 1.052\text{s}$, about 18% smaller. Eighteen per cent is far above any conceivable measurement error in a laboratory stopwatch experiment, so the rolling correction is not optional.

Think Like a Physicist: There is no friction in our final equation, even though we needed friction to make the ball roll in the first place. This is not a contradiction. The static friction at the rolling contact does no work; its point of application is instantaneously at rest, by definition of rolling without slipping. It transmits force between bowl and ball but does not dissipate energy. The energy method, which counts only quantities that change during the motion, is therefore exact. If kinetic friction were present (ball was actually slipping

a little), some energy would be lost each cycle and the amplitude would decrease, and that would be the damping problem of the next section.

REAL Example 35

Mama Clyfar is preparing ugali in the kitchen. While the maize flour is steaming, her son **Arvion** drops a glass marble into a stainless-steel mixing bowl that has been emptied for washing. The bowl has an inner radius of about 18cm and the marble has a radius of about 1.5cm. The marble rolls back and forth across the bottom of the bowl, making a soft clinking sound that fascinates the boy. Estimate the period of the marble's motion, treating the marble as a solid sphere. Then estimate how many oscillations the boy could count in ten seconds before friction noticeably reduced the amplitude.

Solution

We have $R = 0.18\text{m}$, $r = 0.015\text{m}$, and the marble is a solid sphere with $1 + k^2/r^2 = 7/5$. The period is:

$$T = 2\pi \sqrt{\frac{7(R-r)}{5g}} = 2\pi \sqrt{\frac{7(0.18\text{m} - 0.015\text{m})}{5 \times 9.8\text{m/s}^2}} = 0.965\text{s} \approx 1.0\text{s}$$

About one cycle per second. In ten seconds, before damping takes noticeable effect, the boy should count roughly **ten oscillations**. (In practice, the marble's amplitude does decrease, because some kinetic energy is lost to imperfect rolling, vibration of the bowl, and air resistance. But for the first ten seconds the loss is gentle: amplitudes drop perhaps 20% over that time, while the period stays essentially constant. Damping changes amplitudes; it does almost nothing to periods, as we shall verify rigorously in the next section.)

Making Sense of the Answer: *This is an experiment any reader can repeat. Find a hemispherical metal bowl. Drop a marble. Time ten oscillations with a phone stopwatch. If the bowl is roughly 18 cm in radius and the marble roughly 1.5 cm, you should measure ~ 9.7s for ten cycles. If you measure ~ 8.2s, the marble was sliding; if you measure ~ 13s, you used a different bowl or a different marble than you thought. The formula is precise enough to catch your mistake.*

Think Like a Physicist: *Why does the formula know nothing about the marble's mass? Because gravity scales mass exactly: doubling M doubles both the inertia and the restoring weight, and the two cancel; the same cancellation that makes the period of a simple pendulum independent of bob mass. Why does it know nothing about the marble's material (glass, steel, marble proper, plastic)? Because the formula depends only on shape (k^2/r^2) and on the geometric ratio $R-r$, not on density. A glass marble and a steel ball of the same radius and shape category will oscillate identically in the same bowl. This material-independence is one of the most beautiful invariances in mechanics; we shall meet it again in damped oscillations and in resonance.*

HOT Example 36

A solid sphere of mass M and radius r is placed in a bowl of inner radius R and rolls without slipping with period T_0 . Without changing anything else, the solid sphere is replaced by a thin spherical shell of the same outer radius r and the same total mass M .

(a) Show that the new period is $T_1 = T_0\sqrt{25/21}$. (b) Express the percentage increase in the period.

Solution

(a) The two effective-mass factors come from the moments of inertia. For a solid sphere, $I = (2/5)Mr^2$, giving factor $7/5$. For a thin spherical shell, $I = (2/3)Mr^2$, giving factor $1 + 2/3 = 5/3$. The ratio of the new to the old period is therefore:

$$\frac{T_1}{T_0} = \frac{2\pi \sqrt{\frac{5(R-r)}{3g}}}{2\pi \sqrt{\frac{7(R-r)}{5g}}} = \sqrt{\frac{5/3}{7/5}} = \sqrt{\frac{25}{21}}$$

$$T_1 = T_0\sqrt{25/21}$$

(b) Computing the numerical value:

$$\sqrt{25/21} = 1.0911$$

So the percentage increase is:

$$\frac{T_1 - T_0}{T_0} = \frac{T_1}{T_0} - 1 = 1.0911 - 1 = 0.0911 = 9.11\%$$

Making Sense of the Answer: The fact that the ratio T_1/T_0 came out to $\sqrt{25/21}$, a number depending only on k^2/r^2 for the two bodies, is the strongest possible illustration of the universality we discussed earlier. The bowl could be 5cm or 5m in radius; the ball could be 1 mm or 10 cm; the gravity could be Earth's or the Moon's; the period would always increase by the same 9.11% on switching from solid sphere to thin shell.

Think Like a Physicist: The ratio $\sqrt{25/21} = 1.0911$ has a curious quality: it is not a recognisable square root. This is fine for physics, but it warns the student that the universe does not always obey clean Pythagorean ratios. Pick the right reference state, and many of the messy numbers become clean. Pick a different reference, and they need not.

We have now completed the three rigid-body oscillators of this chapter: the compound pendulum, the torsional pendulum, and the rolling oscillator. Each had its own restoring agent (gravity along a lever arm; an elastic twisted wire; gravity through a curved arc, supplemented by rotational inertia), and each gave rise to a clean period formula of the SHM kind. In every case, the body oscillated forever: undamped, lossless, eternal. Real oscillators do not behave like that. The next section confronts that universal complication.

DAMPED OSCILLATIONS

When we built the equation $a = -\omega^2 x$ at the start of this chapter, we took for granted a quiet and very generous assumption: the oscillator we were studying lost no energy to anything. The mass on the spring would oscillate at the same amplitude on its tenth cycle, on its hundredth cycle, and on its millionth. The pendulum, displaced once at noon, would still be sweeping out the same arc at midnight. The U-tube, the floating cylinder, and the ball in the bowl would all be performing tirelessly at their original amplitudes years after we had walked away from them.

This is, of course, a lie.

You have known it was a lie since Kipanga first stopped pushing the swing. The moment Kipute stopped receiving energy from him, the moment her swing was left to its own physics, it began to die. Each pass through the lowest point was a little less ambitious than the last. The metal chain creaked a little less audibly. The arc shrank. After a few minutes the swing was once again hanging straight down, at rest, looking innocent. Whatever motion it had borrowed from Kipanga had been quietly returned to the air, the bearings, and the chain, like a guest tidying up before leaving.

This section is the proper account of what happened. Real oscillators lose energy. We call the loss of energy from an oscillating system **damping**, and we call the resulting motion a **damped oscillation**. The chapter so far has dealt with the idealised, undamped, eternal version. This section brings the chapter back to the world.

We will keep the discussion qualitative. The mathematics of damped oscillation, with its exponential envelopes and its damping coefficient, belongs to a more advanced course. The physics that you must understand for examination is entirely descriptive, and entirely accessible. What we need are clear definitions, three named cases, one good diagram, and the discipline to read a graph correctly. That is a fair half-day's work.

What damping is

In any real oscillator the body that is moving rubs against something, pushes air out of the way, drags through a fluid, or loses energy to internal friction in the spring or wire that supports it. Each of these processes drains a little of the oscillation's mechanical energy and converts it, irreversibly, into the disordered thermal energy of the surrounding medium. The energy is not destroyed (Chapter 4 told us no energy ever is), but it is no longer available to the oscillation. It has migrated into the random motion of air molecules, into a faintly warmer chain at the top of the swing, into a slightly higher temperature in the metal lattice of the spring. The oscillator does not know where its energy went. It only knows it has less, and it adjusts its amplitude downward in response, the way a household budget adjusts after a long Saturday at Kariakoo.

A **damped oscillation** is therefore an oscillation in which the amplitude decreases progressively over time because energy is being lost to the surroundings. An **undamped oscillation** is one in which there is no such loss, so the amplitude remains constant from cycle to cycle. Strictly speaking, no oscillation in the real world is fully undamped. What we call undamped is whatever sits close enough to the ideal that its losses are too small to notice over the timescale we care about. The pendulum of a wall clock is very lightly damped, so over the timescale of a single swing its motion appears nearly undamped. However, over longer timescales, energy losses due to air resistance and friction accumulate. Without a continuous energy supply from the clock's mechanism, the pendulum would gradually lose amplitude and eventually come to rest.

A natural question, and a good one to settle now because students frequently get wrong, is *whether damping changes the period of the oscillation as well as the amplitude*. The honest answer has two parts. For **light damping**, which is the only case that interests us at this level and which actually applies to swings, springs, pendulums, and almost every laboratory oscillator, the period is **approximately unchanged**. The oscillator goes through its cycle at almost exactly the same rate that it would if no damping were present. What damping does, decisively and visibly, is shrink the amplitude of each successive cycle. If you are asked in an examination whether damping affects the period, the careful answer is “*approximately not, for light damping*”. Do not say it has no effect. Say it has approximately no effect, and you will be precise.

What damping definitely does is shrink the amplitude. **That is the thing to remember:** *the amplitude shrinks cycle by cycle, the period barely budges, and the energy that has left the oscillation has gone, by way of friction or air resistance, into the thermal energy of whatever surrounds it.*

Three types of damped oscillation

Different oscillators lose energy at very different rates, and the rate at which energy drains away has a beautiful effect on the kind of motion you observe. There are three qualitatively different cases. You must know all of them, and you must be able to recognise the displacement-time graph of each one on sight. The good news is that the three look completely different on a graph. The bad news is that examiners know this too.

1) Underdamped oscillation

This is the case in which the damping is **light**. Energy leaks out of the system gradually, on a timescale much longer than a single oscillation. The body therefore overshoots its equilibrium position on every pass, swings to the other side, returns, swings back, and oscillates on and on, but with each successive amplitude smaller than the one before. If you sketch the displacement against time you find a sinusoidal oscillation that is being progressively squeezed. Its peaks lie inside a smooth envelope that closes in toward zero. The motion continues until effectively all the energy has been dissipated.

Examples are everywhere. The guitar string dies away after you have plucked it. Kipute’s playground swing slows once Kipanga stops pushing. The pendulum oscillates in air. The metal ruler quivers visibly after you have flicked its end against the edge of the desk. In every one of these cases the body oscillates many times before it comes to rest, and each oscillation is a little smaller than the last. Underdamping is what happens when an oscillator is left to its own devices in a world that is only slightly resistive. That is to say, in our world.

2) Critically damped oscillation

This is the case in which the *damping is exactly heavy enough to suppress oscillation altogether, while still allowing the body to return to equilibrium as quickly as possible*. The body does not overshoot, does not swing back the other way, and does not oscillate at all. It simply moves smoothly to its equilibrium position and stays there. Critical damping is the engineer’s favourite. It gives the fastest possible return to equilibrium that does not involve oscillation.

There are two examples that you must know. The first is the **shock absorber on a car**. After the car drives over a pothole on the Morogoro road, the suspension must return the body of the car to its resting height as quickly as possible without bouncing. A passenger in a car with underdamped suspension would feel the car oscillating up and down for several seconds after every bump, which would be intolerable on a bad road and dangerous on a good one. Shock absorbers are therefore designed to be critically damped. The second example is the **pointer of a moving-coil galvanometer or ammeter**. When you switch the current on, the pointer must swing across to the correct reading and stop there as quickly as possible, so that the experimenter can read the value without having to wait for the needle to settle. An underdamped pointer would swing past the correct reading several times before coming to rest. An overdamped pointer would creep so slowly toward the correct reading that the experimenter would have given up. Critical damping is the compromise that does the job in the shortest time.

3) Overdamped oscillation

This is the case in which the *damping is so heavy that the body returns to equilibrium more slowly than the critical case*. It still does not overshoot, and it does not oscillate, but it takes longer. The damping force is so large that the body is, in effect, always moving against a thick resistance, and inches its way back to equilibrium without ever quite arriving promptly. Overdamping is rarely what you want when speed is the priority, although it is sometimes what you choose deliberately. A **heavy hydraulic door closer** is overdamped on purpose: the door must close gently and slowly enough that nobody is hit on the way out.

The same is true of a **swing door in a hospital ward**, where a rapidly returning door would be a hazard to the patients and a small embarrassment to the architect. **Heavily loaded vehicle suspensions** also tend toward overdamping. The trade-off is a softer ride at the cost of a slower recovery from disturbances.

Three cases, three behaviours: underdamping oscillates with shrinking amplitude, critical damping returns to equilibrium fastest without oscillating, and overdamping returns to equilibrium slowly without oscillating. The same physical system, say, a pointer on a spring, can be made to behave in any of the three ways depending on how much fluid resistance, magnetic braking, or internal friction you arrange around it. This is why an engineer must know all three. The kind of motion is something you choose, not something the system imposes on you. If your engineer has chosen badly, you will know. The market scale will swing, the car will bounce, the needle will dither, and someone will write to the manufacturer.

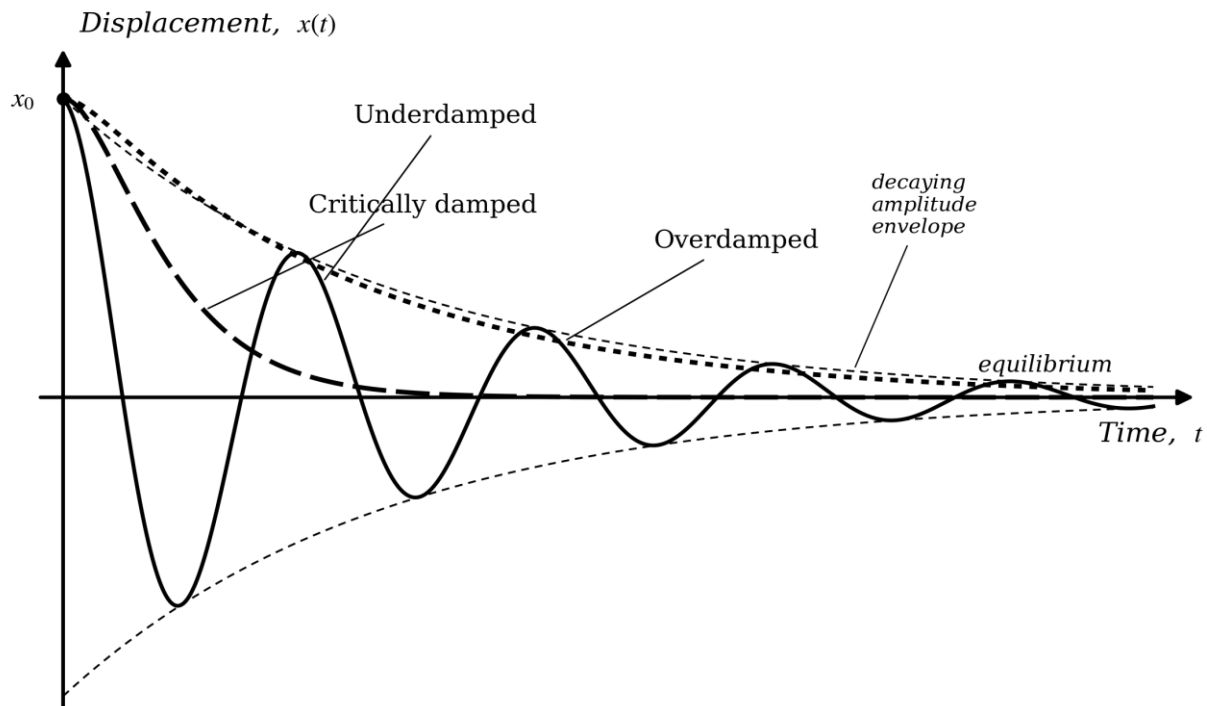


Figure: Three types of damped oscillation, all starting from the same initial displacement x_0 at $t = 0$. The underdamped curve oscillates inside a decaying envelope (dashed); each peak is smaller than the one before. The critically damped curve returns to the equilibrium axis in the shortest possible time without crossing it. The overdamped curve also returns to the equilibrium axis without crossing it, but takes longer than the critically damped curve. Neither the critically damped nor the overdamped curve oscillates. All three curves represent the same kind of physical system; only the strength of the damping has been changed.

Two worked examples now. They have been waiting patiently while the prose did its work.

BINDER Example 37

- Distinguish between damped oscillation and undamped oscillation.
- State and describe the three types of damped oscillation.
- For each type, sketch the displacement–time graph and give one engineering application.
- State whether damping affects the period of oscillation. Justify your answer.

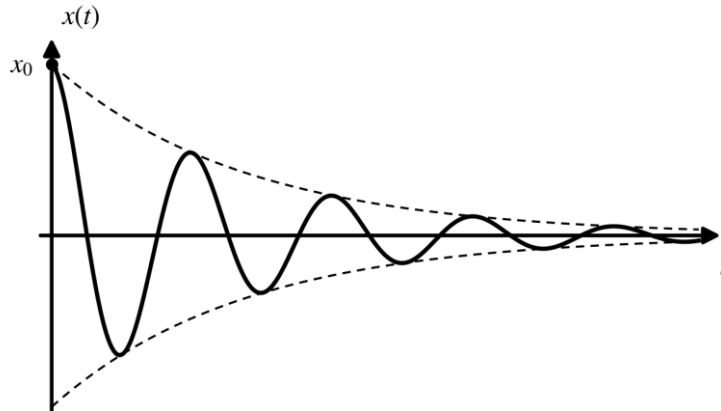
Solution

(a) A damped oscillation is one in which the amplitude decreases progressively over time because energy is lost to the surroundings, typically as thermal energy through friction, air resistance, or viscous drag. An undamped oscillation is one in which no such energy loss occurs, so the amplitude remains constant from cycle to cycle.

(b) The three types are: (1) underdamped, (2) critically damped, and (3) overdamped.

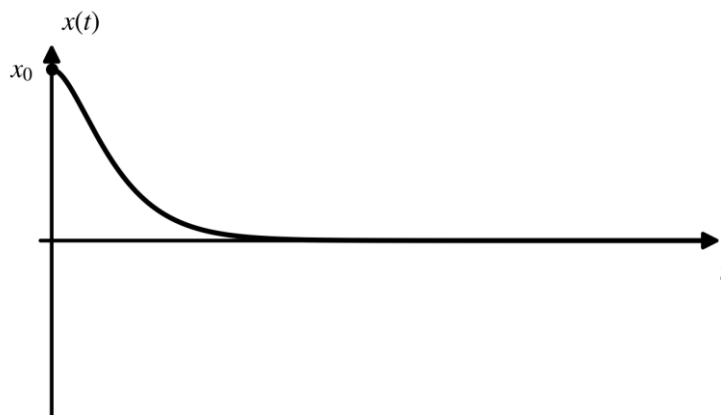
- In an **underdamped** oscillation the damping is light, so the body oscillates many times before coming to rest, with each amplitude smaller than the last.
 - In a **critically damped** oscillation the damping is exactly heavy enough that the body returns to equilibrium in the shortest possible time without oscillating.
 - In an **overdamped** oscillation the damping is heavier still, so the body returns to equilibrium without oscillating but takes longer than the critical case.
- (c) The three displacement–time graphs, are shown below. Each graph starts from the same initial displacement x_0 at $t = 0$, and the equilibrium axis is the horizontal line where $x = 0$.

Underdamped

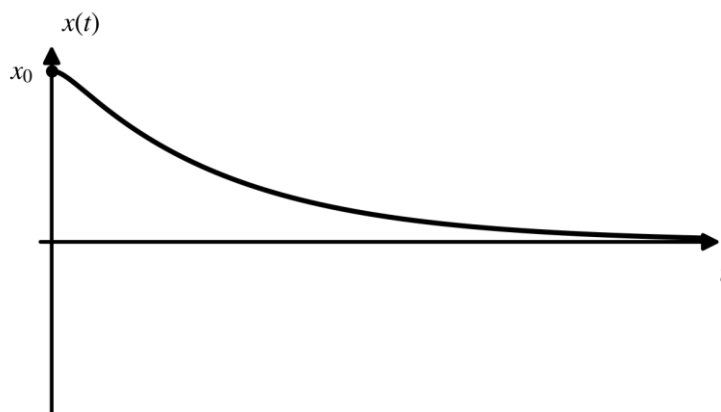


Underdamped. Sinusoidal oscillation with progressively decreasing amplitude. **Engineering example:** a guitar string (its vibration is gradually decaying after being plucked).

Critically damped



Critically damped. Smooth return to equilibrium in the shortest possible time, without crossing the time axis or oscillating. **Engineering example:** a car shock absorber (or the pointer of a moving-coil galvanometer).

Overdamped

Overdamped. Slow smooth return to equilibrium, again without oscillating. **Engineering example:** a heavy hydraulic door closer.

(d) For light damping the period is **approximately unchanged**. Damping decisively reduces the amplitude of each successive cycle, but it does not appreciably alter the rate at which the body completes a cycle. Strictly, the period increases very slightly with damping, but the change is too small to measure for the lightly-damped systems considered at this level.

Making sense of the answer: Each of the four parts is a small standalone paragraph that addresses exactly what was asked, no more. Examiners reward this structure: a clean (a) / (b) / (c) / (d) layout, a precise distinction where one is asked for, a sketch where one is asked for, and a justified statement where justification is asked for. Part (d) is the most discriminating. Students who write “damping does not affect the period” lose the precision. Students who write “approximately unchanged for light damping” earn it. The word “approximately” is small, but it is doing the work of an entire sentence.

Think like a physicist: The two engineering examples for critical damping, the car shock absorber and the galvanometer pointer, are the canonical pair examiners look for. Memorise them as a unit. They both make the same physical demand on the engineer: get to the right answer fast and stop there, with no oscillation. Whether the right answer is the resting height of the car body or the correct value of the current, the engineering principle is the same. Arrive, stop, do not bounce. Politicians could learn from this.

REAL Example 38

A butcher at Kariakoo market places a 2kg cut of beef on a spring scale. The scale pointer swings back and forth several times before settling on a reading. The butcher is annoyed. He has customers waiting.

- Which type of damping does this spring scale exhibit?
- Which type of damping would be ideal for this scale, and why?
- Explain why each of the other two types of damping would make the scale difficult to use in a busy market.

Solution

- The pointer swings back and forth several times before settling, so the scale is **underdamped**.
- The ideal scale would be **critically damped**. The pointer would swing across to the reading corresponding to the weight of the beef, stop there at once, and not oscillate. The butcher would read the value the moment the pointer arrived. Critical damping gives the fastest possible return to the equilibrium reading without overshoot, which is exactly what a busy market scale requires.
- If the scale were **underdamped**, as in part (a), the pointer would oscillate around the true reading, and the butcher would have to wait several seconds for the swings to die away. With customers waiting, this is too slow. If the scale were **overdamped**, the pointer would creep slowly toward the correct reading and take a long time to arrive. The butcher would still be waiting, only now the pointer would be moving with painful steadiness instead of swinging visibly. Both extremes waste time. Only critical damping, which avoids both the oscillation of underdamping and the sluggishness of overdamping, gives a scale that is fit for a busy market.

Making sense of the answer: *The Kariakoo example shows that the engineering choice between the three types of damping is not abstract. A working scale is a real piece of measuring apparatus, and the wrong damping ruins it for its job. The same underlying physics, the equation of motion of a mass on a spring, produces three completely different commercial outcomes depending on how much resistance the manufacturer designs into the system. The butcher's irritation is, at heart, a complaint about an under-engineered damping coefficient.*

Think like a physicist: *When you are next at Kariakoo, or at any shop that uses a mechanical balance, watch the pointer. If it swings, the manufacturer chose underdamping. If it crawls, the manufacturer chose overdamping. If it arrives at the reading at once and stops dead, the manufacturer chose critical damping, and you are looking at a well-engineered scale. The difference between the three is invisible until you put a 2kg piece of beef on the pan and start watching the pointer. Then it is the only thing in the room. The beef, briefly, becomes philosophy.*

This is the natural place to leave damping. We have learned how a real oscillator quietly loses its energy to the surroundings, what the three named cases of damping look like on a graph, and which engineering choice picks which case. What we have not yet asked is the opposite question. *What happens when energy is being put into the oscillator from the outside, by a periodic driving force?* The answer is the subject of the next section. It contains the explanation we owe Kipanga for the swing on Tuesday afternoon, and the explanation history owes itself for a famous bridge that came down in 1940. Both turn out to be the same explanation.

FREE, FORCED OSCILLATIONS AND RESONANCE

This is the chapter's third visit to the playground swing. The first visit looked at the swing alone, with Kipanga off to one side and Kipute hanging straight down and bored. We solved that swing earlier as a simple pendulum, and we found a single number that told us how long each cycle would take. The second visit, in the section just behind us, watched what happened when nobody pushed. The arc shrank, the chain creaked less audibly each time, and eventually the swing returned to its bored equilibrium. That swing was damped.

The third visit is the one we owe Kipanga an explanation for. *Why did pushing at one rate build the swing higher and higher, while pushing at any other rate seemed to fight it?* The answer is the subject of this section. We meet three new vocabulary words. We meet a graph that holds all of them at once. And we keep an old promise: by the end of the section we will know exactly what went wrong on the bridge at Tacoma Narrows in 1940, and exactly why nobody who knew physics was surprised.

We will keep the discussion qualitative. The full mathematics of forced oscillation is the subject of a more advanced course. What you do need is precise definitions, two pairs of factors that you can list off-hand, and the ability to read a resonance curve correctly. We give you each, in order.

Free oscillations

Imagine a swing displaced once and let go. With no further help from anyone, what does the swing do? It returns through its lowest point, climbs the other side, returns again. Each cycle costs the swing a tiny amount of energy through friction at the pivot and through air resistance, and so each cycle is a little smaller than the one before. Eventually, far in the future, the swing is at rest. While it oscillates, it does so at one particular rate, and the rate is set entirely by the swing itself, by how long the chain is and by how strong gravity is at this latitude. A different swing, with a different chain length, would oscillate at a different rate. The local physical properties of the swing fix the rate, and the rate is the same whether you released the swing from a tiny initial displacement or a generous one.

This rate is called the **natural frequency** of the system. It is written f_0 (sometimes ω_0), with the subscript zero reminding us that nobody is driving the system. For a simple pendulum of length l in a gravitational field of strength g , the natural frequency is $f_0 = (1/2\pi)\sqrt{g/l}$. For a mass m on a spring of stiffness k , the natural frequency is $f_0 = (1/2\pi)\sqrt{k/m}$. The two formulas are the same shape: a square root that contains the **restoring agent on top** and the **inertial agent on the bottom**. We will see this same shape repeated in every oscillator we ever meet.

A system that has been displaced once and is then left alone, with no continuing external driving force, is said to undergo **free oscillation**. "Free" here does not mean unrestrained. It means free of any external driver. Damping is normally still present, so the amplitude does still decay, but no outside source is pumping fresh energy in. The defining feature of a free oscillation is therefore the absence of a driver, not the absence of damping.

A natural question now arises. *What factors determine the period of free oscillation of a mechanical system?* The answer comes straight from the formulas above. There are two factors, no more and no less. The **first** is the *inertia of the system*, mass m for a linear oscillator, or moment of inertia I for a rotational one. The **second** is the *restoring force constant*: spring stiffness k for a mass on a spring, or the equivalent for the system in question (mg/l for a pendulum, $2\rho gA$ for a U-tube, and so on).

One more property of free oscillation deserves a sentence. In the absence of damping, the amplitude of a free oscillation is constant, and equals the initial displacement that set the motion going. In the presence of light damping, the amplitude decays gradually as we saw in the previous section, but the period stays approximately the same. The natural frequency is a property of the system, not of the amplitude or of how violently it was displaced.

Forced oscillations

Now the situation Kipanga walked into. A periodic external force is applied to the system continuously: a hand on a swing, a piston on a spring, a wind on a bridge, a needle on a vibrating phone. The driver chooses the frequency at which the force is applied, called the **driving frequency**, and this frequency may be quite different from the natural frequency of the system. The driver can push gently or strongly, slowly or quickly, but the system has no say in the rate at which the pushes arrive.

A system that is being driven this way is said to undergo **forced oscillation**. The word "forced" is the right one: the system is not allowed to oscillate at its own natural frequency. After a brief settling time, called the **transient**, the system gives up and oscillates at the driving frequency, not at its own natural frequency. This is a subtle point and a common one to miss. A forced oscillator does not oscillate at the rate it would prefer. It oscillates at the rate the driver dictates. The natural frequency does not disappear. It hides, and waits to matter.

It matters in the size of the response. The amplitude of the resulting forced oscillation depends on two things. The **first** is *how close the driving frequency is to the natural frequency*. The closer the two frequencies, the larger the amplitude of the forced oscillation. The **second** is the *amount of damping in the system*. The lighter the damping, the larger the amplitude.

Resonance

Resonance is the special case of forced oscillation in which the *driving frequency equals the natural frequency of the system*. At this single frequency, energy is transferred from the driver to the oscillator with *maximum efficiency*, and the *amplitude reaches its maximum value*. Every push from the driver arrives at exactly the moment the system is most receptive to it, and so every push adds, rather than partly cancels, the energy already in the system. The amplitude grows. This is what Kipanga discovered on Tuesday afternoon, although he did not yet have the word for it.

Without damping, the amplitude at resonance would grow without limit. Each cycle would add a fixed quantum of energy and lose none, so the amplitude would climb until something physical tore. In practice, damping always exists, and damping cuts the resonance off at a finite amplitude. The lighter the damping, the higher and the narrower the resonance peak. The heavier the damping, the lower and the broader the peak. The graph that shows all of this on a single set of axes is the resonance curve, and you will be expected to sketch it from memory.

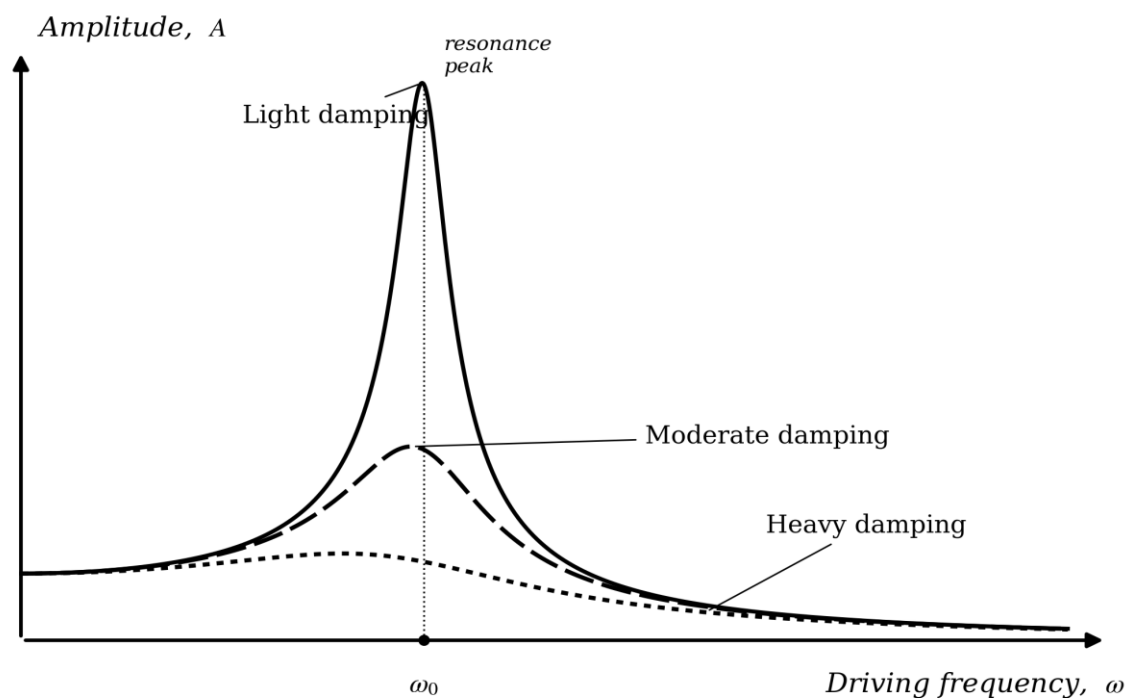


Figure: The resonance curve. Amplitude of forced oscillation A is plotted against driving frequency ω , for three levels of damping. With light damping the curve has a tall, narrow peak at the natural frequency ω_0 ; with moderate damping the peak is shorter and broader; with heavy damping the peak is short, very broad, and may even be slightly displaced to lower frequency. All three curves rise from a small value at very low driving frequencies, peak near ω_0 , and fall away at very high driving frequencies.

Useful resonance

When you want to amplify, resonance is your friend. Engineers and physicists exploit it in at least four places that you will recognise from your own pocket and your own kitchen.

1) Musical instruments

Every instrument with a hollow body, a guitar, a violin, a kora, a marimba, a piano, depends on resonance. The strings (or the bars, or the air column) vibrate at certain natural frequencies, and the body of the instrument is shaped so that it resonates at those same frequencies and amplifies them into something the audience can hear. Without resonance, a plucked guitar string would be barely audible across a quiet room.

2) Tuning a radio

Inside every radio receiver is an electrical oscillator, made of an inductor and a capacitor, whose natural frequency you can adjust by turning the tuning knob (or, in modern devices, by pressing a button). When the natural frequency of this LC circuit matches the broadcast frequency of a particular station, the circuit resonates with the radio waves arriving from that station and ignores the rest. Tuning a radio is, at heart, the engineering of resonance.

3) Microwave ovens

The oven generates electromagnetic waves at a frequency close to one of the natural rotational frequencies of the water molecule. Water molecules in your *sufuria* absorb these waves resonantly, rotate vigorously, and convert the absorbed energy into heat. The food cooks. Most of the rest of your kitchen, lacking water, is left cold.

1) Magnetic resonance imaging (MRI)

The scanner places the patient in a strong magnetic field, which makes hydrogen nuclei in the body precess at a particular frequency. A pulsed radio wave at exactly that frequency causes the nuclei to absorb energy resonantly, and the energy they re-emit a moment later is used to build the image. The scan is, in effect, a controlled resonance experiment performed on every cell of the body at once.

Harmful resonance

When you do not want to amplify, resonance is your enemy. Engineers spend a great deal of design time arranging that natural frequencies do not match the dominant frequencies of any forces the system is likely to experience. The price of forgetting this discipline is steep, as the following examples have, over the years, made clear.

1) Tacoma Narrows Bridge, USA, 1940

A steady wind across the deck of the newly opened bridge happened to drive the deck at a frequency close to one of its torsional modes. The deck twisted with growing amplitude, and on the morning of 7 November 1940 it tore itself apart and collapsed into the strait below. There is film of the collapse. It is one of the most studied engineering failures in history, and it is the reason every modern long-span bridge is designed with explicit attention to its torsional and bending natural frequencies.

2) Buildings in earthquakes

An earthquake is a broad spectrum of seismic frequencies passing through the ground. If a building's natural frequency happens to lie within the dominant band of the earthquake, the building shakes resonantly with growing amplitude until it cracks or falls. Modern earthquake-resistant design tunes a building's natural frequencies far away from the typical earthquake spectrum, and adds heavy dampers (often huge masses suspended on springs near the top of the building) to absorb energy that resonance would otherwise turn into damage.

1) Soldiers crossing footbridges

Marching soldiers walking in unison apply periodic force to the bridge at the rhythm of their step. If that rhythm matches one of the bridge's natural frequencies, the cumulative effect can be enough to set the bridge oscillating dangerously. This is why soldiers are ordered to break step when crossing any footbridge. The order looks superstitious; the physics behind it is not.

1) Engine vibration in vehicles

Every engine produces vibrations at characteristic frequencies set by its rotation rate. Certain panels, exhaust pipes, mirrors, or trim pieces in a poorly designed vehicle have natural frequencies that match those vibrations at common engine speeds. The result is the buzzing rattle that drivers know well, and that returns at exactly the same engine speed every time. The fix, when it can be afforded, is to retune the natural frequencies of the offending parts away from the engine's working range.

The same physics drives the swing higher and brings the bridge down. Resonance is indifferent to consequences. It does what the equations say it will do, and the engineer who knows the equations decides whether resonance helps or hurts. We carry that thought into the worked examples, and into the conversation Kipanga has been waiting all chapter to have.

Two worked examples now. The first is examination practice, dressed in the familiar four-part clothing. The second is a conversation.

BINDER Example 39

- Distinguish between free oscillation and forced oscillation.
- State two factors that determine the period of free oscillation of a mechanical system.
- State two factors that determine the amplitude of a forced oscillation.
- Give one example of resonance that is useful in everyday life and one example that is harmful. Explain the physics in one sentence each.

Solution

(a) A **free oscillation** is one in which the system is displaced once and then left to oscillate under its own restoring force only, with no continuing external driving force. The system oscillates at its own natural frequency. A **forced oscillation** is one in which a periodic external driving force is continuously applied to the system, and the system oscillates at the driving frequency, not at its own natural frequency.

(b) The two factors are:

- the **inertia** of the system (mass m for a linear oscillator, or moment of inertia I for a rotational oscillator), and
- the **restoring force constant** of the system (spring stiffness k , or the equivalent restoring parameter for the system in question).

(c) The two factors are:

- how **close the driving frequency is to the natural frequency** of the system (the closer they are, the larger the amplitude), and
- the **amount of damping** present (the lighter the damping, the larger the amplitude).

(d) **A useful example:** tuning a radio. The natural frequency of an LC circuit inside the receiver is adjusted to match the broadcast frequency of the desired station, so that the circuit resonates with the radio waves from that station and amplifies them while ignoring others.

A harmful example: the Tacoma Narrows Bridge collapse of 1940. A steady wind drove a torsional mode of the bridge deck at its natural frequency; the resonant amplitude grew until the deck tore itself apart and the bridge fell into the water below.

Making sense of the answer: The four parts test the four most frequently examined ideas of this section: the free / forced distinction, the two factors fixing the natural frequency, the two factors fixing the amplitude of the response, and one beneficial and one harmful example of resonance. The structure of the answer mirrors the structure of the question. Examiners reward this. Notice that parts (b) and (c) are answered with two named factors each, no fewer and no more, and that each factor is given in the precise wording the section uses. The candidate who writes "the mass and the springiness" will earn fewer marks than the candidate who writes "the inertia of the system and the restoring force constant". Vague paraphrases lose marks; the section's own phrasing earns them.

Think like a physicist: Notice how part (d) has no calculation. Resonance is a phenomenon that can be tested, and is regularly tested, with no numbers at all. The grade depends entirely on the precision of the example and the precision of the one-sentence explanation. Pick examples whose physics you can explain in one breath. The radio and the Tacoma Narrows Bridge are good for this: each has a one-sentence story, and each is famous enough that the examiner will recognise it without needing context. Resist the temptation to invent exotic examples that you cannot explain. The familiar example with a clean explanation outscores the exotic example with a vague one, every time.

REAL Example 40

It is the same school field, two days later. Kipanga finds Mr. Akilikubwa on the bench by the boundary wall. Kipute is sitting near him, looking impatient with both of them in equal measure. The afternoon is hot. The swing, in the distance, is empty.

Kipanga: "Sir, about the swing on Tuesday. I think I have understood."

Mr. Akilikubwa: "Tell me."

Kipanga: "For the first few minutes, every push I gave was undoing what the swing was already doing. I was pushing while it was rising, then pushing again while it was on its way back. The pushes cancelled. The swing barely grew. Then I waited. I pushed at the moment Kipute crossed the lowest point and was moving away from me. Then I pushed again, at exactly the same moment of the next cycle. And again. The pushes added up. The swing rose."

Mr. Akilikubwa: "And what is the name of that matching?"

Kipanga: "Resonance."

Kipute: "It took him three lessons to remember that word."

Mr. Akilikubwa: "Be patient with him, Kipute. He had to feel it before he could name it. That is the right order, when there is no examiner standing over you."

Kipanga: "But why did the swing build so high? At one point I thought it would loop right over the top bar."

Mr. Akilikubwa: "Because at resonance the energy you put in adds to the energy already in the system, every cycle, with no cancellation. The amplitude grows for as long as you keep pushing at the right frequency, until something stops it. In your case the swing eventually rose so high that the small-angle approximation broke down. The simple-pendulum equations stopped applying. Kipute became briefly airborne."

Kipute: "I forgive him. The physics was beautiful."

Mr. Akilikubwa: "And there is your last lesson. The same resonance that built that swing has, at other times, built more dangerous things. A wine glass cracks when a singer holds the right note. An earthquake

destroys a building whose natural frequency happens to match the dominant tremor. A wind plays the wrong tune across a bridge at the wrong speed, and the bridge accepts the invitation."

Kipanga: *"Tacoma Narrows."*

Mr. Akilikubwa: *"Tacoma Narrows. November 1940. The bridge was lightly damped, the wind was steady, and the wind happened to drive a torsional mode of the bridge deck at exactly the natural frequency of that mode. The deck twisted with growing amplitude until it tore itself apart. There is a film. You have seen it."*

Kipanga: *"Yes, sir."*

Mr. Akilikubwa: *"The wind did not know it was tuning a bridge. The bridge did not know it was singing back. Resonance does not know what it is doing. It is indifferent to consequences. It does the same thing for a swing, for a glass, for a bridge, for the cavity behind the soundboard of a violin. The mathematics is the same. The outcome depends only on what we choose to put in resonance with what."*

Kipute: *"Resonance is indifferent. It does exactly what the physics says."*

Mr. Akilikubwa: *"Exactly. We use it where we want to amplify, and we avoid it where we want to survive. The tuner of a violin and the engineer of a bridge are working on opposite sides of the same equation."*

Mr. Akilikubwa: *"The bridge did not know it was falling."*

The bridge did not know it was falling. The string, as it happens, knows even less. Mr. Akilikubwa's image of the violin and the bridge working on opposite sides of the same equation has one more thing to give us, and we collect it in the next section. So far the oscillations of this chapter have all sat in one place: the swing returned to its bench, the spring returned to its natural length, the bridge twisted around its own deck. None of them travelled. In the next section the oscillations begin to move. We pull a string tight between two fixed points, pluck it, and watch a transverse disturbance run along its length at a perfectly definite speed, set by two perfectly ordinary quantities: how tightly the wire is stretched, and how heavy a metre of it is. The speed has a formula. The formula has a derivation. And the formula is the doorway from oscillation, which has occupied this chapter, into the physics of waves, which will occupy the next.

VELOCITY OF VIBRATION OF A STRING

We are about to close the chapter with a question that has waited politely all the way through it. When you pluck a stretched string, *what determines the speed at which the disturbance runs along the string?* Mr. Akilikubwa's parting line about the violin and the bridge working on opposite sides of the same equation was a hint. We are now in the workshop of the violin, the zeze, and the steel guitar, and we ask the question that every instrument maker has had to answer in the dark, by ear: *how do I make this string vibrate at a particular pitch?*

The answer comes in two halves. The first half, given here, is the speed at which a transverse vibration travels along the string. The second half, given in the next chapter, combines that speed with the boundary condition that the two ends of the string are clamped, and from that combination it pulls out the harmonics, the pitches, and the entire physics of stringed instruments. This section is the doorway. The next chapter is the room behind it.

The section also serves a methodological purpose. We do not derive the wave-on-a-string speed by writing down a wave equation and solving it. That belongs to a more advanced course. We derive it by dimensional analysis, a method that uses nothing but the units of physical quantities to deduce the form of an equation. Dimensional analysis cannot tell you everything. It cannot fix numerical constants. But it can fix the shape of an equation, and for many physical questions the shape is what you need. Learning the method here is, on its own, worth the section.

What the speed could possibly depend on

Before writing any equation, ask physically: what properties of the string could possibly determine *how fast a transverse disturbance travels along it?* Reach for two answers, both of which the body knows before the algebra arrives.

The first is the **tension** T in the string. A more tightly stretched string carries vibrations faster, the way a tightly stretched bedsheet snaps back into place faster than a slack one. Tension provides the restoring force that pulls a displaced segment back toward equilibrium, and a stronger restoring force gives a faster wave.

The second is the **linear mass density** μ , the mass per unit length of the string. A heavier string carries vibrations more slowly. More inertia per metre, more sluggish response. The Greek letter μ is the standard notation, and $\mu = m/L$ where m is the mass of a length L of string.

These two physical arguments are doing real work. They identify, on physical grounds alone, the only two quantities that could plausibly enter the answer. Resist the temptation to add other quantities. The wave speed does not depend on the frequency of vibration, on the amplitude of the wave, on the shape of the wave, on how the wave was launched, on the colour of the string, or on the brand of the guitar. It depends only on the medium that the wave is travelling through, which here means only on tension and linear mass density. A wave's speed is a property of the medium. Hold on to this sentence; it is one of the most important sentences in the whole of wave physics.

Derivation by dimensional analysis

Suppose we believe, on the physical grounds just argued, that the wave speed v depends only on the tension T and the linear mass density μ . Then we may write, very generally,

$$v = k \times T^a \times \mu^b$$

where k is a dimensionless constant of proportionality, and a and b are unknown powers that we are about to determine. Demanding that the dimensions of the two sides match will fix a and b uniquely. The numerical constant k cannot be fixed by dimensions alone, and we will return to that limitation honestly at the end.

Write down the dimensions of each quantity

Speed has dimensions of length over time, so $[v] = L T^{-1}$. Tension is a force, so $[T] = M L T^{-2}$. Linear mass density is mass over length, so $[\mu] = M L^{-1}$.

Substitute these dimensions into the assumed equation

The dimensionless constant drops out and we are left with:

$$L T^{-1} = (M L T^{-2})^a \times (M L^{-1})^b = M^{a+b} \times L^{a-b} \times T^{-2a}$$

Equate the powers of M , L , and T on the two sides

The left-hand side has M^0 , L^1 , and T^{-1} , so:

$$M: a + b = 0 \quad L: a - b = 1 \quad T: -2a = -1$$

From the T equation, $a = 1/2$. From the M equation, $b = -a = -1/2$. The L equation is then $(1/2) - (-1/2) = 1$, which is satisfied automatically. The system is consistent. There is exactly one pair of powers that works.

Substitute back

The wave speed must take the form:

$$v = k \times T^{1/2} \times \mu^{-1/2} = k \times \sqrt{\frac{T}{\mu}}$$

This is as far as dimensional analysis can take us. The dimensionless constant k is not fixed by the method. To fix it we need either a careful experiment with a stretched wire and a stopwatch, or a more advanced derivation from Newton's second law applied to a small element of the string. Either route gives the same answer. Both yield:

$$k = 1$$

and so the speed of a transverse wave on a stretched string is:

$$v = \sqrt{\frac{T}{\mu}}$$

Pause for a sentence on the method itself. Dimensional analysis has given us the correct shape of the equation, with the right exponents on T and μ , by doing nothing more than counting masses and lengths and times. It has not given us the constant $k = 1$; that is an experimental result confirmed independently. Honest physicists state both halves: dimensional analysis fixes the shape, experiment fixes the scale. Students who

believe dimensional analysis is magic come unstuck on the day they meet a problem whose dimensionless constant is not 1.

v is not f: a warning that students must internalise

The result $v = \sqrt{\frac{T}{\mu}}$ tells you **how fast** a transverse disturbance travels along the string. It does not tell you the frequency of vibration, and it does not tell you the pitch. These are different quantities, and confusing them is the most common error students make in the next chapter.

The wave speed v is set by the tension and the linear mass density, both of which are properties of the string itself. The frequency f is set by the boundary conditions: how long the string is, and where it is clamped. A guitarist does not change the wave speed when she plays a different note; the wave speed of her G string at constant tension is the same whether she is playing a melody or sitting silently. What she changes is the length of string that is free to vibrate, and that change in length changes the frequency without changing the wave speed. We will derive the precise relationship in the next chapter; for now, hold the two quantities apart in your mind.

In plain words: v is the speed at which the wave travels along the string. f is the frequency at which the string vibrates. These are different. Tighter strings vibrate at higher pitch because a faster wave fits more cycles per second into the same length, but that is a Chapter 7 argument. Until then: **v depends on T and μ . f depends on the length and the boundary conditions.** Two equations for two questions.

Three worked examples now. The first is single-step substitution, the second a two-step warm-up that traps the careless candidate at the linear-mass-density line, and the third a three-part scaling problem that asks the equation to do real work.

BINDER Example 41

A steel guitar string has a linear mass density of 0.012 kg m^{-1} . It is tuned to a tension of 75 N . Calculate the speed of transverse vibrations along the string.

Solution

Apply the result of this section directly:

$$v = \sqrt{\frac{T}{\mu}} = \sqrt{\frac{75 \text{ N}}{0.012 \text{ kg m}^{-1}}} = 79.1 \text{ m s}^{-1}$$

The speed of transverse waves along the string is approximately 79 m/s .

Making sense of the answer: A guitar string and a wave travelling on it at about 79 m/s , or roughly 285 km/h , may sound impossible until you remember that the wave does not have to travel far. Even a 65 cm string is crossed by such a wave in about 8 milliseconds. The string is plucked, the disturbance reaches the bridge in 8 ms , reflects, runs back to the nut in 8 ms , reflects again, and is therefore making full round trips at a rate of $1/(2 \times 8 \times 10^{-3}) \approx 60$ cycles per second. That figure is suspiciously close to a low musical pitch. The next chapter will turn this rough calculation into the formula for the harmonics of every stringed instrument ever made.

Think like a physicist: Notice the unit-by-unit cancellation. The expression $\text{N} / (\text{kg m}^{-1})$ is, after substituting $\text{N} = \text{kg m s}^{-2}$, simply $\text{m}^2 \text{s}^{-2}$. Square-rooting gives m s^{-1} , which is a speed. The dimensions are right, and they were guaranteed to be right because we built the equation by dimensional analysis. This kind of unit check is your safety net for every numerical question. If the units come out wrong, the answer is wrong, even when the numbers look plausible.

BINDER Example 42

A wire of length 0.60 m has a mass of 4.8 g and is stretched under a tension of 360 N . Find (a) the linear mass density μ , and (b) the speed of transverse vibrations on the wire.

Solution

(a) The wire's mass is given in grams; convert to kilograms before dividing, otherwise the units of the answer will be wrong by a factor of a thousand. $4.8\text{g} = 4.8 \times 10^{-3}\text{ kg}$. Then

$$\mu = \frac{m}{L} = \frac{4.8 \times 10^{-3}\text{ kg}}{0.60\text{m}} = 8.0 \times 10^{-3}\text{ kgm}^{-1}$$

(b) Apply $v = \sqrt{\frac{T}{\mu}}$:

$$v = \sqrt{\frac{360\text{N}}{8.0 \times 10^{-3}\text{ kgm}^{-1}}} = \sqrt{45\,000\text{ m}^2\text{s}^{-2}} = 212\text{ms}^{-1}$$

The wave speed on this wire is 212m/s.

Making sense of the answer: Notice that part (a) was the trap. A candidate who substitutes 4.8g and 0.60m directly into m/L gets $\mu = 8.0\text{kg/m}$, a thousand times too large. The square root then gives a wave speed a thousand times too small (about 6.7m/s instead of 212m/s). The mistake is invisible until you check units: kilograms and metres must be paired, grams and metres must not. Convert grams to kilograms before any division, and the rest of the calculation runs through cleanly.

Think like a physicist: The wire here is heavier per metre than the guitar string of Example 41 ($8.0 \times 10^{-3}\text{kgm}^{-1}$ compared with $1.2 \times 10^{-2}\text{kgm}^{-1}$, about two-thirds the mass per unit length), and it is under almost five times the tension. Both effects raise the wave speed: less inertia per metre, more restoring force per unit displacement. The wave speed climbs from 79m/s on the guitar string to 212m/s on this wire, a factor of about 2.7. Take a moment to check that the factor of 2.7 is consistent with the changes in T and μ via the $v \propto \sqrt{T}$ and $v \propto 1/\sqrt{\mu}$ scalings; if it is, you have understood the equation rather than just used it.

HOT Example 43

A vibrating string of length 0.90m and mass 5.4g carries transverse waves at a speed of 120ms⁻¹. (a) Find the tension in the string. (b) The tension is now increased by 44%. Find the new wave speed. (c) The string is replaced by a new string of the same length and the same tension as in part (a), but with twice the linear mass density of the original. Find the new wave speed.

Solution

(a) Compute the linear mass density first, again converting grams to kilograms:

$$\mu = \frac{m}{L} = \frac{5.4 \times 10^{-3}\text{ kg}}{0.90\text{m}} = 6.0 \times 10^{-3}\text{ kgm}^{-1}$$

Then invert $v = \sqrt{\frac{T}{\mu}}$ to give $T = v^2 \times \mu$:

$$T = (120\text{ms}^{-1})^2 \times 6.0 \times 10^{-3}\text{ kgm}^{-1} = 14\,400 \times 6.0 \times 10^{-3}\text{ N} = 86.4\text{N}$$

(b) The new tension is $T' = 1.44 \times 86.4\text{N} = 124.4\text{N}$. The shortcut is to notice that $v \propto \sqrt{T}$, so multiplying T by 1.44 multiplies v by $\sqrt{1.44} = 1.2$:

$$v' = 1.2 \times 120\text{ms}^{-1} = 144\text{ms}^{-1}$$

Increasing the tension by 44% raises the wave speed only by 20%, not by 44%. The square-root relationship between v and T softens every change.

(c) The new string has the same tension $T = 86.4\text{N}$ but a new linear mass density $\mu' = 2 \times 6.0 \times 10^{-3}\text{ kgm}^{-1} = 1.2 \times 10^{-2}\text{ kgm}^{-1}$. Compute directly:

$$v'' = \sqrt{\frac{86.4\text{N}}{1.2 \times 10^{-2}\text{ kgm}^{-1}}} = \sqrt{7200\text{ m}^2\text{s}^{-2}} = 84.9\text{ms}^{-1}$$

Or, equivalently, use $v \propto 1/\sqrt{\mu}$: doubling μ divides v by $\sqrt{2} \approx 1.414$, giving $120 / \sqrt{2} \approx 84.9\text{ms}^{-1}$, in agreement with the direct calculation.

Making sense of the answer: The two scalings $v \propto \sqrt{T}$ and $v \propto 1/\sqrt{\mu}$ are the most useful corollaries of the formula. They tell you, without any new calculation, how the wave speed changes when you change either T or μ . Internalise them. To double v you must quadruple T , or you must reduce μ to one quarter of its previous value. Tightening a string by a small amount has only a small effect on the wave speed; halving the mass per unit length only raises v by a factor of $\sqrt{2}$. Real instrument makers know this in their hands.

Think like a physicist: Note the discipline of part (b). The slow route is to compute $T' = 1.44 \times 86.4 = 124.4\text{N}$, then μ stays the same, then $v' = \sqrt{124.4/0.006} = 144\text{m/s}$. The fast route is $v' = \sqrt{1.44} \times v = 1.2 \times 120 = 144\text{m/s}$. The two routes give the same answer because they are the same physics seen from two angles. The faster method is preferred under examination conditions because it saves time and reduces the risk of arithmetic errors.

We have arrived at one of the most quietly powerful equations in the chapter. The wave speed on a stretched string depends on exactly two ordinary quantities, the tension and the linear mass density. Dimensional analysis fixed the shape of the equation, experiment fixed the constant, and a handful of substitutions showed how the equation behaves in the workshop. In the next chapter we will combine this speed with the boundary conditions of a clamped string and read off, in a single equation, the entire physics of every stringed instrument ever made: the harmonic frequencies of the guitar, the violin, and the *zeze*.

Before we get there, however, we owe ourselves a short tour. The chapter has produced a handful of equations and eight worked physical systems, and the world has been quietly using all of this for a long time without our help. In the next section we collect six places where the physics of this chapter has earned its keep, from the wristwatch on the desk to the seismometer on the rift, and from the loudspeaker in the phone to the violin in the orchestra pit.

APPLICATIONS OF SIMPLE HARMONIC MOTION

We have spent this chapter chasing a single equation, $a = -\omega^2 x$, into eight different physical systems. We have seen it govern a mass on a spring, a simple pendulum, a U-tube of water, a floating cylinder, a ball in a bowl, a compound pendulum, a torsion wire, and a wave on a stretched string. We close the chapter with a short tour of six places in the world where the same equation has earned its keep. Each example is a working device. Each one depends on a physicist (or a long line of physicists) understanding the equation we have just spent about eighty pages developing. None of the devices know the equation; all of them obey it.

1. Clocks and timing

The oldest application is the easiest to forget. A simple pendulum has a period that depends only on its length and on the local gravitational field, $T = 2\pi\sqrt{\frac{l}{g}}$. A pendulum of length close to 1m (more precisely 0.994m at standard gravity) has a period of exactly 2s. This is the famous "seconds pendulum". It ticks twice per period, once at each extreme of the swing, so the escapement that advances the gears clicks every 1s. The grandfather clock in your aunt's living room is a direct application of $a = -\omega^2 x$: each tick is a half-cycle of a simple pendulum, and the steady tick that keeps the clock running on time is the isochronous property we proved in the simple-pendulum section. The amplitude of the swing decays slightly between winds; the period does not. The clock keeps time.

Modern timekeeping has moved on, but only to a different oscillator. The quartz crystal in a wristwatch is a tiny fork of crystalline silicon dioxide cut to vibrate, by piezoelectric coupling, at exactly 32 768 Hz (which is 2^{15} cycles per second, a number chosen because dividing by two fifteen times, that is 15 successive divisions by 2 gives one cycle per second). The vibration is simple harmonic to extraordinary precision. A digital circuit counts the cycles and ticks the second hand once every 2^{15} of them. A wristwatch costing a few thousand shillings keeps time to about a second a day, which is a million-to-one accuracy in a piece of consumer electronics. The physics is the same physics that governs Kipute's swing.

2. The seismometer

When the ground beneath a seismometer shakes during an earthquake, you might expect the instrument to shake along with it. The clever trick of a seismometer is to use a heavy mass suspended on a long, soft spring or a long pendulum whose natural frequency is far below the frequency of the seismic waves. The mass is, by inertia, almost stationary. The frame of the instrument moves with the ground. The relative displacement between the stationary mass and the moving frame is what the instrument records. A seismometer is therefore

an SHM oscillator deliberately tuned to be a bad responder to the driving frequency, so that the frame's motion (which is the earthquake) shows up cleanly as the displacement we want to measure.

The same physics, with the natural frequency tuned in the opposite direction, is at the heart of the accelerometer in your phone. There the mass is light, the spring is stiff, and the natural frequency is high. The mass moves with the phone, and the relative displacement is small but measurable. Whether the design wants the mass to move with the frame or to lag behind it depends entirely on whether the natural frequency sits above or below the driving frequency, which is the resonance physics of the previous section put to engineering use.

3. Vehicle suspension

Every vehicle suspension is a damped mass-spring oscillator. The mass is the car body. The spring is the steel coil between the wheel and the chassis. The damper is the hydraulic shock absorber that we met in the damping section. After the wheel runs over a pothole on the Morogoro road, the spring stores energy and the body rises; the damper then dissipates that energy into heat as the body returns to its resting height. The engineer's choice of damping coefficient, as we discussed in the damping section, is critical. Underdamped suspension produces the bouncing car that takes seconds to settle. Overdamped suspension produces the harsh ride that transmits every bump straight to the passenger. Critically damped suspension does the job in the shortest time without overshoot, and that is what the design aims for.

The natural frequency of a passenger-car suspension is around 1 Hz to 1.5 Hz, a deliberate choice. Frequencies in this range are slow enough that the human body can absorb them without discomfort. Higher natural frequencies would feel jarring; lower ones would feel like motion sickness. The suspension is, in other words, tuned to the natural frequencies of the human gut. Engineering at its most considerate.

4. The loudspeaker and the microphone

A loudspeaker is forced oscillation made audible. Inside the speaker is a paper or fabric cone glued to a coil of wire that sits in the field of a permanent magnet. When a current oscillating at audio frequency is sent through the coil, the coil experiences a force that varies with the current, and the cone is driven back and forth at the same frequency. The cone, by pushing the air in front of it, sets the air oscillating at that frequency, and the oscillation propagates outward as a sound wave. Every note you have ever heard from a phone, a radio, or a public-address system was produced by a forced simple-harmonic oscillator turning electrical energy into mechanical displacement and then into a sound wave.

A microphone is the same device run backwards. A diaphragm sits in the path of an arriving sound wave. The pressure variations of the wave force the diaphragm into oscillation, and the diaphragm's motion is converted (by induction in a coil, or by capacitance change, or by piezoelectric coupling) into an electrical signal at the same frequency. Loudspeaker and microphone are mechanical inverses of each other, and the same equation, with the same ω , runs in both directions. A physicist who designs one understands the other free of charge.

5. Medical imaging

Two imaging modalities in any modern hospital depend on simple-harmonic physics. The first is ultrasound. A piezoelectric crystal, driven by an oscillating voltage, vibrates as a forced oscillator at frequencies of a few MHz, well above the range of human hearing. The vibrations propagate as ultrasound waves into the body, reflect off boundaries between tissues of different acoustic properties, and return to the same crystal which now acts as a microphone. The timing of the returns is converted into the depth of the reflecting tissue, and the image is built up. The crystal is the same simple-harmonic oscillator we developed in this chapter, only operating at a frequency three orders of magnitude above what any musical instrument could produce.

The second is magnetic resonance imaging, or MRI. We met the principle in the resonance section: hydrogen nuclei in the body precess, in a strong magnetic field, at a particular frequency that depends linearly on the local field strength. By applying gradient magnetic fields and a radio-frequency pulse tuned to the precession frequency, the scanner causes the protons in a thin slice of the body to resonate. The signal that they re-emit a moment later, decaying in amplitude exactly the way the damped oscillations of this chapter do, is used to build the image. An MRI scan is, at its physical heart, a controlled resonance experiment performed on every cell of the patient's body simultaneously, with the decay constant of the signal carrying the contrast information.

6. Musical instruments

Every musical instrument the world has ever made is an application of simple harmonic motion subject to a boundary condition. The string of a guitar is held fixed at both ends, and the standing-wave constraint forces the string to vibrate only at frequencies for which a whole number of half-wavelengths fits between the two clamped ends. Each of those frequencies is a separate simple-harmonic mode. The lowest is the fundamental,

$f_1 = (1/2L)\sqrt{\frac{T}{\mu}}$ using the wave speed $v = \sqrt{\frac{T}{\mu}}$ we derived in the previous section, and the higher modes $f_n = n \times f_1$ are the harmonics that give the instrument its character. We will derive this formula explicitly in the next chapter; for now, accept that every plucked, bowed, or struck string in the world is an assembly of simple-harmonic oscillators superposed on the same string.

Wind instruments work on the same principle but with the air column inside the instrument as the oscillating medium and the open or closed ends of the tube as the boundary conditions. Drums and other percussion instruments use the simple-harmonic modes of a stretched membrane. Even the human voice is an air-column resonator: the vocal tract resonates at frequencies set by the geometry of the throat and mouth, and the voice we hear is a superposition of those resonances on the buzzing of the vocal folds.

Six applications, one equation.

That ends the formal teaching of the chapter. The miscellaneous worked examples that follow do not have the courtesy of treating one section at a time. They will combine ideas across sections, and almost always catching the candidate who has been reading the chapter rather than learning it. Read each one through once before reaching for a pen. Identify which sections it draws on. Then solve.

MISCELLANEOUS WORKED EXAMPLES ON SIMPLE HARMONIC MOTION

Example 44

- Explain, in physical terms, why a child on a swing must time pushes to match the swing's natural rhythm to make the amplitude grow, and explain what would happen if the pushes arrived at random intervals.
- A particle moves so that its acceleration is given by $a = -k_1 \sin(k_2 x)$, where k_1 and k_2 are positive constants. Show that the motion is approximately SHM for small x and find its angular frequency.

Solution

(a) Pushes timed with the swing's natural rhythm act in the same direction as the swing's instantaneous velocity at the moment of each push. Each push therefore does positive work on the swing, adding energy to it on every cycle, and the amplitude builds up. This is resonance: driving frequency equal to natural frequency.

With pushes at random intervals, some coincide with the swing moving towards the pusher (negative work, energy removed) and others with it moving away (positive work, energy added). On average the energy added per cycle is small, and the amplitude does not grow.

(b) For small x : $\sin(k_2 x) \approx k_2 x$. Hence

$$a \approx -k_1 k_2 x$$

which is of the SHM form $a = -\omega^2 x$ with $\omega^2 = k_1 k_2$, giving

$$\omega = \sqrt{k_1 k_2}$$

The approximation, and therefore the SHM, holds only for small x .

Example 45

- Account for the experimental observation that the period of a simple pendulum is independent of the mass of the bob.
- An SHM particle has angular frequency 4 rad/s and amplitude 6cm. Find the time taken to move from $x = 0$ to $x = 3$ cm, and from $x = 3$ cm to $x = 6$ cm. Comment briefly on answers.

Solution

(a) The restoring force on the bob along its arc is $F = -mg \sin \theta \approx -mg \theta$ for small θ . Newton's second law along the arc gives $ma = -mg \theta$, so $a = -g \theta$. The mass m cancels: it appears once in the inertia and once in the gravitational restoring force, since gravity itself acts on the same mass. The acceleration, and therefore the period, depends only on g and the length of the pendulum, not on the bob's mass.

(b) Take $x = A \sin(\omega t)$ with $A = 6\text{cm}$, $\omega = 4\text{ rad/s}$.

From $x = 0$ to $x = 3\text{cm} = A/2$: $\sin(\omega t_1) = 1/2$, so $\omega t_1 = \pi/6$, giving:

$$t_1 = \frac{\pi}{6\omega} = \frac{\pi}{24}\text{ s} = 0.131\text{ s}$$

From $x = 0\text{cm}$ to $x = 6\text{cm} = A$: $\sin(\omega t_2) = 1$, so $\omega t_2 = \pi/2$, giving:

$$t_2 = \frac{\pi}{2\omega} = \frac{\pi}{8}\text{ s}$$

From $x = 3\text{cm}$ to $x = 6\text{cm}$:

$$t = t_2 - t_1 = \frac{\pi}{8}\text{ s} - \frac{\pi}{24}\text{ s} = 0.262\text{ s}$$

The time taken in the first leg ($x = 0$ to $x = 3\text{cm}$) is 0.131 s , and for the second leg ($x = 3\text{cm}$ to $x = 6\text{cm}$) is 0.262 s .

The second leg takes twice as long as the first, because the particle slows as it approaches the extreme position.

Example 46

- Explain why a tuned vehicle suspension uses critical damping rather than light damping for passenger comfort, and what would be the consequence of using light damping.
- Two SHM oscillators have the same amplitude A and the same angular frequency ω , but they differ in phase by $\pi/3$ radians. At the instant the first is at its maximum displacement, find the displacement and the velocity of the second, expressed in terms of A and ω .

Solution

(a) Critical damping returns the suspension to its rest position in the shortest time without oscillating. After the wheel passes a bump, the body of the car rises and is then brought back to its rest height by the damper without bouncing or overshooting. This gives a smooth, quickly-settled ride.

With light damping, the suspension would be underdamped: the car body would oscillate up and down for several cycles after each bump before settling, producing a sea-sickness ride and reducing tyre contact with the road, which is dangerous at speed.

(b) Take the first oscillator as $x_1 = A \sin(\omega t)$, which is at its maximum ($x = A$) when $\omega t = \pi/2$. The second leads or lags the first by $\pi/3$:

$$x_2 = A \sin(\omega t - \pi/3)$$

At $\omega t = \pi/2$ (maximum displacement for the first):

$$x_2 = A \sin(\pi/2 - \pi/3) = A \sin(\pi/6) = A/2$$

Velocity of the second is $v_2 = dx/dt = A\omega \cos(\omega t - \pi/3)$. At $\omega t = \pi/2$:

$$v_2 = A\omega \cos(\pi/2 - \pi/3) = A\omega \cos(\pi/6) = A\omega \times (\sqrt{3}/2)$$

Hence at that instant the second has displacement $A/2$ and velocity $(\sqrt{3}/2)A\omega$, moving toward larger displacement.

Example 47

- Explain why a particle in SHM has its maximum speed at the equilibrium position and zero speed at the extreme positions.
- A 0.20kg mass on a spring executes SHM of amplitude 5cm and period 0.50s . Find the displacement at which the kinetic energy is three times the potential energy.

Solution

(a) The total mechanical energy $E = KE + PE$ is conserved. At the extreme positions, the particle is momentarily at rest before reversing direction, so $KE = 0$ and all the energy is potential: $PE = E$. As the particle moves towards equilibrium, the restoring force does positive work, converting PE into KE . At the equilibrium position the displacement is zero, so $PE = (1/2)kx^2 = 0$, and all the energy is kinetic: $KE = E$. The speed is therefore maximum there.

(b) Total energy: $E = (1/2)kA^2$. At displacement x : $PE = (1/2)kx^2$ and $KE = E - PE = (1/2)k(A^2 - x^2)$.

The condition $KE = 3PE$ gives:

$$\begin{aligned}(1/2)k(A^2 - x^2) &= 3 \times (1/2)kx^2 \\ A^2 - x^2 &= 3x^2 \\ x^2 &= \frac{A^2}{4} \\ x &= \pm \frac{A}{2} = \pm \frac{5}{2} \text{ cm} = \pm 2.5 \text{ cm}\end{aligned}$$

The kinetic energy is three times the potential energy at $x = \pm 2.5 \text{ cm}$ from equilibrium.

Example 48

- Explain how the period of a simple pendulum changes when it is taken from sea level to the top of Mount Kilimanjaro, identifying which physical quantity is responsible for the change.
- A spring of stiffness 400 N/m has an elastic limit of 24 N . A 1.0 kg mass is hung from it and set into vertical SHM. Find the maximum amplitude that can be used without exceeding the elastic limit, and the corresponding total energy of the oscillation. Take $g = 9.8 \text{ m/s}^2$.

Solution

- The period of a simple pendulum is $T = 2\pi\sqrt{\frac{l}{g}}$. The length l of the pendulum is unchanged by relocation.

The acceleration due to gravity g , however, decreases with altitude (the pendulum is further from the centre of the Earth). At the summit of Kilimanjaro (about 5.9 km above sea level), g is smaller than at sea level by roughly 0.2% . A smaller g makes T larger: the pendulum runs slower at the summit. The change is therefore caused by the variation of g with altitude.

- The static extension under the weight of the bob is:

$$x_0 = \frac{mg}{k} = \frac{1.0 \times 9.8}{400} \text{ m} = 0.0245 \text{ m} = 2.45 \text{ cm}$$

The corresponding tension in the spring at equilibrium is $mg = 9.8 \text{ N}$. The elastic limit is reached when the total tension equals 24 N , i.e., when the extension equals $\frac{24 \text{ N}}{400 \text{ N/m}} = 0.060 \text{ m} = 6.0 \text{ cm}$. The maximum further stretch from equilibrium is therefore:

$$A_{\text{max}} = 6.0 \text{ cm} - 2.45 \text{ cm} = 3.55 \text{ cm} = 0.0355 \text{ m}$$

The total energy of the oscillation at this amplitude is

$$E = (1/2)kA_{\text{max}}^2 = 0.5 \times 400 \text{ N/m} \times (0.0355 \text{ m})^2 = 0.252 \text{ J}$$

Maximum amplitude is approximately 3.6 cm , corresponding total energy is 0.25 J .

Example 49

- Explain why a building whose lowest natural frequency is close to the dominant frequency of seismic waves can collapse during an earthquake of moderate intensity, while a much shorter or much taller building nearby remains intact.
- A 5.0 kg block hangs from two identical springs, each of stiffness 200 N/m . Find the period of vertical SHM when the springs are connected (i) in parallel, (ii) in series.

Solution

- During the earthquake the ground oscillates and drives the base of every nearby building at the same frequency. A building whose own natural frequency matches that driving frequency is in resonance: each cycle of the ground transfers energy efficiently into the building's structural oscillation, and the amplitude grows from cycle to cycle. The cumulative stress in the columns and beams eventually exceeds the elastic limit and the building collapses, even though the ground motion itself was moderate.

A much shorter building has a much higher natural frequency, and a much taller one has a much lower natural frequency. In both cases the driving frequency is far from the natural frequency, the resonance condition is not met, and the response amplitude stays small. The buildings shake but do not fail. This is why earthquake-resistant design tunes a building's lowest natural frequency well away from the dominant earthquake band.

(b) Two springs of stiffness $k = 200\text{N/m}$ each.

(i) Parallel: effective stiffness $k_p = 2k = 400\text{N/m}$.

$$T_p = 2\pi \sqrt{\frac{m}{k_p}} = 2\pi \sqrt{\frac{5.0\text{kg}}{400\text{N/m}}} = 0.70\text{s}$$

(ii) Series: effective stiffness $k_s = k/2 = 100\text{N/m}$.

$$T_s = 2\pi \sqrt{\frac{5.0\text{kg}}{100\text{N/m}}} = 1.40\text{s}$$

Example 50

- (a) Explain why a marching column of soldiers is ordered to break step when crossing a long footbridge.
 (b) A grandfather-clock pendulum is made of brass with linear expansion coefficient $\alpha = 1.9 \times 10^{-5} \text{ K}^{-1}$. The clock keeps perfect time at 18°C . The temperature in Dar es Salaam rises to 32°C . Find, in seconds per day, the time the clock loses.

Solution

(a) Marching soldiers strike the bridge at a steady rhythm, applying a periodic vertical force at the frequency of their step (typically $\approx 1.5\text{Hz}$ to 2Hz). A long footbridge has natural frequencies of vertical oscillation that may lie within this same range. If the marching frequency happens to coincide with one of those natural frequencies, the bridge is driven at resonance: each step adds energy, the amplitude grows, and the bridge can sway dangerously or fail. Breaking step makes the foot impacts random in timing, so the cumulative driving force at any single frequency is small, and resonance is avoided.

(b) The period at temperature T depends on the length: $T = 2\pi \sqrt{\frac{l}{g}}$. A small fractional change of length gives

$$\frac{\Delta T}{T} = \frac{1}{2} \times \frac{\Delta l}{l} = \frac{1}{2} \alpha \Delta \theta$$

Substitute $\alpha = 1.9 \times 10^{-5} \text{ K}^{-1}$ and $\Delta \theta = 32 - 18 = 14\text{K}$:

$$\frac{\Delta T}{T} = \frac{1}{2} \times (1.9 \times 10^{-5}) \times 14 = 1.33 \times 10^{-4}$$

This is the fractional increase in the pendulum's period. Each second of true time the clock now records as fewer seconds. Over one day (86400s):

$$\Delta t = 86400\text{s} \times 1.33 \times 10^{-4} = 11.5\text{s}$$

The clock loses about 11.5s per day in the heat.

Example 51

- (a) Explain why a moving-coil galvanometer is designed to be critically damped, while a hydraulic door closer is designed to be overdamped.
 (b) A simple pendulum of length 80cm hangs from the ceiling of a lift. Find its period when the lift (i) ascends with acceleration 2.0m/s^2 , (ii) descends with acceleration 2.0m/s^2 , (iii) is in free fall. Take $g = 9.8\text{m/s}^2$.

Solution

(a) A galvanometer must give a steady reading as soon as the current is switched on. Critical damping is the unique condition that returns the pointer to its correct deflection in the shortest possible time without overshoot or oscillation, so that the experimenter can read the value at once.

A hydraulic door closer must close the door slowly enough that anyone passing through is not struck. Overdamping returns the door to its closed position smoothly but slowly, with no oscillation. Critical damping would shut the door too fast for safe use; light damping would let it swing back and forth and never settle.

(b) Inside the lift, the bob feels an effective gravitational acceleration g_{eff} set by the lift's acceleration.

(i) Lift ascending at $a = 2.0\text{m/s}^2$: $g_{\text{eff}} = g + a = 9.8\text{m/s}^2 + 2.0\text{m/s}^2 = 11.8\text{m/s}^2$.

$$T_1 = 2\pi \sqrt{\frac{0.80\text{m}}{11.8\text{m/s}^2}} = 1.64\text{s}$$

(ii) Lift descending at $a = 2.0\text{m/s}^2$: $g_{\text{eff}} = g - a = 9.8\text{m/s}^2 - 2.0\text{m/s}^2 = 7.8\text{m/s}^2$.

$$T_2 = 2\pi \sqrt{\frac{0.80\text{m}}{7.8\text{m/s}^2}} = 2.01\text{s}$$

(iii) Free fall: $a = g$ downward, so $g_{\text{eff}} = 0$. The bob is weightless inside the freely falling lift; the restoring force vanishes; the period is undefined (formally infinite). The pendulum simply does not oscillate.

Example 52

- (a) Explain why a microwave oven cooks food efficiently while leaving most of the kitchen cold.
 (b) A wooden cylinder of length 25cm and uniform density floats with 18cm submerged in fresh water at 20°C, where $\rho = 1000\text{kg/m}^3$. The water is heated to 60°C, where $\rho = 983\text{kg/m}^3$. Find the percentage change in the period of vertical oscillation of the cylinder.

Solution

(a) The microwave oven generates electromagnetic waves at a frequency close to one of the natural rotational frequencies of the water molecule. Water molecules in the food absorb these waves resonantly, rotate vigorously, and the motion is converted into thermal energy through molecular collisions. The food, which contains water, heats up rapidly. Most of the rest of the kitchen (ceramic plates, glass turntables, walls) contains very little water, has no rotational mode at the oven's frequency, and so does not absorb the waves resonantly. It stays cool.

(b) For a floating cylinder of cross-section A , the equilibrium submerged depth is $h = m/(\rho A)$, and the period of small vertical oscillation is:

$$T = 2\pi \sqrt{\frac{h}{g}} = 2\pi \sqrt{\frac{m}{\rho A g}}$$

Cross-section A and mass m are unchanged by heating; g is unchanged. Only ρ changes. Hence $T \propto 1/\sqrt{\rho}$, and:

$$\frac{T'}{T} = \sqrt{\frac{\rho}{\rho'}} = \sqrt{\frac{1000\text{kg/m}^3}{983\text{kg/m}^3}} = 1.00862$$

The period increases by about 0.86% when the water is heated from 20°C to 60°C.

Example 53

- (a) Explain how a seismometer can record an earthquake without being shaken with the ground that is moving beneath it.
 (b) A uniform metre rule pivoted at distance d from its centre is timed and found to make 25 oscillations in 40s. Take $g = 9.8\text{m/s}^2$ and find both possible values of d .

Solution

(a) A seismometer suspends a heavy mass on a long, soft spring (or a long pendulum) whose natural frequency lies far below the frequencies of seismic waves (typically less than 1Hz for a seismometer). When the ground moves, the frame of the instrument moves with it, but the suspended mass cannot follow at the seismic frequency: its natural response time is much longer than one cycle of the ground motion. By inertia the mass remains nearly stationary in space while the frame oscillates around it. The relative displacement between the (stationary) mass and the (moving) frame is recorded mechanically or electronically; that record is the earthquake.

(b) Period of one oscillation: $T = 40\text{s}/25 = 1.60\text{s}$. For a uniform rod of length $L = 1.0\text{m}$, the moment of inertia about the centre is $I_{\text{cm}} = ML^2/12$, giving radius of gyration about the centre $k = L/\sqrt{12}$ m and $k^2 = 1/12$ m².

The period of compound pendulum period is:

$$T = 2\pi \sqrt{\frac{k^2 + d^2}{gd}}$$

Squaring and rearranging gives a quadratic in d :

$$d^2 - \frac{T^2 g}{4\pi^2} d + k^2 = 0$$

Coefficient: $T^2 g / (4\pi^2) = (1.60s^2 \times 9.8m/s^2) / (4\pi^2) = 0.6355m$.

$$d^2 - 0.6355d + 0.0833 = 0$$

$$d_1 = 0.451m, \quad d_2 = 0.185m$$

Both pivot points lie within the metre rule (0 to 0.5m from the centre on either side), and either choice gives the same period $T = 1.60s$.

Example 54

- (a) Explain why a quartz crystal in a wristwatch keeps time more accurately than a mechanical pendulum clock.
- (b) A solid sphere of mass 2.0kg and radius 5.0cm is suspended by a wire and executes torsional oscillation with period 1.5s. The sphere is replaced by a thin uniform ring of the same mass and the same radius, suspended by the same wire about a diameter. Find the new period.

Solution

(a) A quartz crystal is a piezoelectric solid that vibrates at a single, sharply-defined natural frequency set by the crystal's mass, geometry, and elastic stiffness. Because the restoring force in a quartz crystal is interatomic and almost free of friction, the damping is extremely light, the resonance peak is correspondingly sharp, and the natural frequency is essentially insensitive to ordinary changes in temperature, mounting, or driving voltage.

A mechanical pendulum, by contrast, depends on its length and on local g ; the length expands and contracts with temperature, the pivot wears, and the air drag varies with humidity and altitude. Consequently, a mechanical pendulum clock is less accurate than a quartz crystal.

(b) Period of a torsional pendulum: $T = 2\pi \sqrt{\frac{I}{c}}$, where I is the moment of inertia of the suspended body about the wire and c is the torsional constant of the wire. The wire is the same in both cases, so $T \propto \sqrt{I}$.

For the solid sphere about a diameter through its centre: $I_{\text{sph}} = (2/5)MR^2$. For a thin uniform ring about a diameter: $I_{\text{ring}} = (1/2)MR^2$. Hence:

$$\frac{T_{\text{ring}}}{T_{\text{sph}}} = \sqrt{\frac{I_{\text{ring}}}{I_{\text{sph}}}} = \sqrt{\frac{(1/2)}{(2/5)}} = \frac{\sqrt{5}}{2}$$

$$T_{\text{ring}} = (\sqrt{5}/2) \times 1.5s = 1.68s$$

The new period is approximately 1.68s.

Example 55

- (a) Explain why a vehicle's exhaust pipe sometimes rattles loudly at one particular engine speed but is silent above and below it.
- (b) A steel wire of length 80cm and mass 4.0g is fixed at both ends. A small disturbance launched at one end takes 4.0ms to reach the other end. Find the tension in the wire.

Solution

(a) Each loose part of the exhaust system (a panel, a clamp, a heat shield) has its own natural frequency of mechanical vibration, set by its mass and the stiffness of its mounting. The engine produces vibration at a frequency proportional to its rotational speed. At one specific engine speed, the engine's driving frequency

happens to coincide with the natural frequency of the loose part, and the two enter resonance: the part oscillates at large amplitude and rattles audibly. Above or below that speed the driving frequency is detuned from the natural frequency, the response amplitude collapses to a small value, and the rattle disappears.

(b) Wave speed:

$$v = \frac{L}{t} = \frac{0.80\text{m}}{4.0 \times 10^{-3} \text{ s}} = 200\text{m/s}$$

Linear mass density:

$$\mu = \frac{m}{L} = \frac{4.0 \times 10^{-3} \text{ kg}}{0.80\text{m}} = 5.0 \times 10^{-3} \text{ kg/m}$$

From $v = \sqrt{\frac{T}{\mu}}$, solving for tension:

$$T = v^2 \mu = (200\text{m/s})^2 \times 5.0 \times 10^{-3} \text{ kg/m} = 200\text{N}$$

The tension in the wire is 200N.

Example 56

- (a) Explain how a vehicle's shock absorbers protect passengers from road bumps, and identify the type of damping they exhibit.
- (b) A thin uniform rod of length 1.2m and mass 0.5kg is pivoted at its end and allowed to swing as a compound pendulum of small amplitude. Take $g = 9.8\text{m/s}^2$ and find (i) its period, (ii) the length of the equivalent simple pendulum.

Solution

(a) A shock absorber is a hydraulic damper connected in parallel with the suspension spring. When the wheel passes over a bump, the spring first stores the impact energy as elastic potential energy; the shock absorber then converts that stored energy into thermal energy of the hydraulic fluid as the spring relaxes. The result is that the body of the vehicle returns to its rest height in the shortest possible time without oscillating. Shock absorbers are designed to operate close to **critical damping**; this is the damping condition that achieves the fastest non-oscillatory return to equilibrium and gives the smoothest ride.

(b) For a uniform rod of length L pivoted at its end, the moment of inertia about the pivot (by the parallel-axis theorem) is

$$I_O = (1/12)ML^2 + M(L/2)^2 = (1/3)ML^2$$

The centre of mass is at distance $d = L/2$ from the pivot. The compound pendulum period is:

$$T = 2\pi \sqrt{\frac{I_O}{Mgd}} = 2\pi \sqrt{\frac{(1/3)ML^2}{Mg(L/2)}} = 2\pi \sqrt{\frac{2L}{3g}}$$

(i) Substitute $L = 1.2\text{m}$, $g = 9.8\text{m/s}^2$:

$$T = 2\pi \sqrt{\frac{2 \times 1.2\text{m}}{3 \times 9.8\text{m/s}^2}} = 1.80\text{s}$$

(ii) Equivalent simple pendulum length L_{eq} has the same period: $T = 2\pi \sqrt{\frac{L_{\text{eq}}}{g}}$. Comparing with $T = 2\pi \sqrt{\frac{2L}{3g}}$:

$$L_{\text{eq}} = \frac{2}{3}L = \frac{2}{3} \times 1.2\text{m} = 0.80\text{m}$$

Period is approximately 1.80s, equivalent simple pendulum length is 0.80m.

Example 57

- (a) Explain why a particle in uniform circular motion can be regarded as the superposition of two SHM oscillations along perpendicular diameters.
- (b) A particle moves uniformly around a circle of radius 5cm at 4 rev/s. Find the amplitude, the period, and the maximum velocity of the projection of this motion on a diameter of the circle.

Solution

(a) Let the particle move on a circle of radius A with angular speed ω , and let the centre be the origin. Its position vector is

$$(x, y) = (A \cos(\omega t), A \sin(\omega t))$$

The x-coordinate is $A \cos(\omega t)$, which satisfies $d^2x/dt^2 = -\omega^2 x$: SHM along the x-axis with amplitude A , angular frequency ω . The y-coordinate is $A \sin(\omega t)$, which is SHM along the y-axis with the same A and ω , but a quarter-cycle out of phase with x . The actual motion is the vector sum of these two perpendicular SHMs; the resultant traces out the circle. Hence circular motion is exactly the superposition of two perpendicular SHMs of equal amplitude and equal frequency, ninety degrees out of phase.

(b) The projection of the motion on a diameter is one of the two perpendicular SHMs above.

Amplitude: $A = 5\text{cm}$ (the radius of the circle). Period: $T = 1/f = 1/4\text{ s} = 0.25\text{s}$.

Maximum speed of the projection equals the speed of the parent particle around the circle:

$$v_{\max} = \omega A = 2\pi f \times A = 2\pi \times 4/\text{s} \times 0.05\text{m} = 1.26\text{m/s}$$

Amplitude 5cm, period 0.25s, maximum velocity 1.26m/s.

Example 58

- (a) Explain why the period of free oscillation of a mass on a spring is the same regardless of the size of the initial displacement, while the period of a simple pendulum is independent of amplitude only for small angles.
- (b) A glass U-tube of uniform cross-section contains water of total column length 80cm. The water is gently displaced and released. Find the period of the resulting oscillation. Take $g = 9.8\text{m/s}^2$.

Solution

(a) For a mass on a spring, the restoring force is exactly $F = -kx$ for all displacements within the spring's elastic range. The acceleration is therefore exactly $a = -(k/m)x = -\omega^2 x$ with $\omega^2 = k/m$ depending only on the spring stiffness and the mass. Since ω is independent of x , so is the period: amplitude does not enter. The motion is exactly SHM at any amplitude up to the elastic limit.

For a simple pendulum, the restoring force along the arc is $F = -mg \sin\theta$, which is not linear in θ . Only when θ is small can we approximate $\sin\theta \approx \theta$, giving $a \approx -g\theta$ and hence SHM with period $T = 2\pi \sqrt{\frac{L}{g}}$. For larger angles $\sin\theta < \theta$, the restoring acceleration is weaker than the SHM model predicts, and the period grows with amplitude. Isochronism is therefore exact for the spring but only approximate for the pendulum.

(b) For a U-tube of uniform cross-section, displacing the liquid by y on one side raises the other by y , giving a height difference of $2y$. The restoring force on the column is $F = -2\rho g A y$; the moving mass is $m = \rho A L$ where L is the total liquid length. Hence

$$a = F/m = -(2g/L)y$$

$$\omega^2 = 2g/L$$

$$T = 2\pi \sqrt{\frac{L}{2g}} = 2\pi \sqrt{\frac{0.80\text{m}}{2 \times 9.8\text{m/s}^2}} = 2\pi \times 0.2020\text{s} = 1.27\text{s}$$

The period of the oscillation is approximately 1.27s.

That ends the chapter's worked-out teaching. What follows is Digging Deeper Exercise 6, the chapter's final hand-off. Pen ready, paper ready, mind clear. The chapter ends when you complete it.

DIGGING DEEPER EXERCISE 6

EXERCISE 6A: BINDER QUESTIONS

Question 1

Kipanga and Kipute argue about the relationship between simple harmonic motion and uniform circular motion. Kipanga claims that a particle moving in a circle of radius A with constant angular speed ω is itself executing simple harmonic motion, since its position varies sinusoidally. Kipute insists that the circular motion is not SHM at all, and that only the projection on a diameter is. Resolve the disagreement, identifying clearly what condition the acceleration must satisfy for motion to be SHM, and showing whether the circular motion meets it.

Question 2

The displacement of a body executing SHM is written $y = A \sin(\omega t + \varepsilon)$. A student claims that the phase constant ε has no physical meaning, since one is free to choose the moment $t = 0$. Identify the error in this claim, and account for what ε actually fixes about the motion once two oscillators with the same A and ω are observed together.

Question 3

Kipanga writes the speed of a particle in SHM as $v = \omega \sqrt{A^2 - y^2}$, and Kipute writes it as $v = A\omega \cos(\omega t + \varepsilon)$. Each insists that the two formulas describe different physical situations. Account for the fact that the two formulas are equivalent, and explain in physical terms why the speed is maximum at $y = 0$ and zero at $y = \pm A$, even though it is at the extremes that the magnitude of the acceleration is greatest.

Question 4

Watching a body in SHM at the moment when its displacement is at the positive extreme, a student reasons: *"the body has stopped at this instant; therefore it has no acceleration at this instant either."* Identify the error in this reasoning, and account for the fact that the body, although momentarily stationary, has acceleration of the largest magnitude it ever has during the motion.

Question 5

Over one complete cycle of SHM, the time-averaged kinetic energy and the time-averaged potential energy of the oscillator are observed to be equal, each being half of the total energy $\frac{1}{2}kA^2$. A student concludes that, since the two averages are equal, the kinetic and potential energies must be equal at every instant. Identify the error in this conclusion, and account in physical terms for the equality of the two averages.

Question 6

A vertical mass-spring system has the same period of SHM, $T = 2\pi \sqrt{\frac{m}{k}}$, whether it is set up at sea level or on top of Mt. Kilimanjaro, where the gravitational field is appreciably weaker. A student concludes from this that gravity plays no part in the motion at all. Identify the error in this conclusion, and account in physical terms for the disappearance of g from the period formula.

Question 7

Kipanga is told that the simple pendulum is "isochronous" and concludes that any pendulum, swung through any amplitude, must have the period $T = 2\pi \sqrt{\frac{l}{g}}$. Kipute insists that this is only an approximation and that the period must in fact grow as the amplitude grows. By referring to the form of the restoring acceleration $a = -g \sin \theta$, and to the approximation by which it is replaced in deriving the formula, account for the fact that Kipute is right and identify the angular range within which Kipanga's claim is acceptable.

Question 8

A uniform metre stick is pivoted at a series of points, the distance h of the pivot from the centre of mass being varied through the experiment. The period of small oscillations is found to fall as h is increased from zero, to reach a minimum, and then to rise again as h approaches the end of the stick. By referring to the period of a compound pendulum,

$$T = 2\pi \sqrt{\frac{k^2 + h^2}{gh}}$$

(where k is the radius of gyration about the centre of mass), account for the appearance of a minimum period.

Question 9

Two everyday systems are observed. The shock absorber on a daladala settles back to rest very quickly after a bump in the **Nyerere** road, with no visible oscillation. A swing door, after being pushed open and released, swings shut, swings half-open, swings shut, and only after several diminishing swings comes to rest. By referring to the three regimes of damped oscillations and to the relative size of the damping coefficient compared with the natural angular frequency ω_0 , account for the difference and identify which regime each system is designed to operate in.

EXERCISE 6B: REAL QUESTIONS

Question 10

The shock absorber on a daladala plying the Mbezi-Kariakoo route, when new, returns the vehicle to its rest height after a single bump in the road with no visible oscillation. Six months later, after long use, the suspension is observed to oscillate up and down two or three times after every bump before settling. Identify the regime of damped motion in which the suspension operates in each case (when new and when worn), and account in physical terms for the change.

Question 11

The brass bell at a primary school in Mtwara, when struck once at break time, rings with a clear note that fades steadily over about ten seconds before falling silent. Account in physical terms for the steady fading, and explain why a heavier bell of the same shape and material would ring for longer.

Question 12

A washing machine in a hostel in Dodoma is observed to shake the floor violently when its drum is being driven through one particular spin speed during start-up, but to be smooth at every other speed, including the final one. By referring to the natural frequency of the floor-machine system and to the forcing frequency applied by the rotating drum, account in physical terms for the speed-specific shaking.

Question 13

A driver on the Dar–Morogoro highway notices that her vehicle shakes most uncomfortably at exactly 80km/h, but smoothly at 70km/h and at 90km/h. The road has expansion joints set at a regular interval along its length. By identifying the forcing frequency that the joints impose on the suspension, and by comparing it with the suspension's natural frequency, account in physical terms for the speed-specific shaking.

Question 14

A platoon of police cadets practising drills near the Selander Bridge in Dar es Salaam is told by their instructor to "break step" before crossing. By referring to the forced oscillation of the bridge under a periodic footfall, and to what happens when the footfall frequency matches the bridge's natural frequency, account in physical terms for the safety concern that the instruction addresses.

Question 15

A pendulum clock made by a Dar es Salaam watchmaker keeps perfect time at sea level on the coast. The same clock, carried by a friend up to a hotel on the slope of Mt. Kilimanjaro at about 3000m altitude, is observed to run slow by about one minute per day. By referring to the period formula of the simple pendulum, account in physical terms for the loss of time, and indicate the physical adjustment the watchmaker must make to restore correct timekeeping.

Question 16

The seconds pendulum in Mr. Mushu's physics laboratory has been knocked accidentally and is now keeping time slightly fast. To return it to correct time, the technician adjusts the position of the bob along the rod. State in which direction (up or down along the rod) the bob must be moved, and account in physical terms for the resulting change in period.

Question 17

A spring scale in use by a tomato seller at Makoroboi market in Mwanza shows a steady reading when 2kg of tomatoes is placed on its pan. When the seller jostles the pan, the pointer is observed to oscillate up and down briefly about the steady reading before coming to rest. By referring to the spring-mass system, account in physical terms for the brief oscillation, and identify the period of the pointer's motion in terms of the spring constant of the scale and the mass on the pan.

Question 18

Two mattresses of the same brand in a Kahama guesthouse are observed to differ when a guest sits down: a softer mattress lets the guest oscillate up and down at a slow tempo, with the rise and fall plainly visible, while a firmer mattress of the same outer dimensions barely lets the guest perceive any motion before it settles. By referring to the period of a spring-mass system, account in physical terms for the difference in tempo, and indicate which of the two mattresses has the higher spring constant.

Question 19

A nylon clothesline stretched tight between two posts in a courtyard, when struck with a stick, sends a barely-visible pulse from one post to the other; the same line, when allowed to sag loose, sends a slow visible wave when struck. Account in physical terms for the difference in propagation speed of the two pulses.

Question 20

A heavy mango hanging from a slender branch in Bagamoyo is observed, after the branch is pulled aside and released, to swing back and forth approximately periodically before coming to rest. By referring to the form of the restoring force on the branch-mango system and to the condition for SHM, account in physical terms for why the motion approximates SHM, and identify a specific circumstance under which the approximation would fail.

EXERCISE 6C: HOT QUESTIONS**Question 21**

A 2.5kg block hangs in equilibrium from a vertical spring whose natural length is 0.50m and which stretches to a total length of 0.65m under the block. The block is then displaced 0.05m below equilibrium and released from rest. Find the position of the block, relative to the equilibrium position, 1.0s after release, and the maximum kinetic energy of the motion that follows.

Question 22

A simple pendulum has a period of 2.000s at sea level, where $g = 9.80\text{m/s}^2$. The same pendulum, when taken to a high-altitude site, is observed to lose 30 seconds per day. The length of the pendulum may be taken to be unchanged. Find the value of g at the new site.

Question 23

A particle executes SHM with amplitude 0.04m and frequency 2.0Hz. Find the displacement at which its kinetic energy equals its potential energy, and the time taken for it to move from $x = 0$ to $x = A/2$, taking the particle to be moving in the positive direction at $x = 0$.

Question 24

A spring of natural length 0.20m stretches to 0.24m when a 0.5kg mass is hung from it in equilibrium. The same spring is then connected end-to-end (in series) with a second, identical spring, and the same 0.5kg mass is hung from the lower end of the combination. Find the period of small vertical oscillations of the mass on the series combination.

Question 25

A pendulum clock has a steel rod that beats seconds ($T = 2.000\text{s}$) at 20°C . The clock is moved to a workshop where the temperature is 35°C . The linear thermal expansion coefficient of steel is $\alpha = 1.2 \times 10^{-5} \text{K}^{-1}$. Find by how many seconds the clock runs fast or slow per day, and state which.

Question 26

A steel wire of length 2.0m, cross-sectional area 0.50mm^2 , and Young's modulus $2.0 \times 10^{11} \text{ N/m}^2$ hangs vertically from a fixed support. A 4.0kg mass attached to the lower end is given a small downward displacement and released. Find the period of the resulting vertical oscillations.

Question 27

A horizontal nylon string of length 1.50m and total mass 0.030kg is stretched between two posts in a courtyard. A short transverse pulse, generated by a flick at one post, travels to the other post in 0.10s. Find the tension in the string.

Question 28

The displacement of a particle of mass 0.20kg executing SHM is given by $x = 0.05\sin(4\pi t + \pi/3)$, with x in metres and t in seconds. Find the first positive time at which the displacement equals $+0.025\text{m}$, the velocity of the particle at that instant, and the total mechanical energy of the motion.

Question 29

Two simple pendulums, of lengths 1.00m and 1.21m, are released simultaneously from small angles on the same side of the vertical. Find the time at which they next return simultaneously to their initial positions, both moving in the same direction as at the moment of release.

Question 30

A 0.40kg mass on a horizontal spring of stiffness 25N/m executes SHM on a frictionless surface with amplitude 0.10m. Find the kinetic energy of the mass when its displacement from equilibrium is $x = 0.06\text{m}$, and the displacement at which its kinetic energy is twice its potential energy.

Question 31

A boat at a fixed mooring on Lake Victoria moves vertically through a total range of 1.20m (peak to peak), with a period of 4.0s. Treating the motion as SHM, find the maximum vertical acceleration of the boat and the time taken for it to rise from its lowest point to a position 0.30m above the mean position.

Question 32

A uniform wire of length 0.60m and total mass 1.2g is stretched horizontally between two clamps with a tension of 80N. Find the time taken by a transverse pulse to make a round trip from one clamp to the other and back. If the wire is now replaced by a uniform wire of the same length and the same tension but with twice the mass per unit length, find the new round-trip time.

Chapter 7

WAVE MOTION

INTRODUCTION

The morning light on Lake Victoria in Mwanza had the apologetic quality of a guest who arrives too early to a wedding and is not sure where to sit. A bus had deposited the Form Five biology class on a study tour along the lakeshore. The biology teacher had promised fish. There were no fish. There was a great deal of lake, a great deal of mist, and Kipanga, who had been told there would be fish.

After forty minutes of looking at the same patch of water, Kipanga did what bored teenagers have done at lake shores since lakes were invented. He picked up a stone and he dropped it.

The ripples spread outward in widening rings. He watched them go, mildly impressed that a single dropped stone could cause a small geometric event of this much beauty for free. Mr. Akilikubwa, who had not been on the trip officially but had attached himself to it the way physics teachers attach themselves to anything involving water, materialised at Kipanga's elbow. He had with him a small dry leaf which he placed, with the unhurried solemnity of a magician introducing his most famous trick, on the surface of the lake about a metre away from where the ripples were heading.

They both watched. The leaf rose. The leaf fell. The leaf rose. The leaf fell. The ripples passed underneath it and continued on toward the open lake. The leaf, having bobbed, stayed exactly where it had been placed.

Kipanga: *"The leaf didn't move."*

Mr. Akilikubwa: *"Correct."*

Kipanga: *"But the wave moved."*

Mr. Akilikubwa: *"Also correct. Then what exactly moved?"*

Before Kipanga could answer, Mr. Akilikubwa picked up a second stone and threw it further out. A second pattern of ripples bloomed and spread. The two patterns met somewhere in the middle of the spreading region. Where two crests crossed, a taller crest briefly stood up. Where a crest met a trough, the water momentarily went flat, for a single instant, as if the lake had paused to think. Then the two patterns passed through one another, untroubled, untouched, and continued outward. Each ring kept its shape. Each ring kept its speed. Neither one knew, after the meeting, that the other had ever existed.

Mr. Akilikubwa: *"Energy moved outward, Kipanga. Disturbance moved outward. What did NOT move is the water itself. The water molecules merely oscillated up and down, passing the disturbance along to their neighbours like a message whispered down a line of schoolchildren. When the message has gone, the children remain, in their seats, in their original places. That is a wave. And the two ripple patterns added together where they met (displacements summed, not waters) and then walked away from each other unchanged. This is **superposition**, and it is the most important principle in the whole of wave physics. We will spend a chapter on it."*

Kipanga: *"All this from one stone?"*

Mr. Akilikubwa: *"All this from one stone. And from a leaf that refused to travel with it. The leaf has just told you, free of charge, that a wave carries energy without carrying matter. The two ripple patterns have just told you that waves can pass through each other and survive. You have witnessed two of the four most important facts about waves before breakfast. The other two are coming. The third is that every point on a wave is doing simple harmonic motion, of the kind discussed in the previous chapter. The fourth is that a wave is a single equation, and that equation is the only equation you need to memorise."*

He picked up a stick and wrote in the wet sand, in handwriting that would have failed an art examination but would have passed a physics one:

$$y(x, t) = A \sin(\omega t - kx)$$

"The whole chapter," he said, "is about earning that equation. You will earn the right to write it down. You will derive it. You will read parameters off it the way you read a clock. You will use it to predict the speed of seismic waves through Tanzanian rock and the tension in a guitar string and the pitch of the zeze your cousin plays at weddings. You will also, twice in this chapter, watch it become a different equation, a stationary wave, which is what happens when a wave meets its own reflection and has to share the room with it. By the end of these eighty-odd pages you will know why a leaf does not travel with a ripple, why two

speakers can make some seats in a hall louder and some quieter, why a guitar string an octave higher is half as long, and why the railway worker on the TAZARA line presses his ear to the rail. All of it from that."

He pointed at the equation in the sand. A small wave came up the shore, lapped politely at the writing, considered it, and erased it without comment.

Mr. Akilikubwa: "The lake is also a teacher. It does not believe in keeping notes!"

Welcome to Chapter 7. The previous six chapters spent themselves on matter, with chapter 6, the most recent, taught you the language of repetition: simple harmonic motion, the rhythm at the bottom of every smooth potential well in the universe.

This chapter is the first to be about the spatial extension of that rhythm. SHM, until now, has been a story about a single point: one mass, one spring, one bob. A wave is what happens when SHM gets shared. One particle is doing simple harmonic motion. Its neighbour is doing the same simple harmonic motion, but a fraction of a second later. Its neighbour's neighbour is doing the same simple harmonic motion, but two fractions of a second later. And so on, all the way along the medium. The result, viewed from above and frozen at one instant, is a sinusoidal pattern in space. Viewed at one fixed point and watched in time, it is a sinusoidal oscillation. Both descriptions are the same wave. Both descriptions live inside the equation Mr. Akilikubwa wrote in the sand.

We also have a promise to keep. At the end of Chapter 6, in the last full section before the chapter ended, we derived the speed of a transverse vibration on a stretched string by dimensional analysis:

$$v = \sqrt{\frac{T}{\mu}}$$

We told you, with a tip of the hat, that the next chapter would do something with that result. Here is the next chapter, and here is the something. We will take that single line for the wave speed, combine it with the requirement that the two ends of a guitar string are tied down so they cannot move, and out will fall the entire physics of stringed instruments: every harmonic, every overtone, every reason a *kora* and a *zeze* can be tuned by ear without an electronic tuner ever leaving its case. The room behind that doorway is this chapter.

And we will need, more often than is comfortable, the language of phase from circular motion in the first book of this book series, **Foundations of Physics**. Phase is the angular bookkeeping of where a particle is in its cycle, and a wave is built out of particles whose phases differ from each other by amounts proportional to how far apart they are. If the reference circle that produced the SHM equation feels rusty, take ten minutes before the next section and refresh it. The calculations here will not be hard, but they will be relentless about asking you to subtract one phase from another. A student who is fluent in phase finds this chapter delightful. A student who is shaky in phase finds it punishing. The choice is yours and it costs ten minutes.

Two other things to flag before we start. First, this chapter makes a sharp and important distinction between two different velocities. One is the velocity at which the wave pattern travels through the medium. The other is the velocity at which an individual particle moves up and down at right angles to the medium. Confusing the two is the single most common error in this entire chapter. Second, this chapter contains two equations that look almost identical but mean different things, namely:

$$y = A \sin(\omega t - kx) \quad \text{and} \quad y = (2A \cos kx) \sin \omega t$$

The first describes a wave that is travelling. The second describes a wave that is standing still while vibrating in place. Telling them apart is half the chapter's examination work, and we will spend a careful section on it.

That is the chapter. By the time you finish this chapter, you should be able to look at the surface of any pool of water disturbed by anything, and read the physics off it the way a musician reads a score. Let us begin.

WHAT A WAVE IS: ENERGY WITHOUT MATTER

Mr. Akilikubwa's lakeside lecture has done its job. We know, before any equation, two non-trivial facts about a wave: *it carries energy through space, but it does not carry the medium with it*. The water in front of the leaf at the start of the morning is the same water in front of the leaf at the end. The leaf has bobbed politely on the spot while the disturbance has cleared off into the open lake without it. Whatever a wave is, then, it is not a moving object. It is something more interesting: a moving **pattern** of disturbance, hosted by a medium whose particles agree, by the laws of mechanics, to take turns moving up and down so that the pattern can pass.

Here is the formal definition. A **wave** is a disturbance that propagates through a medium (or, in the case of light, through empty space), transferring energy from one place to another without the bulk transport of the medium itself. The medium does its work locally, by passing the disturbance along, and then it returns to where it started. The disturbance keeps going. Energy moves. Matter does not.

The point of this section is to make sure that, before we write down the equation of a wave, you have a firm physical picture of three separate things: *what is moving*, *what is not moving*, and *what we mean by every word we are about to use to describe both*. We will start with the two physical kinds of wave that nature offers, then build the small vocabulary that the rest of the chapter will live inside.

Transverse and longitudinal: two ways for a particle to dance

Imagine a long horizontal rope held by two people, one at each end. One person flicks her wrist sharply upward and then back to rest. A pulse runs along the rope toward the other person. The pulse travels horizontally. Each tiny segment of rope, however, has only moved vertically: it has gone up, and then back down, and is now where it was. The **direction of disturbance** (*vertical*) is at *right angles* to the **direction of propagation** (*horizontal*). A wave with this geometry is called a **transverse wave**. *Water surface waves*, *the waves on a guitar string*, and *all electromagnetic waves* (light, radio, X-rays) are transverse.

Now imagine instead a long spring. One person pushes one end of the spring sharply along its length and pulls back. A region of compression travels down the spring toward the other person. The pulse travels along the spring. Each coil of the spring, however, has only moved along the spring: it has been squeezed forward into its neighbour and then bounced back. The **direction of disturbance** is *the same as* the **direction of propagation**. A wave with this geometry is a **longitudinal wave**. *Sound waves in air*, *sound in water*, the waves your eardrum receives at every *Bongo Flewa* song, are all longitudinal. So are *seismic P-waves*, which we will meet again at the end of this chapter when we use them to locate earthquakes.

Two pieces of advice that will pay you back many times in this chapter. First, longitudinal waves are harder to draw than transverse waves, because the disturbance is along the propagation axis and overlaps the picture itself. The standard trick is to draw the longitudinal wave as a sequence of compressions (tightly packed coils) and rarefactions (loosely spaced coils), with a graph of **displacement** of each coil from its rest position drawn separately above the spring. Both pictures describe the same wave. The graph is the trick that lets you compare a longitudinal wave to a transverse one on equal terms. Second, even though sound is longitudinal, when you read its equation it will look exactly like the equation of a transverse wave. The mathematics does not care about the geometry; it cares about the displacement. We will exploit this in the next section to write a single equation that describes both kinds at once.

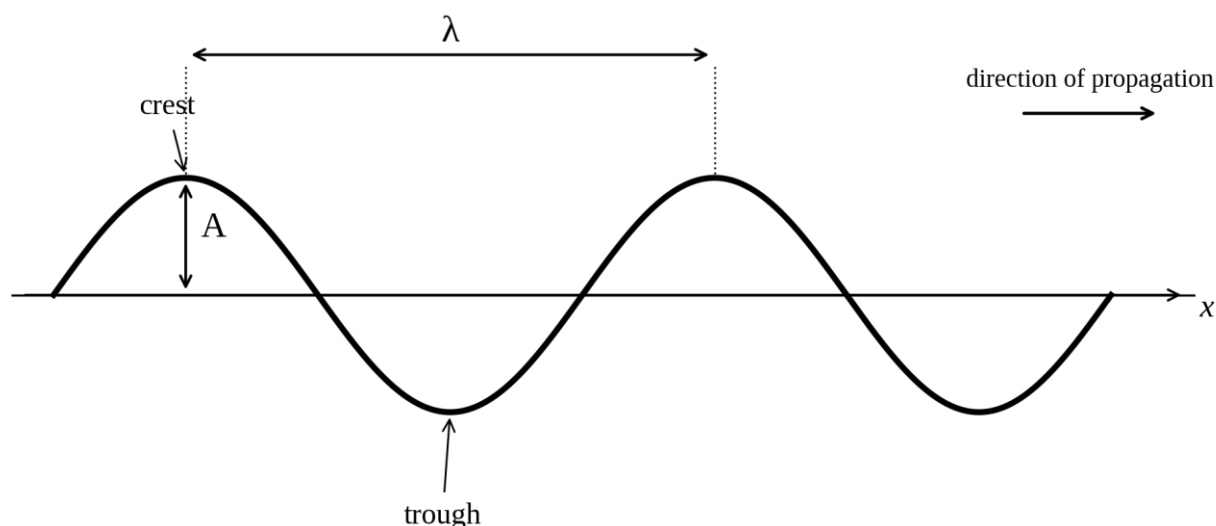
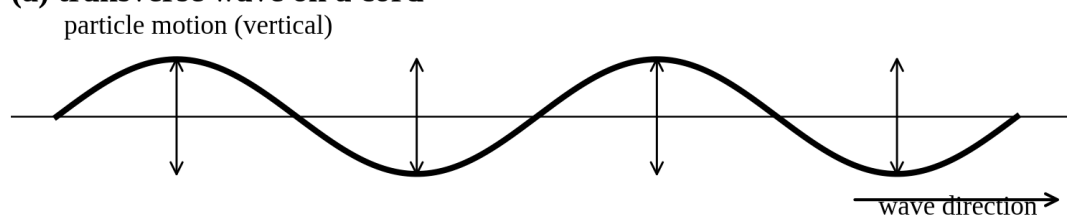
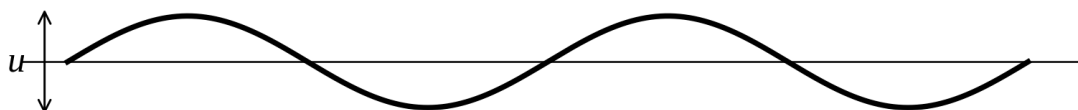


Figure: A transverse wave. The medium oscillates vertically (the y direction). The wave propagates horizontally (the x direction). A is the amplitude (peak displacement). λ is the wavelength (distance between consecutive crests, or between any two consecutive points in identical phase).

(a) transverse wave on a cord



(b) displacement of each coil from its rest position



(c) longitudinal wave on a spring

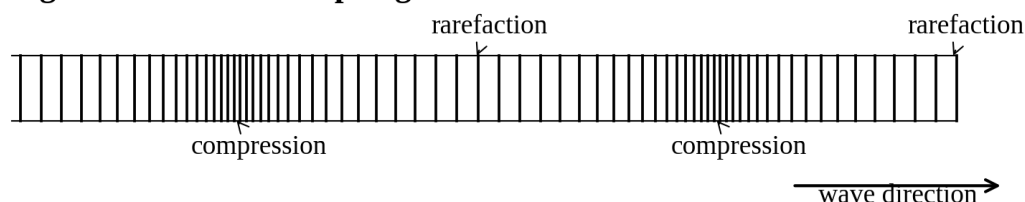


Figure: Transverse versus longitudinal. (a) A transverse wave drawn on a stretched cord; particles move vertically while the wave moves horizontally. (b) The displacement of each coil of a spring from its rest position, plotted against position along the spring. (c) The same longitudinal wave drawn as the spring itself, with compressions where the displacement-curve slope is steeply negative and rarefactions where it is steeply positive.

The wave vocabulary

Almost every question in this chapter, in any examination paper you will ever sit, will ask you to read or to compute one of the following six numbers. They are the wave's coordinates, in the same sense that latitude and longitude are a city's coordinates. If you are fluent in them, the chapter will feel light. If they are slippery, the chapter will be exhausting. Two minutes of careful attention now buys an enormous amount of comfort later.

Amplitude. The amplitude, written **A**, is the maximum displacement of any particle of the medium from its equilibrium position during the wave's passage. For a wave on water, **A** is the height of a crest above the still surface. For a wave on a guitar string, **A** is half the distance between the highest and lowest points of the vibrating string. The amplitude is always a positive number, and it has the units of length: metres or centimetres or, for a sound wave, micrometres.

Wavelength. The wavelength, written λ (Greek lambda), is the distance between any two consecutive points in identical phase. It may be: between two consecutive crests of a wave, between two consecutive troughs, or between any two consecutive points of the medium that are doing the same thing at the same time. For a **longitudinal wave**, the **wavelength** is the distance between two consecutive compressions. The wavelength is measured in metres.

Period. The period, written **T**, is the time taken for any one particle of the medium to complete a single full oscillation: up, down, and back to the start. The period is also the time the wave pattern takes to advance by exactly **one wavelength**. The period is measured in seconds.

Frequency. The frequency, written **f**, is the number of full oscillations completed by a particle per second. It is the reciprocal of the period:

$$f = \frac{1}{T}$$

The frequency is measured in hertz, where 1Hz means one oscillation per second. A piano A above middle C has a frequency of 440Hz. The hum of a Tanzanian power line is at 50Hz. The cosmic microwave background, which is light left over from the early universe, peaks at about 160GHz. Frequencies span an enormous range; periods are their inverses and span the inverse range.

Angular frequency. The angular frequency, written ω (Greek omega), is the frequency expressed in radians per second rather than oscillations per second. Since one full oscillation is 2π radians,

$$\omega = 2\pi f = \frac{2\pi}{T}$$

The angular frequency is measured in rad/s. *Why bother with two different ways of saying "frequency"?* Because the equation of a wave, like the equation of SHM in Chapter 6, contains a sine. The argument of a sine is an angle. Angles are measured in radians. So the natural variable for the time part of a wave's equation is ω , not f . The two are related by a factor of 2π and you must be fluent in both.

Wave number. The wave number, written **k**, is the spatial counterpart of the angular frequency. Where ω tells you how many radians of oscillation pass per second, **k** tells you *how many radians of oscillation fit into a metre*. Since one full wavelength is 2π radians of phase,

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

The wave number is measured in radians per metre, or simply in m^{-1} . A wave with a long wavelength (for example, an ocean swell, $\lambda \approx 100\text{m}$) has a small wave number ($k \approx 0.063\text{rad/m}$). A wave with a short wavelength (a fruit-bat echolocation call along the shores of Lake Tanganyika, $\lambda \approx 1\text{cm}$) has a large wave number ($k \approx 628\text{rad/m}$). The wave number is the spatial pulse-rate of a wave.

Wave speed. The sixth number is **v**, the wave speed: *the speed at which the pattern of disturbance travels through the medium*. The wave speed is **not** the speed of any particle. The wave speed is determined by the medium, the same way the speed of a transverse vibration on a string is determined by the tension and the linear mass density. From the geometry of the wave, the wave speed is related to the wavelength and the period by:

$$v = \frac{\lambda}{T} = f\lambda = \frac{\omega}{k}$$

All three forms are useful. The first (λ/T) is the most physical: in one period, the wave moves forward by one wavelength, so its speed is wavelength per period. The second ($f\lambda$) is the same thing reorganised so the experimentalist's two natural measurements (frequency and wavelength) appear directly. The third (ω/k) is the form in which v emerges naturally from the wave equation we will derive in the next section. They are the same equation in three coats.

Wave behaviour at a boundary: frequency is preserved, wavelength is not

One physical fact about waves is so important, and so frequently misunderstood by students, that we will state it now and refer back to it in three later sections. When a wave crosses from one medium into another (a sound wave going from air into water, an ultrasound pulse moving from fat into muscle, a seismic wave moving from sedimentary rock into harder basement rock), three things in general can happen to it. Some of the wave reflects off the boundary. Some of the wave transmits through. And the transmitted part travels at a different speed in the new medium, because the wave speed depends on the medium and the new medium has different elastic and inertial properties.

What does **not** change at a boundary is the wave's frequency. The reason is physical, not mathematical. The source on one side of the boundary is producing the wave at a particular rate, say 440 oscillations per second. Each oscillation must, for continuity, push exactly one corresponding oscillation across the boundary into the new medium. If the new medium produced more oscillations per second than the old medium delivered, oscillations would be appearing from nowhere; if it produced fewer, oscillations would be disappearing without trace. Neither is allowed by the conservation of disturbance count. The rate at which oscillations cross the boundary must be the same on both sides. Therefore the frequency is the same on both sides.

The wave speed, however, **does** change. And since $v = f\lambda$ with f kept constant, the wavelength has no choice but to scale in proportion to the speed. If the wave speed doubles in the new medium, the wavelength doubles. If it halves, the wavelength halves. This rule will resurface in the very last section of the chapter, when an ultrasound transducer pushes a 5MHz wave from fat (where it travels at about 1450m/s) into muscle (where it travels at about 1580m/s); we will use frequency invariance and $v = f\lambda$ to work out how the wavelength changes across the boundary, and the radiologist's image-resolution argument will follow from that. For now, simply remember the rule. If somebody asks what changes when a wave crosses from one medium into another, the truthful answer is: *the speed changes; the wavelength changes in step; the frequency does not.*

Two worked examples now. They are kinder than they look, but only the second one is keeping a straight face about it.

BINDER Example 1

On a calm morning at the southern end of Lake Tanganyika, a fisherman watches the small wooden float on his line bob steadily up and down. The crests of the surface waves on the lake are 2.4m apart, and the float completes 3 oscillations every 2 seconds. Find the wavelength, frequency, period, angular frequency, wave number, and wave speed of the surface waves.

Solution

The wavelength is given to us directly: it is the distance between consecutive crests.

$$\lambda = 2.4\text{m}$$

The frequency is the number of complete oscillations per second. The float completes 3 in 2 seconds, so:

$$f = \frac{3}{2\text{s}} = 1.5\text{Hz}$$

The period is the time for one full oscillation, which is the reciprocal of the frequency:

$$T = \frac{1}{f} = \frac{1}{1.5\text{Hz}} = 0.667\text{s}$$

The angular frequency is the frequency in radians per second:

$$\omega = 2\pi f = 2\pi \times 1.5\text{Hz} = 3\pi \text{ rad/s} = 9.42\text{rad/s}$$

The wave number is 2π divided by the wavelength:

$$k = \frac{2\pi}{\lambda} = \frac{2\pi}{2.4\text{m}} = 2.62\text{rad/m}$$

And the wave speed comes from $v = f\lambda$:

$$v = f\lambda = 1.5\text{Hz} \times 2.4\text{m} = 3.6\text{m/s}$$

Making sense of the answer: A surface-water wave on a calm lake travelling at 3.6m/s, or about 13km/h, is exactly what a casual eye would call "a gentle ripple cantering toward the shore". The number is right. So is its bookkeeping cousin: 1.5 oscillations per second is what a fisherman watching his float would describe as "bobbing comfortably". Six numbers, all of them describing the same wave, all of them computed from two pieces of given data: the crest spacing and the count of bobs in 2 seconds. That is the entire chapter's calculation pattern in miniature: read off the easy numbers, use $v = f\lambda$ or $\omega = 2\pi f$ or $k = 2\pi/\lambda$ to fill in the rest, and watch the units cancel cleanly.

Think like a physicist: Notice what just happened. The fisherman gave us only two numbers: the spacing between consecutive crests, and the rate at which his float bobbed. From those two numbers we extracted six. This is not a free lunch; it is the deep structure of a wave. A monochromatic wave has only two truly independent pieces of information, one belonging to space (any one of λ or k) and one belonging to time (any one of T , f , or ω). Everything else, including the wave speed itself, is derived.

BINDER Example 2

A sound wave of wavelength λ travels in a first medium with velocity v . It enters a second medium in which the velocity is $4v$. Determine the wavelength of the sound wave in the second medium.

Solution

The source frequency is unchanged across the boundary; conservation of oscillation count requires the same number of oscillations per second on both sides. Call that frequency f . In medium 1,

$$v = f\lambda \Rightarrow f = \frac{v}{\lambda}$$

In medium 2, the wave speed is $4v$ and the same frequency f applies. Calling the new wavelength λ' , the basic relation gives:

$$4v = f\lambda' = \left(\frac{v}{\lambda}\right)\lambda' \Rightarrow \lambda' = 4\lambda$$

So the wavelength in the second medium is four times the wavelength in the first.

Making sense of the answer: Notice that the answer is decided not by clever algebra but by a single physical commitment: the source dictates the rate; the medium dictates the speed. Wavelength is a derived quantity, $\lambda = v/f$, and it does whatever the speed and frequency tell it to do.

Think like a physicist: There is a beautiful continuity argument hidden inside this calculation. Imagine a long row of disturbance peaks marching toward the boundary in medium 1. They arrive at the boundary at a steady rate, f peaks per second. On the far side, they must depart at exactly the same rate, f peaks per second, for the same reason that cars on a road cannot vanish at a checkpoint. But on the far side they are spreading out four times as fast. Therefore on the far side the spacing between consecutive peaks must be four times as large: $\lambda' = 4\lambda$. The wavelength is just the geometric record of how far one peak has gone before the next one is born; if the peaks are now travelling four times as quickly, each peak runs four times as far before its successor turns up. This is the calculation done with hands instead of algebra, and the two views must agree. They do.

With those two examples we have caught up, in vocabulary if not yet in equations, with the language the rest of the chapter will speak. We know what a wave is. We know its six descriptive numbers. We know that two of those numbers, one belonging to space and one belonging to time, are enough to pin down the rest. We know what happens when a wave crosses a boundary. What we do not yet have is the actual equation of a wave: the one Mr. Akilikubwa wrote in the wet sand and the lake then erased. The next section finally writes that equation, derives it from things you already know about simple harmonic motion, and teaches you to read its parameters off the page the way a musician reads notes off a stave. The chapter has, until now, been about words. From here on it will be about a single sentence in those words.

THE PROGRESSIVE WAVE EQUATION

So far we have spoken about a wave the way an ornithologist speaks about a bird that has flown overhead: in admiring generalities. We know what a wave does, what it carries, and what it does not carry. We know its six descriptive numbers. What we do not yet have is the equation that puts all of this on paper. The point of this section is to write that equation, derive it from things you already know, and then teach you to read it the way a musician reads a score.

The equation we are after is the same one Mr. Akilikubwa wrote in the wet sand at the start of the chapter. It will turn out to be a sentence in three pieces of vocabulary, all of which you already speak fluently: the amplitude A from this section's predecessor, the angular frequency ω from Chapter 6, and the wave number k from a few pages ago. The derivation rests on one physical observation: a wave is simple harmonic motion shared, with a delay, between every particle of the medium. Once we have the equation, every wave problem in the chapter becomes one of two things: read off its parameters, or build it from given parameters. There is no third option, and there is no third equation.

From SHM to a travelling wave: a derivation in two steps

Picture a long horizontal rope or a long stretched string, with one end held at the origin $x = 0$ by a person who can move that end up and down at will. Suppose she moves the end up and down in simple harmonic motion. From Chapter 6, we know that the displacement of that end, written y , is a sinusoidal function of time:

$$y(0, t) = A \sin(\omega t)$$

Here A is the amplitude of her hand, ω is its angular frequency, and the bracket $y(0, t)$ emphasises that this expression describes the displacement at the location $x = 0$ at any time t . So far this is just SHM with a label. Nothing new.

Now the new ingredient. The string is elastic. When the source end (at $x = 0$) moves up, it pulls its immediate neighbour up too, but only after a tiny delay. The neighbour pulls its neighbour, also after a delay. The motion runs along the string. By the time the disturbance reaches a point at distance x from the source, a finite time has elapsed. If the wave travels at speed v , that time is exactly the distance divided by the speed:

$$\text{delay} = \frac{x}{v}$$

So the particle at position x is doing the same simple harmonic motion as the source, but lagging it by a time of x divided by v . Whatever the source was doing at time t , the particle at x is doing at time $t + x/v$; equivalently, what the particle at x is doing right now (time t) is what the source was doing at the earlier time $t - x/v$. Substituting the earlier time into the SHM expression:

$$y(x, t) = A \sin\left(\omega\left(t - \frac{x}{v}\right)\right)$$

Distribute the angular frequency through the bracket:

$$y(x, t) = A \sin\left(\omega t - \frac{\omega}{v}x\right)$$

And the cluster ω/v has a name we have already met. The wave **number** k is, by definition, *the angular frequency divided by the wave speed*:

$$k = \frac{\omega}{v} \quad \text{so} \quad v = \frac{\omega}{k}$$

Substituting back, we have the equation that the chapter has been earning the right to write:

$$y(x, t) = A \sin(\omega t - kx)$$

This is the displacement, at any time t and any position x , of a sinusoidal wave travelling along the string in the positive x -direction with amplitude A , angular frequency ω , and wave number k . The minus sign in front of kx is the wave's signature. It says: *the further along the string we look, the further behind the source the particle is in its cycle*. That phase lag, growing in step with x , is what makes the disturbance pattern shift to the right as time advances.

If the source is instead at the far end and the wave travels in the negative x -direction, the same derivation runs through with the opposite sign of the delay, and the equation reads:

$$y(x, t) = A \sin(\omega t + kx)$$

So the sign of kx is the wave's compass: a **minus** sign means a wave travelling toward **positive** x ; a **plus** sign means a wave travelling toward **negative** x . Read the sign first; everything else falls into place.

Reading the equation's parameters at a glance

The single most common type of question in this chapter, in any examination paper you will ever sit, will hand you a wave equation in the standard form and ask you to extract its parameters. The trick is to recognise that $y = A\sin(\omega t - kx)$ is a recipe in which every parameter sits in plain sight. You read them off the way you read the time off a clock face, by knowing where each number lives.

Given $y = A\sin(\omega t - kx)$, the parameters of the wave are:

amplitude = A (the coefficient outside the sine)

ω = coefficient of t ,

k = coefficient of x

$$T = \frac{2\pi}{\omega}, \quad f = \frac{\omega}{2\pi}, \quad \lambda = \frac{2\pi}{k}, \quad v = \frac{\omega}{k} = f\lambda$$

Six parameters, five identities, one minute of work. There are no judgement calls and no algebraic tricks. Practise the extraction enough times that it becomes mechanical, because in an examination the marks will go to whoever does it without thinking.

Two velocities, one wave: keeping wave speed and particle speed apart

Before any worked examples, we must spend a careful minute on the conceptual error that this chapter is most ruthless about catching. The equation $y = A\sin(\omega t - kx)$ describes two motions at once.

The first is the motion of the wave pattern itself, which slides rigidly to the right at speed $v = \omega/k$.

The second is the motion of any individual particle of the string, which oscillates up and down without travelling anywhere on average.

These are different things. Their speeds are different numbers. The chapter that mistakes the two will fail.

To find the velocity of an individual particle at position x , we differentiate y with respect to t while holding x fixed. The particle's position depends on time, so its velocity is dy/dt :

$$v_p = \frac{dy}{dt} = A\omega\cos(\omega t - kx)$$

This is also a sinusoid, but it is a sinusoid of a particle's velocity in time, not the wave's velocity. Its maximum value, achieved when the particle is passing through its equilibrium position, is:

$$v_{p,\max} = A\omega$$

Compare this to the wave speed. The wave speed $v = \omega/k$ is determined entirely by the medium (its tension, its linear mass density, its elasticity, and so on). It does not depend on the amplitude A . It does not depend on where on the string you look or on what the particle there happens to be doing. It is a property of the highway, not of any one car driving on it.

The maximum particle speed $v_{p,\max} = A\omega$, on the other hand, depends on the amplitude. Make the wave taller and every particle on it moves faster up and down, even though the wave pattern itself slides past at exactly the same v as before. Doubling the amplitude doubles the maximum particle speed; it does not change the wave speed at all. The two are separate physical quantities, related to each other only by the equation that governs the wave, and asking for one when the question wants the other is the most reliable way to lose marks in this chapter. Examiners know this, and they place traps accordingly.

Warning: the two velocities of a wave

The **wave speed** $v = \omega/k$ is the speed at which the wave pattern moves along the medium. It is a property of the medium and does not depend on the amplitude.

The **particle velocity** $v_p = dy/dt$ is the speed at which an individual particle is moving up and down at right angles to the medium. It depends on time, on position, and on the amplitude. Its maximum value is $v_{p,\max} = A\omega$.

These are completely different quantities. They have completely different values. Confusing them is the most common error in the chapter. In every numerical problem, ask first: *am I being asked for the speed of the pattern, or the speed of a particle?*

The differential equation that every progressive wave obeys

There is one further fact about $y = A\sin(\omega t - kx)$ that is worth establishing now, because it underpins half of the rest of the chapter and all of optics in Chapter 8. The displacement function we have written down is not just a static picture; it is a solution to a particular partial differential equation, called the wave equation:

$$\frac{\partial^2 y}{\partial t^2} = v^2 \frac{\partial^2 y}{\partial x^2}$$

Read this equation in plain English. On the left is how fast the displacement of a particle is accelerating in time. On the right is how sharply the wave profile is curving in space, scaled by the square of the wave speed. The equation says these two are equal, always. In a stretched string, this equation is just Newton's second law applied to a tiny segment of string, with restoring force from the tension on either side and inertia from the segment's linear mass density. We will not derive it from Newton here; we will simply check that $y = A\sin(\omega t - kx)$ is a solution.

Differentiate y twice with respect to t , holding x fixed:

$$\begin{aligned}\frac{\partial y}{\partial t} &= A\omega \cos(\omega t - kx) \\ \frac{\partial^2 y}{\partial t^2} &= -A\omega^2 \sin(\omega t - kx) = -\omega^2 y\end{aligned}$$

And twice with respect to x , holding t fixed:

$$\begin{aligned}\frac{\partial y}{\partial x} &= -Ak \cos(\omega t - kx) \\ \frac{\partial^2 y}{\partial x^2} &= -Ak^2 \sin(\omega t - kx) = -k^2 y\end{aligned}$$

Take the ratio of the two second derivatives:

$$\frac{\frac{\partial^2 y}{\partial t^2}}{\frac{\partial^2 y}{\partial x^2}} = \frac{-\omega^2 y}{-k^2 y} = \frac{\omega^2}{k^2} = \left(\frac{\omega}{k}\right)^2 = v^2$$

Hence:

$$\frac{\partial^2 y}{\partial t^2} = v^2 \frac{\partial^2 y}{\partial x^2}$$

So $y = A\sin(\omega t - kx)$ does indeed satisfy the wave equation, with the wave speed coming out as $v = \omega / k$. The arithmetic is honest. Two differentiations expose the deep structure: a wave is a function whose curvature in space is exactly proportional to its acceleration in time, with the proportionality constant being the square of the wave speed.

For now, give theory a short holiday and let us sharpen our understanding with six worked examples. None of them are unkind, but the third one quietly pretends to be a parameter-extraction problem and is in fact a particle-velocity problem in disguise. Read the question twice.

BINDER Example 3

A transverse wave is described by the equation $y(x,t) = 8\sin(200\pi t - 4\pi x)$ mm, where t is in seconds and x is in metres. Find the amplitude, angular frequency, wave number, period, frequency, wavelength, and wave speed.

Solution

Compare the equation directly with the standard form $y = A\sin(\omega t - kx)$.

$$A = 8.0\text{mm} = 8.0 \times 10^{-3}\text{m}$$

$$\omega = 200\pi \text{ rad/s}$$

$$k = 4\pi \text{ rad/m}$$

Now compute the period, frequency, wavelength, and wave speed from the standard relations.

$$T = \frac{2\pi}{\omega} = \frac{2\pi}{200\pi \text{ rad/s}} = 0.010\text{s}$$

$$f = \frac{1}{T} = \frac{1}{0.010\text{s}} = 100\text{Hz}$$

$$\lambda = \frac{2\pi}{k} = \frac{2\pi}{4\pi \text{ rad/m}} = 0.50\text{m}$$

$$v = \frac{\omega}{k} = \frac{200\pi \text{ rad/s}}{4\pi \text{ rad/m}} = 50\text{m/s}$$

Making sense of the answer: A 50m/s transverse wave on a stretched wire is well within the range of physical realism, somewhere between a guitar string at modest tension and a clothesline carrying a hand pulse. A 100Hz oscillation is a low musical note, well below the lower edge of human hearing if it were a sound. An 8mm amplitude is a generously plucked string; you would see it. Notice the bookkeeping: every one of the six numbers came from comparing the given equation against $y = A\sin(\omega t - kx)$ and applying $T = 2\pi/\omega$, $f = 1/T$, $\lambda = 2\pi/k$, $v = \omega/k$. There were no judgement calls. This is the engine of every parameter-extraction problem in the chapter.

Think like a physicist: Look once more at the wave speed: $v = \omega/k$. We could equally well have computed it from $v = f\lambda = (100\text{Hz})(0.50\text{m}) = 50\text{m/s}$. Both routes give the same answer because the four numbers ω , k , f , λ are not independent: $\omega = 2\pi f$ and $k = 2\pi/\lambda$ pin them together. There are only two genuinely independent pieces of information in this wave (one belonging to time, one belonging to space), and once you have either pair (T , λ), (f , λ), or (ω , k), every other parameter follows. This is the lesson Example 1 hinted at in the previous section, now made unambiguous: a monochromatic wave has two degrees of freedom, not six.

BINDER Example 4

A horizontal stretched elastic string of length 3.0m and mass 12kg is subjected to a tension of 1.6N. A transverse wave of frequency 40Hz is propagated down the string. Determine the distance between successive crests of this wave.

Solution

The distance between successive crests is the wavelength. We need v first, then λ from $v = f\lambda$. The linear mass density is

$$\mu = \frac{m}{L} = \frac{12\text{kg}}{3.0\text{m}} = 4.0\text{kg/m}$$

The wave speed for a transverse wave on a stretched string was derived at the end of Chapter 6 as:

$$v = \sqrt{\frac{T}{\mu}} = \sqrt{\frac{1.6\text{N}}{4.0\text{kg/m}}} = 0.632\text{m/s}$$

The wavelength then follows from $v = f\lambda$:

$$\lambda = \frac{v}{f} = \frac{0.632\text{m/s}}{40\text{Hz}} = 0.0158\text{m} \approx 1.6\text{cm}$$

Making sense of the answer: A heavy string under a tiny tension carries waves slowly, and a slow wave at a moderate frequency has a short wavelength. The numbers obey the right physics: $\mu = 4\text{kg/m}$ is unusually large (most laboratory strings are a thousand times lighter per metre), $T = 1.6\text{N}$ is unusually small (most laboratory strings are tensioned to tens or hundreds of newtons), so a wave speed of 0.6m/s is exactly the order of magnitude expected. The 1.6cm wavelength is what you would see if you happened to film the string in slow motion. The arithmetic is unforgiving, but it is also unambiguous: get μ right, get v right, get λ from $v = f\lambda$.

Think like a physicist: This problem hands you four numbers (L , m , T , f) and asks you for one (λ). It looks unbalanced, but it is not. The four given numbers split into two functional pairs. The pair (L , m) gives μ . The pair (T , μ) gives v . Then (v , f) gives λ . Each pair contributes one of the equation links of the chapter: $\mu = m/L$, $v = \sqrt{T/\mu}$, $v = f\lambda$. Three links, three steps, one answer. Examination problems often look as if they are giving you more information than you need; they are not. Every given number has a job. If a number seems irrelevant, look for a chain of equations you have not noticed yet.

BINDER Example 5

Write the equation of a progressive wave travelling in the positive x -direction with amplitude 0.030m, frequency 80Hz, and wave speed 40m/s.

Solution

This is the reverse of Example 3: from given physical parameters, build the wave equation. The standard form is $y = A \sin(\omega t - kx)$. We need ω and k .

$$\omega = 2\pi f = 2\pi \times 80\text{Hz} = 160\pi \text{ rad/s}$$

$$\lambda = \frac{v}{f} = \frac{40\text{m/s}}{80\text{Hz}} = 0.50\text{m}$$

$$k = \frac{2\pi}{\lambda} = \frac{2\pi}{0.50\text{m}} = 4\pi \text{ rad/m}$$

Substituting $A = 0.030\text{m}$, $\omega = 160\pi \text{ rad/s}$, and $k = 4\pi \text{ rad/m}$ into $y = A \sin(\omega t - kx)$:

$$y(x, t) = 0.030 \sin(160\pi t - 4\pi x) \text{ m,}$$

with t in seconds and x in metres.

Making sense of the answer: Read the equation back to confirm. Coefficient of t is $160\pi \text{ rad/s}$; dividing by 2π gives 80Hz . Coefficient of x is $4\pi \text{ rad/m}$; 2π divided by it gives 0.50m . Wave speed is the ratio $160\pi/4\pi = 40\text{m/s}$. All three given numbers come back out unchanged. The amplitude was carried straight across. The minus sign in front of kx encodes "travelling in the positive x -direction", which is what the question asked for. Every check passes.

Think like a physicist: Construction problems like this one are how you debug your own understanding. If you can build a wave equation from physical parameters, you understand what each symbol means; if you cannot, the symbols are still words you have not internalised. Keep practising the round trip in both directions. Extraction problems like Example 3 take you from the equation to the physics; construction problems like this one take you from the physics back to the equation. They are the same skill viewed from opposite sides, and an honest physicist owns both.

BINDER Example 6

A transverse wave on a string is given by $y(x, t) = 0.040 \sin(10\pi t - 2\pi x) \text{ m}$, where t is in seconds and x is in metres. (a) Find the speed of the wave. (b) Find the velocity of the particle at $x = 0.125\text{m}$ at time $t = 0.050\text{s}$. (c) Find the maximum particle speed anywhere on the string. (d) Explain physically why the answers to (a) and (b) are different.

Solution

Comparing with $y = A \sin(\omega t - kx)$ we read off $A = 0.040\text{m}$, $\omega = 10\pi \text{ rad/s}$, $k = 2\pi \text{ rad/m}$.

(a) The wave speed is the ratio ω/k :

$$v = \frac{\omega}{k} = \frac{10\pi \text{ rad/s}}{2\pi \text{ rad/m}} = 5.0\text{m/s}$$

(b) The particle velocity is the time derivative of y at fixed x :

$$v_p = \frac{\partial y}{\partial t} = A\omega \cos(\omega t - kx)$$

First evaluate the cosine argument at the chosen point. With $x = 0.125\text{m}$ and $t = 0.050\text{s}$,

$$\omega t - kx = (10\pi \text{ rad/s})(0.050\text{s}) - (2\pi \text{ rad/m})(0.125\text{m}) = 0.5\pi - 0.25\pi = \frac{\pi}{4} \text{ rad}$$

The cosine of $\pi/4$ is $\sqrt{2}/2$. Substituting:

$$v_p(0.125\text{m}, 0.050\text{s}) = A\omega \cos\left(\frac{\pi}{4}\right) = (0.040\text{m})(10\pi \text{ rad/s})\left(\frac{\sqrt{2}}{2}\right) = 0.20\pi\sqrt{2} \text{ m/s} \approx 0.889\text{m/s}$$

(c) The particle speed is maximised when the cosine is ± 1 , giving:

$$v_{p,\text{max}} = A\omega = (0.040\text{m})(10\pi \text{ rad/s}) = 0.40\pi \text{ m/s} \approx 1.26\text{m/s}$$

So at the chosen point and instant the particle is moving up at about 0.889m/s , which is $\sqrt{2}/2$ (about 71%) of its maximum particle speed of 1.26m/s . Different particles, at other positions and instants, will be moving at different speeds between zero and the maximum, depending only on the cosine of their particular phase $\omega t - kx$.

(d) The wave speed and the particle speed are different physical quantities. The wave speed $v = 5.0\text{m/s}$ is the speed at which the disturbance pattern moves to the right along the string; it is determined by the medium. The particle velocity $v_p \approx 0.889\text{m/s}$ is the speed at which the particle at $(x = 0.125\text{m}, t = 0.050\text{s})$ is moving

up at that instant; it is determined by the equation that governs the wave evaluated at that one point and one time.

Understand: The two numbers describe different physical things: one is a property of the highway (the medium), the other is a property of one car driving on the highway at one moment (one particle at one phase).

Making sense of the answer: Notice the order of magnitude. The wave pattern is moving along the string at 5.0m/s, while the particle at our chosen point is wiggling upward at about 0.89m/s, somewhat below its own maximum of 1.26m/s. The disturbance is racing past the particle. This is the right qualitative picture: a wave is a fast-moving pattern hosted by particles each of which is doing comparatively modest oscillation in place. Different particles, at other positions or instants, will be moving at different speeds, between zero and the maximum, but never at the wave speed itself. If the wave moved at the same speed as the particles, it would not be a wave at all; it would be a rigid translation of the medium.

Think like a physicist: There is an extreme case worth keeping at the back of your mind. The maximum particle speed $v_{p,max} = A\omega$ is proportional to the amplitude. If you crank the amplitude up high enough, you can in principle make the maximum particle speed exceed the wave speed $v = \omega/k$. When that happens, the small-amplitude approximation behind $y = A\sin(\omega t - kx)$ breaks down, and the wave starts to behave non-linearly: it can steepen, it can break, and it can shed energy through means we have not yet introduced. For ordinary waves at ordinary amplitudes we are far from this regime. The inequality $A\omega < v$ is the silent assumption that makes the linear theory work. For our wave, $A\omega = 1.26\text{m/s}$ sits comfortably below $v = 5.0\text{m/s}$, so the linear theory holds.

REAL Example 7

On the morning of 7 June, a seismograph operated by the Tanzania Meteorological Authority near Dodoma records the ground displacement of a small earthquake originating beneath the East African Rift. The recorded displacement is well-fit by the equation $y(x,t) = 0.0030 \sin(3.0t - 0.0060x)$ m, where t is in seconds and x is in metres. From the equation alone, find (a) the amplitude of ground shaking, (b) the wave speed, (c) the wavelength, and (d) the frequency.

Solution

Compare with $y = A\sin(\omega t - kx)$. The amplitude appears as the coefficient outside the sine; the angular frequency is the coefficient of t ; the wave number is the coefficient of x .

$$A = 0.0030\text{m} = 3.0\text{mm}$$

$$\omega = 3.0 \text{ rad/s}$$

$$k = 0.0060 \text{ rad/m}$$

(a) The amplitude is 3.0mm.

(b) The wave speed is the ratio:

$$v = \frac{\omega}{k} = \frac{3.0 \text{ rad/s}}{0.0060 \text{ rad/m}} = 500\text{m/s}$$

(c) The wavelength comes from $\lambda = 2\pi/k$:

$$\lambda = \frac{2\pi}{k} = \frac{2\pi}{0.0060 \text{ rad/m}} \approx 1047\text{m} = 1.047\text{km}$$

(d) The frequency comes from $f = \omega/(2\pi)$:

$$f = \frac{\omega}{2\pi} = \frac{3.0 \text{ rad/s}}{2\pi} = 0.48\text{Hz}$$

Making sense of the answer: A ground-shaking amplitude of 3mm is small, the kind of tremor that rattles cups and is felt by people sitting still but not by people walking; it is consistent with a small to moderate East African Rift event recorded at moderate epicentral distance. A wave speed of 500m/s is squarely in the range expected for seismic surface waves through Tanzanian sedimentary cover; this is roughly half the P-wave speed in solid rock, which is what we would expect for a shallow crustal disturbance. The wavelength of about a kilometre and the frequency of about half a hertz both fit the picture: long, slow undulations of the kind a seismograph records and a casual observer feels as a gentle swaying of buildings rather than a sharp jolt.

Think like a physicist: The three numbers (0.0030, 3.0, 0.0060) and the implicit minus sign are an extraordinary compression of physical information. From three numbers we extracted the depth of the disturbance (amplitude), the rate of the shaking (frequency), the spatial scale of the disturbance

(wavelength), and the speed of its travel through the rock (wave speed). The seismograph trace is, in one literal sense, a sentence in the language of $y = A \sin(\omega t - kx)$, and the language is rich enough that a single recorded equation tells the whole story. This is why physicists insist on writing waves as equations rather than as pictures. The picture is beautiful; the equation is computable.

HOT Example 8

A transverse wave on a string is described by $y(x,t) = 0.050 \sin(100\pi t - 5\pi x)$ m, where t is in seconds and x is in metres. The string has linear mass density $\mu = 4.0 \times 10^{-3}$ kg/m. (a) Find all six wave parameters (A , ω , k , T , f , λ) and the wave speed. (b) Find the tension in the string. (c) Find the velocity of the particle at $x = 0.10$ m at time $t = 0.010$ s. (d) Show that y satisfies the wave equation $\partial^2 y / \partial t^2 = v^2 \cdot \partial^2 y / \partial x^2$.

Solution

(a) Comparing with $y = A \sin(\omega t - kx)$:

$$A = 0.050 \text{ m}, \quad \omega = 100\pi \text{ rad/s}, \quad k = 5\pi \text{ rad/m}$$

$$T = \frac{2\pi}{\omega} = \frac{2\pi}{100\pi \text{ rad/s}} = 0.020 \text{ s}$$

$$f = \frac{1}{T} = 50 \text{ Hz}$$

$$\lambda = \frac{2\pi}{k} = \frac{2\pi}{5\pi \text{ rad/m}} = 0.40 \text{ m}$$

$$v = f\lambda = (50 \text{ Hz})(0.40 \text{ m}) = 20 \text{ m/s}$$

(b) The tension follows from $v = \sqrt{T/\mu}$ rearranged:

$$T = \mu v^2 = (4.0 \times 10^{-3} \text{ kg/m})(20 \text{ m/s})^2 = 1.6 \text{ N}$$

(c) The particle velocity is dy/dt at fixed x :

$$v_p = A\omega \cos(\omega t - kx)$$

At $x = 0.10$ m, $t = 0.010$ s, the argument of the cosine is

$$\omega t - kx = (100\pi \text{ rad/s})(0.010 \text{ s}) - (5\pi \text{ rad/m})(0.10 \text{ m}) = \pi - 0.5\pi = 0.5\pi \text{ rad}$$

$$\cos(0.5\pi) = 0$$

$$v_p = A\omega \times 0 = 0 \text{ m/s}$$

That particle, at that instant, is momentarily at rest. It is at one extreme of its oscillation, about to reverse direction.

(d) The wave equation we are checking is

$$\frac{\partial^2 y}{\partial t^2} = v^2 \cdot \frac{\partial^2 y}{\partial x^2}$$

We check this by differentiating the given equation $y = 0.050 \sin(100\pi t - 5\pi x)$ m directly, in three small stages: first the time side, then the space side, then the comparison.

The time side

Differentiate y once with respect to t , holding x fixed:

$$\frac{\partial y}{\partial t} = 0.050 \times 100\pi \times \cos(100\pi t - 5\pi x) = 5\pi \cos(100\pi t - 5\pi x) \text{ m/s}$$

Differentiate once more:

$$\frac{\partial^2 y}{\partial t^2} = -5\pi \times 100\pi \times \sin(100\pi t - 5\pi x) = -500\pi^2 \sin(100\pi t - 5\pi x) \text{ m/s}^2$$

Now notice that $0.050 \sin(100\pi t - 5\pi x)$ is exactly y . We can therefore pull the factor 0.050 back inside the sine to expose y :

$$\frac{\partial^2 y}{\partial t^2} = -(100\pi)^2 \times y = -10000\pi^2 y$$

The space side

Now differentiate y twice with respect to x , holding t fixed:

$$\frac{\partial y}{\partial x} = 0.050 \times (-5\pi) \times \cos(100\pi t - 5\pi x) = -0.25\pi \cos(100\pi t - 5\pi x)$$

$$\frac{\partial^2 y}{\partial x^2} = (-0.25\pi) \times (-5\pi) \times (-\sin(100\pi t - 5\pi x)) = -1.25\pi^2 \sin(100\pi t - 5\pi x)$$

Pull the 0.050 back inside the sine again to expose y:

$$\frac{\partial^2 y}{\partial x^2} = -(5\pi)^2 \times y = -25\pi^2 y$$

The comparison

Substitute both results into the wave equation:

The left-hand side is;

$$\frac{\partial^2 y}{\partial t^2} = -10000\pi^2 y$$

The right-hand side is;

$$v^2 \times \frac{\partial^2 y}{\partial x^2} = v^2 \times (-25\pi^2 y) = -25\pi^2 v^2 y$$

The two sides agree provided that

$$10000\pi^2 = 25\pi^2 v^2$$

Solve for v:

$$v^2 = \frac{10000}{25} = 400, \text{ so } v = 20\text{m/s.}$$

This matches the wave speed $v = 20\text{m/s}$ found in part (a). The wave equation is satisfied, and the wave speed that falls out of the differential equation is precisely the same wave speed we read directly from the ratio ω/k of the original equation. The arithmetic closes the loop.

Making sense of the answer: All three parts of this problem fit together. The single given equation tells us everything about the wave (part a) and, combined with the linear mass density of the string, lets us deduce the tension (part b). The particle velocity (part c) zooms in on one particle at one instant and finds it momentarily at rest, exactly the point in its oscillation where its kinetic energy is least and its potential energy is greatest.

Think like a physicist: Look at the structure of part (d). Two derivatives in time gave a factor of $-(100\pi)^2 = -\omega^2$. Two derivatives in space gave a factor of $-(5\pi)^2 = -k^2$. The wave equation says these two are related by a factor of v^2 . Squaring $v = \omega/k$ gives $v^2 = \omega^2/k^2$, and the ratio of the second derivatives reproduces this exactly: $10000\pi^2 \div 25\pi^2 = 400 = v^2$. This is more than a calculation; it is the reason the wave equation has the form it does. Acceleration in time and curvature in space are both proportional to y, with proportionality constants ω^2 and k^2 , and their ratio fixes the wave speed. If you ever forget the wave equation, write down any sinusoidal wave you happen to know, differentiate twice each way, take the ratio, and the wave equation falls out of the algebra without your having to memorise it. This is what physicists mean when they say a structural equation is its own mnemonic.

Eight worked examples, six wave parameters, and one master equation. We have done the heavy lifting: the equation $y = A \sin(\omega t - kx)$ is no longer a mystery to be unpacked but a sentence whose vocabulary you can read at a glance. What we have not yet done is something that, on the surface, sounds simpler than anything in this section. We have not yet compared two waves. When two waves meet, what matters is not the value of either of their phases in isolation but the difference between them. The next section is about that difference.

PHASE, PHASE DIFFERENCE, AND WAVE INTENSITY

In the previous section we learned to read the equation of one wave. We can extract its parameters, build it from given physical quantities, and prove that it satisfies the wave equation. The next natural question, and the one this section answers, is *what to do when there are two of them*. Two waves on the same string. Two ripples meeting on Lake Victoria. Two voices arriving at the same ear from two different speakers. The numerical comparison of one wave to another rests almost entirely on a single concept: *the difference in phase between them*. Master phase difference and you have unlocked, in one sitting, the calculation that drives Young's double-slit experiment, the colours of an oil film on a puddle, the entire physics of antenna

arrays, and (in less glamorous form) the noise-cancelling headphones a Form Five student might wear on the bus to school.

This section also picks up the second important property of a wave that the previous section postponed: the energy a wave carries. Every wave has an amplitude, and every amplitude implies a definite quantity of energy per unit volume of the medium. When that energy spreads outward from a point source through three-dimensional space, it dilutes in a strict geometrical pattern called the inverse-square law. We will derive that pattern, see what it implies about how the amplitude falls off with distance, and use it to read the loudness of any sound source from any distance.

Two ideas in this section, therefore: **phase difference** and **intensity**. They are linked by a common ancestor, the equation $y = A \sin(\omega t - kx)$, which tells you both where in the cycle each particle is and how much energy each particle is carrying.

Phase: the angular bookkeeping of where you are in the cycle

Take the wave equation we earned in the previous section,

$$y = A \sin(\omega t - kx)$$

and look closely at the argument inside the sine, namely $\omega t - kx$. This argument has a name: it is called the **phase** of the wave at position x at time t , often written ϕ . **Phase is an angle, measured in radians, that tells you exactly where in its cycle the particle at (x, t) is at that moment.**

- A phase of 0 means the particle is at its equilibrium position and moving upward.
 - A phase of $\pi/2$ means the particle is at its highest point.
 - A phase of π means it is back at equilibrium, but now moving downward.
 - A phase of $3\pi/2$ means it is at its lowest point.
 - After a phase advance of 2π the particle returns to where the cycle started and the whole story repeats.
- So a wave's equation is really a bookkeeping system: every particle of the medium has its own phase clock, and the equation $y = A \sin(\omega t - kx)$ is the master rule that tells you what each clock reads at each instant.

Two observations follow immediately.

- 1) The phase of a single particle (held at a fixed x) advances steadily in time at a rate ω . Time t increases, kx is constant, so $\phi = \omega t - kx$ grows linearly.
- 2) At any single instant in time, two particles at different positions read different phases, because their kx values are different. This second observation is the entire content of phase difference, which is what the next subsection is about.

Phase difference between two points on the same wave

Pick two specific points on the same wave at the same instant t . Call them point 1 (at position x_1) and point 2 (at position x_2). Each has its own phase at that instant:

$$\phi_1 = \omega t - kx_1, \quad \phi_2 = \omega t - kx_2$$

The **phase difference** between point 2 and point 1 is, by definition, the difference of their phases:

$$\Delta\phi = \phi_1 - \phi_2 = k(x_2 - x_1) = k\Delta x \text{ (where } \Delta x \text{ is the distance separating the two points)}$$

The ωt cancels because it is shared by both points: a phase difference is taken at the same instant. The remaining piece, $k\Delta x$, is the entire result. Substituting $k = 2\pi/\lambda$:

$$\Delta\phi = \frac{2\pi}{\lambda} \Delta x$$

This is the master equation of this section. **In words:** the phase difference between two points on a wave is 2π radians per wavelength of separation. Two points exactly one wavelength apart have a phase difference of 2π (which means they are doing the same thing at the same time, just one cycle apart). Two points exactly half a wavelength apart have a phase difference of π (which means one is doing the exact opposite of the other, called **antiphase**). Two points exactly a quarter wavelength apart have a phase difference of $\pi/2$ (which means one is at a crest while the other is crossing equilibrium).

Three landmark cases worth committing to memory:

- $\Delta\phi = 0$ (in phase, doing the same thing at the same instant)
- $\Delta\phi = \pi$ (in antiphase, doing the exact opposite)
- $\Delta\phi = 2\pi$ (in phase again, separated by one full wavelength)

Any phase difference between these landmarks describes a state somewhere between full agreement and full disagreement. The mathematics treats them all on the same footing, but physically the in-phase and antiphase cases are the most important, because they are the cases in which the next section's discussion of superposition produces the cleanest results.

Path difference and phase difference: two faces of the same idea

The quantity Δx in the master equation has a special name of its own. It is called the **path difference** between the two points. Path difference is measured in metres; phase difference is measured in radians. They are two ways of describing the same physical separation, related by the bookkeeping rule:

$$\text{phase difference} = \frac{2\pi}{\lambda} \times \text{path difference}$$

Why introduce the same idea twice with two different names? Because in some contexts the natural way to describe a separation is in physical units of length (a path difference of 0.85m), and in other contexts the natural way is in angular units of phase (a phase difference of π radians). The two are interchangeable, and we will move freely between them. In Chapter 8, when we calculate the bright and dark fringes in Young's double-slit experiment, the natural quantity will be a path difference of one wavelength or half a wavelength. The same equation we just wrote will turn each path difference into the corresponding phase difference, and the phase difference will then determine whether two light waves arriving at a given point on a screen reinforce each other or cancel each other out.

Hold on to the equation. It is small, and it is everywhere.

The energy a wave carries: intensity and the inverse-square law

We have spent this chapter so far on the kinematics of a wave: where each particle is, how fast it is moving, what its phase is. We now turn to dynamics. A wave carries energy. The amount of energy depends on the amplitude in a particular way, and when that energy spreads through space from a point source it dilutes in a particular geometric pattern. Two equations cover this whole subsection.

We start with the energy. From Chapter 6, the total mechanical energy of any simple harmonic oscillator is

$$E = \frac{1}{2} m \omega^2 A^2$$

Every particle on a wave is doing simple harmonic motion at angular frequency ω with amplitude A , so every particle carries an energy proportional to A^2 . When we accumulate the energies of all the particles in a unit volume of the medium, the energy density of the wave (energy per cubic metre) is also proportional to A^2 . And when we ask how much of this energy crosses a unit area per unit time, the same proportionality survives: the **intensity** of a wave, written I , is proportional to A^2 :

$$I \propto A^2$$

Intensity has the SI units watts per square metre (W/m^2), the same units as solar flux. It is the density of power flow through space carried by the wave, and the proportionality with A^2 has remarkable physical consequences. Doubling the amplitude of a wave quadruples its intensity. Halving the amplitude quarters the intensity. The square law is unforgiving and is the reason a softly-strummed guitar is, in energy terms, dramatically less loud than a vigorously strummed one, even though the amplitude looks only modestly larger to the eye.

Now consider a **point source** of waves: a small object that radiates energy outward equally in all directions. A small loudspeaker, a small lamp, a small earthquake epicentre. By symmetry the wave fronts emanating from such a source are concentric spheres. If the source emits total power P (in watts), and the medium does not absorb energy as the wave passes through it, then the entire power P must cross every spherical surface centred on the source, however far away. The surface area of a sphere of radius r is $4\pi r^2$, so the intensity at distance r is

$$I = \frac{P}{4\pi r^2}$$

This relationship is called the **inverse-square law** of intensity. The **intensity at distance r** is the source's total power divided by the surface area of a sphere of radius r . Doubling the distance multiplies the area by 4 and divides the intensity by 4. Tripling the distance multiplies the area by 9 and divides the intensity by 9. Move twice as far from a uniform sound source and the loudness, in energy terms, drops to a quarter of what it was.

Now combine the two ideas. Intensity is proportional to A^2 , and intensity falls as $1/r^2$. So A^2 falls as $1/r^2$, which means amplitude itself falls as $1/r$. Doubling the distance halves the amplitude (not quarters it!) but quarters the intensity. The two distance dependencies are different, and the difference is a square root. This is the trap students fall into most often. Memorise the two rules side by side:

$$I \propto \frac{1}{r^2} \quad \text{and} \quad A \propto \frac{1}{r}$$

Or in ratio form, between two different distances r_1 and r_2 from the same source:

$$\frac{I_1}{I_2} = \left(\frac{r_2}{r_1}\right)^2 \quad \text{and} \quad \frac{A_1}{A_2} = \frac{r_2}{r_1}$$

The ratio of intensities is the square of the inverse ratio of distances; the ratio of amplitudes is just the inverse ratio of distances. The third worked example below is built precisely on this distinction.

Three worked examples now. Pen ready

BINDER Example 9

A progressive wave is described by the equation $y = 0.04\sin(200\pi t - (20/17)\pi x)$ m, where t is in seconds and x is in metres. Find the phase difference between (a) the points at $x = 0.25$ m and $x = 1.10$ m, and (b) the points at $x = 0$ and $x = \lambda/4$.

Solution

Comparing the equation with the standard form $y = A\sin(\omega t - kx)$, we read off:

$$\omega = 200\pi \text{ rad/s}, \quad k = \frac{20}{17} \pi \text{ rad/m}$$

(a) The path difference between the two given points is

$$\Delta x = x_2 - x_1 = 1.10 \text{ m} - 0.25 \text{ m} = 0.85 \text{ m}$$

Multiply by k to get the phase difference:

$$\Delta\phi = k\Delta x = \left(\frac{20}{17} \pi \text{ rad/m}\right) \times (0.85 \text{ m}) = \pi \text{ rad}$$

(b) When the path difference is exactly $\lambda/4$, the phase difference is

$$\Delta\phi = \frac{2\pi}{\lambda} \times \frac{\lambda}{4} = \frac{\pi}{2} \text{ rad}$$

The wavelength λ cancels, so the answer does not depend on what λ actually is. Two points separated by a quarter of a wavelength are always $\pi/2$ radians out of phase, on any wave whatsoever.

Making sense of the answer: In part (a), the two specific points $x = 0.25$ m and $x = 1.10$ m are exactly half a wavelength apart on this particular wave. Verify it: the wavelength is $\lambda = 2\pi/k = 2\pi \div ((20/17)\pi) = 17/10 = 1.7$ m, so half a wavelength is 0.85 m, matching Δx exactly. Two points half a wavelength apart are in antiphase, which is what $\Delta\phi = \pi$ says. The cleverness of the question is that the numbers 0.25 , 1.10 , and $20/17$ were chosen so that the fractions cancel without leaving a numerical residue. Examiners do this on purpose. Trust the cancellation; do not try to convert to decimals halfway through. In part (b), the answer falls out without computing λ at all, because the same wavelength appears in numerator and denominator. The lesson: when a path difference is given as a fraction of λ , the corresponding phase difference is the same fraction of 2π . No arithmetic required.

Think like a physicist: Phase difference is a dimensionless angle, but path difference is a length. The conversion between them, $\Delta\phi = (2\pi/\lambda)\Delta x$, is the only equation in this section that you must memorise. Notice its structure: the right-hand side multiplies a length (Δx) by an inverse length ($2\pi/\lambda$ has units of rad/m), and the result is an angle (in radians). Units are doing real work here. If you ever forget the formula, restore it by asking: **how many radians of phase fit into one wavelength?** Answer: 2π . So per metre of separation, the phase advances by $2\pi/\lambda$. Multiply by the actual separation Δx to get the actual phase difference. You have just rederived the formula on the spot, and that is the kind of confidence that survives an examination room.

REAL Example 10

A small wooden buoy on Lake Victoria, used by a Mwanza fisherman to mark the position of his fishing nets, rises and falls in simple harmonic motion with a period of 2.0 s as the lake's surface waves pass through it. A second buoy of identical design floats 5.0 m away in the direction the waves are travelling. The fisherman notices that the second buoy reaches its highest point exactly 0.5 s after the first buoy reaches its own highest point. From this single observation, find (a) the phase difference between the oscillations of the

two buoys, (b) the wave speed and the wavelength of the surface waves, and (c) verify the phase difference from part (a) using the master equation $\Delta\phi = (2\pi/\lambda)\Delta x$ with the wavelength from part (b).

Solution

(a) Both buoys oscillate at the same period $T = 2.0\text{s}$, but the second buoy lags the first by a time $\Delta t = 0.5\text{s}$. A particle's phase advances by 2π in one full period, so in a time Δt it advances by the fraction $\Delta t/T$ of 2π . Therefore the second buoy lags the first in phase by:

$$\Delta\phi = \frac{\Delta t}{T} \times 2\pi = \frac{0.5\text{s}}{2.0\text{s}} \times 2\pi = \frac{\pi}{2} \text{ rad}$$

(b) The wave is moving at a speed such that whatever it does at the first buoy is repeated 0.5s later at the second buoy, 5.0m downstream. So the wave covers 5.0m in 0.5s:

$$v = \frac{\Delta x}{\Delta t} = \frac{5.0\text{m}}{0.5\text{s}} = 10\text{m/s}$$

The wavelength then follows from $\lambda = vT$:

$$\lambda = vT = (10\text{m/s})(2.0\text{s}) = 20\text{m}$$

(c) Substitute the path difference $\Delta x = 5.0\text{m}$ and the wavelength $\lambda = 20\text{m}$ into the master equation:

$$\Delta\phi = \frac{2\pi}{\lambda} \Delta x = \frac{2\pi}{20\text{m}} \times 5.0\text{m} = \frac{\pi}{2} \text{ rad}$$

Same answer. The two routes: one through time delay and period, the other through path difference and wavelength; agree exactly, as they must, because they are describing the same physical lag in two different units.

Making sense of the answer: The relationship between time delay and phase difference is a natural one. A particle's phase advances by 2π in one period T , so in any time Δt it advances by $(\Delta t/T) \cdot 2\pi$. Half a period of delay corresponds to π radians (antiphase). A quarter period corresponds to $\pi/2$ (which is what we found here). A full period corresponds to 2π (back in phase). The two buoys 5m apart on Lake Victoria are reading the same wave at different stages of the same cycle, and 5m happens to be exactly a quarter of the 20m wavelength. The numbers in this problem are chosen so the phase difference works out clean; in any real measurement the fisherman would more likely find a non-textbook fraction.

Think like a physicist: Notice what the fisherman did **not** need to do. He did not measure the wavelength directly (no ruler stretched over the lake's surface). He did not measure the wave speed directly (no tracked crest with a stopwatch). He measured a time delay (0.5s), a separation (5.0m), and his own buoy's bobbing period (2.0s). From those three numbers and the master equation $\Delta\phi = (2\pi/\lambda)\Delta x$, plus the relation $\lambda = vT$, he extracted both the phase difference and every other parameter of the wave. This is the same two-degrees-of-freedom point we made earlier: a wave is fixed by two pieces of information, and once you have any two (here, period and wave speed deduced from the time-and-distance pair), every other parameter follows. The fisherman is doing applied wave physics with a stopwatch and a pole, and the bookkeeping closes without his ever being told the answer.

HOT Example 11

A small loudspeaker radiates sound uniformly in all directions, with a total acoustic power output of 80W. Find (a) the intensity of the sound at a distance of 3.0m from the speaker, (b) the intensity at a distance of 6.0m from the speaker, (c) the ratio of the two intensities, and (d) the ratio of the corresponding amplitudes of the sound waves.

Solution

The speaker radiates uniformly, so the inverse-square law applies. At a distance r from a point source of total power P , the intensity is $I = P/(4\pi r^2)$.

(a) At $r_1 = 3.0\text{m}$:

$$I_1 = \frac{P}{4\pi r_1^2} = \frac{80\text{W}}{4\pi(3.0\text{m})^2} = 0.707 \text{ W/m}^2$$

(b) At $r_2 = 6.0\text{m}$:

$$I_2 = \frac{P}{4\pi r_2^2} = \frac{80\text{W}}{4\pi(6.0\text{m})^2} = 0.177 \text{ W/m}^2$$

(c) The ratio of intensities can be computed from the explicit values, or directly from the inverse-square law:

$$\frac{I_1}{I_2} = \left(\frac{r_2}{r_1}\right)^2 = \left(\frac{6.0\text{m}}{3.0\text{m}}\right)^2 = (2)^2 = 4$$

So the intensity at 3.0m is four times the intensity at 6.0m.

(d) The ratio of amplitudes follows from $I \propto A^2$. If the intensity ratio is 4, then the amplitude ratio is the square root of that:

$$\frac{A_1}{A_2} = \sqrt{\frac{I_1}{I_2}} = \sqrt{4} = 2$$

Equivalently, $A \propto 1/r$, so the amplitude at 3.0m is twice the amplitude at 6.0m.

Making sense of the answer: The intensity at 3m is about 0.7 W/m², which is loud but realistic for a 1m-distance public-address speaker tested in an open field. By 6m the intensity has dropped fourfold to about 0.18 W/m², still loud but noticeably softer. A casual listener would describe the change as a reduction of about half in loudness, which on the decibel scale corresponds to roughly 6 dB, exactly the figure that an inverse-square doubling of distance produces. Everything is consistent.

Think like a physicist: The trap in this example is part (d). A student who has memorised "the intensity ratio is the inverse square of the distance ratio" may, while distracted, also write "the amplitude ratio is the inverse square of the distance ratio". This is wrong by a factor of 2 in the wrong direction. The correct rule is that amplitude scales as $1/r$, not $1/r^2$. The two rules differ by a square root because intensity is proportional to amplitude squared. Doubling the distance quarters the intensity but only halves the amplitude. Memorise the two scalings as a pair, and remember which one is which by the energy bookkeeping: it is energy that spreads over the spherical surface $4\pi r^2$, so it is energy density (intensity) that falls as $1/r^2$, and amplitude (which encodes the displacement, not the energy) follows the square root.

Phase difference, path difference, and intensity. Three new ideas; one master equation $\Delta\phi = (2\pi\lambda)\Delta x$; one inverse-square law $I = P/(4\pi r^2)$; and one square-root subtlety that converts between intensity and amplitude. With these in hand, we are ready to ask the question that has been hovering in the background since the first stone hit Lake Victoria: *when two waves arrive at the same point, what does the medium actually do?* The next section answers that question with a single principle, the principle of superposition, and uses it to derive the constructive and destructive cases that are the bedrock of every interference calculation in physics. Phase difference is what we have just learned to compute. Superposition is what tells us what to do with the answer.

SUPERPOSITION OF WAVES

We have arrived at the most important principle in this chapter, and arguably in the next one as well. Phase difference, the subject of the previous section, gave us a way to compare two waves at two points. The principle of superposition, the subject of this section, tells us what to do with that comparison: *when two or more waves arrive at the same point at the same time, what does the medium actually do?* The answer is so simple that students sometimes mistrust it on first encounter, and so powerful that it underpins every interference, diffraction, and resonance calculation in the rest of the chapter and in the whole of optics in Chapter 8.

Mr. Akilikubwa illustrated the principle on Lake Victoria, by throwing a second stone into the water beside the first one. Where the two ripple patterns met, two of his observations are now ready to be promoted to working principles.

Observation one: the disturbances added at the meeting point. Where two crests crossed, a taller crest briefly stood up; where a crest met a trough, the water momentarily went flat.

Observation two: after the crossing, the two patterns continued outward unchanged. Each ring kept its shape and its speed; neither ring carried away any record of having met the other.

Both observations come from a single physical statement, which we are about to write down formally and use over and over again.

The principle of superposition

The principle states: *When two or more waves pass through the same point in a medium simultaneously, the resultant displacement of the medium at that point is the algebraic sum of the displacements that each wave would produce there individually.* Mathematically, if y_1 is the displacement that the first wave would

produce on its own, y_2 the displacement the second wave would produce on its own, and so on, then the displacement actually produced by all the waves together is:

$$y = y_1 + y_2 + y_3 + \dots$$

That is the entire principle. Three things about it are worth pinning down before we begin to use it.

First, the addition is algebraic. A positive displacement and a negative displacement of the same magnitude cancel; two positive displacements reinforce; two negative displacements also reinforce, in the negative direction. The principle does not say that crests add and troughs subtract; it says that displacements add. The signs sort themselves out automatically.

Second, the principle holds only in linear media; media whose response to a small disturbance is proportional to the size of the disturbance, the way a Hooke's-law spring responds linearly to a small displacement. Almost every wave you will study in this book lives in a linear medium: a stretched string at modest amplitude, a sound wave in air at speech-level intensities, a light wave in vacuum or glass. There are exceptions (for instance the shock wave at the front of a supersonic jet), and for them the simple addition does not work, but those exceptions are not in the syllabus and do not affect any calculation here.

Third, and remarkably, the principle implies that waves do not interact with each other. They only interact with the medium they are passing through. Two waves crossing the same point both deposit their displacements there, but neither wave is altered by the encounter. After the crossing, each wave continues exactly as it would have if the other had never existed. This is why the two ripple patterns on Lake Victoria walked away from one another unchanged: the water at the meeting point was bobbing up and down by the algebraic sum of the two ripples' separate demands, and once the two patterns had passed through, the two demands were no longer being made at the same point, and the water reverted to obeying each ripple independently.

Constructive and destructive interference

Apply the principle now to the simplest non-trivial case: two waves of the same frequency, the same amplitude, and the same direction of travel, arriving at the same point with a definite phase difference $\Delta\phi$. Call them:

$$y_1 = A\sin(\omega t) \quad \text{and} \quad y_2 = A\sin(\omega t + \Delta\phi)$$

Their sum, by superposition, is:

$$y = y_1 + y_2 = A(\sin(\omega t) + \sin(\omega t + \Delta\phi))$$

Apply the sum-to-product identity $\sin \alpha + \sin \beta = 2 \sin((\alpha+\beta)/2) \cos((\alpha-\beta)/2)$ with $\alpha = \omega t$ and $\beta = \omega t + \Delta\phi$. The half-sum is $\omega t + \Delta\phi/2$, and the half-difference is $-\Delta\phi/2$ (whose cosine is the same as $\cos(\Delta\phi/2)$ because cosine is even). The result is

$$y = \left(2A \cos\left(\frac{\Delta\phi}{2}\right) \right) \sin\left(\omega t + \frac{\Delta\phi}{2}\right)$$

This is itself a sinusoidal wave at the same frequency ω . The new amplitude, sitting outside the sine, is the bracket

$$\text{resultant amplitude} = 2A \left| \cos\left(\frac{\Delta\phi}{2}\right) \right|,$$

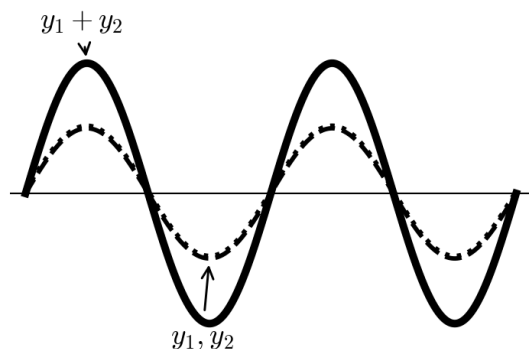
which depends only on $\Delta\phi$. Two limiting cases dominate the rest of this chapter and all of Chapter 8.

Constructive interference: When $\Delta\phi = 0, 2\pi, 4\pi, \dots$ (an even multiple of π), the cosine is ± 1 and the resultant amplitude is $2A$, twice the amplitude of either wave alone. Two waves in phase, doing the same thing at the same time, reinforce each other completely. The medium at the meeting point oscillates with amplitude $2A$.

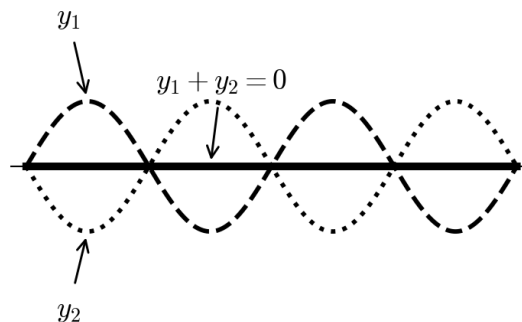
Destructive interference: When $\Delta\phi = \pi, 3\pi, 5\pi, \dots$ (an odd multiple of π), the cosine is 0 and the resultant amplitude is also 0. Two waves in antiphase, doing exact opposites, cancel each other completely. The medium at the meeting point does not oscillate at all.

For unequal amplitudes the same algebra runs through unchanged. The constructive case becomes $A_1 + A_2$ and the destructive case becomes $|A_1 - A_2|$. For equal amplitudes the constructive sum is $2A$ and the destructive sum is 0. The figure summarises both.

(a) constructive interference



(b) destructive interference

**Conditions:**

Constructive: $\Delta\phi = 0, 2\pi, 4\pi, \dots$ resultant amplitude $= A_1 + A_2$

Destructive: $\Delta\phi = \pi, 3\pi, 5\pi, \dots$ resultant amplitude $= |A_1 - A_2|$

(For equal amplitudes, constructive gives $2A$ and destructive gives 0 .)

Figure: Two-wave superposition. (a) Constructive: two in-phase waves (dashed and dotted) of equal amplitude superpose to give a wave of double amplitude (solid). (b) Destructive: two antiphase waves (dashed and dotted) superpose to give zero (solid). The conditions on $\Delta\phi$ for each case are stated below.

What superposition does NOT do

It is worth a single paragraph to stress what the principle of superposition does **not** change.

- It does not destroy waves.
- It does not slow them down.
- It does not deflect them.

Two waves that meet at a point part company afterwards as if the meeting had never happened. This independence of waves is what makes interference patterns stable: *a single point in space, illuminated by two coherent light sources, settles into a definite condition (bright or dark or somewhere between) determined entirely by the phase difference at that point, and stays in that condition as long as the two sources keep their relative phase fixed.* Move the listener (or the screen) and the phase difference changes; move them back and the original condition returns. Superposition is reversible, geometrical, and humane: it adds; it never breaks.

Four worked examples follow.

BINDER Example 12

Two waves arriving at the same point in a medium are described by $y_1 = 5\sin(\omega t + 2\pi/3)$ m and $y_2 = 5\sin(\omega t - \pi/3)$ m. (a) Find the resultant displacement $y = y_1 + y_2$, and state the resultant amplitude. (b) Verify your part-(a) result independently by computing the phase difference $\Delta\phi$ between the two waves and identifying the resulting interference type. (c) The amplitude of the second wave is now changed from 5m to 4m, with everything else (including its phase) unchanged. Find the new resultant amplitude.

Solution

(a) Apply the sum-to-product identity $\sin \alpha + \sin \beta = 2 \sin((\alpha+\beta)/2) \cos((\alpha-\beta)/2)$ with $\alpha = \omega t + 2\pi/3$ and $\beta = \omega t - \pi/3$. The half-sum and half-difference are

$$\frac{\alpha + \beta}{2} = \omega t + \frac{\pi}{6}, \quad \frac{\alpha - \beta}{2} = \frac{\pi}{2}$$

Putting this back into the identity:

$$\begin{aligned}
 y &= 5 \left(\sin \left(\omega t + \frac{2\pi}{3} \right) + \sin \left(\omega t - \frac{\pi}{3} \right) \right) \\
 &= 5 \times 2 \sin \left(\omega t + \frac{\pi}{6} \right) \cos \left(\frac{\pi}{2} \right) = 10 \sin \left(\omega t + \frac{\pi}{6} \right) \times 0 = 0
 \end{aligned}$$

So the resultant displacement at every instant is zero. The resultant amplitude is 0.

(b) Compute the phase difference between the two given waves:

$$\Delta\phi = \left(\omega t + \frac{2\pi}{3} \right) - \left(\omega t - \frac{\pi}{3} \right) = \frac{2\pi}{3} + \frac{\pi}{3} = \pi \text{ rad}$$

A phase difference of π places the two waves in antiphase. The constructive-and-destructive bookkeeping of this section says that two equal-amplitude waves in antiphase produce a resultant amplitude of $|A - A| = 0$. This matches part (a) exactly. The two routes (direct algebraic summation and the phase-difference shortcut) agree, as they must.

(c) The phase difference is unchanged at $\Delta\phi = \pi$ (we changed only the amplitude of the second wave, not its phase). The two waves are still in antiphase, so the resultant amplitude is $|A_1 - A_2|$:

$$\text{resultant amplitude} = (5\text{m} - 4\text{m}) = 1\text{m}$$

With the small amplitude mismatch, the cancellation is no longer perfect. The medium at the meeting point now oscillates with amplitude 1m, one-fifth of the larger of the two incoming waves.

Making sense of the answer: Two non-trivial waves, each of amplitude 5m and oscillating at angular frequency ω , conspire to produce a perfectly motionless medium at the point where they meet. This is the most striking consequence of superposition. Energy is not destroyed (the two waves carry energy past the meeting point unchanged), but at this one point in space the medium is held perfectly still by two competing pulls of exactly equal magnitude and opposite direction. Part (c) then breaks the perfection: a 1m mismatch in amplitude leaves a 1m residual oscillation. Perfect cancellation requires both the antiphase condition **and** an exact amplitude match. The first condition is easy to satisfy in the laboratory; the second is hard.

Think like a physicist: The sum-to-product identity is doing real work in part (a), but you do not actually need it to get the answer. The phase-difference calculation in part (b) tells you $\Delta\phi = \pi$ directly, and from that you know without further algebra that two equal-amplitude waves in antiphase cancel. The full sum-to-product computation is the rigorous way; the phase-difference recognition is the fast way. Practise both. In examination conditions the fast way will save minutes; in physics conditions, the rigorous way will protect you from arithmetic slips. Part (c) shows that the phase-difference shortcut also handles the unequal-amplitude case for free: the destructive-interference rule reads $|A_1 - A_2|$, no extra algebra required. The honest physicist owns both routes and uses each when it earns its keep.

REAL Example 13

A pair of noise-cancelling headphones works by detecting incoming environmental noise with a small microphone, generating an electronic copy of the noise, inverting that copy in phase by π radians, and playing the inverted copy through the speaker into the listener's ear. The incoming noise and the cancelling wave then meet at the eardrum and add by superposition. Suppose the incoming noise on a city bus is well-modelled as a progressive wave of amplitude $A = 0.040\text{mm}$ at the eardrum, and the cancelling wave generated by the headphones has amplitude $A' = 0.035\text{mm}$ with a relative phase shift of exactly π radians.

- Using the principle of superposition, explain why a phase shift of π radians produces silence when the two amplitudes are equal.
- Find the residual amplitude at the eardrum when $A = 0.040\text{mm}$ and $A' = 0.035\text{mm}$.
- Express the residual as a percentage of the incoming amplitude, and convert this to a percentage of the original intensity.

Solution

(a) Let the incoming noise at the eardrum be $y_1 = A\sin(\omega t)$. A wave with the same frequency but phase-shifted by π radians is $y_2 = A'\sin(\omega t + \pi)$. Using the trigonometric identity $\sin(\theta + \pi) = -\sin(\theta)$, the cancelling wave can be rewritten as $y_2 = -A'\sin(\omega t)$. By superposition, the total displacement of the eardrum is:

$$y_{\text{total}} = y_1 + y_2 = A\sin(\omega t) - A'\sin(\omega t) = (A - A')\sin(\omega t)$$

The resultant amplitude is therefore $|A - A'|$. When $A = A'$ the difference is zero and the eardrum experiences no displacement at all, resulting in complete silence.

(b) With the given numbers,

$$\text{residual amplitude} = (0.040\text{mm} - 0.035\text{mm}) = 0.005\text{mm}$$

(c) Express the residual as a percentage of the incoming amplitude:

$$\text{residual fraction of } A = \frac{0.005\text{mm}}{0.040\text{mm}} = \frac{1}{8} = 12.5\%$$

Since intensity is proportional to amplitude squared, the residual intensity is the square of the residual amplitude fraction:

$$\text{residual fraction of } I = (0.125)^2 = 0.0156 = 1.56\%$$

So the headphones reduce the intensity of the incoming noise to about 1.6% of its original level (which corresponds to about 18 decibels of attenuation).

Making sense of the answer: Real noise-cancelling headphones achieve roughly 20 to 30 decibels of attenuation in their target frequency band, so 18 decibels from this calculation is realistic. The lesson buried in the calculation is that perfect noise cancellation requires both a phase shift of exactly π **and** an amplitude match exactly equal to the incoming amplitude. The phase shift on its own is necessary but not sufficient. Engineers building these devices spend most of their effort on the amplitude-matching half of the problem, because the phase inversion is electronically easy but the amplitude match requires real-time measurement and adjustment as the user moves through changing acoustic environments.

Think like a physicist: This example is the simplest non-trivial application of superposition, and it captures the entire architecture of every interference calculation in this chapter and in Chapter 8. Two waves arrive at the same point. Each contributes its own sinusoidal displacement. The resultant is the algebraic sum. When the two waves are in antiphase and have equal amplitudes, the sum is zero. When they are in phase and have equal amplitudes, the sum is double. Everything else, including the loudness pattern in a concert hall and the bright fringes in Young's experiment, is just this same calculation done with different specific phase differences and amplitude ratios. Master this small example and you have the structural insight that drives the rest of the chapter and an entire chapter of optics in Chapter 8.

REAL Example 14

Two identical loudspeakers, each emitting a pure sinusoidal tone of frequency 680Hz, are placed 3.0m apart on a school playing field at Miono Secondary, both connected to the same source so they emit in phase. Kipute walks slowly along a straight line parallel to the line joining the speakers, at a distance of 5.0m from each speaker. As she walks she hears the loudness of the tone alternate between distinct "loud" zones and distinct "quiet" zones.

- Using the principle of superposition, explain qualitatively why the loudness varies as she walks.
- At the midpoint between the two speakers (the point exactly equidistant from both), what does she hear, and why?
- She moves a small distance until the path difference from her position to the two speakers equals $\lambda/2$. What does she hear at that position, and why?

Solution

(a) Each speaker emits a sound wave of identical frequency and identical amplitude. The two waves arrive at the listener's ear at the same instant, but having travelled, in general, two different distances. The path difference Δx (the difference between those two distances) translates into a phase difference $\Delta\phi = (2\pi/\lambda)\Delta x$ at the listener's ear. Where Δx is zero or a whole number of wavelengths, $\Delta\phi$ is an even multiple of π and the two waves arrive in phase: superposition gives constructive interference and the listener hears the tone loudly. Where Δx is half a wavelength or any odd multiple of half a wavelength, $\Delta\phi$ is an odd multiple of π and the two waves arrive in antiphase: superposition gives destructive interference and the listener hears the tone faintly or not at all. As the student walks, her path difference changes continuously, and the loudness alternates between the constructive maxima and the destructive minima.

(b) At the midpoint, the listener is equidistant from both speakers, so the path difference is zero. The two waves arrive in phase, the resultant amplitude is the sum of their individual amplitudes (here $2A$, since they are identical), and the loudness is at its constructive maximum. The midpoint is therefore one of the loudest positions on her path.

(c) When the path difference is $\lambda/2$, the phase difference is:

$$\Delta\phi = \frac{2\pi}{\lambda} \times \frac{\lambda}{2} = \pi \text{ rad}$$

The two waves arrive in antiphase. With equal amplitudes, the resultant amplitude is zero and (in a perfectly absorbing field with no reflections) the listener hears silence. In a real outdoor field the cancellation is not quite perfect, because the two speakers are not point sources and the listener's ear is not a single point, but the loudness drops to a clear minimum that any human ear can pick out as a quiet zone.

Making sense of the answer: This is exactly what happens in any open-air concert at the National Stadium in Dar es Salaam, in any school assembly with two PA speakers running the same microphone, and in every public-address system that has more than one speaker. The room (or field) divides itself into a pattern of loud and quiet zones, set by the geometry of the speakers and the wavelength of the sound. Concert hall designers spend serious engineering effort on placing speakers so that no listener sits permanently in a quiet zone for the dominant frequencies of speech and music.

Think like a physicist: Notice the architecture of the argument. The path-difference equation $\Delta\phi = (2\pi/\lambda)\Delta x$, established in the previous section, is the engine that turns geometry (the listener's position) into phase difference. Then the principle of superposition, established in this section, turns phase difference into amplitude. The two equations chain together to predict, from the speaker spacing, the wavelength, and the listener's position, exactly how loud the tone will be at every point on the field. This same chain of reasoning underwrites the entire physics of interference, in any wave at any wavelength. If you can do the calculation here for sound, you can do it for light in Chapter 8.

HOT Example 15

Two waves of identical amplitude A , identical angular frequency ω , and identical wave number k travel along the same string in opposite directions. The wave moving in the positive x -direction is $y_1 = A\sin(\omega t - kx)$; the wave moving in the negative x -direction is $y_2 = A\sin(\omega t + kx)$. (a) Find the resultant displacement $y = y_1 + y_2$. (b) Identify the type of wave that the resultant describes. (c) Find the displacement at the point $x = \lambda/6$ at the instant t when $\sin(\omega t) = 1$.

Solution

(a) Apply superposition. The resultant is the sum:

$$y = y_1 + y_2 = A\sin(\omega t - kx) + A\sin(\omega t + kx)$$

Use $\sin \alpha + \sin \beta = 2 \sin((\alpha+\beta)/2) \cos((\alpha-\beta)/2)$ with $\alpha = \omega t - kx$ and $\beta = \omega t + kx$. The half-sum is ωt ; the half-difference is $-kx$ (whose cosine is $\cos(kx)$ because cosine is even). The resultant becomes:

$$y = 2A \sin(\omega t) \cos(kx) = (2A \cos kx) \sin \omega t$$

(b) The resultant is not a travelling wave. A travelling wave's argument always contains the combination $\omega t \pm kx$, with t and x bound together inside a single sine. Here the spatial dependence ($\cos kx$) and the temporal dependence ($\sin \omega t$) appear in separate factors. The wave does not advance through the medium. Each particle of the medium oscillates in time at angular frequency ω , with an amplitude $2A \cos(kx)$ that depends only on its location x . This kind of wave is called a **standing wave** (or **stationary wave**). (It will be the subject of the next two sections of this chapter.)

(c) At $x = \lambda/6$, the spatial argument is:

$$kx = \frac{2\pi}{\lambda} \times \frac{\lambda}{6} = \frac{\pi}{3}$$

So $\cos(kx) = \cos(\pi/3) = 1/2$. At the instant when $\sin(\omega t) = 1$, the displacement at $x = \lambda/6$ is

$$y = \left(2A \times \frac{1}{2}\right) \times 1 = A$$

So the particle at $x = \lambda/6$ reaches a peak displacement of A , which is half the maximum standing-wave amplitude of $2A$. It is not at a node (the cosine factor is non-zero there), and it is not at an antinode (the cosine factor is not ± 1 there). It is at an intermediate amplitude.

Making sense of the answer: The algebra has just produced a standing wave out of two travelling waves. Read this carefully: the principle of superposition, the sum-to-product identity, and the simple fact that two waves can travel in opposite directions on the same medium have together delivered the most physically interesting object in the chapter. Every guitar string, every organ pipe, every cantilever bridge swaying in

the wind, every microwave oven cavity, and every electron's standing pattern around a hydrogen nucleus is a special case of this result. The next two sections take this expression apart and use it to derive the harmonics of stringed instruments and the boundary-condition rules that select which standing waves an instrument can support.

Think like a physicist: Notice the structural difference between $(2A \cos kx) \sin \omega t$ and $A \sin(\omega t - kx)$. The second is a travelling wave: t and x appear together inside a single function. The first is a standing wave: t and x appear in separate functions, multiplied together. This separation of variables, in mathematics, is a deep structural fingerprint, and the physics fingerprint matches: the travelling-wave shape moves through the medium without changing, and the standing-wave shape stays in place while changing in amplitude. Two superficially similar equations describe two completely different physical objects. The arithmetic that produced the standing wave from two opposite-direction travelling waves is honest: the sum-to-product identity is the only step, and it is the algebraic shadow of the physical fact that two opposite-going waves on the same line can fold themselves into a single oscillation locked in place. Hold this calculation in your head; the next section's derivation of node and antinode positions falls out of it without further work.

Superposition is now in our toolkit. We know how to add two waves at the same point, what the result looks like for special phase differences (constructive $2A$, destructive 0), and what the result looks like when the two waves are travelling in opposite directions on the same medium (a standing wave). The next section uses superposition for a slightly different special case: two waves of nearly the same frequency, travelling in the same direction. The result there is not a steady interference pattern but a slow rhythmic rise and fall of loudness called beats, and the physics of beats is the foundation of every musician's tuning ear and every stationary-engine vibration calibration.

BEATS

In the previous section we used the principle of superposition on two waves of the same frequency. The two waves either reinforced (constructive) or cancelled (destructive), depending on the fixed phase difference between them, and the result was a steady pattern in space. We turn now to the slightly different and altogether more musical case: two waves of nearly but not quite the same frequency, travelling in the same direction, arriving at the same point. The result is no longer a steady pattern. It is a *periodic rise and fall of loudness* called **beats**, which every guitar tuner, piano technician, and physics student with two slightly mistuned notes has heard. Beats are how musicians tune by ear, how engineers detect tiny frequency differences in machinery, and how a Form Five student can learn, with no equipment more sophisticated than a tuning fork and a guitar string, that two frequencies separated by 1Hz announce themselves as one full cycle of loud-quiet-loud per second.

This section does three things. It describes what beats sound like, derives the beat-frequency formula in two complementary ways (algebraic sum-to-product, and direct cycle-counting), and sets up the worked examples that come at the end. The arithmetic is tiny. The physics is large.

What a beat actually is

Imagine two pure tones being sounded at the same point in your ear. Tone 1 has frequency f_1 ; tone 2 has frequency f_2 ; both have the same amplitude A ; they arrive at the eardrum at the same time, and they superpose by the principle of the previous section. If f_1 and f_2 are very different (say one at 440Hz and another at 880Hz), you hear the two tones as separate notes; the brain has no trouble identifying them. But if f_1 and f_2 are very close (say one at 440Hz and one at 442Hz), the brain cannot resolve them. It interprets the superposition as a single tone whose loudness rises and falls with a slow rhythmic pulse. The **pulse rate**, which is *how many loudness peaks the ear hears per second*, is the **beat frequency**, written f_b . For our 440 and 442 example, the beat frequency turns out to be 2Hz, meaning two loud-quiet-loud cycles per second. The result is the kind of slow throbbing that anyone who has tuned a guitar knows.

Beats are not an illusion. They are the algebraic consequence of the principle of superposition. The next subsection writes them down formally.

The sum-to-product derivation

Let the two arriving waves be

$$y_1 = A \sin(2\pi f_1 t) \quad \text{and} \quad y_2 = A \sin(2\pi f_2 t)$$

By superposition, the total displacement of the eardrum at time t is

$$y = y_1 + y_2 = A(\sin(2\pi f_1 t) + \sin(2\pi f_2 t))$$

Apply the sum-to-product identity $\sin \alpha + \sin \beta = 2 \sin\left(\frac{\alpha + \beta}{2}\right) \cos\left(\frac{\alpha - \beta}{2}\right)$ with $\alpha = 2\pi f_1 t$ and $\beta = 2\pi f_2 t$. The half-sum is $\pi(f_1 + f_2)t$ and the half-difference is $\pi(f_1 - f_2)t$. Substituting:

$$y = 2A \cos(\pi(f_1 - f_2)t) \sin(\pi(f_1 + f_2)t)$$

Read this expression with care, because it has two factors and they are doing two different jobs.

The second factor, $\sin(\pi(f_1 + f_2)t)$, is a fast oscillation at the average frequency $\frac{f_1 + f_2}{2}$. For two close frequencies, this average is essentially the same as either f_1 or f_2 . It is the tone the ear hears.

The first factor, $2A \cos(\pi(f_1 - f_2)t)$, is a slow oscillation at half the difference frequency. It plays the role of a **time-varying amplitude** or **envelope** for the fast oscillation. As time advances, the envelope swings between $+2A$ (when the cosine is $+1$, both waves are momentarily in phase, full constructive interference) and $-2A$ (when the cosine is -1 , both waves are again momentarily in phase but on the negative half of the cycle), passing through zero (when the cosine is zero, the two waves are antiphase, full destructive interference) twice per envelope cycle.

Now the subtle step that produces the beat-frequency formula. The ear does not respond to the signed amplitude of the envelope. It responds to the loudness, which is the squared amplitude. In a single envelope cycle the cosine factor goes through the values $+2A, 0, -2A, 0, +2A$; the squared amplitude therefore goes through $4A^2, 0, 4A^2, 0, 4A^2$. Two loudness peaks per envelope cycle. Therefore the loudness pulse rate is twice the envelope frequency:

$$f_b = 2 \times \frac{|f_1 - f_2|}{2} = |f_1 - f_2|$$

This is the beat-frequency formula. **In words:** the **beat frequency** is the absolute value of the difference between the two superposing frequencies. Two notes one hertz apart give one beat per second. Ten hertz apart give ten beats per second. A hundred hertz apart give a hundred beats per second, which is too fast for the ear to resolve as beats and is heard instead as a third tone (the so-called difference tone of music theory).

The same result by counting cycles

There is a second derivation of the beat-frequency formula that uses no trigonometric identities, only the simple observation that two oscillators of different periods drift in and out of phase at a definite rate. Suppose, without loss of generality, that $f_1 > f_2$ and that the two waves happen to start in phase at $t = 0$. Wave 1 has period $T_1 = \frac{1}{f_1}$; wave 2 has period $T_2 = \frac{1}{f_2}$. Wave 1 oscillates a little faster than wave 2.

After some time T_b , the two waves are once again exactly in phase. During that interval, wave 1 has completed some integer number n of full cycles, and wave 2 has completed exactly $n - 1$ full cycles (it has fallen behind by one cycle). Both completed-cycle counts must equal the same time interval T_b :

$$T_b = nT_1 = (n - 1)T_2$$

Eliminate n by dividing the two equations. From the first, $n = \frac{T_b}{T_1}$; from the second, $n - 1 = \frac{T_b}{T_2}$. Subtracting,

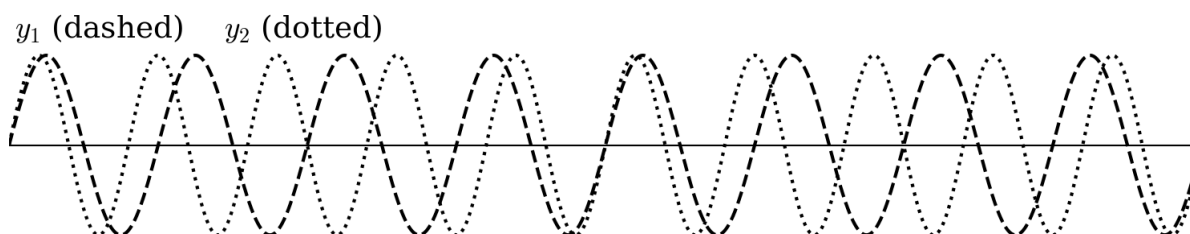
$$1 = T_b \left(\frac{1}{T_1} - \frac{1}{T_2} \right) = T_b(f_1 - f_2)$$

The reciprocal of T_b is the beat frequency:

$$f_b = \frac{1}{T_b} = f_1 - f_2$$

Restoring the absolute-value sign for the case $f_2 > f_1$ gives back the same formula as before. The two derivations agree, as they must: they are different paths through the same physics. The sum-to-product route gives you the envelope and the squared-amplitude reasoning that explains why the loudness pulses at the difference frequency rather than at half of it. The cycle-counting route gives you a quick mental shortcut: the two notes drift past each other at a rate of $|f_1 - f_2|$ cycles per second, and that drift rate **is** the beat rate.

two waves of slightly different frequency



their superposition: a slow loud-quiet-loud envelope (beats)

$$\text{envelope } \pm 2A \cos[\pi(f_1 - f_2)t]$$

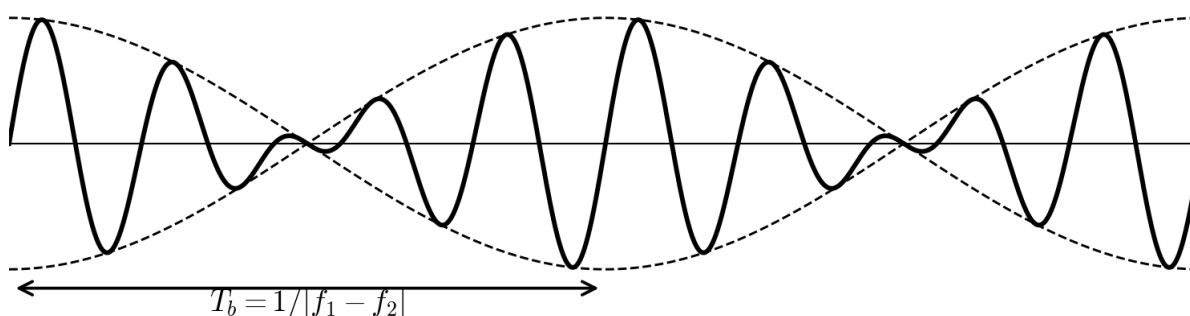


Figure: Beat formation. Top: two waves of slightly different frequency, drawn dashed and dotted. Bottom: their superposition (solid), bounded by the slow envelope $\pm 2A \cos[\pi(f_1 - f_2)t]$ (dashed). One full beat period $T_b = 1/|f_1 - f_2|$ is marked. Within T_b the loudness rises from quiet to loud back through quiet.

Before the ideas drift away, let us anchor them using two worked examples.

BINDER Example 16

Two tuning forks, P and Q, sounded together produce 6 beats per second. Fork P is in unison with a 40cm length of a sonometer wire, and fork Q is in unison with a 32cm length of the same wire under the same tension. Find the frequencies of the two forks.

Solution

The frequency of the fundamental on a sonometer wire under fixed tension is inversely proportional to the vibrating length: a wire half as long has twice the frequency. So with a fixed wire and fixed tension,

$$f \propto \frac{1}{L}$$

Fork P is in unison with the 40cm length and Fork Q with the 32cm length, so the ratio of their frequencies is the inverse ratio of their lengths:

$$\frac{f_P}{f_Q} = \frac{L_Q}{L_P} = \frac{32\text{cm}}{40\text{cm}} = \frac{4}{5}$$

So $f_P = \frac{4}{5}f_Q$. The shorter wire (32cm) sounds at the higher frequency, so $f_Q > f_P$. The beat-frequency formula combined with the unknown frequencies gives the second equation:

$$f_Q - f_P = 6\text{Hz}$$

Substitute $f_P = \frac{4}{5}f_Q$ into this:

$$f_Q - \frac{4}{5}f_Q = \frac{1}{5}f_Q = 6\text{Hz}$$

So $f_Q = 30\text{Hz}$, and $f_P = \frac{4}{5} \times 30\text{Hz} = 24\text{Hz}$.

Making sense of the answer: Two unknown frequencies, two equations: one from the geometry of the sonometer (the ratio of the lengths) and one from the beat measurement (the difference of the frequencies). Two equations in two unknowns close the system without further information. The beat measurement is the part of the problem that closes it; without the 6Hz beats, the geometry alone would only tell us the ratio of the two frequencies, not the absolute values. This is exactly how a piano tuner converts the audible 6 beats per second between a tuning fork and a string into a numerical correction to the string's tension. The beats turn a comparison into a measurement.

Think like a physicist: Notice the structural lesson. Two pieces of physical information were given: the geometric relationship $f \propto 1/L$ (which lives in the chapter on stationary waves) and the beat-frequency relationship $f_b = |f_1 - f_2|$ (which is brand new in this section). Neither one on its own would have solved the problem; both together did. The chapter is now mature enough that examination questions will routinely require you to chain results from two or three of its sections in a single calculation. The discipline is to look at the given information and ask, before you reach for any equation: what physical relationship does this measurement give me, and which other section's relationship will I need to combine with it? Once you have that mental list of two or three equations, the algebra is mechanical.

REAL Example 17

The school music room at Miono Secondary, Saturday afternoon. Kipanga is tuning his guitar's high E string against a 440Hz tuning fork, and Kipute and Mr. Akilikubwa have stopped by to listen. The fork is sounded; Kipanga plucks the string; the room fills with eight clear beats per second. Kipanga reaches for the tuning peg.

Kipanga: "Eight beats. So the string is wrong by eight hertz. I know it is below the fork because the string sounds lower; let me tighten it."

Kipute: "Wait. You don't actually know it is below the fork. The beats only tell you the absolute difference."

Kipanga: "Then how do I know which way to turn the peg?"

Kipute: "You don't know yet. You will know in two seconds. Tighten the peg slightly. If the beats slow, you were below the fork and you are moving toward it. If the beats speed up, you were above and you are moving away from it. Either way, the string answers the question for you within one cycle of the beat."

Mr. Akilikubwa: "This is what physicists call breaking the absolute-value sign. The beat-frequency formula has a $|\cdot|$ in it, and the $|\cdot|$ hides one bit of information, the sign of $f_{\text{string}} - f_{\text{fork}}$. A small perturbation reveals that bit."

Kipanga tightens the peg. The beats slow to six per second, then four, then two, then one, and then, at one particular angle of the peg, the room goes quiet. The string and the fork sing together as if they were the same instrument. Kipanga grins. Kipute nods. Mr. Akilikubwa goes back to his book.

From this dialogue, answer the following:

- From the original eight beats per second, find the two possible frequencies of the guitar string before any tightening took place.
- The beats slow as the string is tightened. From this single observation, determine which of the two possibilities in (a) was the actual starting frequency, and explain your reasoning physically.
- Find the frequency of the string at the moment of silence, and explain physically why "silence" is the equation $f_b = 0$ made audible.

Solution

(a) The beat-frequency formula gives the magnitude of the difference between the string's frequency and the fork's frequency:

$$f_b = |f_{\text{string}} - f_{\text{fork}}| = 8\text{Hz}$$

With $f_{\text{fork}} = 440\text{Hz}$, the absolute-value sign hides whether the string is below or above the fork. Both options satisfy the equation:

$$f_{\text{string}} = 440\text{Hz} \pm 8\text{Hz} = 432\text{Hz or } 448\text{Hz}$$

(b) Tightening a string raises its tension, which raises its frequency. So as Kipanga tightens, f_{string} increases. The observation is that the beats slow as he tightens, which means $|f_{\text{string}} - 440|$ is decreasing, which means f_{string} is approaching 440 from one side. There are two cases. If the string started at 432Hz, raising it moves it toward 440 and the beats decrease, consistent with what Kipanga heard. If the string started at 448Hz, raising it moves it away from 440 and the beats would increase, inconsistent with what he heard. Therefore the string was at 432Hz at the start.

(c) At the moment of silence, the beat frequency is zero:

$$f_b = |f_{\text{string}} - 440\text{Hz}| = 0$$

This requires $f_{\text{string}} = 440\text{Hz}$ exactly.

The string and the fork now sound at identical frequencies; they superpose constructively at every instant; they reinforce each other into a single steady tone with no envelope. The silence Kipanga heard is not absence of sound. It is absence of **variation** in the loudness. The note is still there, twice as loud as either source alone, but it does not pulse, because there is nothing for it to pulse against.

Making sense of the answer: Every musician who has ever tuned by ear has solved exactly this physics problem, with no algebra and no awareness of the equation $f_b = |f_1 - f_2|$. They listen for the silence, and the silence is the equation. Kipate's argument in the dialogue is what physicists call a **perturbation method**: the ambiguity (above or below the fork?) cannot be resolved by listening, but it can be resolved by listening combined with one small action whose direction you control. The same logic underwrites every measurement of a tiny frequency shift, from the Doppler-shift radar speed gun to the gravitational redshift of an atomic clock at the top of a tall building. A small known perturbation breaks the absolute-value sign and tells you the sign of the unknown.

Think like a physicist: Mr. Akilikubwa's phrase "breaking the absolute-value sign" is worth keeping. Many physical measurements give you the magnitude of a signed quantity but not its sign. The beat-frequency formula is the textbook example. So is the Doppler shift formula for an unknown source moving along your line of sight. So is the phase difference at a single point in an interferometer. The measurement device hands you $|x|$; the missing bit of information is the **sign function** ($\text{sgn}(x)$); a deliberate perturbation reveals $\text{sgn}(x)$ within one beat period or one fringe shift. This is a pattern, and once you spot it you find it everywhere in experimental physics. The guitar tuner is doing it without knowing it is doing it. You, who have just read this dialogue, are doing it on purpose.

Beats are now in our toolkit. Two waves of nearly equal frequency, superposing in the same direction, give a slow rhythmic envelope at the difference frequency, and that envelope is what the ear hears as beats. The next section moves to the third and final special case of superposition that this chapter cares about: two waves of **equal** frequency, equal amplitude, but travelling in **opposite** directions on the same medium. Their superposition will not produce a steady interference pattern, and it will not produce beats. It will produce something stranger and more important than either: a wave whose shape is fixed in space and whose amplitude oscillates in time. We have already seen the algebra (it appeared in HOT Example 15 of the previous section as a foreshadowing). The next section gives that algebra a name, a physical interpretation, and a string of consequences that govern every musical instrument in this chapter.

STATIONARY WAVES: FORMATION AND CHARACTERISTICS

In the previous section we superposed two waves of nearly equal frequency travelling in the same direction and got beats. We now do something seemingly minor and physically enormous: we superpose two waves of **exactly equal** frequency, equal amplitude, and equal wavelength, but travelling in **opposite** directions on the same medium. The two waves $y_1 = A\sin(\omega t - kx)$ and $y_2 = A\sin(\omega t + kx)$ superpose, and the sum-to-product identity gives:

$$y = (2A \sin(kx)) \cos(\omega t)$$

(or equivalently, with a different phase choice for the source, $y = (2A \cos(kx)) \sin(\omega t)$). This is not a wave that travels. It is a wave that stays in place while oscillating in amplitude. The shape is fixed in space; only the size of the displacement at each fixed point oscillates in time. This object is called a **stationary**

wave or **standing wave**. This section gives that object its physical interpretation, identifies the special points called **nodes** and **antinodes**, and lays out how a stationary wave differs from the progressive waves we have been studying so far.

The next section will use the same equation to compute tensions in strings. The section after that will derive the harmonic frequencies of every musical instrument made of a string fixed at both ends. Both of those calculations rest on the picture this section draws.

How a stationary wave actually forms in a string

The mathematics asks for two waves of identical frequency and amplitude travelling in opposite directions on the same medium. The physics offers two ways to arrange that.

The first way is to drive the medium from both ends with two synchronised oscillators of identical frequency and identical amplitude, the left source sending a wave rightward and the right source sending a wave leftward. The two waves meet in the middle and superpose. The arrangement is rare in practice (it requires very careful synchronisation of the two sources) and is mostly a theoretical idealisation.

The second way, which is far more common and is how every musical string in the universe actually does this, is to drive the medium from one end and let the wave reflect from the other end. A wave travelling rightward hits a fixed boundary, inverts in phase, and reflects back leftward. The reflected wave has the same frequency, the same amplitude (in an idealised medium with no dissipation), and the same wavelength as the incident wave; it just runs the other way. Incident and reflected superpose. If the source frequency is chosen so the boundary conditions are satisfied (we will sharpen this in the next section), the result is a stable stationary wave that rings indefinitely. Pluck a guitar string and what you have set up, in less than a millisecond, is exactly this two-wave reflection picture, with hundreds of round trips per second between the two fixed ends.

In either physical arrangement, the algebraic result is the same: a function of x multiplied by a function of t , separately. That separation of variables, with no $\omega t \pm kx$ appearing inside a single sine, is the algebraic fingerprint of a stationary wave. It is the physical fingerprint too: time and space are no longer bound together inside the wave's argument.

Nodes and antinodes: what they are and where they live

Read the equation $y = (2A \sin(kx)) \cos(\omega t)$ carefully. The factor $2A \sin(kx)$ depends only on position. It tells us, at each x , how big the oscillation will be. The factor $\cos(\omega t)$ depends only on time. It tells us, at any given instant, what fraction of that maximum oscillation we are currently seeing. Two factors, two completely separate jobs.

Look first at the spatial factor. There are particular points on the string where $\sin(kx)$ is exactly zero. At those points, the displacement y is zero for all time, regardless of what the cosine factor is doing. These are the points where the string never moves. They are called **nodes** and we write them with the letter **N**. From $\sin(kx) = 0$, the nodes occur at $kx = 0, \pi, 2\pi, 3\pi, \dots$, that is, at:

$$x = 0, \frac{\lambda}{2}, \lambda, \frac{3\lambda}{2}, 2\lambda, \dots$$

Adjacent nodes are therefore $\frac{\lambda}{2}$ apart. Half a wavelength, not a whole one. **Memorise this.** *The space between two consecutive points where the string never moves is half the wavelength of the constituent travelling waves, not the full wavelength.*

Now look at the points where $\sin(kx)$ reaches its maximum magnitude of ± 1 . At those points, the spatial factor multiplies the cosine by the largest possible amount; the displacement at those points oscillates between $+2A$ and $-2A$ as the cosine cycles. These are the points of maximum oscillation. They are called **antinodes** and we write them with the letter **A**. From $\sin(kx) = \pm 1$, the antinodes occur at $kx = \pi/2, 3\pi/2, 5\pi/2, \dots$, that is, at:

$$x = \frac{\lambda}{4}, \frac{3\lambda}{4}, \frac{5\lambda}{4}, \dots$$

Adjacent antinodes, like adjacent nodes, are $\frac{\lambda}{2}$ apart. The distance from a node to its nearest antinode is:

$$\frac{\lambda}{4}$$

a quarter of a wavelength. The N–A–N–A–N... pattern marches along the string with this fixed quarter-wavelength spacing, and the pattern is fixed: the same x position is a node now, ten seconds from now, and ten years from now, as long as the wave persists.

One important physical consequence of the node geometry. At a node, the medium never moves; therefore no energy is transported across a node, because energy transport requires both a force and a velocity, and a particle at a node has zero velocity at every instant. Energy in a stationary wave is stored locally, oscillating back and forth between kinetic (when the segment is moving past equilibrium) and potential (when the segment is at maximum displacement). It does not travel along the string the way energy in a progressive wave does. This is the most important physical difference between a stationary wave and a progressive wave, and it is the first item on the list that comes next.

How a stationary wave differs from a progressive wave

The algebra of a stationary wave is $y = (2A \sin(kx)) \cos(\omega t)$. The algebra of a progressive wave is $y = A \sin(\omega t - kx)$. Five physical differences fall out of comparing those two equations, and each difference is a consequence of the algebra rather than something to be memorised in isolation. Read each as: see what the equation does, and read off the physics.

1. Energy

The progressive wave's argument $\omega t - kx$ advances steadily as the disturbance moves rightward, and energy is carried along with it; a wave train passing through a fixed point delivers a steady stream of energy per second. The stationary wave has no such directed transport. Its argument never combines t and x inside a single sine; the time and space dependences are factored apart. There is no propagation direction, and there is no net energy transport. Energy is stored, locally, between consecutive nodes.

2. Amplitude

In the progressive wave, every particle of the medium oscillates with the same amplitude A . The amplitude is constant in space. In the stationary wave, the amplitude depends on position: it is $|2A \sin(kx)|$, which varies from zero (at the nodes) to $2A$ (at the antinodes). Two particles in different parts of the same stationary wave oscillate at different sizes, and a third particle exactly at a node does not oscillate at all.

3. Phase

In the progressive wave, the phase of a particle depends continuously on its position; two particles at different x values are at different stages of the same cycle, and the phase difference between them grows linearly with their separation. In the stationary wave, all the particles between two adjacent nodes share the same phase: they all rise together to maximum displacement, all fall back through equilibrium together, all reach minimum together. The two segments on either side of a node are antiphase: while one segment is up, the other is down. There are exactly two phases in the entire stationary wave (separated by π), and every particle is in one of them.

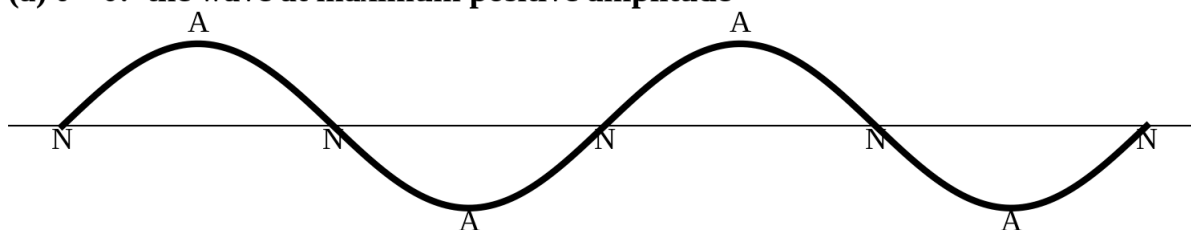
4. Wavelength

In the progressive wave, the distance between consecutive crests, or between any two consecutive points doing the same thing, is the wavelength λ . In the stationary wave, the distance between consecutive nodes is $\frac{\lambda}{2}$, half the wavelength of the constituent travelling waves. A common student mistake is to read the visible spatial period of the standing pattern (one node-to-node distance) as the wavelength of the wave; it is not. The wavelength is twice that distance.

5. Waveform

In the progressive wave, the entire pattern of crests and troughs slides bodily along the medium at speed $v = \frac{\omega}{k}$. In the stationary wave, no part of the pattern moves. The shape of the string at three different times in the figure (a) $t = 0$, (b) $t = \frac{T}{4}$, and (c) $t = \frac{T}{2}$ looks the same in horizontal location but different in vertical scale: the nodes are always at the same x values, the antinodes are always at the same x values, and only the displacement at each fixed x rises and falls in time.

(a) $t = 0$: the wave at maximum positive amplitude



(b) $t = T/4$: the wave is momentarily flat (every particle at zero)



(c) $t = T/2$: the wave at maximum negative amplitude (mirror of (a))

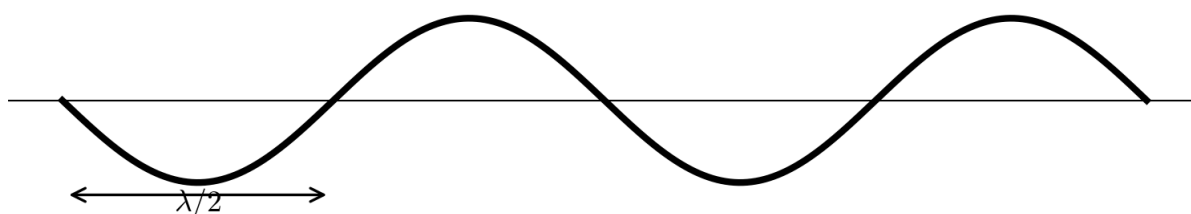


Figure: A stationary wave at three time instants. (a) At $t = 0$ the wave is at maximum positive displacement; nodes (N) and antinodes (A) are labelled. (b) At $t = T/4$ every particle is momentarily at zero. (c) At $t = T/2$ the wave is at maximum negative displacement, the mirror of (a). The N and A positions are the same in all three panels: only the size of the displacement changes with time. The distance between two adjacent nodes is $\lambda/2$.

With nodes and antinodes now simmering nicely on the string, let us serve them through two worked examples and enjoy the flavour of physics in action.

BINDER Example 18

A stationary wave is observed on a stretched string. Adjacent nodes on the string are 0.30m apart. A particle at the centre of one of the antinodes is observed to oscillate at frequency 50Hz, and the maximum displacement at that antinode is 4.0cm. Find: (a) the wavelength λ of the constituent travelling waves whose superposition produced this stationary wave, (b) the distance from any node to its nearest antinode, (c) the wave speed of the constituent travelling waves, and (d) the amplitude A of each of the two constituent travelling waves.

Solution

(a) The distance between adjacent nodes on a stationary wave equals $\frac{\lambda}{2}$. So:

$$\frac{\lambda}{2} = 0.30\text{m}, \quad \text{and} \quad \lambda = 2 \times 0.30\text{m} = 0.60\text{m}$$

(b) The distance from any node to its nearest antinode is half the node-to-node distance, equivalently a quarter of the wavelength:

$$\text{node} - \text{to} - \text{antinode distance} = \frac{\lambda}{4} = \frac{0.60\text{m}}{4} = 0.15\text{m}$$

(c) The wave speed comes from the standard relation $v = f\lambda$:

$$v = f\lambda = (50\text{Hz})(0.60\text{m}) = 30\text{m/s}$$

(d) The maximum displacement at any antinode of the stationary wave is $2A$, where A is the amplitude of each of the two constituent travelling waves. So

$$2A = 4.0\text{cm}, \quad \text{and} \quad A = 2.0\text{cm}$$

Each of the two travelling waves whose superposition produced this stationary wave has an amplitude of 2.0cm. Together, when their crests align, they produce the 4.0cm maximum at the antinode.

Making sense of the answer: All four numbers fall out of the stationary wave's geometry. The 0.30m node-to-node distance gives λ by doubling. The 0.15m node-to-antinode distance follows from λ by halving. The 30m/s wave speed follows from $v = f\lambda$. And the 2.0cm constituent amplitude follows from the antinode being twice as tall as a single travelling wave's crest. None of these calculations required the equation $y = (2A \sin(kx)) \cos(\omega t)$ to be written down explicitly; the geometric facts of a stationary wave (adjacent nodes are $\frac{\lambda}{2}$ apart, antinodes oscillate at amplitude $2A$) were enough. The next section will use the equation directly to compute tensions; this section is the geometric warm-up.

Think like a physicist: Notice the structural lesson. A stationary wave is a wave whose properties have collapsed into spatial and temporal fingerprints that can be read separately. Geometry alone (without dynamics) gives you wavelength, node positions, antinode positions, and the constituent-wave amplitude. Dynamics enters through $v = f\lambda$ once a frequency is specified. The split is exactly the split between the two factors of the equation: $2A \sin(kx)$ lives in space and is set by the geometry of the boundary; $\cos(\omega t)$ lives in time and is set by whoever is driving the source. This factorisation is what makes a stationary wave such a useful structure for analysis: each factor can be examined and measured on its own, without the other.

REAL Example 19

The physics laboratory at Miono Secondary, Wednesday afternoon. Mr. Akilikubwa has clamped a horizontal string of length 1.2m at both ends and connected one end to an electromagnetic vibrator. He has carefully tuned the driver frequency to one particular value at which the string settles into a single clean standing pattern: the two ends are perfectly still, the middle of the string sweeps up and down with a visible amplitude, and the rest of the string falls smoothly between those extremes. Kipanga and Kipute are watching.

Kipanga: "The ends are not moving at all. The middle is moving a lot. Why does the string know where the ends are?"

Kipute: "It does not know anything. The ends are clamped. A clamped point has zero displacement always, and that is the same condition as a node. The string has no choice: at every instant the ends are at zero, so the standing pattern that lives on it must have nodes there. The middle is wherever the equation puts the next antinode after a node at one end."

Kipanga: "And the smooth shape between?"

Kipute: "Is $2A \sin(kx)$, with k set so that $\sin(kL) = 0$ too, since the far end is also a node. The boundary at one end is a forced condition the wave cannot escape. The boundary at the other end is the same condition, applied at $x = L$. Both ends being nodes selects which standing patterns can live on this string, and the simplest one is what you are looking at."

Mr. Akilikubwa: "Boundary conditions are the third partner in every standing-wave problem in this chapter. The wave equation provides the family of possible solutions; the boundary conditions select which member of the family actually shows up. Without boundary conditions a string would have no harmonics, no music, and no physics worth a Form Five examination paper."

Use the demonstration to answer:

- The string is 1.2m long, fixed at both ends, and is observed to vibrate with one antinode at its midpoint and no other antinodes. Find the wavelength of the standing wave.
- Mr. Akilikubwa's driver is set to 50Hz. Find the speed of the transverse waves on the string.
- Mr. Akilikubwa now doubles the driver frequency to 100Hz. The string's tension and linear mass density are unchanged, so the wave speed is unchanged. Find the new wavelength of the standing wave on the string and the locations along the string of all nodes and antinodes in the new pattern.

Solution

- The two ends at $x = 0$ and $x = L = 1.2\text{m}$ are nodes. The midpoint at $x = \frac{L}{2} = 0.6\text{m}$ is an antinode. The distance from a node to the nearest antinode is $\frac{\lambda}{4}$, so:

$$\frac{\lambda}{4} = 0.6\text{m}, \text{ and } \lambda = 2.4\text{m}$$

(b) From $v = f\lambda$:

$$v = (50\text{Hz})(2.4\text{m}) = 120\text{m/s}$$

(c) At 100Hz with $v = 120\text{m/s}$ unchanged, the new wavelength is

$$\lambda' = \frac{v}{f'} = \frac{120\text{m/s}}{100\text{Hz}} = 1.2\text{m}$$

The boundary condition still requires nodes at both ends. Adjacent nodes are $\frac{\lambda'}{2} = 0.6\text{m}$ apart. So in a 1.2m string we fit exactly two such intervals, which means three nodes in total: at $x = 0$, $x = 0.6\text{m}$, and $x = 1.2\text{m}$. The antinodes lie midway between consecutive nodes, so they are at $x = 0.3\text{m}$ and $x = 0.9\text{m}$. The pattern is N A N A N, with two antinodes instead of the previous one.

Making sense of the answer: Doubling the driver frequency, with the wave speed unchanged, halved the wavelength and doubled the number of antinodes that fit between the two fixed ends. The string now hosts a more complex standing pattern, but the rule that selects which patterns can live there is unchanged: the two ends are forced to be nodes, and any wavelength compatible with that constraint is permitted. The smallest permitted wavelength fits one antinode (part a). The next smallest fits two antinodes (part c). The pattern continues: the next mode would fit three antinodes at $\lambda'' = \frac{2L}{3} = 0.8\text{m}$, at frequency 150Hz; the mode after that fits four, at 200Hz; and so on. We will systematise this counting in the next sections of the chapter.

Think like a physicist: The deep lesson is Mr. Akilikubwa's. The wave equation has a family of solutions, all of the form $y = (2A \sin(kx)) \cos(\omega t)$, parameterised by which k you pick. Without boundary conditions, every k would be allowed and the string could ring at any frequency. The boundary conditions select the discrete set of allowed k values for which the chosen ends are nodes. That selection is what turns a continuum of possibilities into a discrete spectrum of harmonics, and that discrete spectrum is what gives a guitar a recognisable tone, a zeze a recognisable tone, and your own larynx a definite vowel rather than a continuous noise. Boundary conditions, applied to the wave equation, are the reason music exists.

The picture is now in place. Stationary waves form when two opposite-direction travelling waves of equal frequency and amplitude superpose on the same medium; their displacement is $y = (2A \sin(kx)) \cos(\omega t)$, with nodes spaced $\frac{\lambda}{2}$ apart and antinodes a quarter of a wavelength from each node. Energy is stored, not transported. Five differences from progressive waves emerge as direct consequences of the equation. The next section makes that equation the protagonist: it derives both the cosine form and the sine form, explains which form a particular boundary condition selects, and shows how to extract the wavelength, frequency, wave speed, and tension of a stretched string directly from a given stationary wave equation. The arithmetic ahead is short. The bookkeeping is exact.

THE STATIONARY WAVE EQUATION

The previous section drew the picture. Two travelling waves of equal frequency, equal amplitude, and equal wavelength pass through one another in opposite directions on the same medium, and the result is a wave whose shape stays put while every point oscillates in time. Nodes are the points that never move. Antinodes are the points that move by twice the constituent amplitude. The picture is complete; the equation is now ready to do work.

This section makes the equation the protagonist. It derives the equation algebraically from the superposition of two oppositely directed travelling waves, presents the two equivalent forms a stationary wave can take, shows how to read all the wave parameters off a given equation, and culminates in the calculation of greatest practical importance: extracting the tension of a stretched string directly from its stationary wave equation. The arithmetic ahead is straightforward. What matters is the order of the steps and what each step extracts.

Deriving the equation by superposition

Consider two transverse travelling waves of identical amplitude A , identical angular frequency ω , and identical wavenumber k , but moving in opposite directions along the same string. Take the first wave to be moving in the $+x$ direction and the second in the $-x$ direction:

$$y_1 = A \sin(\omega t - kx), \quad y_2 = A \sin(\omega t + kx)$$

By the principle of superposition the resultant displacement is $y = y_1 + y_2$, that is:

$$y = A \sin(\omega t - kx) + A \sin(\omega t + kx)$$

The two sines on the right have the same amplitude and add to a product by the standard sum-to-product identity:

$$\sin \alpha + \sin \beta = 2 \sin\left(\frac{\alpha + \beta}{2}\right) \cos\left(\frac{\alpha - \beta}{2}\right)$$

With $\alpha = \omega t - kx$ and $\beta = \omega t + kx$, the half-sum is $\frac{\alpha + \beta}{2} = \omega t$ and the half-difference is $\frac{\alpha - \beta}{2} = -kx$. Substitute these gives:

$$y = 2A \sin(\omega t) \cos(-kx)$$

The cosine is an even function, so $\cos(-kx) = \cos(kx)$. The result is the equation we will live with for the rest of the chapter:

$$y = (2A \cos(kx)) \sin(\omega t)$$

Two factors. The factor $2A \cos(kx)$ depends only on position and not on time. The factor $\sin(\omega t)$ depends only on time and not on position. The product separates space from time completely, and that separation is the algebraic fingerprint of a stationary wave. No $\omega t - kx$ combination appears anywhere; nothing travels.

Two equivalent forms: which one shows up depends on the boundary

The derivation above started with sources whose phases were chosen so that at $t = 0$ both waves were in their pure $A \sin$ form. That phase choice produced the cosine spatial factor $2A \cos(kx)$ and the sine temporal factor $\sin(\omega t)$. A different phase choice for the same physical setup produces the equally valid form:

$$y = (2A \sin(kx)) \cos(\omega t)$$

Both forms describe stationary waves. They differ only in where the nodes sit. In the cosine spatial form, $\cos(kx)$ is zero where $kx = \frac{\pi}{2}, \frac{3\pi}{2}, \frac{5\pi}{2}, \dots$, that is at $x = \frac{\lambda}{4}, \frac{3\lambda}{4}, \frac{5\lambda}{4}, \dots$. The **point $x = 0$ is an antinode** in this form. In the sine spatial form, $\sin(kx)$ is zero at $x = 0, \frac{\lambda}{2}, \lambda, \frac{3\lambda}{2}, \dots$, and the **point $x = 0$ is itself a node**.

Which form actually shows up on a real string is decided by the physical boundary at $x = 0$. A string clamped at $x = 0$ cannot move there, so its displacement is zero at that point at every instant. The only form that satisfies this is the sine spatial form, $y = (2A \sin(kx)) \cos(\omega t)$, which gives $y(0, t) = 0$ identically. A string with a free end at $x = 0$ (one end loose, the other clamped, as in some pipe-end analogues we will meet in Chapter 9) would by contrast select the cosine spatial form, which puts an antinode at the free end. The mathematics offers both forms; the boundary condition picks the one that fits.

Reading the parameters off a given equation

Whenever a stationary wave equation is given to us, the procedure for extracting the wave parameters is mechanical. Suppose we are handed:

$$y(x, t) = (2A \sin(kx)) \cos(\omega t)$$

with specific numbers in place of $2A$, k , and ω . The four parameters fall out in a fixed order:

- 1) **Read the wavenumber k off the spatial argument and convert to wavelength.** The number multiplying x inside the trigonometric function is k , in radians per metre. The wavelength is:

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

- 2) **Read the angular frequency ω off the temporal argument and convert to frequency.** The number multiplying t inside the trig function is ω , in radians per second. The frequency is:

$$f = \frac{\omega}{2\pi}$$

- 3) **Compute the wave speed of the constituent travelling waves.** Either $v = f\lambda$ or, more directly,

$$v = \frac{\omega}{k}$$

- 4) **Read the constituent amplitude A off the leading factor.** The maximum amplitude at any antinode is the leading factor $2A$, so the amplitude of each of the two travelling waves whose superposition produced this stationary wave is half of that, which is A .

Finding the tension of a stretched string from the equation

In Chapter 6 (simple harmonic motion of a stretched string), we established that the speed of a transverse wave on a stretched string of linear mass density μ under tension T is:

$$v = \sqrt{\frac{T}{\mu}}$$

Square both sides:

$$v^2 = \frac{T}{\mu}, \quad \text{so} \quad T = \mu v^2$$

This is the closing step of the whole procedure. Once we have read k and ω off the equation and computed $v = \omega/k$, the tension follows from $T = \mu v^2$ provided the linear mass density μ is known. If only the total mass m and the length L of the string are given, the linear mass density is $\mu = \frac{m}{L}$, and the calculation continues as before.

That is the whole arithmetic of this section, end to end:

$$\text{equation} \rightarrow k \rightarrow \lambda \rightarrow \omega \rightarrow f \rightarrow v = \frac{\omega}{k} \rightarrow T = \mu v^2$$

Read the wavenumber, read the angular frequency, divide them to get the wave speed, multiply the speed squared by the linear mass density. Six lines, one tension. The worked examples below make the procedure automatic.

So before the spatial and temporal factors of the equation tangle any further, let us straighten them out using a few worked examples.

BINDER Example 20

A stationary wave on a long stretched string has the equation

$$y(x, t) = 6\cos(3x)\sin(4t) \quad \text{cm},$$

with x measured in metres and t in seconds. (a) State the type of wave the equation describes and give the algebraic feature of the equation that confirms your identification. (b) Find the wavelength, frequency, wave speed of the constituent travelling waves, and the amplitude of each of those constituent travelling waves. (c) Find the positions of the first four nodes of the stationary wave on the positive x axis. (d) Write down the equations of the two component progressive waves whose superposition produced this stationary wave.

Solution

(a) **Stationary wave.** The equation has the form $y = (2A \cos(kx)) \sin(\omega t)$. The position factor $\cos(3x)$ and the time factor $\sin(4t)$ are multiplied together rather than appearing as a single combined argument $\omega t \pm kx$. This product structure is the algebraic fingerprint of a **stationary wave**. Equivalently, points x where $\cos(3x) = 0$ remain at zero displacement for all t : these are the nodes, and their fixed positions in space confirm that nothing is travelling.

(b) Read the parameters off the equation in the standard order.

Wavenumber: $k = 3 \text{ rad/m}$, so:

$$\lambda = \frac{2\pi}{k} = \frac{2\pi}{3} \text{ m} \approx 2.09 \text{ m}$$

Angular frequency: $\omega = 4 \text{ rad/s}$, so:

$$f = \frac{\omega}{2\pi} = \frac{4}{2\pi} \text{ Hz} = \frac{2}{\pi} \text{ Hz} \approx 0.637 \text{ Hz}.$$

Wave speed:

$$v = \frac{\omega}{k} = \frac{4}{3} \text{ m/s} \approx 1.33 \text{ m/s}.$$

Constituent amplitude: the leading factor is $2A = 6.0 \text{ cm}$, so:

$$A = 3.0 \text{ cm}.$$

(c) The nodes are the points where the position factor vanishes:

$$\cos(3x) = 0 \quad \Rightarrow \quad 3x = \frac{\pi}{2}, \frac{3\pi}{2}, \frac{5\pi}{2}, \frac{7\pi}{2}, \dots$$

Solve for x and keep the first four positive values:

$$x = \frac{\pi}{6} \text{ m}, \frac{\pi}{2} \text{ m}, \frac{5\pi}{6} \text{ m}, \frac{7\pi}{6} \text{ m} \quad \approx 0.524 \text{ m}, 1.571 \text{ m}, 2.618 \text{ m}, 3.665 \text{ m}.$$

Adjacent nodes are $\frac{\pi}{3} \text{ m} \approx 1.047 \text{ m}$ apart. This is exactly $\frac{\lambda}{2}$, as it must be for a stationary wave.

(d) The two component progressive waves are obtained by undoing the sum-to-product identity. Both travel with amplitude $A = 3.0 \text{ cm}$, angular frequency $\omega = 4 \text{ rad/s}$, and wavenumber $k = 3 \text{ rad/m}$, in opposite directions:

$$y_1 = 3\sin(4t - 3x) \text{ cm}, \quad y_2 = 3\sin(4t + 3x) \text{ cm}.$$

Their sum is exactly the given stationary wave equation, as the derivation at the start of the section confirms.

Making sense of the answer: Every parameter of the stationary wave was extracted by inspection of two coefficients, k and ω . The amplitude required reading the leading factor and dividing by two. The node positions required nothing but solving $\cos(kx) = 0$. The component travelling waves required only writing back what the sum-to-product identity had hidden. The lesson is that a stationary wave equation hides nothing once you know which coefficient lives in space and which lives in time. The factor $\cos(3x)$ is the spatial fingerprint, locking the node positions in place at $x = \frac{\pi}{6} \text{ m}, \frac{\pi}{2} \text{ m}, \dots$. The factor $\sin(4t)$ is the temporal fingerprint, telling every non-node point of the string to oscillate at $\omega = 4 \text{ rad/s}$. The two fingerprints share no variable, and that is what lets us read them independently.

Think like a physicist: Note carefully what the cosine spatial form is telling us about the boundary at $x = 0$. At $x = 0$ the equation gives $y(0, t) = 6.0 \sin(4t) \text{ cm}$: the point at the origin oscillates with the maximum amplitude. The origin of this string is an antinode, not a node. If the string were clamped at the origin we would not be looking at this equation; we would be looking at the sine spatial form. The form of the equation, in other words, is itself a piece of physical information about how the string is supported at $x = 0$. Whenever you meet a stationary wave equation, the very first reflex should be to ask which form it is in and what that form is telling you about the boundary. The arithmetic comes second; the boundary lives in the equation itself.

REAL Example 21

On the Kigamboni Bridge in Dar es Salaam, an engineer has fitted a strain-gauge sensor to one of the steel stay cables. After the morning rush of cargo lorries crosses the deck, the sensor records a transverse vibration on a particular cable described by:

$$y(x, t) = 0.8 \cos\left(\frac{\pi x}{6}\right) \sin(40\pi t) \quad \text{cm},$$

with x measured along the cable in metres from the deck-end anchor and t in seconds. The cable's linear mass density is $\mu = 12 \text{ kg/m}$ and its total free length between the two anchors is $L = 18 \text{ m}$.

- Identify the type of wave the equation describes and state two algebraic features of the equation that confirm your identification.
- Find the wavelength, frequency, and wave speed of the transverse vibration on the cable.
- Find the tension in the cable.
- The cable is 18 m long. How many half-wavelengths fit between the two anchors? Examine whether the equation's prediction at $x = 0$ is consistent with that end being a rigid anchor, and explain what the engineer should conclude.

Solution

(a) The equation has the form $y = (2A \cos(kx)) \sin(\omega t)$. Two features confirm that this is a **stationary wave**.

First, the position factor and the time factor appear as a product, not as a single combined argument $\omega t \pm kx$: there is no travelling phase.

Second, the equation predicts fixed points (nodes) where $\cos\left(\frac{\pi x}{6}\right) = 0$, which remain at zero displacement for all t : such fixed nodes cannot exist on a progressive wave.

(b) Read the parameters off in the usual order.

Wavenumber: $k = \frac{\pi}{6} \text{ rad/m}$, so:

$$\lambda = \frac{2\pi}{k} = \frac{2\pi}{\pi/6} \text{ m} = 12 \text{ m}$$

Angular frequency: $\omega = 40\pi \text{ rad/s}$, so:

$$f = \frac{\omega}{2\pi} = \frac{40\pi}{2\pi} \text{ Hz} = 20 \text{ Hz}.$$

Wave speed:

$$v = \frac{\omega}{k} = \frac{40\pi}{\pi/6} \text{ m/s} = 240 \text{ m/s}$$

(c) The cable obeys $v = \sqrt{\frac{T}{\mu}}$, so $T = \mu v^2$:

$$T = (12 \text{ kg/m})(240 \text{ m/s}^2)^2 = 6.91 \times 10^5 \text{ N}$$

(d) Compare the cable length to the wavelength:

$$\frac{L}{\lambda/2} = \frac{18 \text{ m}}{6 \text{ m}} = 3.$$

Three half-wavelengths fit exactly between the two anchors.

Now examine the boundary at $x = 0$. The equation gives $y(0, t) = 0.8 \cos(0) \sin(40\pi t) = 0.8 \sin(40\pi t) \text{ cm}$: the point at the deck-end anchor oscillates with the full antinode amplitude of 0.8 cm. But a properly anchored stay cable should have zero displacement at its anchor: that point is fixed to the deck and cannot move. The equation predicts an antinode at the anchor, while the physical setup demands a node there. The two are inconsistent.

The engineer should conclude that the deck-end anchor is not behaving as a rigid fix during the recorded vibration. There is play at that anchor, perhaps a worn bushing or a loosened socket allowing the cable end to move freely. The cable's vibration mode has therefore become close to that of a free-fixed system rather than the design-intended fixed-fixed system, which is a structural concern requiring inspection of the deck-end anchor before the next rush of traffic.

Making sense of the answer: The arithmetic on this cable runs in five lines: $k \rightarrow \lambda$, $\omega \rightarrow f$, $v = \omega/k$, $T = \mu v^2$. Each line takes one substitution. The numerical answer $T = 6.91 \times 10^5 \text{ N}$ is large, but it is the correct order of magnitude for a major stay cable carrying a span of bridge deck. Where the worked example earns its keep is part (d). The equation form has built-in information about the boundaries, and reading that information off the equation is part of what makes the calculation engineering rather than algebra. An antinode at $x = 0$ instead of a node is not a mathematical curiosity; it is a measurement that asks for action, because rigid anchors and antinodes are not compatible. A bridge engineer who does not check this part is just doing arithmetic.

Think like a physicist: There is a deeper lesson hiding in this example. The equation form (sine versus cosine in the spatial factor) is itself a measurement. The numerical coefficients k , ω , and the leading amplitude tell us the wavelength, frequency, speed, and tension. But which trigonometric function appears in the spatial factor tells us the boundary condition that the wave is satisfying at $x = 0$. A stationary wave on a fixed-end cable should be in the sine spatial form; the cosine spatial form means the cable is behaving as if it had a free end at $x = 0$. When the form contradicts the design, the design is wrong, or the cable is wrong, or the anchor is wrong. The equation cannot lie, because the equation came from the sensor; the

only thing left to question is the cable's claimed boundary condition. This is how a physicist or an engineer reads a stationary wave equation: the parameters are arithmetic, but the form is diagnostic.

HOT Example 22

A stretched string of length $L = 1.5\text{m}$ and mass $m = 3.0 \times 10^{-2}\text{ kg}$ carries a stationary wave described by:

$$y(x, t) = 0.06 \sin\left(\frac{2\pi x}{3}\right) \cos(120\pi t) \quad \text{m,}$$

with x in metres and t in seconds. (a) Find the wavelength, frequency, and wave speed of the transverse vibration on the string. (b) Calculate the tension in the string. (c) Find which harmonic of the string is excited.

Solution

(a) Wavenumber: $k = \frac{2\pi}{3}\text{ rad/m}$, so

$$\lambda = \frac{2\pi}{k} = \frac{2\pi}{2\pi/3}\text{ m} = 3\text{m}.$$

Angular frequency: $\omega = 120\pi\text{ rad/s}$, so

$$f = \frac{\omega}{2\pi} = \frac{120\pi}{2\pi}\text{ Hz} = 60\text{Hz}.$$

Wave speed:

$$v = \frac{\omega}{k} = \frac{120\pi}{2\pi/3}\text{ m/s} = 180\text{m/s}.$$

(b) The linear mass density of the string is

$$\mu = \frac{m}{L} = \frac{3.0 \times 10^{-2}\text{ kg}}{1.5\text{m}} = 0.020\text{ kg/m}.$$

From $v = \sqrt{\frac{T}{\mu}}$ we have $T = \mu v^2$:

$$T = (0.020\text{ kg/m})(180\text{ m/s})^2 = 648\text{N}.$$

(c) The string is fixed at both ends, so both ends must be nodes. The sine spatial form $\sin\left(\frac{2\pi x}{3}\right)$ puts a node at $x = 0$. At $x = L = 1.5\text{m}$:

$$\sin\left(\frac{2\pi \times 1.5}{3}\right) = \sin(\pi) = 0,$$

so the far end is also a node, exactly as required. The number of half-wavelengths between the two ends is

$$n = \frac{L}{\lambda/2} = \frac{1.5\text{m}}{1.5\text{m}} = 1,$$

so the string is vibrating in its **first harmonic**, that is, the fundamental mode.

Making sense of the answer: The chain $k \rightarrow \lambda \rightarrow \omega \rightarrow f \rightarrow v = \omega/k \rightarrow T = \mu v^2$ runs the question from start to finish in six lines, and every line is a single substitution. The wave speed $v = 180\text{m/s}$ is well within the range expected for a stretched musical string under tension. The tension $T = 648\text{N}$ corresponds to about 66 kg of pull on a 1.5m string of mass 30g, large enough that the string is plausibly a steel piano-style wire rather than a guitar nylon. The fundamental wavelength on a 1.5m fixed-fixed string is $\lambda = 2L = 3\text{m}$, which matches the value the equation gave us; the equation and the boundary conditions agree on the mode. That self-consistency check is what tells us we have read the equation correctly: a fixed-fixed string of length $L = 1.5\text{m}$ cannot host a wavelength of 3m in any mode other than the first harmonic.

Think like a physicist: This is exactly how a stationary wave equation wants to be read. It hands you k and ω ; you produce λ , f , and v by elementary division; the string's mass and length give you μ ; the formula

$T = \mu v^2$ delivers the tension. The harmonic is identified by counting half-wavelengths against the string length. Six steps. No memorisation, no tricks, no hidden constants. The reason this question is HOT is not that any single step is hard but that all six must be performed in order, with the right physical relation invoked at each step. Mastering this sequence by repetition is what we have done in this section, and what every well-set examination paper will continue to test.

The equation $y = (2A \sin(kx)) \cos(\omega t)$ is now machinery. We have derived it by superposition; we have shown that the cosine and sine spatial forms are equivalent up to a phase, with the boundary condition selecting between them; we have read off the wavelength, frequency, wave speed, and constituent amplitude from a given equation in a fixed order; and we have closed the loop on stretched strings by computing the tension from the equation and the linear mass density. The next section takes the same equation, applies the boundary condition that the two ends of a string are fixed, and discovers the harmonic spectrum that gives every stringed musical instrument its identity. The arithmetic that follows will rest squarely on the procedure made automatic in this section.

VIBRATIONS OF STRINGS: HARMONICS AND THE SONOMETER

The previous section showed how to read every parameter of a stationary wave off its equation, and how to compute the tension of a stretched string from those parameters. We turn now to the question that turned the stretched string into the foundation of musical instruments and acoustic measurement: *what frequencies, exactly, can a string fixed at both ends be made to vibrate at?* The answer is one of the most elegant in classical physics. Out of a continuum of mathematical possibilities, the boundary condition that the two ends are fixed selects only a discrete sequence of frequencies. Each frequency is an integer multiple of a single fundamental, and the spacing between successive allowed frequencies is the fundamental itself.

Three short calculations make this discrete spectrum explicit. From the spectrum follow three universal laws of stringed instruments, expressing how the fundamental depends on length, tension, and linear mass density. Every guitar, *kora*, *zeze*, violin, and piano on Earth obeys those three laws to the letter; the same laws govern the sonometer in the laboratory. Once the laws are written down, the rest of the section is exercise: given any three of the four quantities (length, tension, linear mass density, harmonic frequency), the fourth follows from a single line of arithmetic.

From the boundary condition to the harmonic spectrum

Take a uniform stretched string of length L , clamped rigidly at both ends. The two fixed ends are at $x = 0$ and $x = L$. From the previous section, the stationary wave equation that has a node at $x = 0$ is the sine spatial form:

$$y(x, t) = (2A \sin(kx)) \cos(\omega t)$$

The boundary condition at the far end requires that $y(L, t) = 0$ for all t , which forces the spatial factor to vanish at $x = L$:

$$\sin(kL) = 0.$$

The sine function is zero only at integer multiples of π , so the boundary condition selects

$$kL = n\pi, \quad n = 1, 2, 3, \dots$$

The integer n is called the **harmonic number**. The case $n = 0$ is excluded because it gives $k = 0$ and a zero wave. Negative integers give the same physical solutions as their positive counterparts, and so are not listed separately. The wavenumber that is allowed in the n -th harmonic is therefore

$$k_n = \frac{n\pi}{L}, \quad \text{so} \quad \lambda_n = \frac{2\pi}{k_n} = \frac{2L}{n} \quad \text{or} \quad n = \frac{L}{\left(\frac{\lambda_n}{2}\right)}$$

This is the central geometric statement of the section: *in the n -th harmonic, exactly n half-wavelengths fit between the two clamped ends.* The fundamental ($n = 1$) fits one half-wavelength, so $\lambda_1 = 2L$. The second harmonic ($n = 2$) fits two half-wavelengths, so $\lambda_2 = L$. The third harmonic ($n = 3$) fits three, so $\lambda_3 = 2L/3$. And so on. The following figure, shows the first three.

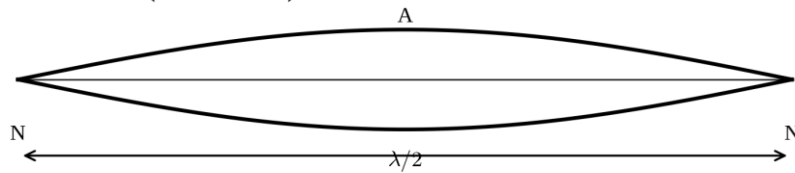
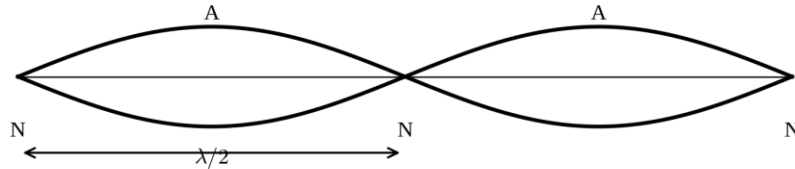
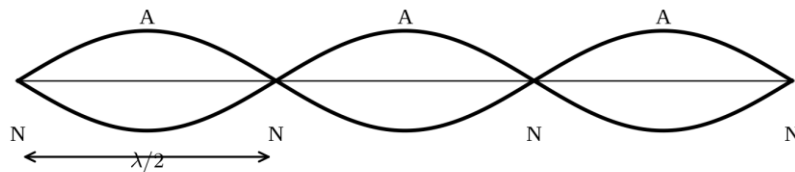
(a) 1st harmonic (fundamental): $L = \lambda/2$ (b) 2nd harmonic: $L = \lambda$ (c) 3rd harmonic: $L = 3\lambda/2$ 

Figure: The first three harmonic patterns on a string fixed at both ends. (a) 1st harmonic (fundamental): one half-wavelength fits between the two clamps, so $L = \lambda/2$. There are two nodes (one at each end) and one antinode (in the middle). (b) 2nd harmonic: two half-wavelengths fit, so $L = \lambda$. Three nodes, two antinodes. (c) 3rd harmonic: three half-wavelengths fit, so $L = 3\lambda/2$. Four nodes, three antinodes. In every harmonic the distance between adjacent nodes is $\lambda/2$.

The wave speed on the string is fixed by its tension and linear mass density alone, so it is the same for every harmonic. That common wave speed times the wavelength gives the frequency:

$$f_n = \frac{v}{\lambda_n} = \frac{nv}{2L}, \quad n = 1, 2, 3, \dots$$

Substituting $v = \sqrt{\frac{T}{\mu}}$ gives the master formula of the section:

$$f_n = \frac{n}{2L} \sqrt{\frac{T}{\mu}}, \quad n = 1, 2, 3, \dots$$

In particular, the fundamental frequency is:

$$f_1 = \frac{1}{2L} \sqrt{\frac{T}{\mu}}$$

and every higher harmonic is an integer multiple of it: $f_n = nf_1$. This integer-multiple structure is what makes a vibrating string a **harmonic** oscillator and not merely a periodic one. It is what allows a single bowed or plucked string to sound a definite musical pitch rather than an indefinite noise: the human ear hears the fundamental, while the ear and the brain together register the higher harmonics as the timbre that distinguishes a violin from a guitar from a zeze.

The three laws of string vibration

Read off the master formula three times, each time holding two of the three quantities (L , T , μ) fixed and varying the third.

Law 1 (length): with T and μ fixed, $f_1 \propto \frac{1}{L}$. A shorter string vibrates at a higher fundamental frequency. Halving the length doubles the fundamental: this is the rule a guitarist uses every time they press a string halfway along its length to play one octave higher.

Law 2 (tension): with L and μ fixed, $f_1 \propto \sqrt{T}$. A tighter string vibrates at a higher fundamental frequency, but only as the square root of the tension. Doubling the tension multiplies the fundamental by $\sqrt{2} \approx 1.41$, not by 2. This is why tuning a guitar requires fine adjustment: a small change in tension produces a small change in frequency.

Law 3 (linear mass density): with L and T fixed, $f_1 \propto \frac{1}{\sqrt{\mu}}$. A lighter string vibrates at a higher fundamental frequency. This is why the high strings of a guitar or piano are made of thin wire: lower μ gives higher frequency at the same length and tension.

All three laws together explain every stringed instrument ever made. To raise the pitch of a string at fixed material and fixed tension, shorten it (a violinist's left hand). To raise the pitch at fixed length and fixed material, tighten it (a tuning peg). To make a string play a low note without making it too long or too floppy, increase its mass per unit length (the wound copper of a piano bass string).

Adjacent resonances and the fundamental

A useful corollary of $f_n = nf_1$ concerns the spacing between consecutive harmonics. The difference between any two adjacent allowed frequencies on the same string is:

$$f_{n+1} - f_n = (n + 1)f_1 - nf_1 = f_1.$$

That is, **the difference between any two adjacent resonant frequencies of a fixed-fixed string is the fundamental frequency**. This corollary is exceedingly useful in the laboratory. If a sonometer string is observed to resonate at two adjacent driver frequencies, their numerical difference is the fundamental of the string, immediately, without any reference to length, tension, or mass density.

With the harmonic spectrum in hand and the three laws written down, let us put them through their paces using a few worked examples.

BINDER Example 23

A guitar string of length $L = 0.65\text{m}$ has total mass $m = 3.6\text{g}$ and is stretched to a tension $T = 480\text{N}$. Find: (a) its linear mass density μ , (b) the speed of transverse waves on the string, and (c) the fundamental frequency of the string.

Solution

(a) The linear mass density is mass per unit length:

$$\mu = \frac{m}{L} = \frac{3.6 \times 10^{-3} \text{ kg}}{0.65\text{m}} = 5.5 \times 10^{-3} \text{ kg/m}$$

(b) The speed of transverse waves follows from the string formula:

$$v = \sqrt{\frac{T}{\mu}} = \sqrt{\frac{480 \text{ N}}{5.5 \times 10^{-3} \text{ kg/m}}} \approx 294 \text{ m/s}$$

(c) The fundamental frequency of a string fixed at both ends is:

$$f_1 = \frac{v}{2L} = \frac{294 \text{ m/s}}{2 \times 0.65\text{m}} \approx 226 \text{ Hz}$$

Making sense of the answer: Three substitutions, three answers: μ from mass and length; v from μ and T ; f_1 from v and L . The fundamental frequency 226Hz lies just above A3 (220Hz), the open A pitch on the violin and the twelfth-fret pitch of a guitar's A string. The number is squarely in the audible musical range, and it confirms that a string of mass 3.6g, length 65cm, and tension 480N produces a tone in the A region of the musical scale. Each of the three steps used exactly one and only one physical relation: the definition of linear mass density, the wave speed on a stretched string, and the fundamental frequency of a fixed-fixed string. There is no shortcut that skips a step; there is no step that needs more than one relation.

Think like a physicist: Notice the order in which the quantities appeared. The mass and length came together to give μ , a property of the string itself that does not depend on how the string is being used. The tension entered next, combining with μ to give v , a property of waves on this particular string at this

particular tension. The length then re-entered (this time as a boundary), combining with v to select f_1 , a property of the entire fixed-fixed system. The same string at the same tension would have a different fundamental if it were clamped at different points; the boundary condition is what turns a wave speed into a frequency. Hold this distinction. It is the difference between a property of the medium (which v is) and a property of the medium-plus-boundary system (which f_1 is).

BINDER Example 24

A sonometer wire of length $L = 60\text{cm}$ and total mass $m = 12\text{g}$ is observed to carry a transverse wave of wavelength 4.0cm at a frequency of 720Hz . Find the tension in the wire.

Solution

Linear mass density:

$$\mu = \frac{m}{L} = \frac{0.012\text{ kg}}{0.60\text{m}} = 0.020\text{kg/m}$$

Wave speed from $v = f\lambda$:

$$v = (720\text{ Hz}) \times (0.040\text{ m}) = 28.8\text{m/s}$$

Tension from $T = \mu v^2$:

$$T = (0.020\text{ kg/m}) \times (28.8\text{ m/s})^2 = 16.6\text{N}$$

Making sense of the answer: The wave speed and tension followed from $v = f\lambda$ and $T = \mu v^2$ without any reference to which harmonic the wire happens to be vibrating in. The given wavelength 4.0cm fits $L/(\lambda/2) = 0.60\text{m}/0.020\text{m} = 30$ times into the wire, so the wire is in fact in its 30th harmonic, a high mode but a perfectly legitimate one. The tension calculation does not care about that, because tension and wave speed are properties of the wire itself, independent of which standing pattern is excited on it. A sonometer in this configuration is a routine laboratory setup: a tension of about 17N applied to a wire of mass density 20g/m supports waves at modest speed, and any of the discrete frequencies $f_n = nf_1$ can be excited by an external driver tuned to that frequency.

Think like a physicist: This is the **inverse** of the previous example. There, the unknowns were derived properties (wave speed, fundamental frequency) computed from the given physical setup (length, mass, tension). Here, one of those derived properties (the tension) is the unknown, and the input is a measurement of an actual standing pattern on the wire (its wavelength and frequency). The wave speed serves as the bridge: $v = f\lambda$ reads it off the observed pattern, and $T = \mu v^2$ converts it back to the underlying tension.

BINDER Example 25

A steel piano wire of length $L = 0.50\text{m}$ has mass $m = 4.0\text{g}$ and is held under tension $T = 600\text{N}$. (a) Find the fundamental frequency of the wire. (b) Find the frequency of the third harmonic. (c) Find the highest harmonic of the wire whose frequency is at most 12000Hz .

Solution

Common preliminaries:

$$\mu = \frac{m}{L} = \frac{0.0040\text{ kg}}{0.50\text{m}} = 8.0 \times 10^{-3}\text{ kg/m},$$

$$v = \sqrt{\frac{T}{\mu}} = \sqrt{\frac{600\text{ N}}{8.0 \times 10^{-3}\text{ kg/m}}} = 273.9\text{ m/s}.$$

(a) Fundamental frequency:

$$f_1 = \frac{v}{2L} = \frac{273.9\text{ m/s}}{2 \times 0.50\text{ m}} \approx 274\text{Hz}$$

(b) Third harmonic: $f_3 = 3f_1$:

$$f_3 = 3 \times 274\text{ Hz} = 822\text{Hz}$$

(c) The highest harmonic with frequency at most 12000Hz is the largest integer n satisfying,

$$nf_1 \leq 12000\text{Hz}$$

$$n \leq \frac{12000\text{ Hz}}{274\text{ Hz}} = 43.8, \quad \text{so} \quad n_{\text{max}} = 43$$

The highest harmonic is 43

Checking: The frequency of that harmonic is:

$$f_{43} = 43 \times 274\text{Hz} = 11782\text{Hz} \text{ which is very close but less than the } 12000\text{Hz} \text{ limit as required.}$$

Making sense of the answer: The wire's fundamental at about 274Hz lies near C₄# (277Hz), just above middle C. The third harmonic at about 822Hz is in the higher range a piano routinely sounds. The 43rd harmonic at 11.8kHz is a thin, high overtone, and audible to most young ears (the upper limit of human hearing is roughly 16-20kHz). A piano wire does not actually sound at any single one of these frequencies in isolation; struck, it sounds in a superposition of many of them, with the relative amplitudes determining the timbre that distinguishes a piano from a harpsichord and a harpsichord from a guitar.

Think like a physicist: Part (c) is a discrete counting problem dressed up as physics: divide a frequency limit by a fundamental, take the floor of the quotient, and that integer is the answer. Notice that $n \times f_1$ is a strictly increasing arithmetic sequence with common difference f_1 , so once the fundamental is known the location of every harmonic on the frequency axis is fixed. The 43rd harmonic at 11.8kHz illustrates how rapidly a string can populate the upper end of the audible range: a wire whose fundamental is below middle C can produce overtones extending nearly to the edge of human hearing.

BINDER Example 26

A string of length $L = 3.20\text{m}$, fixed at both ends, is observed to resonate at two adjacent driver frequencies, 170Hz and 204Hz, with no resonance for any driver frequency in between. Find: (a) the fundamental frequency of the string, and (b) the speed of transverse waves on the string.

Solution

(a) For a string fixed at both ends, the allowed frequencies are $f_n = nf_1$, and adjacent allowed frequencies differ by exactly the fundamental:

$$f_{n+1} - f_n = f_1$$

Therefore

$$f_1 = 204\text{Hz} - 170\text{Hz} = 34\text{Hz}$$

(b) The wave speed follows from $f_1 = v/(2L)$, solved for v :

$$v = 2L \times f_1 = 2 \times 3.20\text{m} \times 34\text{Hz} = 217.6\text{m/s}$$

Making sense of the answer: The fundamental frequency and the wave speed both fell out of one number, the difference between two consecutive resonances, and the string length, with no reference to mass or tension. This is what makes the adjacent-resonance method useful in the laboratory: f_1 can be extracted from the resonance pattern alone, without weighing the string or measuring its tension. Once f_1 is known, the wave speed follows immediately from the length, and (if desired) the tension follows from $T = \mu v^2$ once the linear mass density is independently known. Note also which harmonic numbers correspond to the observed resonances: $170\text{Hz} = 5 \times 34\text{Hz}$ and $204\text{Hz} = 6 \times 34\text{Hz}$, so the observed pair is the 5th and 6th harmonics of the string.

Think like a physicist: The three quantities L , T , and μ combine in the master formula to determine the fundamental frequency, but a single resonance frequency alone does not tell us which harmonic it is. Two adjacent resonances do: their difference is the fundamental, and the ratio of either to that fundamental is the harmonic number. This is the structural lesson. The harmonic spectrum is fixed by L , T , and μ ; reading f_1 off the spectrum requires either knowing those three quantities directly, or sampling the spectrum at two adjacent points. The latter is often the easier route in practice.

REAL Example 27

The physics laboratory at Miono Secondary, Friday afternoon. Mr. Akilikubwa walks in carrying two stringed instruments: a *kora*, the West African harp with twenty-one strings stretched between a long neck and a calabash gourd, and a *zeze*, a traditional Tanzanian one-string fiddle with a goat-skin resonator. He sets both on his desk and writes three relations on the board:

$$f_1 \propto \frac{1}{L}, \quad f_1 \propto \sqrt{T}, \quad f_1 \propto \frac{1}{\sqrt{\mu}}$$

Mr. Akilikubwa: "Every stringed instrument that has ever moved a human being to tears (*kora*, *zeze*, violin, piano, guitar, harp, oud) was operating under exactly those three relationships. The one-string instrument I have here in my left hand and the twenty-one string instrument I have in my right hand differ in nothing else of physical importance. Their twenty-two strings, taken together, are twenty-two boundary-value problems.

Each one obeys the same master formula. Each one selects a different fundamental by setting L , T , and μ to particular values that the maker chose deliberately."

Kipute: "Sir, when a guitarist presses a string halfway along its length and plucks it, what does the string do?"

Kipanga (without prompting): "It plays the second harmonic. The pressed-down finger forces a node at the midpoint, so the effective length of the freely vibrating segment is $L/2$. By the first law, f_1 of the shortened segment is twice the original f_1 . The pitch jumps up one octave."

Mr. Akilikubwa: "He is learning. Notice what just happened, Kipute. Kipanga did not invoke the second harmonic of the original string. He invoked the first harmonic of a new, shorter string. The two are mathematically the same frequency in this particular case, but conceptually they are not the same thing. The original string's second harmonic is what you would hear if you plucked the original string and forced a node at the midpoint by holding it there gently with one finger while plucking with another. The shortened string's fundamental is what you hear when you press the finger down hard enough to clamp the string against the fret, terminating the original string and creating a new, shorter, fixed-fixed string. The numerical answer is the same; the physical setup is different."

Kipute: "And the kora's twenty-one strings, sir? They are all the same length on the instrument; the maker did not change L ."

Mr. Akilikubwa: "Correct. The kora maker varies μ and T , not L . A traditional kora uses fishing-line strings of varying gauge: the bass strings are thicker (higher μ), the treble strings are thinner (lower μ), and each one is tuned to its desired fundamental by adjusting the leather tuning rings on the neck (varying T). The third law, $f_1 \propto 1/\sqrt{\mu}$, is what makes a low note possible on a string of fixed length without making the string impossibly long: increase μ instead. The piano maker does the same: the wound copper of a bass string raises μ enough that a low fundamental can be produced from a string short enough to fit inside the piano case."

Use the dialogue to answer the following questions:

- The fundamental frequency of a particular kora bass string is 110Hz. The string is replaced by another string of the same length and tension but with twice the linear mass density. Find the new fundamental frequency.
- The high-pitched first string of a guitar plays $f_1 = 330\text{Hz}$ when open. The guitarist presses the string against the seventh fret, which shortens the vibrating segment to $2/3$ of its original length. Assuming T and μ are unchanged, find the new fundamental frequency.

Solution

- (a) By the third law, $f_1 \propto 1/\sqrt{\mu}$. Doubling μ multiplies f_1 by $1/\sqrt{2}$:

$$f_{1,\text{new}} = \frac{110\text{Hz}}{\sqrt{2}} = 77.8\text{Hz}$$

- (b) By the first law, $f_1 \propto 1/L$. Reducing L to $2L/3$ multiplies f_1 by $3/2$:

$$f_{1,\text{new}} = \frac{3}{2} \times 330\text{Hz} = 495\text{Hz}$$

Making sense of the answer: Both calculations used a single law each, applied to a single change of variable, with the other two variables held fixed. (a) Doubling the linear mass density lowers the fundamental by the factor $1/\sqrt{2}$, from 110Hz to about 78Hz: a kora bass string given a heavier replacement plays a lower note. (b) Pressing the high E string at the seventh fret shortens the vibrating segment to two-thirds of its open length, and the inverse-length law gives a fundamental of 495Hz: a perfect fifth above the open string's 330Hz. The factor 3:2 between the two frequencies is the famous **just intonation** interval that the ancient Pythagoreans first identified, and the same factor that every guitar's seventh fret produces. The maths and the music agree.

Think like a physicist: Mr. Akilikubwa's emphasis is the right one. The three laws each hold **only when the other two variables are fixed**. The kora maker who replaces a string with a heavier one of the same gauge but a different alloy is changing μ alone (Law 3). The guitarist who presses a fret is changing L alone (Law 1). The tuner who turns a peg is changing T alone (Law 2). Multiple simultaneous changes do not obey any

one of the three simple proportionalities; they obey the master formula $f_1 = \frac{1}{2L} \sqrt{\frac{T}{\mu}}$ and require the master formula's full machinery. The three laws are the partial derivatives of that master formula in disguise; treating them as more than that is a recipe for arithmetic that does not match what the instrument actually does.

HOT Example 28

A stretched string clamped at both ends is sounded together with a tuning fork while the string is held under tension $T_1 = 245\text{N}$. The combination produces beats at 15 beats per second. The tension on the string is then increased smoothly to $T_2 = 320\text{N}$, and at that tension the string and the fork are heard to be in unison: no beats. Find the frequency of the tuning fork.

Solution

Let f_1 denote the string's fundamental frequency at T_1 , f_2 its fundamental frequency at T_2 , and F the (constant) frequency of the fork. By the second law of string vibration, with L and μ fixed, $f_1 \propto \sqrt{T}$, so:

$$\frac{f_2}{f_1} = \sqrt{\frac{T_2}{T_1}} = \sqrt{\frac{320\text{N}}{245\text{N}}} = \sqrt{\frac{64}{49}} = \frac{8}{7}$$

At T_2 the string is in unison with the fork, so $f_2 = F$.

At T_1 the beat condition is $|f_1 - F| = 15\text{Hz}$.

The string's frequency rose with the tension to meet the fork's frequency, so f_1 was below F , and we drop the absolute value:

$$F - f_1 = 15\text{Hz}$$

Substitute $F = f_2 = \frac{8}{7}f_1$:

$$\frac{8}{7}f_1 - f_1 = 15\text{ Hz}, \quad \text{so} \quad f_1 = 105\text{ Hz}.$$

Therefore:

$$F = \frac{8}{7} \times 105\text{Hz} = 120\text{Hz}$$

The frequency of the tuning fork is 120Hz.

Making sense of the answer: The fork frequency is 120Hz, near the lower end of the bass-baritone vocal range. The string's fundamental rose from 105Hz at T_1 to 120Hz at T_2 , a 15Hz increase that matches the original 15Hz beat frequency, as it must. The arithmetic worked out cleanly because the two tensions were in the ratio $320:245 = 64:49$, whose square root is the rational number $8:7$. Real-world tunings rarely give such clean ratios; we have chosen them here so that the structure of the calculation, and not the arithmetic, is the lesson.

Think like a physicist: The crucial physical step in this example is not the use of $f_1 \propto \sqrt{T}$, that is straightforward. The crucial step is deciding which side of the unison the string was on at T_1 . The beat frequency 15Hz tells us how far apart the string and the fork were in absolute value, but not which one was higher. The physical clue is the direction of the tension change. Tension was **increased** between T_1 and T_2 , which **raised** the string's frequency. If raising the string's frequency brought it into unison with the fork, the fork must have been **higher** than the string at the start. That single sentence resolved the ambiguity in the absolute value and let us drop it. Always look for the physical clue that selects between the algebraic possibilities; algebra alone will not do it.

HOT Example 29

A monochord (a single-string instrument with a movable bridge for changing the vibrating length) is to be built so that its open-string fundamental is $f_1 = 220\text{Hz}$. The wire available for the string has linear mass density $\mu = 2.5 \times 10^{-3}\text{ kg/m}$ and a maximum safe operating tension $T_{\text{max}} = 450\text{N}$, above which the wire is at risk of snapping.

- At a tension $T = 300\text{N}$, find the speed of transverse waves on the wire.
- For $T = 300\text{N}$, find the string length L that gives the required fundamental frequency.
- For a vibrating length $L = 40\text{cm}$, find the tension that gives the required fundamental frequency.

- (d) Is the choice $L = 40\text{cm}$ within the safe operating range of the wire?
 (e) What is the longest vibrating length that gives the required fundamental, given the maximum safe tension T_{max} ?

Solution

- (a) Wave speed at $T = 300\text{N}$:

$$v = \sqrt{\frac{T}{\mu}} = \sqrt{\frac{300\text{ N}}{2.5 \times 10^{-3}\text{ kg/m}}} \approx 346\text{ m/s}$$

- (b) Length for $f_1 = 220\text{Hz}$ at $v \approx 346\text{ m/s}$:

$$L = \frac{v}{2f_1} = \frac{346\text{ m/s}}{2 \times 220\text{ Hz}} = 0.787\text{m}, \quad (\text{about } 78.7\text{cm})$$

- (c) Tension for $L = 0.40\text{ m}$ at $f_1 = 220\text{ Hz}$. First the wave speed:

$$v = 2L \times f_1 = 2 \times 0.40\text{m} \times 220\text{Hz} = 176\text{m/s}$$

Then the tension from $T = \mu v^2$:

$$T = (2.5 \times 10^{-3}\text{ kg/m}) \times (176\text{ m/s})^2 = 77.4\text{ N}$$

- (d) Compare the required tension with the maximum:

$$77.4\text{ N} < T_{\text{max}} (450\text{ N})$$

Hence the choice is well within the safe operating range.

The tension required by $L = 40\text{ cm}$ uses only about 17% of the wire's strength.

- (e) Maximum wave speed at the maximum tension:

$$v_{\text{max}} = \sqrt{\frac{T_{\text{max}}}{\mu}} = \sqrt{\frac{450\text{N}}{2.5 \times 10^{-3}\text{ kg/m}}} \approx 424\text{ m/s}.$$

Maximum vibrating length at T_{max} for the same fundamental:

$$L_{\text{max}} = \frac{v_{\text{max}}}{2f_1} = \frac{424\text{ m/s}}{2 \times 220\text{ Hz}} = 0.964\text{m}, \quad (\text{about } 96.4\text{cm})$$

Making sense of the answer:

The design space is bounded. To play a 220Hz fundamental on this wire, the vibrating length must be between zero and about 96cm; below 96cm the required tension is below the wire's safe limit, and above 96cm the required tension exceeds it and the wire is at risk of snapping. The numerical answers show how design freedom is exercised: a short string (40cm) needs a low tension (77N); a medium string (79cm) needs a medium tension (300N); the longest possible string (96cm) needs the maximum safe tension (450N). The designer chooses based on instrument size, the desired feel of the strings under the player's fingers, and the safety margin.

Think like a physicist: The fundamental frequency depends on three quantities (L, T, μ) combined into one master formula. With μ fixed by the choice of wire and f_1 fixed by the design specification, the designer has two free parameters left, L and T , and one constraint. That leaves a one-dimensional family of solutions: every (L, T) pair on a particular curve in the (L, T) plane is a valid design. The maximum-tension constraint truncates that curve at one end; the minimum-tension constraint (the wire must still be taut) would truncate it at the other. The designer picks any point on the surviving curve and builds an instrument. This is how engineering design actually proceeds: the physics gives the family of possibilities; the engineer picks one based on practical considerations the physics does not encode.

The harmonic spectrum of a fixed-fixed string is a discrete sequence $f_n = nf_1$, with $f_1 = \frac{1}{2L}\sqrt{\frac{T}{\mu}}$, and every higher harmonic an integer multiple of the fundamental. Three laws follow by holding two of the three quantities fixed and varying the third. Every stringed musical instrument and every sonometer in every laboratory operates within those laws. The next section generalises the wave-speed formula beyond strings to all elastic media, and asks what limits the speed of a transverse wave on a real wire of finite tensile strength.

VELOCITY OF MECHANICAL WAVES IN DIFFERENT MEDIA

The previous section gave a single formula, $v = \sqrt{\frac{T}{\mu}}$, for the speed of a transverse wave on a stretched string.

The formula is the launching point for a far broader statement that holds for every elastic medium in classical mechanics. The structure of the formula is what generalises: *a square root, with an **elastic property** of the medium in the **numerator** and an **inertial property** of the medium in the **denominator***. Strings, solids, and liquids all obey it; only the specific elastic and inertial properties change from medium to medium. This section presents the generalised formula, applies it to longitudinal waves in solids and liquids, and asks what physical limit caps the transverse wave speed on a real wire of finite tensile strength.

Gases require one further refinement (an adiabatic factor that Newton missed and Laplace later supplied), and they are postponed to Chapter 9 where sound in air will be treated in full. This section confines itself to the media where the elastic property is straightforward: **tension** on a *string*, **Young's modulus** in a *solid*, and **bulk modulus** in a *liquid*.

The unified wave-speed formula

In every elastic medium that supports a mechanical wave, the speed of that wave is the square root of an elastic property divided by an inertial property of the medium:

$$v = \sqrt{\frac{\text{elastic property}}{\text{inertial property}}}$$

The structure is a piece of physical bookkeeping. The **numerator** *quantifies how stiffly the medium resists deformation*: the **larger the elastic property**, the larger the restoring force the medium exerts when displaced from equilibrium, and the **faster a disturbance propagates**. The **denominator** *quantifies how heavy the medium is per unit volume (or per unit length, in the case of a string)*: the **larger the inertia**, the more sluggish the medium's response to any restoring force, and the **slower the wave propagates**. The square root captures the way these two competing tendencies enter the equation of motion: every elastic medium obeys a wave equation in which the wave speed enters as the square root of the ratio.

Three particular cases cover the bulk of this chapter and the next:

1) Transverse waves on a stretched string

Elastic property: the *tension* T (force restoring the string toward its equilibrium line). Inertial property: the *linear mass density* μ (mass per unit length). The wave speed is:

$$v = \sqrt{\frac{T}{\mu}}$$

2) Longitudinal waves in a solid rod

Elastic property: *Young's modulus* Y (force per unit area per unit fractional extension along the rod). Inertial property: the *volume mass density* ρ (mass per unit volume). The wave speed is:

$$v = \sqrt{\frac{Y}{\rho}}$$

3) Longitudinal waves in a liquid

Elastic property: the *bulk modulus* K (pressure required to compress the liquid by unit fractional volume). Inertial property: the *volume mass density* ρ . The wave speed is:

$$v = \sqrt{\frac{K}{\rho}}$$

All three are special cases of the same square-root rule. Once a student has internalised the rule, the only thing left to learn for any new medium is which elastic property and which inertial property apply: from there the wave speed follows by one substitution.

The maximum transverse wave speed on a real wire

A real stretched wire cannot be tightened indefinitely. Beyond a certain tensile stress (force per unit cross-sectional area), the wire's atomic bonds yield and the wire snaps. The maximum stress a wire can sustain without breaking is its **ultimate tensile stress**, denoted $\sigma_{\max}$. This finite limit places a finite upper bound on the wave speed that the wire can carry, and the bound has a particularly elegant form.

Take a uniform wire of cross-sectional area A , linear mass density μ , and volume mass density ρ . The two densities are connected by $\mu = \rho A$: a length ℓ of wire has volume $A\ell$ and mass $\rho A\ell$, so its mass per unit length is ρA . The maximum tension is the maximum stress times the area: $T_{\max} = \sigma_{\max} A$. Substituting both into $v = \sqrt{\frac{T}{\mu}}$:

$$v_{\max} = \sqrt{\frac{T_{\max}}{\mu}} = \sqrt{\frac{\sigma_{\max} A}{\rho A}} = \sqrt{\frac{\sigma_{\max}}{\rho}}$$

The cross-sectional area A cancels exactly. The maximum transverse wave speed depends only on the wire's material, through $\sigma_{\max}$ and ρ , and not on its thickness. A thicker wire of the same material can take more tension and is also heavier per unit length; the two effects scale identically with A and cancel. This cancellation is the elegant structural fact that closes the section: *for any given material there is a single maximum wave speed, set by its ultimate tensile stress and its density, and that speed cannot be exceeded by any wire made of that material no matter what gauge it is.*

With the unified formula now stated for any elastic medium, the worked examples can begin.

BINDER Example 30

A long uniform steel rod is struck on one end with a hammer. Find the speed of the longitudinal disturbance that propagates along the rod, given that steel has Young's modulus $Y = 2.0 \times 10^{11} \text{ Pa}$ and density $\rho = 7800 \text{ kg/m}^3$.

Solution

For a longitudinal wave in a solid rod, the elastic property is Young's modulus and the inertial property is the volume mass density:

$$v = \sqrt{\frac{Y}{\rho}} = \sqrt{\frac{2.0 \times 10^{11} \text{ Pa}}{7800 \text{ kg/m}^3}} = 5.06 \times 10^3 \text{ m/s}$$

That is approximately 5060 m/s, or about 15 times the speed of sound in air at room temperature (340 m/s).

Making sense of the answer: The arithmetic is one substitution into $v = \sqrt{\frac{Y}{\rho}}$. The result, about 5060 m/s, is in close agreement with measured values for the speed of longitudinal sound in steel rods, which experiment places between 5000 m/s and 5200 m/s depending on the specific alloy. The reason steel structures ring so loudly when struck (a steel rail clanged with a hammer, the side of a steel ship, the bell of a railway crossing) is that the longitudinal wave travels through the steel with this very high speed and re-radiates as audible air-pressure oscillations at the surface of the steel. The sharpness of the ring follows from the speed: a struck steel object delivers its acoustic energy to the air in a brief pulse, not in the slow diffusion characteristic of less stiff materials.

Think like a physicist: Compare this calculation with the string-tension calculation of the previous section. The structure is identical: square root, elastic property over inertial property, one substitution. Only the names of the properties have changed. For the string the elastic property was a force (T); for the solid it is a force per unit area per unit strain (Y). For the string the inertial property was a mass per unit length (μ); for the solid it is a mass per unit volume (ρ). The dimensions of the ratio still work out to those of speed squared, and the square root recovers the speed. Whenever a new elastic medium appears, the question to ask is which elastic property restores it and which inertial property loads it. From those two answers the wave speed follows automatically.

REAL Example 31

On the Tanzania-Zambia (TAZARA) railway line, before the era of two-way radios and satellite tracking, track-walkers and trackside workers used a centuries-old technique to detect an approaching train long before they could hear it through the air: they pressed one ear to the steel rail and listened. The vibration carried by the steel arrived noticeably earlier than the sound carried by the air, giving the listener several seconds of warning to step clear. Given that:

Steel: $Y = 2.0 \times 10^{11} \text{Pa}$, $\rho = 7800 \text{kg/m}^3$. Speed of sound in air: $v_{\text{air}} = 340 \text{m/s}$, $K_{\text{air}} = 1.4 \times 10^5 \text{Pa}$, $\rho_{\text{air}} = 1.2 \text{kg/m}^3$

- Calculate the speed of the longitudinal vibration that travels through the steel rail.
- A train is 1.5km up the line from a worker pressing his ear to the rail. How much earlier does the vibration through the steel reach the worker than the sound through the air?
- Explain physically, using the unified wave-speed formula, why sound travels so much faster in steel than in air.

Solution

(a) From the previous example, $v_{\text{steel}} = \sqrt{\frac{Y}{\rho}} \approx 5060 \text{m/s}$.

(b) Distance $d = 1500 \text{m}$. Time through the steel:

$$t_{\text{steel}} = \frac{d}{v_{\text{steel}}} = \frac{1500 \text{m}}{5060 \text{m/s}} = 0.296 \text{s}$$

Time through the air:

$$t_{\text{air}} = \frac{d}{v_{\text{air}}} = \frac{1500 \text{m}}{340 \text{m/s}} = 4.41 \text{s}$$

The advance warning is the difference:

$$\Delta t = t_{\text{air}} - t_{\text{steel}} \approx 4.41 \text{s} - 0.30 \text{s} = 4.1 \text{seconds}.$$

(c) The unified formula is $v = \sqrt{\frac{\text{elastic}}{\text{inertia}}}$. Steel's Young's modulus is approximately $2 \times 10^{11} \text{Pa}$; the bulk modulus of air at atmospheric pressure is approximately $1.4 \times 10^5 \text{Pa}$. Steel is therefore about 1.4×10^6 times stiffer than air as an elastic medium. Steel's density is 7800kg/m^3 ; air's density at room temperature is about 1.2kg/m^3 . Steel is therefore about 6500 times denser than air. The ratio inside the square root, (elastic/inertia), is therefore about $\frac{1.4 \times 10^6}{6500} \approx 220$ times larger for steel than for air. The wave speed scales with the square root of that ratio, so the speed in steel is about $\sqrt{220} \approx 15$ times the speed in air. Steel's overwhelming stiffness advantage easily defeats its density disadvantage, and hence the sound waves runs faster in steel.

Making sense of the answer: The 4-second warning is real and was decisive in the days before mechanical track sensors. A worker bent over the rail with his ear pressed to the steel learned of the approaching train roughly 4.1 seconds before the same train would have been audible through the air at the same distance, and that interval was the difference between safety and disaster. The ratio of speeds, 15:1, is the structural answer: steel is so much stiffer than air that even though it is also much denser, the elastic-over-inertial ratio comes out far in favour of steel.

Think like a physicist: This is one of the few cases in physics where two factors point in opposite directions but one factor wins by a wide enough margin that the qualitative answer is obvious. Steel is stiffer than air; this raises the wave speed. Steel is denser than air; this lowers the wave speed. The two effects compete inside the square root. In some materials the competition is genuinely close, and the answer requires a careful numerical calculation. Always check both factors when comparing wave speeds in two media.

HOT Example 32

A wire is made of a metal alloy with ultimate tensile stress $\sigma_{\text{max}} = 5.0 \times 10^8 \text{Pa}$ and density $\rho = 8000 \text{kg/m}^3$. The wire has a cross-sectional area $A = 2.0 \times 10^{-6} \text{m}^2$.

- Find the maximum transverse wave speed the wire can carry without snapping, using the formula

$$v_{\text{max}} = \sqrt{\frac{\sigma_{\text{max}}}{\rho}}.$$

- Find the maximum tension and the linear mass density of the wire.

- (c) Compute the wave speed from $v = \sqrt{\frac{T}{\mu}}$ using the values from part (b), and verify that the answer agrees with part (a).

Solution

- (a) Maximum transverse wave speed:

$$v_{\max} = \sqrt{\frac{\sigma_{\max}}{\rho}} = \sqrt{\frac{5.0 \times 10^8 \text{ Pa}}{8000 \text{ kg/m}^3}} = 250 \text{ m/s}$$

- (b) Maximum tension and linear mass density:

From the definition of tensile stress, $T_{\max} = \sigma_{\max} A$:

$$T_{\max} = (5.0 \times 10^8 \text{ Pa}) \times (2.0 \times 10^{-6} \text{ m}^2) = 1000 \text{ N}$$

From $\mu = \rho A$:

$$\mu = (8000 \text{ kg/m}^3) \times (2.0 \times 10^{-6} \text{ m}^2) = 0.016 \text{ kg/m}$$

- (c) Verify by direct substitution into the string formula:

$$v = \sqrt{\frac{T}{\mu}} = \sqrt{\frac{1000 \text{ N}}{0.016 \text{ kg/m}}} = 250 \text{ m/s}$$

The two computations agree exactly, as the algebra of the section guarantees they must.

Making sense of the answer: *The two computations agree because the cancellation of A between numerator and denominator is exact, not approximate. Algebraically, $\frac{\sigma_{\max} A}{\rho A} = \frac{\sigma_{\max}}{\rho}$, so the cross-sectional area drops out completely and the maximum wave speed is a property of the material alone. A thicker wire of the same alloy has the same maximum transverse wave speed as a thinner wire: the thicker wire can take more tension (because $T_{\max} \propto A$), but it is also heavier per unit length (because $\mu \propto A$), and the two effects scale together so as to leave the wave speed unchanged.*

Think like a physicist: *This kind of cancellation is not a coincidence. Whenever a physical quantity is the ratio of two extensive properties (properties that scale with the size of the system), the scaling factors cancel and the quantity becomes intensive (a property of the material alone, independent of system size). Tension is extensive; mass is extensive; their ratio under the square root, divided by the cross-sectional area in numerator and denominator, leaves an intensive ratio of stress to density.*

The unified wave-speed formula $v = \sqrt{\frac{\text{elastic property}}{\text{inertial property}}}$ governs every mechanical wave in every elastic medium, with strings, solids, and liquids each providing one specialisation. Real wires of finite tensile strength have a finite maximum transverse wave speed determined by the material alone, $v_{\max} = \sqrt{\frac{\sigma_{\max}}{\rho}}$, a result whose elegance comes from the exact cancellation of the cross-sectional area. Gases supply one further refinement that Newton missed and Laplace corrected, and they belong to Chapter 9 on sound.

The next section brings together the results we have developed and shows how they apply to real-world situations and everyday life.

APPLICATIONS OF WAVE MOTION

So far, the chapter has focused on building the theory of waves through definitions, derivations, and worked examples. This section now turns from the page to the world beyond it. It asks what these results actually do once they leave the classroom and enter real life. The purpose of the final section is therefore not to introduce new physics, but to reveal how remarkably far the physics already in hand can reach.

1. Musical instruments

Every stringed instrument that has ever been built (the kora, the zeze, the violin, the guitar, the piano, the harp, the oud) is a collection of fixed-fixed strings tuned by the maker to particular fundamentals. The harmonic spectrum $f_n = nf_1$, with $f_1 = \frac{1}{2L} \sqrt{\frac{T}{\mu}}$, determines the pitch the string sounds. The three laws (length, tension, linear mass density) determine how the maker varies the strings to produce a chosen range of pitches: a piano lengthens the bass strings and weights them with copper winding, a violin shortens its strings and uses higher tension, a guitar uses six strings of varying mass density tuned to fixed pitches. None of this is discretionary; it is fixed by the master formula and the boundary condition that both ends of a string under a player's finger are nodes.

Two further pieces complete the picture. First, when a string is plucked or bowed it does not vibrate at any single one of the allowed harmonics; it vibrates in a superposition of many of them, with the relative amplitudes set by where and how it is excited. The mixture of harmonics, called the **timbre**, is what tells a listener at the back of a concert hall that a violin is playing rather than a guitar even when the two are sounding the same fundamental. Second, the same boundary-condition selection that picks out the harmonics on a string also picks out the resonant modes of an air column inside an organ pipe, an open or closed flute, or a clarinet. The full treatment of pipe modes belongs to Chapter 9; the pattern (a discrete spectrum of allowed frequencies, set by boundary conditions on the standing wave) is identical to the one this chapter has worked out for strings.

2. Seismology and the structure of the Earth

When an earthquake releases its energy near the surface of the Earth, the disturbance travels outward through the planet as a mechanical wave. Two distinct kinds of wave propagate. The **primary** wave (P-wave) is longitudinal: the rock oscillates back and forth along the direction of propagation, the way a sound wave does in air. The **secondary** wave (S-wave) is transverse: the rock oscillates perpendicular to the direction of propagation, the way a wave does on a guitar string. Each obeys the unified speed formula $v = \sqrt{\frac{\text{elastic}}{\text{inertia}}}$, but with different elastic moduli in the numerator. The P-wave's restoring force comes from a combination of the rock's bulk modulus and its shear modulus; the S-wave's restoring force comes from the shear modulus alone. The mathematical consequence is that P-waves always travel faster than S-waves in the same rock.

Two consequences follow that have shaped the science of geophysics. First, when an earthquake's waves arrive at a distant seismograph, the P-wave arrives first and the S-wave arrives later, and the gap between their arrival times reads back the distance to the epicentre: the longer the gap, the further the source. With three or more seismographs at known locations, the position of the epicentre is fixed by triangulation. Second, S-waves cannot propagate through a liquid, because a liquid has no shear strength. The fact that S-waves are absent from seismograph records on the side of the Earth opposite a major earthquake is what told twentieth-century geophysicists that the Earth has a liquid outer core. The diagnosis was made entirely from the absence of a kind of wave whose physics this chapter has already worked out.

3. Medical ultrasound imaging

A medical ultrasound scanner sends short pulses of progressive longitudinal waves into the body, at frequencies between roughly 1MHz and 20MHz, and listens for the echoes. Each tissue boundary (skin to fat, fat to muscle, muscle to bone, the wall of an organ, the wall of a foetal heart) reflects part of the wave back toward the transducer. The pulse-echo timing converts directly to depth: a wave that returns Δt seconds after it was emitted has travelled $v\Delta t$ through tissue and back, so the reflecting boundary is at depth $d = v\Delta t/2$, with v the speed of sound in soft tissue (about 1540m/s, set by the unified formula with the bulk modulus and density of water-rich biological tissue).

Two design choices are dictated by the physics of the chapter. The first is the choice of frequency. Higher frequency means shorter wavelength means finer spatial resolution; but higher frequency also means greater absorption in the tissue and shorter penetration depth. Obstetric scans of a foetus several centimetres deep

use frequencies near the lower end of the range; superficial scans of a thyroid or a tendon a few millimetres deep use frequencies near the upper end. The second is the duration of the pulse: short pulses give better depth resolution, but a pulse can be no shorter than a few wave periods, so the depth resolution is fundamentally limited by the wavelength. Both trade-offs are direct consequences of $v = f\lambda$ applied to a medium whose absorption increases with frequency.

4. Resonance in bridges and buildings

Every elastic structure (a footbridge, a road bridge, a tall building, a chimney, a guitar's wooden body) has a discrete spectrum of natural frequencies of vibration, set by the same kind of boundary-condition selection that gives a string its harmonics. The structure's lowest natural frequency is called its fundamental; the higher natural frequencies are its overtones. If the structure is driven by an external force whose frequency matches one of these natural frequencies, the amplitude of the structure's response builds up dramatically through resonance, and the resulting oscillation can be large enough to damage or destroy the structure.

Foot traffic across a footbridge can drive the bridge's lateral mode if the footfalls happen to land at the natural frequency, with one well-known modern example being the synchronised lateral sway of a London footbridge on its opening day in the year 2000. Marching regiments are routinely instructed to break step when crossing a small bridge for exactly this reason. Wind blowing across a tall building can drive its bending modes through vortex shedding at frequencies determined by the wind speed and the building's cross-section. Modern civil engineering specifies a structure's natural frequencies during design, identifies the loading conditions that could excite each mode, and adds damping or stiffening so that no excitable mode lies near a frequency the building or bridge is likely to encounter in service. The physics is the same boundary-value problem that selects the harmonics of a string, scaled up from a few millimetres of wire to several hundred metres of steel.

5. Active noise cancellation

Noise-cancelling headphones, the cabin of a modern aircraft, and certain car engine compartments use a direct application of the principle of superposition. A microphone outside the listener's ear samples the unwanted sound. A small electronic processor inverts the sampled waveform (shifting its phase by π) and a tiny loudspeaker in the ear cup re-emits the inverted wave in real time. The original wave and the inverted wave superpose at the listener's eardrum; their displacements sum to zero, exactly the destructive interference we worked out in the chapter on superposition. The listener hears silence (or near silence) where there would otherwise have been engine rumble.

The technique works best for low-frequency steady noise: the inversion-and-replay loop has to complete in less than a fraction of a wavelength, and shorter wavelengths leave the loop too little time to catch up with the incoming wave. High-pitched sudden sounds, like a baby's cry or a screech of brakes, are not well cancelled and remain audible. The physics is unmistakable: superposition gave us **constructive interference** when two waves added with the same phase and **destructive interference** when they added with opposite phase. The headphones simply manufacture the second condition deliberately.

Having tasted each section on its own, we are now ready for the combination plate: miscellaneous worked examples that bring the whole chapter together on a single tray.

MISCELLANEOUS WORKED EXAMPLES ON WAVE MOTION

Example 33

- Account for the observation that water surface waves are classified as transverse, while sound waves travelling through the same water are longitudinal.
- A long horizontal spring is laid on a smooth bench and one end is moved back-and-forth along the spring's own length, sending a train of longitudinal waves along it. The compressions in the wave train are observed to be 0.40m apart, and a particular compression is timed to travel 6.0m along the spring in 0.50s. Find the wave speed, the frequency, the period, the angular frequency and the wave number of the wave.

Solution

(a) The classification depends on the direction in which the molecules of the medium are forced to oscillate, not on the medium itself. A water surface wave is held together by gravity acting on the displaced surface; the surface has to rise and fall for the restoring force to act, so each surface molecule oscillates vertically while the disturbance propagates horizontally, so the disturbance is perpendicular to the propagation and the wave is therefore transverse. A sound wave in the same body of water is held together by the bulk elasticity

of water; compressing the water momentarily generates a restoring force that drives the molecules along the line of compression, so each molecule oscillates along the direction of propagation, so the disturbance is parallel to the propagation and the wave is therefore longitudinal. The two waves coexist in the same water without confusion because they exploit two independent restoring mechanisms.

(b) Wavelength is the distance between consecutive compressions, $\lambda = 0.40\text{m}$.

Wave speed:

$$v = \frac{6.0\text{m}}{0.50\text{s}} = 12\text{m/s}$$

Frequency:

$$f = \frac{v}{\lambda} = \frac{12\text{m/s}}{0.40\text{m}} = 30\text{Hz}$$

Period:

$$T = \frac{1}{f} = \frac{1}{30\text{Hz}} = 0.033\text{s}$$

Angular frequency:

$$\omega = 2\pi f = 2\pi(30\text{Hz}) = 60\pi\text{rad/s} \approx 188\text{rad/s}$$

Wave number:

$$k = \frac{2\pi}{\lambda} = \frac{2\pi}{0.40\text{m}} = 5\pi\text{rad/m} \approx 15.7\text{rad/m}$$

Example 34

- (a) A leaf placed on the surface of a still pond bobs up and down without travelling along with the ripples that pass beneath it. Account for this in physical terms.
- (b) A wave is described by $y(x, t) = 0.005\sin(2\pi x - 100\pi t)\text{m}$, with x in metres and t in seconds. State the direction of propagation and find the wavelength, the period, the wave speed and the maximum particle speed of the wave.

Solution

(a) A wave is a propagating disturbance, not a flow of matter. The water beneath the leaf executes a small simple-harmonic up-and-down motion as each ripple passes; the molecules return to where they started by the time the ripple has gone. The leaf, riding on the water, shares only that local up-and-down motion. The energy and the geometric pattern of the disturbance travel outward, but the water (along with the leaf) remains on average in place.

(b) The argument $2\pi x - 100\pi t = -(100\pi t - 2\pi x)$, so $y = -0.005\sin(100\pi t - 2\pi x)\text{m}$. Comparing with $y = A\sin(\omega t - kx)$, the wave travels in the $+x$ direction with $\omega = 100\pi\text{rad/s}$ and $k = 2\pi\text{rad/m}$.

Wavelength:

$$\lambda = \frac{2\pi}{k} = \frac{2\pi}{2\pi\text{rad/m}} = 1.0\text{m}$$

Period:

$$T = \frac{2\pi}{\omega} = \frac{2\pi}{100\pi\text{rad/s}} = 0.020\text{s}$$

Wave speed:

$$v = \frac{\omega}{k} = \frac{100\pi\text{rad/s}}{2\pi\text{rad/m}} = 50\text{m/s}$$

Maximum particle speed:

$$v_{p,\text{max}} = A\omega = (0.005\text{m})(100\pi\text{rad/s}) = 0.5\pi\text{m/s} \approx 1.57\text{m/s}$$

Example 35

- (a) A diver swimming a few metres beneath the surface of a calm lake hears almost nothing of the loud conversation between two friends standing in a boat directly above her. Account for this observation.
- (b) A sound wave of frequency 1.50kHz travels through fresh water at 1500m/s . It then crosses a clear boundary into a thick layer of mud at the bottom of the lake, where its speed is 1100m/s . Find the wavelength of the wave in the water, the wavelength in the mud and the percentage change in wavelength.

Solution

(a) Sound moving from air into water crosses a boundary across which the elastic and inertial properties of the medium change abruptly: the product of density and wave speed (the acoustic impedance) is roughly four thousand times larger in water than in air. A boundary with such a large impedance mismatch reflects almost all the incident energy back into the originating medium and transmits only about 0.1% of it across. The diver therefore receives a tiny fraction of the air-borne sound, which is why she hears almost nothing.

(b) The source frequency is preserved across the boundary; the medium changes the speed and the wavelength adjusts accordingly.

Wavelength in water:

$$\lambda_1 = \frac{v_1}{f} = \frac{1500\text{m/s}}{1500\text{Hz}} = 1.00\text{m}$$

Wavelength in mud:

$$\lambda_2 = \frac{v_2}{f} = \frac{1100\text{m/s}}{1500\text{Hz}} = 0.733\text{m}$$

Percentage change:

$$\frac{\lambda_2 - \lambda_1}{\lambda_1} \times 100\% = \frac{0.733\text{m} - 1.00\text{m}}{1.00\text{m}} \times 100\% = -26.7\%$$

The wavelength decreases by about 26.7% on crossing the boundary into mud.

Example 36

(a) Account for the experimental observation that the colour of a beam of light passing from air into a glass of water does not change, even though both its speed and its wavelength decrease.

(b) Light of wavelength 600nm in vacuum enters a slab of glass of refractive index 1.50. Find the wavelength and the frequency of the light inside the glass. Take $c = 3.00 \times 10^8\text{m/s}$.

Solution

(a) The colour seen by the eye is determined by the frequency of the light, not by its wavelength. Frequency is the rate at which the source emits oscillations of the electromagnetic field; conservation of oscillation count requires the same rate to emerge from any boundary the wave crosses. The medium can change the speed of the wave, and the wavelength is forced to scale in step ($\lambda = v/f$), but the frequency, and hence the colour, is preserved.

(b) Frequency in vacuum (preserved on entering the glass):

$$f = \frac{c}{\lambda_0} = \frac{3.00 \times 10^8\text{m/s}}{600 \times 10^{-9}\text{m}} = 5.00 \times 10^{14}\text{Hz}$$

Speed of light in the glass:

$$v_{\text{glass}} = \frac{c}{n} = \frac{3.00 \times 10^8\text{m/s}}{1.50} = 2.00 \times 10^8\text{m/s}$$

Wavelength inside the glass:

$$\lambda_{\text{glass}} = \frac{v_{\text{glass}}}{f} = \frac{2.00 \times 10^8\text{m/s}}{5.00 \times 10^{14}\text{Hz}} = 4.00 \times 10^{-7}\text{m} = 400\text{nm}$$

The frequency is unchanged at $5.00 \times 10^{14}\text{Hz}$ and the wavelength inside the glass is 400nm.

Example 37

(a) Distinguish between the wave velocity and the particle velocity of a transverse mechanical wave, indicating which depends on the amplitude of the wave and which does not.

(b) A transverse wave on a horizontal nylon rope is described by $y(x, t) = 0.030\sin(120\pi t - 4\pi x)\text{m}$, with x in metres and t in seconds. Find the wavelength, the wave speed, the displacement of the particle at $x = 0.50\text{m}$ at the instant $t = \frac{1}{240}\text{s}$, and the velocity of that same particle at the same instant.

Solution

(a) The wave velocity $v = \omega/k$ is the speed at which the wave pattern (each crest, each trough) advances through the medium. It is fixed by the elastic and inertial properties of the medium and does not depend on

the amplitude. The particle velocity $v_p = \partial y / \partial t$ is the instantaneous speed at which a chosen particle of the medium oscillates about its own equilibrium position; its maximum value $v_{p,\max} = A\omega$ is directly proportional to the amplitude. Doubling the amplitude doubles the maximum particle speed but leaves the wave speed unchanged.

(b) Comparing $y = 0.030\sin(120\pi t - 4\pi x)\text{m}$ with $y = A\sin(\omega t - kx)$:

$$A = 0.030\text{m}, \quad \omega = 120\pi\text{rad/s}, \quad k = 4\pi\text{rad/m}$$

Wavelength:

$$\lambda = \frac{2\pi}{k} = \frac{2\pi}{4\pi\text{rad/m}} = 0.50\text{m}$$

Wave speed:

$$v = \frac{\omega}{k} = \frac{120\pi\text{rad/s}}{4\pi\text{rad/m}} = 30\text{m/s}$$

Phase at $x = 0.50\text{m}$, $t = \frac{1}{240}\text{s}$:

$$\omega t - kx = (120\pi)\left(\frac{1}{240}\right) - (4\pi)(0.50) = \frac{\pi}{2} - 2\pi = -\frac{3\pi}{2}$$

Displacement:

$$y = (0.030\text{m})\sin\left(-\frac{3\pi}{2}\right) = (0.030\text{m})(+1) = 0.030\text{m}$$

The particle is at the upper extreme of its oscillation.

Particle velocity:

$$v_p = A\omega\cos(\omega t - kx) = (0.030\text{m})(120\pi\text{rad/s})\cos\left(-\frac{3\pi}{2}\right) = (3.6\pi\text{m/s})(0) = 0$$

The particle is momentarily at rest at the upper turning point of its motion.

Example 38

(a) Show, by direct substitution, that the function $y(x, t) = A\sin(\omega t - kx)$ is a solution of the one-dimensional wave equation $\frac{\partial^2 y}{\partial t^2} = v^2 \frac{\partial^2 y}{\partial x^2}$, and identify the wave speed v in terms of ω and k .

(b) A transverse wave on a stretched wire of linear mass density $\mu = 5.0 \times 10^{-3}\text{kg/m}$ is described by $y(x, t) = 0.010\sin(80\pi t - 8\pi x)\text{m}$, with x in metres and t in seconds. Find the tension in the wire and the maximum acceleration of any particle of the wire.

Solution

(a) Differentiate $y = A\sin(\omega t - kx)$ twice with respect to t , holding x fixed:

$$\frac{\partial y}{\partial t} = A\omega\cos(\omega t - kx), \quad \frac{\partial^2 y}{\partial t^2} = -A\omega^2\sin(\omega t - kx)$$

Differentiate twice with respect to x , holding t fixed:

$$\frac{\partial y}{\partial x} = -Ak\cos(\omega t - kx), \quad \frac{\partial^2 y}{\partial x^2} = -Ak^2\sin(\omega t - kx)$$

Substitute into the wave equation:

$$-A\omega^2\sin(\omega t - kx) = v^2(-Ak^2\sin(\omega t - kx))$$

The equality holds for all x and t provided $\omega^2 = v^2k^2$, that is

$$v = \frac{\omega}{k}$$

So $y = A\sin(\omega t - kx)$ is a solution of the wave equation with wave speed equal to the ratio of the angular frequency to the wave number.

(b) Comparing $y = 0.010\sin(80\pi t - 8\pi x)\text{m}$ with $y = A\sin(\omega t - kx)$:

$$A = 0.010\text{m}, \quad \omega = 80\pi\text{rad/s}, \quad k = 8\pi\text{rad/m}$$

Wave speed:

$$v = \frac{\omega}{k} = \frac{80\pi\text{rad/s}}{8\pi\text{rad/m}} = 10\text{m/s}$$

Tension:

$$T = \mu v^2 = (5.0 \times 10^{-3} \text{ kg/m})(10 \text{ m/s})^2 = 0.50 \text{ N}$$

Maximum particle acceleration:

$$a_{\max} = A\omega^2 = (0.010 \text{ m})(80\pi \text{ rad/s})^2 = (0.010)(6400\pi^2) \text{ m/s}^2 = 64\pi^2 \text{ m/s}^2 \approx 632 \text{ m/s}^2$$

Example 39

- (a) Account in physical terms for the fact that two particles separated by an integer number of wavelengths along a progressive wave are always in phase, regardless of the magnitude of the separation.
- (b) A transverse wave of wavelength 1.20m and period 0.40s travels along a long stretched string. Find the phase difference between two points on the string separated by 0.30m at the same instant, and the phase difference between the displacements of the same point at two instants separated by 0.05s.

Solution

(a) The phase difference between two points separated by distance Δx on the same wave is $\Delta\phi = k\Delta x = (2\pi/\lambda)\Delta x$. When $\Delta x = n\lambda$ for integer n , the phase difference is $\Delta\phi = 2\pi n$, which is an integer number of full cycles. A sinusoidal function repeats exactly after every full cycle, so the two particles return to identical positions and identical velocities at every instant. The agreement is independent of the magnitude of Δx because only the fractional cycle counts; an integer number of wavelengths contributes no fractional cycle.

(b) **Spatial** phase difference:

$$\Delta\phi = \frac{2\pi}{\lambda} \Delta x = \frac{2\pi}{1.20 \text{ m}} (0.30 \text{ m}) = \frac{\pi}{2} \text{ rad}$$

Temporal phase difference:

$$\Delta\phi = \frac{2\pi}{T} \Delta t = \frac{2\pi}{0.40 \text{ s}} (0.05 \text{ s}) = \frac{\pi}{4} \text{ rad}$$

Example 40

- (a) Account for the experimental observation that two coherent loudspeakers driven in phase produce regions in front of them that are permanently silent, regardless of the volume.
- (b) Two coherent loudspeakers radiate notes of frequency 680Hz in still air, where the speed of sound is 340m/s. At a point P the path difference from the two loudspeakers is 0.25m. State, with reasons, whether P is a region of constructive or of destructive interference.

Solution

(a) Coherent sources radiate waves whose phase difference at any point depends only on the path difference from the two sources to that point. At a point where the path difference is an odd number of half-wavelengths, the two waves arrive in antiphase, the displacements cancel exactly at every instant, and the resultant amplitude is zero. The cancellation is an algebraic identity between two sinusoids of equal amplitude and opposite sign; it persists no matter how large the volume of either source, since louder sources cancel just as completely as faint ones.

(b) Wavelength of the radiated sound:

$$\lambda = \frac{v}{f} = \frac{340 \text{ m/s}}{680 \text{ Hz}} = 0.50 \text{ m}$$

Phase difference at P:

$$\Delta\phi = \frac{2\pi}{\lambda} \Delta x = \frac{2\pi}{0.50 \text{ m}} (0.25 \text{ m}) = \pi \text{ rad}$$

Since $\Delta\phi = \pi \text{ rad}$ (equivalently, the path difference is exactly half a wavelength), the two waves arrive at P exactly out of phase. Their displacements cancel at every instant, and P is a region of destructive interference, that is, a permanent silent region in front of the speakers.

Example 41

- (a) Account for the fact that $y = A\sin(\omega t - kx)$ represents a wave travelling in the $+x$ direction, while $y = A\sin(\omega t + kx)$ represents one travelling in the $-x$ direction.
- (b) A point source of acoustic power 12W radiates uniformly in still air. Find the intensity of the sound at 4.0m from the source, the distance from the source at which the intensity has fallen to 1.0W/m^2 , and the ratio of the displacement amplitudes at the two distances.

Solution

- (a) A point of constant phase on the wave $y = A\sin(\omega t - kx)$ satisfies $\omega t - kx = \text{constant}$.

Differentiating with respect to time gives $\omega - k \frac{dx}{dt} = 0$, so:

$$\frac{dx}{dt} = \frac{\omega}{k} = +v$$

The point of constant phase therefore moves in the direction of increasing x : the wave pattern slides toward $+x$. Repeating the same calculation for $y = A\sin(\omega t + kx)$ gives $\frac{dx}{dt} = -\frac{\omega}{k} = -v$, so the wave pattern slides toward $-x$. The sign in front of kx is the wave's direction-of-travel indicator.

- (b) Intensity at $r_1 = 4.0\text{m}$:

$$I_1 = \frac{P}{4\pi r_1^2} = \frac{12\text{W}}{4\pi(4.0\text{m})^2} = \frac{12\text{W}}{64\pi\text{m}^2} = 5.97 \times 10^{-2}\text{W/m}^2$$

Distance at which $I_2 = 1.0 \times 10^{-3}\text{W/m}^2$:

$$r_2^2 = \frac{P}{4\pi I_2} = \frac{12\text{W}}{4\pi(1.0 \times 10^{-3}\text{W/m}^2)} = 955\text{m}^2,$$

$$r_2 = \sqrt{955\text{m}^2} = 30.9\text{m}$$

Since $I \propto A^2$, $A \propto 1/r$, hence

$$\frac{A_1}{A_2} = \frac{r_2}{r_1} = \frac{30.9\text{m}}{4.0\text{m}} = 7.7$$

The amplitude at 4.0m is about 7.7 times the amplitude at 30.9m .

Example 42

- (a) Account for the observation that, when two stereo loudspeakers play the same note in phase, a listener positioned exactly between them hears a louder note than a listener standing close to either speaker alone.
- (b) Two waves of equal frequency and amplitudes 6.0cm and 4.0cm arrive at the same point with a phase difference of $\frac{2\pi}{3}\text{rad}$. Find the amplitude of the resultant.

Solution

(a) A listener exactly between the two coherent speakers receives the two waves with zero path difference and therefore zero phase difference. The two displacements add at every instant, giving a resultant of amplitude $A_1 + A_2$, which is twice the amplitude received from either speaker alone (when the two amplitudes are equal). Since intensity goes as the square of amplitude, the listener at the midpoint perceives roughly four times the intensity of one speaker alone, which is appreciably louder. A listener close to one speaker is essentially in the field of that speaker only and receives a single amplitude A_1 .

(b) For two waves of amplitudes $A_1 = 6.0\text{cm}$ and $A_2 = 4.0\text{cm}$ with phase difference $\Delta\phi = \frac{2\pi}{3}\text{rad}$, the resultant amplitude is given by:

$$A_R^2 = A_1^2 + A_2^2 + 2A_1A_2\cos(\Delta\phi)$$

Substituting:

$$A_R^2 = (6.0\text{cm})^2 + (4.0\text{cm})^2 + 2(6.0\text{cm})(4.0\text{cm})\cos\frac{2\pi}{3}$$

$$A_R^2 = 36\text{cm}^2 + 16\text{cm}^2 + 2(24\text{cm}^2)\left(-\frac{1}{2}\right) = 52\text{cm}^2 - 24\text{cm}^2 = 28\text{cm}^2$$

$$A_R = \sqrt{28\text{cm}^2} = 2\sqrt{7}\text{cm} \approx 5.3\text{cm}$$

Example 43

- (a) Account for the fact that two musical notes of slightly different frequency, when sounded together, are heard as a single note whose loudness rises and falls in a slow, regular rhythm.
- (b) A piano tuner sounds a tuning fork of frequency 256Hz together with one of the strings of a piano and counts 12 beats in 4s. He then loads the string very lightly with a small piece of wax and the beat rate falls to 2Hz. Find the original frequency of the piano string.

Solution

(a) When two close-frequency notes sound together their displacements at the listener's ear superpose. The two waves drift slowly into and out of phase: at moments when they are in phase the resultant amplitude is $A_1 + A_2$ (loud) and at moments when they are in antiphase the resultant amplitude is $|A_1 - A_2|$ (faint or silent). The cycle of loudness repeats at a frequency equal to the difference of the two source frequencies, $f_b = |f_1 - f_2|$. These regular pulsations of loudness are the beats.

(b) Initial beat frequency:

$$f_b = \frac{12}{4\text{s}} = 3\text{Hz}$$

The piano string is therefore at $f_{\text{string}} = 256\text{Hz} \pm 3\text{Hz}$, i.e., 253Hz or 259Hz.

Loading the string with wax increases its linear mass density. From the third law of vibration of a stretched string, $f \propto 1/\sqrt{\mu}$, so loading lowers the string's frequency. After waxing the beat rate has fallen from 3Hz to 2Hz, so the string's frequency must have moved closer to the fork.

If the original frequency was 259Hz, lowering it slightly produces a string at 258Hz and a beat of 2Hz, which is consistent with observation.

If the original frequency was 253Hz, lowering it would produce a string at 252Hz or below and a beat of 4Hz or more, which is inconsistent with observation.

The original frequency of the piano string is $f_{\text{string}} = 259\text{Hz}$.

Example 44

- (a) Account for the experimental observation that loading a slightly mistuned guitar string with a small piece of plasticine can either remove beats with a reference fork or make them faster, depending on which side of the fork frequency the string was on.
- (b) A sonometer wire of length 25.0cm is stretched between two bridges under fixed tension. When the wire is sounded together with a tuning fork of frequency 384Hz, five beats per second are heard. The wire is shortened, with no change in tension, until the beats vanish. Find the new length of the wire.

Solution

(a) The beat frequency between the fork and the string is $f_b = |f_{\text{fork}} - f_{\text{string}}|$. Loading the string with plasticine increases its mass per unit length, so by the third law of vibration its frequency falls. If the string was originally above the fork, lowering it reduces $|f_{\text{fork}} - f_{\text{string}}|$, so the beat frequency falls and may pass through zero (silence) before rising again. If the string was originally below the fork, lowering it increases $|f_{\text{fork}} - f_{\text{string}}|$, so the beat frequency rises immediately. The direction of the change in beat rate therefore reveals on which side of the fork the string was tuned.

(b) Initial beat frequency $f_b = 5\text{Hz}$, so the original string frequency is

$$f_{\text{old}} = 384\text{Hz} \pm 5\text{Hz} = 379\text{Hz or } 389\text{Hz}.$$

Shortening the wire raises its frequency (first law of vibration, $f \propto 1/L$). The shortening removes the beats, i.e., it raises the string to 384Hz. The original frequency must therefore have been below 384Hz, namely $f_{\text{old}} = 379\text{Hz}$.

Applying $f \propto 1/L$ at constant tension and mass per unit length:

$$\frac{L_{\text{new}}}{L_{\text{old}}} = \frac{f_{\text{old}}}{f_{\text{new}}} = \frac{379\text{Hz}}{384\text{Hz}},$$

$$L_{\text{new}} = (25.0\text{cm}) \left(\frac{379\text{Hz}}{384\text{Hz}} \right) = 24.67\text{cm}$$

The wire must be shortened to 24.67cm for the beats with the fork to vanish.

Example 45

- (a) Account in physical terms for the fact that a stretched string fixed at both ends vibrates only at a discrete set of frequencies, and not at any frequency intermediate between two consecutive members of the set.
- (b) A string of length 1.20m fixed at both ends has linear mass density $\mu = 4.0 \times 10^{-3}\text{kg/m}$. Find the fundamental frequency of the string when the tension is 250N, and find the new tension required for the string to resonate at 400Hz in its fundamental mode.

Solution

(a) The transverse displacement at each fixed end is permanently zero, so any standing wave on the string must have a node at every support. The simplest such wave fits half a wavelength between the supports; the next fits one full wavelength; the next fits one-and-a-half; and so on. Only those wavelengths satisfying $L = n\lambda/2$ for positive integer n produce a stationary pattern that is consistent with both boundary conditions. The corresponding frequencies $f_n = nv/(2L)$ form a discrete set; intermediate frequencies cannot establish a self-consistent stationary pattern and decay quickly.

(b) Wave speed at the original tension:

$$v = \sqrt{\frac{T}{\mu}} = \sqrt{\frac{250\text{N}}{4.0 \times 10^{-3}\text{kg/m}}} = 250\text{m/s}$$

Fundamental frequency:

$$f_1 = \frac{v}{2L} = \frac{250\text{m/s}}{2(1.20\text{m})} = 104.2\text{Hz}$$

From the second law of vibration, $f_1 \propto \sqrt{T}$ at fixed L and μ , so

$$\frac{T_{\text{new}}}{T_{\text{old}}} = \left(\frac{f_{\text{new}}}{f_{\text{old}}} \right)^2 = \left(\frac{400\text{Hz}}{104.2\text{Hz}} \right)^2 = 14.74$$

$$T_{\text{new}} = 14.74 \times 250\text{N} = 3.69 \times 10^3\text{N}.$$

Example 46

- (a) Account for the observation that bowing a sonometer wire at exactly its midpoint excites a much purer note than bowing it close to one of the bridges.
- (b) A sonometer wire of length 1.00m fixed at both ends is excited in its third harmonic mode. Three small paper riders rest on the wire at distances 0.250m, 0.333m and 0.500m from one of the bridges. State, with reasons, which of the riders are thrown off the wire and which remain at rest.

Solution

(a) A vibrating fixed-fixed wire can support only those harmonics that have a node at every point at which the wire is forced. The midpoint $L/2$ is a node for every even harmonic ($n = 2, 4, 6, \dots$) and an antinode for every odd one. Bowing the midpoint forces it to oscillate, which is incompatible with a node there, so every even harmonic is suppressed and only the odd ones survive. The lowest odd harmonic, namely the fundamental, dominates and the sound is comparatively pure. Bowing close to one bridge, where the wire is almost a node of the fundamental but almost an antinode of every higher harmonic, excites all the harmonics, and the resulting note carries a rich complement of overtones.

(b) For the third harmonic on a string of length $L = 1.00\text{m}$ fixed at both ends:

$$\lambda = \frac{2L}{3} = \frac{2(1.00\text{m})}{3} = 0.667\text{m}.$$

Nodes occur at $x = 0, \frac{L}{3}, \frac{2L}{3}, L = 0, 0.333\text{m}, 0.667\text{m}, 1.000\text{m}$.

Antinodes occur at $x = \frac{L}{6}, \frac{L}{2}, \frac{5L}{6} = 0.167\text{m}, 0.500\text{m}, 0.833\text{m}$.

Rider at $x = 0.250\text{m}$: lies between a node at 0 and an antinode at 0.167m , so the wire there has a non-zero displacement. The rider is thrown off.

Rider at $x = 0.333\text{m}$: lies exactly on a node, so the wire at this point does not move at all. The rider remains at rest.

Rider at $x = 0.500\text{m}$: lies exactly on an antinode, where the wire has its largest displacement. The rider is thrown off vigorously.

Example 47

- Account in physical terms for the experimental observation that two steel piano wires of the same length, material and tension produce different fundamental notes when their diameters differ.
- A steel piano wire of diameter 0.80mm and tension 600N produces a fundamental note of frequency 220Hz . The wire is replaced by a second steel wire of the same length under the same tension but of diameter 0.96mm . Find the fundamental frequency of the new wire.

Solution

(a) The fundamental frequency of a stretched wire is $f_1 = \frac{1}{2L} \sqrt{\frac{T}{\mu}}$, where the linear mass density of a uniform wire of circular cross-section is $\mu = \rho \pi d^2/4$. For wires of the same material, ρ is fixed and a thicker wire has a larger μ . At equal length and tension, a larger μ gives a smaller wave speed $v = \sqrt{T/\mu}$ and therefore a smaller fundamental frequency. Two wires that are otherwise identical sound at different pitches purely because their diameters set different masses per unit length.

(b) For wires of the same material, $\mu \propto d^2$. At constant tension and length, $f_1 \propto 1/\sqrt{\mu} \propto 1/d$. Hence

$$\frac{f_{\text{new}}}{f_{\text{old}}} = \frac{d_{\text{old}}}{d_{\text{new}}} = \frac{0.80\text{mm}}{0.96\text{mm}} = \frac{5}{6}$$

$$f_{\text{new}} = \frac{5}{6} (220\text{Hz}) = 183.3\text{Hz}$$

Example 48

- Distinguish between a progressive wave and a stationary wave with respect to the transport of energy along the medium and the variation of amplitude with position along the medium.
- The displacement of a stationary wave on a stretched string is given by $y(x,t) = 0.040\sin(5\pi x)\cos(60\pi t)\text{m}$, with x in metres and t in seconds. Find the amplitude of each constituent travelling wave, the wavelength, the speed of those travelling waves, the position of the antinode nearest to $x = 0$ and the displacement of that antinode at the instant $t = 0.025\text{s}$.

Solution

(a) A progressive wave transports energy continuously through the medium in the direction in which it travels; a stationary wave does not transport any net energy along the medium, energy being exchanged locally between kinetic and potential forms in each cell between consecutive nodes.

Every particle of a progressive wave oscillates with the same amplitude as every other particle. In a stationary wave the amplitude varies sinusoidally with position: it is zero at the nodes (spaced $\lambda/2$ apart) and is a maximum equal to twice the amplitude of either constituent travelling wave at the antinodes (located midway between consecutive nodes).

(b) Compare $y = 0.040\sin(5\pi x)\cos(60\pi t)\text{m}$ with $y = 2A\sin(kx)\cos(\omega t)$:

$$2A = 0.040\text{m}, \quad k = 5\pi\text{rad/m}, \quad \omega = 60\pi\text{rad/s}.$$

Amplitude of each constituent travelling wave:

$$A = \frac{0.040\text{m}}{2} = 0.020\text{m}$$

Wavelength:

$$\lambda = \frac{2\pi}{k} = \frac{2\pi}{5\pi\text{rad/m}} = 0.40\text{m}$$

Wave speed:

$$v = \frac{\omega}{k} = \frac{60\pi \text{ rad/s}}{5\pi \text{ rad/m}} = 12 \text{ m/s}$$

Antinodes are where $\sin(kx) = \pm 1$, i.e., $kx = (2n + 1)\pi/2$ for integer $n \geq 0$. The nearest to $x = 0$ corresponds to $n = 0$:

$$x = \frac{\pi/2}{k} = \frac{\pi/2}{5\pi \text{ rad/m}} = 0.10 \text{ m}$$

Displacement of that antinode at $t = 0.025 \text{ s}$:

$$y = (0.040 \text{ m}) \sin(5\pi \times 0.10) \cos(60\pi \times 0.025) = (0.040 \text{ m}) \sin \frac{\pi}{2} \cos \frac{3\pi}{2} = (0.040 \text{ m})(1)(0) = 0.$$

The antinode is momentarily passing through its equilibrium position, where it is moving with maximum speed.

Example 49

- (a) A steel wire on a stringed instrument tuned in a cool morning sounds noticeably flat by midday, even though no physical alteration is made to the instrument. Account for this observation.
- (b) A steel wire of length 0.65 m , cross-sectional area $8.0 \times 10^{-7} \text{ m}^2$, density 7800 kg/m^3 , Young's modulus $2.0 \times 10^{11} \text{ N/m}^2$ and linear thermal-expansion coefficient $\alpha = 1.2 \times 10^{-5} \text{ K}^{-1}$ is rigidly clamped between two fixed supports and tuned at 20°C to a fundamental frequency of 200 Hz . Find the fundamental frequency of the wire when its temperature has risen to 35°C .

Solution

(a) When the temperature rises, the wire tries to expand longitudinally because of thermal expansion. The supports are rigid and prevent that expansion, so the tension that the supports must apply to keep the wire of unchanged length is reduced by an amount $\Delta T = EA\alpha \Delta\theta$. The fundamental frequency of a stretched wire varies as $f_1 \propto \sqrt{T}$, so a smaller tension gives a smaller fundamental frequency: the wire goes flat.

(b) Linear mass density:

$$\mu = \rho A = (7800 \text{ kg/m}^3)(8.0 \times 10^{-7} \text{ m}^2) = 6.24 \times 10^{-3} \text{ kg/m}$$

Initial tension from $f_1 = \frac{1}{2L} \sqrt{\frac{T_1}{\mu}}$:

$$T_1 = 4\mu L^2 f_1^2 = 4(6.24 \times 10^{-3} \text{ kg/m})(0.65 \text{ m})^2 (200 \text{ Hz})^2 = 421.8 \text{ N}$$

Drop in tension caused by the temperature rise of $\Delta\theta = 35^\circ \text{C} - 20^\circ \text{C} = 15 \text{ K}$:

$$\Delta T = EA\alpha \Delta\theta = (2.0 \times 10^{11} \text{ N/m}^2)(8.0 \times 10^{-7} \text{ m}^2)(1.2 \times 10^{-5} \text{ K}^{-1})(15 \text{ K}) = 28.8 \text{ N}$$

New tension:

$$T_2 = T_1 - \Delta T = 421.8 \text{ N} - 28.8 \text{ N} = 393.0 \text{ N}$$

New fundamental frequency:

$$f_2 = f_1 \sqrt{\frac{T_2}{T_1}} = (200 \text{ Hz}) \sqrt{\frac{393.0 \text{ N}}{421.8 \text{ N}}} \approx 193 \text{ Hz}$$

The wire has dropped from 200 Hz to about 193 Hz , a flattening of roughly 7 Hz , in agreement with the qualitative observation in part (a).

Example 50

- (a) Account for the fact that a transverse pulse generated at the lower (free) end of a long heavy hanging rope speeds up as it travels upward toward the fixed support.
- (b) A uniform rope of length 5.0 m and mass 2.0 kg hangs vertically from a fixed support, with its lower end free. Find the time taken for a small transverse pulse to travel from the free lower end to the support. Take $g = 9.8 \text{ m/s}^2$.

Solution

(a) Each element of the rope must support the weight of the rope hanging beneath it. At a height y above the free end, the tension equals the weight of the segment of length y :

$$T(y) = \mu gy,$$

where μ is the linear mass density of the rope. The transverse wave speed at that height is

$$v(y) = \sqrt{\frac{T(y)}{\mu}} = \sqrt{gy},$$

which depends only on g and on the height y above the free end. As the pulse climbs, y grows and so does $v(y)$: the pulse speeds up because the tension supporting it grows.

(b) The pulse travels with the position-dependent speed $v(y) = \sqrt{gy}$. The time to travel from $y = 0$ to $y = L$ is

$$t = \int_0^L \frac{dy}{v(y)} = \frac{1}{\sqrt{g}} \int_0^L y^{-1/2} dy = \frac{1}{\sqrt{g}} [2\sqrt{y}]_0^L = 2\sqrt{\frac{L}{g}}$$

Substituting $L = 5.0\text{m}$ and $g = 9.8\text{m/s}^2$:

$$t = 2\sqrt{\frac{5.0\text{m}}{9.8\text{m/s}^2}} = 1.43\text{s}$$

The pulse takes about 1.43s to climb the rope from the free end to the support.

Example 51

- (a) Account for the fact that every particle of the medium through which a sinusoidal progressive wave passes executes simple harmonic motion about its undisturbed position.
- (b) A wave on a stretched string is given by $y(x, t) = 0.020\sin(40\pi t - \pi x)\text{m}$, with x in metres and t in seconds. For the particle at $x = 0.50\text{m}$, find the period of its motion, the maximum speed, the maximum acceleration, and the total mechanical energy per unit mass.

Solution

(a) For a sinusoidal travelling wave $y = A\sin(\omega t - kx)$, the displacement of the particle at any fixed position x_0 is $y(t) = A\sin(\omega t - kx_0)$. The acceleration of that particle is

$$\frac{\partial^2 y}{\partial t^2} = -A\omega^2 \sin(\omega t - kx_0) = -\omega^2 y$$

This is the defining equation of simple harmonic motion: acceleration directed toward, and proportional to, the displacement from equilibrium. Every particle of the medium is therefore a simple harmonic oscillator with the same angular frequency ω and the same amplitude A as every other particle, differing only in the constant phase shift $-kx_0$.

(b) From $y = 0.020\sin(40\pi t - \pi x)\text{m}$ we read off amplitude $A = 0.020\text{m}$ and angular frequency $\omega = 40\pi\text{rad/s}$, with a constant phase shift at $x = 0.50\text{m}$ that does not change any of the quantities asked for.

Period:

$$T = \frac{2\pi}{\omega} = \frac{2\pi}{40\pi\text{rad/s}} = \frac{1}{20}\text{s} = 0.050\text{s}$$

Maximum speed:

$$v_{p,\max} = A\omega = (0.020\text{m})(40\pi\text{rad/s}) = 0.8\pi\text{m/s} \approx 2.51\text{m/s}$$

Maximum acceleration:

$$a_{\max} = A\omega^2 = (0.020\text{m})(40\pi\text{rad/s})^2 = (0.020)(1600\pi^2)\text{m/s}^2 = 32\pi^2\text{m/s}^2 \approx 316\text{m/s}^2$$

Total mechanical energy per unit mass of the particle:

$$\frac{E}{m} = \frac{1}{2}\omega^2 A^2 = \frac{1}{2}(40\pi\text{rad/s})^2(0.020\text{m})^2 = \frac{1}{2}(1600\pi^2)(4.0 \times 10^{-4})\text{J/kg} = 0.32\pi^2\text{J/kg} \approx 3.16\text{J/kg}$$

Example 52

- (a) Account in physical terms for the experimental observation that a soprano holding a single sustained note close to a wine glass can shatter the glass, while singing a different note at the same volume close to the same glass produces no effect.
- (b) A wine glass has a natural mechanical frequency of 720Hz when struck. The minimum sound intensity at the glass required to shatter it is 12W/m^2 . A soprano's voice radiates 0.020W of acoustic power uniformly in all directions. Find the maximum distance from her mouth at which the glass can be shattered.

Solution

(a) A wine glass is a lightly damped mechanical oscillator with a sharp natural frequency, set by its mass and stiffness. When the singer holds a note exactly equal to that natural frequency, the glass is driven at resonance: each cycle of the sound wave delivers a small additional store of energy in step with the glass's own oscillation, and the amplitude of the glass's flexure builds up cycle by cycle until the strain in the glass exceeds its breaking strength. When the singer sings a different note the driving frequency is detuned from the natural frequency; the energy delivered per cycle is on average small (some cycles in phase, some out), the response amplitude saturates at a small value, and the glass merely vibrates faintly.

- (b) For a point source the intensity at a distance r is

$$I = \frac{P}{4\pi r^2}$$

Setting I equal to the shattering threshold:

$$r^2 = \frac{P}{4\pi I} = \frac{0.020\text{W}}{4\pi(12\text{W/m}^2)} = 1.33 \times 10^{-4}\text{m}^2$$

$$r = \sqrt{1.33 \times 10^{-4}\text{m}^2} = 0.0115\text{m} = 1.15\text{cm}$$

The glass can be shattered only if it is held within about 1.15cm of the singer's mouth, which is why operatic glass-shattering demonstrations are always performed with the glass held very close to the speaker.

With twenty miscellaneous worked examples now complete, the picture should feel nicely connected. It is time to stretch our understanding a little further. Welcome to the Digging Deeper Exercise!

DIGGING DEEPER EXERCISE 7

EXERCISE 7A: BINDER QUESTIONS

Question 1

Account in physical terms for the fact that the speed at which a progressive wave travels through a given medium does not depend on the amplitude of the wave.

Question 2

Kipanga writes the maximum speed of a particle of a transverse wave as $v_{p,\max} = A\omega$, while Kipute writes it as $v_{p,\max} = 2\pi fA$. Each student insists that the two formulas describe different physical quantities. Account for the fact that the two formulas are equivalent.

Question 3

Account in physical terms for the fact that the frequency of a progressive wave is preserved when the wave crosses a boundary between two different media.

Question 4

Show that the displacement amplitude of a wave radiating uniformly outward from a point source in open space falls as $1/r$ at a distance r from the source.

Question 5

Two coherent sources radiating into a hall produce regions of permanent silence by destructive interference. Account in physical terms for the fate of the energy delivered by the two sources at and around such a silent point.

Question 6

A student claims that the formula $A_R = 2A\cos(\Delta\phi/2)$ for the resultant amplitude of two superposing waves of the same frequency holds only when the two waves have equal amplitudes. Account in physical terms for this restriction.

Question 7

Classify each of the wave equations $y_1 = A\sin(\omega t - kx)$, $y_2 = 2A\sin(kx)\cos(\omega t)$ and $y_3 = A\sin(\omega t + kx)$ as describing a travelling or a stationary wave, giving reasons in each case.

Question 8

Two students argue about the energy of a stationary wave on a stretched string fixed at both ends. Kipanga claims that, since a stationary wave does not transport energy along the string, the string carries no kinetic energy. Kipute claims that every particle of the string is in motion (except at the nodes), so kinetic energy must be present. Resolve the disagreement.

Question 9

Account in physical terms for the experimental observation that the fundamental frequency of a stretched string fixed at both ends decreases as the linear mass density of the string is increased, with the length of the string and the tension in it both held constant.

Question 10

A student observing a stretched string vibrating in its third harmonic claims that, since the frequency is now three times that of the fundamental, every particle of the string must be moving three times as fast as it would be moving if the string were vibrating in the fundamental at the same antinode amplitude. Indicate, with reasons, whether this claim is correct.

Question 11

Explain in physical terms why a thicker uniform metal rod, made of the same material as a thinner one, transmits a longitudinal wave at the same speed as the thinner rod.

EXERCISE 7B: REAL QUESTIONS**Question 12**

A leaf placed on the calm surface of Lake Victoria, when a fisherman drops a stone a metre away from it, is observed to bob up and down repeatedly without travelling along with the ripples that pass beneath it. Account in physical terms for what is moving and what is not.

Question 13

A radiologist at the Bugando Hospital in Mwanza directs a 5MHz ultrasound pulse from the soft tissue of a patient's abdomen into the underlying muscle, where the wave travels at a higher speed than in fat. Account in physical terms for the changes in the wave's parameters at the boundary.

Question 14

A villager in Shinyanga, awaiting the arrival of his herd in the evening, presses his ear to the dry ground and hears the rumble of the cattle hooves long before the sound reaches him through the open air. Account in physical terms for why pressing the ear to the ground gives an earlier warning.

Question 15

Two coherent loudspeakers, mounted at opposite ends of the stage of the Karimjee Hall in Dar es Salaam and driven from the same amplifier at 680Hz, produce a steady note. A patron walking slowly along a line parallel to the stage notices that the loudness rises and falls at regular intervals. Account in physical terms for the variation.

Question 16

A guitarist tuning his instrument at a wedding rehearsal in Unguja sounds his fifth string together with a 440Hz reference fork and counts six beats per second. He tightens the string slightly and the beat rate falls to two beats per second. Identify, with reasons, the original frequency of the fifth string.

Question 17

A guitarist at a school concert in Mtwara presses a finger firmly against the twelfth fret of his guitar string and finds that the resulting note sounds exactly an octave above the open-string note. Deduce the position of the twelfth fret along the vibrating length of the string.

Question 18

A student in Chakechake, watching the cables of an electricity pylon swing sideways during a strong wind, notices that one cable in particular vibrates with very large amplitude at one specific wind speed but is comparatively still at all other wind speeds. Account in physical terms for the speed-specific vibration.

Question 19

A washing-line stretched tightly between two posts in a courtyard at Songea, when struck with a stick, sends a sharp transverse pulse to the other post in a fraction of a second. The same line, when its tension is reduced to a slack droop, sends a slow visible wave when struck. Account in physical terms for the difference in propagation speed.

Question 20

Two tuning forks, both nominally of frequency 440Hz, are sounded together in a music classroom in Tanga. Instead of a steady single note, the class hears a slow rhythmic rise and fall in loudness, repeating roughly twice each second. Account in physical terms for this rhythm.

Question 21

A pedestrian crossing the Bridge during a strong wind notices that the long stay-cables sing with a clearly audible note when the wind reaches one particular speed but are silent above and below it. Account in physical terms for why the singing is heard only at one particular wind speed.

EXERCISE 7C: HOT QUESTIONS**Question 22**

A transverse wave on a stretched string is described by $y(x, t) = 0.060\sin(50\pi t - 2\pi x)\text{m}$, with x in metres and t in seconds. Determine the wavelength, the period and the wave speed, and calculate the maximum speed and the maximum acceleration of any particle of the string.

Question 23

A point source radiates sound at an acoustic power of 4.0W uniformly in still air. Calculate the intensity of the sound at distances of 5.0m and 10.0m from the source, and the ratio of the displacement amplitudes at the two distances.

Question 24

Two waves of equal amplitude 4.0cm travel in the same direction along a stretched rope. Their displacements at a fixed point are $y_1 = 4.0\sin(\omega t + 5\pi/6)\text{cm}$ and $y_2 = 4.0\sin(\omega t + \pi/6)\text{cm}$. Determine the amplitude of the resultant and the phase angle of the resultant relative to y_1 .

Question 25

Two coherent loudspeakers are placed 1.50m apart and emit notes of frequency 680Hz in still air, where the speed of sound is 340m/s . A point P is observed at a distance of 4.00m from the first loudspeaker and 5.50m from the second. Compute the wavelength of the sound and the path difference at P , and state, with reasons, whether P lies in a region of constructive or of destructive interference.

Question 26

A tuning fork of frequency 512Hz is sounded together with a stretched sonometer wire under a tension of 90N . Eight beats per second are heard. The tension is then increased until the beats vanish. Determine, with reasons, the original frequency of the wire, and calculate the new tension at which the beats vanish.

Question 27

A stretched wire of length 0.80m , fixed at both ends and of linear mass density $\mu = 5.0 \times 10^{-3}\text{kg/m}$, is observed to vibrate with two antinodes between its supports at a frequency of 240Hz . Calculate the speed of a transverse wave along the wire, and the tension in the wire.

Question 28

A stretched wire of length 1.20m , fixed at both ends and held under a tension of 200N , has linear mass density $5.0 \times 10^{-3}\text{kg/m}$. Determine the wave speed and the fundamental frequency, and find the highest harmonic of the wire that can be heard by a listener whose upper limit of audible frequency is 20kHz .

Question 29

A copper wire has tensile strength $\sigma_{\max} = 2.2 \times 10^8\text{N/m}^2$ and density $\rho = 8900\text{kg/m}^3$. Show that the maximum speed at which a transverse wave can travel along the wire depends only on $\sigma_{\max}$ and ρ , and use the result to compute the maximum wave speed and the maximum tension that may be applied to a wire of cross-sectional area $1.5 \times 10^{-6}\text{m}^2$.

Question 30

An aluminium rod of length 50m has Young's modulus $Y = 7.0 \times 10^{10}\text{N/m}^2$ and density $\rho = 2700\text{kg/m}^3$. Calculate the speed of a longitudinal wave along the rod and the time taken by a pulse to travel from one end of the rod to the other.

Question 31

A stationary wave on a stretched string is given by $y(x, t) = 0.030\sin(4\pi x)\cos(80\pi t)\text{m}$, with x in metres and t in seconds. Determine the amplitude of each constituent travelling wave, the wavelength, the wave speed, and the position of the antinode nearest to $x = 0$. Calculate also the displacement of that antinode at the instant $t = 0.020\text{s}$.

Chapter 8

PHYSICAL OPTICS

INTRODUCTION

The morning assembly at Miono Secondary had run a little long because the head teacher had a great many announcements to make. By the time it was over the sun had moved high enough to send a sharp white shaft through the louvre windows and on to the back row of desks. Kipanga had brought a CD to school for reasons he could no longer remember, and the disc was lying face-up on the wooden top in front of him. As the light struck the reflective surface, the CD did something the sun, by itself, had never been able to do. It threw a fan of colour across the desk: red on the outside, then orange, yellow, green, blue, and a violet so faint it could only be seen out of the corner of one eye. The colours were not painted on the disc. They had not been there a moment before. They appeared with the light, shifted with the disc, and disappeared the moment Kipanga moved his head.

He stared at the colours for long enough that Kipute, sitting next to him, leaned across to see what was so interesting. Then Mr. Akilikubwa, who had been standing at the doorway of the form-five classroom waiting for the assembly to disperse, spotted the boys hunched over the desk and walked across without comment.

Kipanga: *"It looks like a prism."*

Mr. Akilikubwa: *"It is not a prism, Kipanga."*

Kipanga: *"But it is splitting white light into colours."*

Mr. Akilikubwa: *"It is doing exactly that. It is not, however, doing it the way a prism does it. A prism splits white light by refraction; the speed of light in glass depends on the colour, so different colours bend by different amounts. There is no glass here, and there is no refraction. What is doing the work is the surface of the disc itself, which has been ruled with about six hundred twenty-five microscopic tracks per millimetre during manufacture. You are holding a diffraction grating, and you have been carrying it in your school bag all term without knowing it. By the time this chapter is over, you will be able to compute the angle at which red comes out, and the angle at which violet comes out, and the spacing between successive tracks on this disc, all by ear and entirely from your head."*

He paused, picked up the disc, and tilted it under the light.

Mr. Akilikubwa: *"This is the first of three phenomena we are going to spend a chapter on. The second is the reason a soap bubble shows rainbow colours just before it bursts, and the reason a thin film of petrol left on a Mwanza puddle does the same a few minutes after the rain stops. The third is the reason your aunt's polaroid sunglasses kill the glare from the surface of Lake Victoria, and the reason an ordinary pair only darkens it. All three are wave phenomena, and all three need an explanation that the ray model of light, useful as it has been to us so far, simply cannot supply. Welcome to the part of optics where light remembers it is a wave."*

Welcome to Chapter 8. The previous chapter taught you what a wave is, what a wave equation looks like, and what happens when two waves meet at a point. This chapter takes those three pieces of knowledge and applies them, one after the other, to light. The result is a small library of phenomena that the ray optics of your earlier school years could not even pose, let alone solve. We are about to give a wave-mechanical account of why a CD shows rainbows; why a soap bubble displays a band of colour just before it bursts; why two slits in a screen, illuminated by a single colour of light, produce a striped pattern of bright and dark fringes on a wall behind them; why a single slit, illuminated by the same light, produces a paler, broader, but still striped pattern of its own; why a curved lens placed on a flat glass plate shows a target of dark and bright concentric rings under monochromatic light; and why some of the light reflected off a window pane is polarised while the light coming directly from the sun above is not.

These phenomena go by three collective names. Two waves of light, taking different paths from the same source and meeting at the same point on a screen, produce **interference**. A single wave of light, passing an obstacle whose size is comparable with its wavelength, produces **diffraction**. And the underlying transverse character of light, which the previous chapter mentioned and the next chapter will exploit, has the consequence that the orientation of the electric field along the direction of propagation can be selected; a wave whose electric field lives in a single plane is called **plane-polarised**, and the process of making such a wave is called **polarisation**. The chapter is divided into three blocks, one for each phenomenon, plus a closing block on what physical optics is good for in the world.

Three pieces of Chapter 7 we will need over and over. The first is the principle of superposition: when two waves arrive at the same point of the medium, the resultant displacement at that point is the algebraic sum of the displacements that each wave would produce on its own. The second is the relation between path difference and phase difference, namely that a path difference Δx , for waves of wavelength λ , corresponds to a phase difference $\Delta\phi$ given by

$$\Delta\phi = \frac{2\pi}{\lambda} \Delta x$$

This is the equation we will use to predict where two waves will reinforce and where they will cancel. The third is the inverse-square law for the intensity of a spherical wave, which we will need when we ask why light from independent lamps cannot give sustained interference, and why a laser pointer can.

Two ideas are entirely new in this chapter and have to be installed before the rest will work. The first is **coherence**, *a property of two sources that determines whether their waves can produce a stable interference pattern at all*. Two sources are coherent if they maintain a fixed phase relationship. Two sources are incoherent if their relative phase wanders around the cycle on a timescale much shorter than your eye can follow. Most natural light sources are incoherent, which is the reason interference fringes are not the first thing you see when you walk into a sunlit room. The trick of every interference experiment in this chapter is to take a single coherent source and split it into two paths, so that whatever phase the source happens to be in at any instant, both paths share it. The two paths then recombine at a screen, and the only thing that determines where the fringes fall is the geometry of the apparatus.

The second new idea is the **wavefront**, *the locus of points in identical phase*, together with the **wavelet**, *the secondary spherical wave that every point on a wavefront sends out*. **Huygens' principle**, which we are about to develop in the very next section, says that *the next wavefront is the envelope of all those wavelets*. With wavefronts and wavelets in hand we can derive the law of reflection and the law of refraction without using any rays at all, which is a satisfying proof that the wave model contains the ray model, rather than the other way around. Wavelets are also the foundation of every diffraction calculation later in the chapter.

A short warning before we start. This chapter contains four equations that look almost identical and mean different things, namely:

$$y_n = \frac{n\lambda D}{d}$$

Young's slits, position of the n -th bright fringe.

$$2t = n\lambda$$

Air wedge, condition for the n -th dark fringe.

$$D_n^2 = 4n\lambda a$$

Newton's rings, diameter of the n -th dark ring.

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

Diffraction grating, n -th principal maximum.

The first describes positions of bright spots on a screen behind two slits. The second describes the thickness of an air film at which destructive interference occurs in reflected light. The third describes the diameters of dark rings between a curved lens and a flat plate. The fourth describes the angles at which a diffraction grating sends its bright spectra. They share an integer n and a wavelength λ , and they all follow from the same path-difference logic, which is why they look alike on the page. Telling them apart is two thirds of the chapter's arithmetic, and the rest is doing the algebra after you have decided which one applies. We will spend a little time in each section showing why each equation is the right one for its own geometry, and where the others would lead a student astray if borrowed by accident.

One last orientation. **Optics** is the part of physics where the wave model of light buys back its admission ticket. In your O-level physics studies, when you traced rays through lenses and mirrors and built images on paper with a ruler, you got remarkably far without ever asking what light is. By the end of this chapter you will know why that worked, and where it stops working. It works whenever the apertures and the obstacles are large compared with λ , the wavelength of light, which for visible light lies between roughly four hundred nanometres and seven hundred nanometres. It stops working the moment the apertures and the obstacles shrink to that scale, which is the moment interference and diffraction take over and the simple straight-line picture has to give way to something richer. The richer picture is what the rest of this chapter delivers. It is also why, in many of the questions you will meet in your examinations, the items that draw most heavily on this chapter are also the items that pay the highest mark per minute of work invested. Read this chapter once. Then read it again with a pencil. Then put a CD on a desk under the school light and check, with your own eyes, that the physics is true. Let us begin.

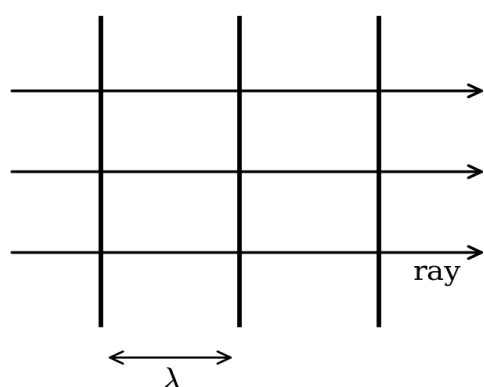
WAVEFRONTS AND HUYGENS' PRINCIPLE

Picture for a moment the surface of Lake Victoria after Mr. Akilikubwa's stone has dropped in. The ripples spread out as widening circles, neatly evenly spaced, all centred on the place where the stone hit the water. From above, what you see is not a single moving thing but a family of expanding circles. Now look at one of those circles, just one, and notice that every point on it is doing the same thing at the same instant. A crest is everywhere on the same circle; a trough is everywhere on the next-but-one circle, half a wavelength behind it; the place where the water surface is rising at exactly the still-water level is a third circle a quarter of a wavelength further behind. Each of these surfaces is a snapshot of one phase, captured around the lake at one instant. Each is what we are about to call a wavefront.

A **wavefront** is the locus of all points of a wave that are in identical phase at the same instant. The crest is a wavefront; the trough is a wavefront; the rising-zero-crossing is a wavefront. There are infinitely many wavefronts inside any wave, packed one against the next like layers of an onion. We usually draw only those that are exactly one wavelength apart, which gives the picture you have in your head from O-level: a series of equally spaced lines or rings, each one wavelength apart from the next.

Two pictures of wavefronts will dominate the rest of this chapter. The first is the **plane wavefront**: a flat surface, like a clean sheet of A4 paper, advancing forward without changing shape. Light arriving from the sun, by the time it has crossed 150 million kilometres to your eyes, has wavefronts that are essentially planes; the source is so far away that you cannot detect the curvature of the wavefronts on a benchtop with the unaided eye. A laser pointer also produces, very nearly, plane wavefronts, which is one of the reasons it will appear in this chapter so often. The second picture is the **spherical wavefront**: a sphere expanding outward, like the surface of an inflating balloon, centred on a single point source. The headlamp of a daladala bus on Sokoine Drive at midnight produces spherical wavefronts, at least until the light hits a wall and starts to do other things which we will discuss in due course.

(a) plane wavefronts



(b) spherical wavefronts

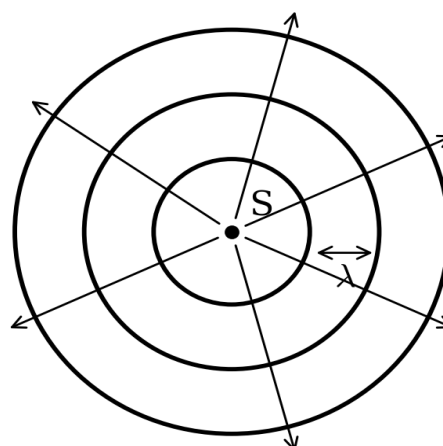


Figure: Plane and spherical wavefronts. Wavefronts are drawn as solid lines; rays, the perpendicular bookkeeping arrows, show the direction in which the wave is travelling.

Notice in the figure above the small arrows that we have drawn perpendicular to the wavefronts. These arrows point in the direction the wave is travelling, and we call them **rays**. A ray is therefore **not** a fundamental physical object; it is just a perpendicular bookkeeping mark on a wavefront, telling us which way the wavefront is advancing. The geometric optics you met in earlier years drew nothing but rays and got remarkably far. Wave optics will draw both wavefronts and rays, and it is the wavefronts that will carry the new physics.

Why did the ray-only picture succeed for so long without ever talking about wavefronts? Because for ordinary optical apparatus the wavelength of light is so much smaller than the apertures and the obstacles around it that nothing interesting happens at the scale of one wavelength. Wavefronts then advance like a tightly packed deck of playing cards, and the only thing worth tracking is which way the deck is moving, namely the ray. Wave optics, as a useful subject, begins exactly where this approximation breaks. We will hit that breaking point twice in this chapter: once when an aperture shrinks to the order of λ , and once again when two waves recombine after a brief separation. Both events demand wavefronts; neither can be solved

by rays alone. So we install the wave machinery now and rent it back to ourselves over and over for the rest of the chapter.

The single physical idea that will let us do this is what Christiaan Huygens, a Dutch physicist, proposed in 1678. He suggested that any wavefront could be replaced, with no loss of physical content, by a swarm of tiny secondary sources scattered over its surface. Each secondary source then sends out its own little spherical wavelet, the wavelets travel forward at the speed of the wave, and the next wavefront is whatever shape is tangent to all those wavelets a moment later. A wavefront, in other words, generates the next wavefront by sub-contracting the work to a million small spheres. Here is the principle stated cleanly.

Huygens' principle: *Every point on a wavefront acts as a source of secondary spherical wavelets that travel forward at the same speed as the wave. The next position of the wavefront, after a short time Δt , is the surface tangent to all those wavelets at that later instant.*

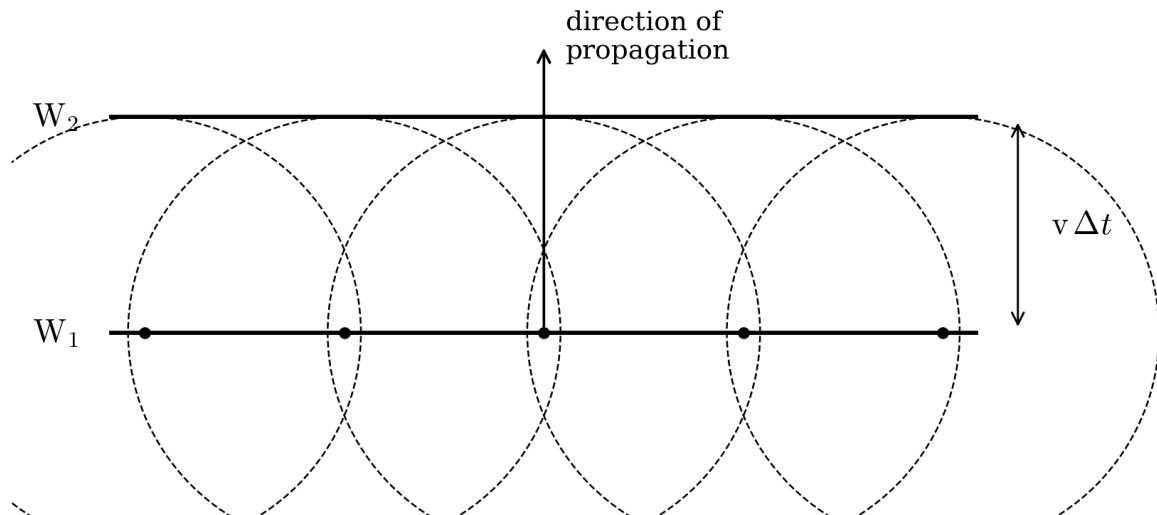


Figure: The Huygens construction. Five secondary sources are placed on the wavefront W_1 ; after a time Δt each has emitted a spherical wavelet of radius $v \Delta t$; the new wavefront W_2 is the line tangent to the forward edge of all the wavelets.

Three observations about the principle deserve a moment before we put it to work.

First, the secondary wavelets are spherical because the medium around them is the same in every direction; the wavelet does not know which way the wavefront was originally going, so it spreads in all directions equally. The original direction of the wave is recovered only when the envelope is taken: the envelope on the forward side is the new wavefront, and the would-be envelope on the backward side, mathematically, would correspond to a wavefront moving back into the wave that has already passed by. Huygens did not have an explanation for why we should ignore the backward envelope. The explanation came later, with Maxwell's equations, and amounts to saying that the backward wavelets cancel each other out leaving only the forward envelope. We will accept this on physical grounds for now.

Second, the principle is **scale-free**. It does not care whether the wave is a sound wave in air, a ripple on Lake Victoria, or a beam of red laser light. The same construction handles all of them, with only the speed v changing.

Third, and most importantly for what is to come, the principle is a **prediction machine**. Given the wavefront now, it tells us the wavefront a moment later. It is the wave-mechanical sister of Newton's second law of motion: $F = ma$ told you what happens to a particle next; Huygens tells you what happens to a wave next. From this single statement we will pull out *the law of reflection, Snell's law of refraction, the Young's double-slit fringe pattern, the single-slit diffraction pattern, the diffraction grating equation, and the Newton's rings formula*. Six results, one principle, no exceptions. It is the bargain of the chapter, and it costs only one figure on a piece of paper.

Let us spend the principle on something the student already believes, by way of a sanity check: *a flat wavefront travelling through a uniform medium continues forward as a flat wavefront*. This is the simplest possible test, and if Huygens fails it we have no business going any further with the chapter.

Take a horizontal flat wavefront W_1 at time t . Decorate it with five secondary sources, evenly spaced. After a short time Δt , each source has emitted a wavelet of radius $v\Delta t$. The envelope of these five circles, on the forward side, is a horizontal straight line a distance $v\Delta t$ above W_1 . That is the new wavefront W_2 . It is flat, it is parallel to W_1 , and it has advanced by $v\Delta t$. The wave model contains the ray model exactly as the previous section promised. Tick.

Sanity check passed. The principle now needs to earn its keep. We will spend it on two laws the student already knows but has never derived: **the law of reflection** and **Snell's law of refraction**. By the time we finish, both laws will be understood in a way that no amount of memorisation could ever achieve.

Reflection from Huygens

The figure below shows a plane wavefront AB approaching a flat reflecting surface XY at an angle. The wavefront is drawn at the very instant the leading point A first touches the surface. At that same instant, the trailing end of the wavefront, point B , is still in the air, with some travel left to do before it too reaches the surface.

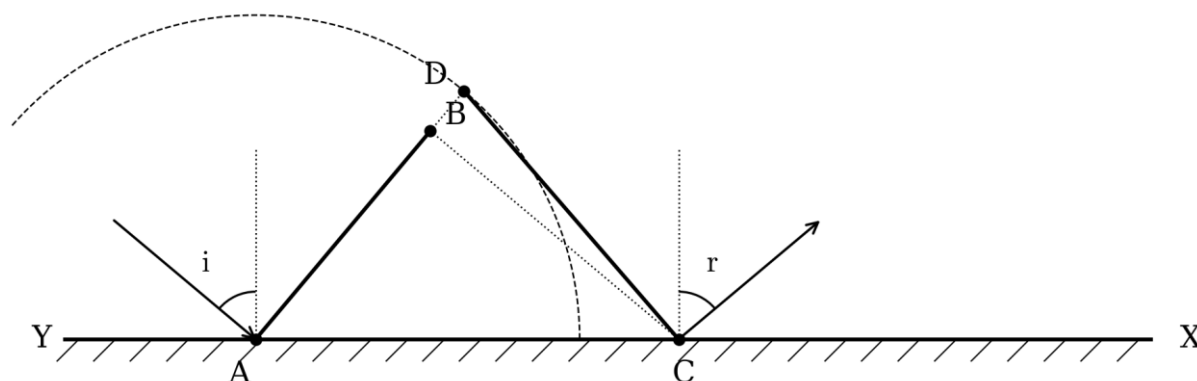


Figure: Reflection by the Huygens construction. AB is the incident wavefront at the moment A first touches the surface. The wavelet emitted from A grows to radius BC during the time the trailing edge of the wavefront takes to travel from B to C . The reflected wavefront CD is tangent to the wavelet at D .

Two physical facts go into the construction, and they are worth saying out loud before we touch any algebra.

The first fact is **timing**. The trailing point B reaches the surface at C only after a time $\Delta t = BC/v$, where v is the speed of the wave in the medium above the surface and BC is the distance from B to C measured along the path the trailing edge of the wavefront has yet to travel. During this same time interval, the wavelet that started at A has been spreading; its radius at the moment B arrives at C is $v\Delta t = BC$. So the wavelet at A has radius BC exactly when B has just arrived at C .

The second fact is the **boundary condition**. The reflecting surface is impenetrable; the wavelets cannot propagate into it. The new wavefront, therefore, is the envelope of secondary wavelets on the side of the surface that the original wave came from.

Now the construction. Draw a circle of radius BC centred at A . This is the wavelet that started when A first touched the surface, drawn at the same later instant when B has arrived at C . From C , draw the tangent to that circle. The tangent meets the circle at some point D . The reflected wavefront at this instant is the line CD .

The geometry now does the work for us. Triangle ABC is right-angled at B , with the angle at A equal to the angle of incidence i (the angle between the incident wavefront and the surface, equivalently the angle between the incident ray and the normal). Triangle ACD is right-angled at D , because CD is tangent to the circle and AD is the radius drawn to the tangent point. Both triangles share the hypotenuse AC . Therefore:

$$\sin i = \frac{BC}{AC} = \frac{AD}{AC} = \sin r$$

where r is the angle the reflected wavefront CD makes with the surface XY , and we have used $AD = BC$ because both equal the wavelet radius. The two sines are equal, so the two angles are equal. The angle of incidence equals the angle of reflection. Done.

That is the law of reflection, derived in three lines from the principle of a Dutchman who lived three hundred and fifty years ago. Notice what we did and did not need: we needed the wave model to give us the wavelets, we needed a tangent-and-radius right-angle, and we needed two right triangles to share a hypotenuse. We did not need to memorise anything. The phenomenon now belongs to you in a way that *just memorise* $i = r$ never could.

Refraction from Huygens: Snell's law

The same construction repeats for refraction, with one new ingredient: the speed of the wave in the second medium is different from the speed in the first. Call them v_1 and v_2 . As we will see in a moment, it is the change in speed that produces the change in direction. Speed is the cause; bending is the symptom.

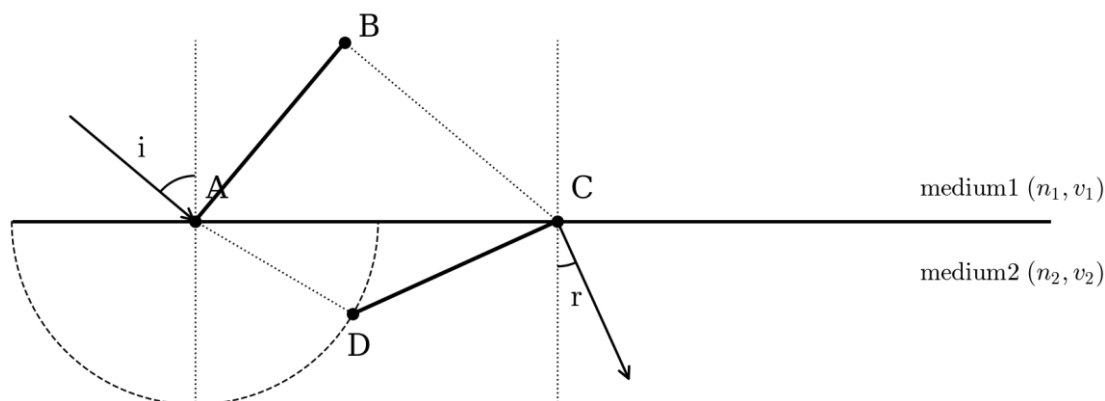


Figure: Refraction by the Huygens construction. The wavelet from A propagates into medium 2 with speed v_2 , so it has reached only a radius $(v_2/v_1) \cdot BC$ by the time the trailing edge of the wavefront has arrived at C in medium 1. The refracted wavefront CD is tangent to the wavelet at D in medium 2.

The figure above shows a plane wavefront AB approaching a flat boundary between two media at an angle. As before, A is the leading point of the wavefront that first touches the boundary; B is the trailing point still in medium 1. While B travels the remaining distance BC in medium 1, taking a time $\Delta t = \frac{BC}{v_1}$, the wavelet that started at A is propagating into medium 2 at the new speed v_2 . By the time B has arrived at C in medium 1, the wavelet at A has reached the radius AD given by:

$$AD = v_2 \Delta t = \frac{v_2}{v_1} BC$$

Draw, then, a circle of this radius centred at A but on the medium-2 side of the boundary. From C draw the tangent to that circle; the tangent meets the circle at some point D. The refracted wavefront at the instant B arrives at C is the line CD.

Two right-angled triangles, sharing the hypotenuse AC. In triangle ABC, $\sin i = \frac{BC}{AC}$. In triangle ACD, $\sin r = \frac{AD}{AC}$, and $AD = \frac{v_2}{v_1} BC$. Dividing the two sines,

$$\frac{\sin i}{\sin r} = \frac{v_1}{v_2}$$

Two practical reorganisations of this result. First, define the **refractive index** of medium 1 by $n_1 = c/v_1$, and of medium 2 by $n_2 = c/v_2$, where c is the speed of light in vacuum. Then:

$$n_1 \sin i = n_2 \sin r$$

which is Snell's law in the form your final examination will use. Second, observe that in passing from medium 1 to medium 2 the wavelength scales as $\frac{\lambda_2}{\lambda_1} = \frac{v_2}{v_1}$, while the frequency does not change at all. The frequency cannot change at a boundary because oscillations cannot be created or destroyed there, a result we already proved at the end of Chapter 7. The refracted wavefronts in the diagram, drawn one wavelength apart, are spaced more closely if the wave has slowed down ($n_2 > n_1$), and more widely apart if it has sped

up. The bend, the speed change, and the wavelength change are not three different things; they are three views of the same one thing.

Two worked examples will now exercise the new machinery, one numerical and one explanatory, while the ideas are still fresh in the head.

BINDER Example 1

Light from a laser pointer enters a basin of clean water from above, making an angle of 50° with the vertical. Calculate the angle the refracted beam makes with the vertical inside the water. The refractive index of water is 1.33.

Solution

Choose the vertical as the normal to the water surface. The angle of incidence in air is $i = 50^\circ$. The angle of refraction in water, the unknown we want, is r . Treat air as vacuum for this purpose, so $n_1 = 1.00$, and $n_2 = 1.33$ for water. Apply Snell's law,

$$n_1 \sin i = n_2 \sin r$$

Rearrange for $\sin r$:

$$\sin r = \frac{n_1}{n_2} \sin i = \frac{1.00}{1.33} \times \sin 50^\circ = (0.7519) \times (0.7660) = 0.5760$$

$$r = \sin^{-1}(0.5760) = 35.2^\circ$$

The refracted beam therefore makes an angle of 35.2° with the vertical inside the water.

Making Sense of the Answer: *The refraction angle (35.2°) is smaller than the incidence angle (50°), so the beam bent towards the normal as it entered the denser medium. That is exactly what the Huygens construction told us to expect: the wavelets in water are slower than the wavelets in air, so the side of the wavefront that hit the water first slowed down first and dragged the rest of the wavefront after it, like one wheel of a daladala bus dropping into a soft patch on the road and pulling the whole bus a little to the side. The same physics, the same answer. The laser pointer, sadly, will not work better in air for having been so wet.*

Think Like a Physicist: *Whenever you compute a refraction angle and the question gives you the refractive indices, write Snell's law in the form $n_1 \sin i = n_2 \sin r$ before you do anything else. Many marks have been lost over the years by students who wrote the inverse form by mistake and then refused to notice that their answer placed the refracted ray bending the wrong way. The cure is to draw a quick sketch with the normal, mark which medium is denser, and check at the end whether the ray bent towards or away from the normal. Bending towards the normal means going into a denser medium. Bending away means going into a less dense one. The arithmetic should agree with the geometry, every single time.*

REAL Example 2

On a clear afternoon on the Dodoma–Singida highway, the tarmac shimmers in the distance like a sheet of standing water. As the bus approaches, the puddle retreats; by the time the bus is over the spot where the puddle was, the spot is dry. Explain, using wavefronts and the dependence of light speed on the medium, why the driver sees water that is not there. Identify what physical quantity varies with height above the road and how that variation produces the illusion.

Solution

The starting point is the layer of air just above the tarmac. Several hours of midday sun have heated the road, the road has heated the air sitting on it, and the temperature of the air decreases smoothly with height for the first one or two metres above the surface. Density of a gas falls as its temperature rises, and a less dense gas has a lower refractive index. The refractive index of air therefore increases smoothly with height above the road. There is no sharp boundary anywhere; the index changes continuously with height.

Now turn to the wavefront. A ray of light coming from the sky and travelling almost parallel to the road carries a wavefront whose lower edge sits in the hotter, lower-index, higher-speed air, and whose upper edge sits in the cooler, higher-index, lower-speed air. By the Huygens construction we have just used twice, the lower edge of the wavefront advances faster than the upper edge during every short interval Δt . The wavefront tilts as it travels: the ray, perpendicular to the wavefront, bends upward, away from the road.

The driver's eye receives this bent ray. The eye, like every eye, assumes that light has travelled in a straight line from its source. It projects the incoming ray backward along that straight line to a virtual source

somewhere on the road in front of the bus, and reads off that virtual source as the origin of the light. The virtual source is a piece of the sky, apparently reflected off the tarmac: a sky-blue puddle, complete with shimmering edges where the temperature gradient is unsteady.

Why does the puddle retreat as the bus approaches? The geometry that places the virtual sky-image on the road in front of the bus depends on the angle at which the bent ray reaches the driver's eye, and that angle depends on the driver's distance from the apparent puddle. As the bus moves forward, the geometry shifts; the spot of road on which the virtual image lands shifts forward by the same amount; and the puddle slides forward in step with the bus. There is never any water; only a continuous refractive index pretending to be one.

The physical quantity that varies with height is the refractive index of the air. The variation is smooth, with no sharp boundary. The variation produces the illusion because the lower part of every wavefront crossing the gradient travels faster than the upper part, the wavefronts tilt upward, the rays curl up and away from the road, and the eye, refusing to believe in curving rays, projects them back to a virtual source on the tarmac.

Making Sense of the Answer: *This is the wave-mechanical account of a mirage, and the only physics it spends is what we already had two pages ago: speed depends on the medium, wavefronts move with that speed, and a wavefront whose two ends move at different speeds tilts. The same picture explains why the sun is geometrically below the horizon for nearly half a minute during which we still see it (the atmosphere bends the rays around the curve of the Earth), why stars appear to twinkle (random fluctuations of the index with altitude), and why a graded-index optical fibre can carry your data across a continent (the index is engineered to vary smoothly across the fibre's cross-section, so that all rays arrive in step). Whenever a medium varies smoothly, divide it in your head into infinitely many infinitely thin layers of constant index, apply Snell's law at each interface, and take the limit. This is what calculus looks like in optics.*

Think Like a Physicist: *When you encounter a phenomenon that looks like reflection without a reflector, check first whether the medium itself has a smoothly varying refractive index. Mirages on a hot road, the apparent flattening of the setting sun, the bending of starlight at low altitude, and the guiding of light along a graded-index fibre are all variations on a single theme. The mathematics is always the same: the wavefront is tangent to a swarm of secondary wavelets, the wavelets travel faster where the index is lower, and the wavefront curls toward the higher-index region. No new physics; only the same physics, paid out continuously instead of in lumps.*

We have now spent the principle three times. Once on a sanity check (a flat wavefront stays flat), once on the law of reflection, and once on Snell's law of refraction. The reward is that none of these laws is a memorised slogan any more. They follow, in three lines each, from the same physical statement that every point on a wavefront is a tiny secondary source. The next partition of this chapter asks the deeper question. *What happens when the wavefront meets an obstacle whose size is comparable with one wavelength?* That is the question Huygens answers most spectacularly, and it is where physical optics finally separates itself from geometric optics for good.

COHERENCE AND THE CONDITIONS FOR INTERFERENCE

Some weeks after the assembly-hall encounter with the CD, Kipute had decided to verify, on his own, the most surprising claim Mr. Akilikubwa had made all term. The teacher had said, repeatedly and with the enthusiasm of a man certain that he was about to be doubted, that two beams of light coming from the same source by two different paths can add up to darkness in some places and to double brightness in others. Kipute did not believe it. So one Saturday afternoon, with the school empty and the laboratory door unlocked, he borrowed two identical school torches, pointed them at a dark wall, and looked very hard for the alternating dark and bright stripes that ought to appear where the two beams overlapped. He saw, instead, a single uniform circle of light, slightly brighter than either torch alone. He looked harder. The circle stayed uniform. He moved the torches closer together, then further apart, dimmed them, brightened them, walked across the room. The wall stayed uniformly lit. After half an hour, he gave up.

Mr. Akilikubwa, who happened to be returning a borrowed laptop to the Headmaster's office, paused at the laboratory door.

Kipute: *"Sir, you said two beams of light could give bright and dark stripes."*

Mr. Akilikubwa: *"I did say that, Kipute. Repeatedly."*

Kipute: *"I cannot make it happen with two torches."*

Mr. Akilikubwa: “You will not, ever, with two torches. You could spend your entire life pointing torches at walls and you will not see fringes. Not because the physics is wrong, but because the torches are not the right kind of source. The previous chapter laid out the principle waves obey when they meet, which is superposition, but it left out the conditions sources must satisfy for the resulting pattern to be visible to a human eye. The most important of those conditions has a name we have not yet used in the book, and the name is **coherence**. Two torches, as you have just discovered, are **not** coherent. Two slits illuminated by the same lamp are coherent. A laser pointer, which we have not yet seen but soon will, is so coherent that two of its echoes off opposite walls of a corridor will give you fringes between themselves on a third wall behind your back. The whole of this section is about explaining what coherence is, why two ordinary lamps do not have it, and what physicists do about it. Bring your torches with you to next Monday's lesson, Kipute. We are going to use them as the negative example.”

Phase difference: A quick recap

Before we install coherence we owe ourselves one minute on what was already proved in the previous chapter. The principle of superposition told us that when two waves arrive at the same point of a medium their displacements add. The relation between the path-difference Δx of the two paths and the phase-difference $\Delta\phi$ of the two waves at that point is:

$$\Delta\phi = \frac{2\pi}{\lambda} \Delta x$$

Constructive interference occurs when $\Delta\phi$ is an even multiple of π (so the path difference is a whole number of wavelengths). Destructive interference occurs when $\Delta\phi$ is an odd multiple of π (path difference of half-wavelengths). At a fixed point in space, with two fixed paths, the value of $\Delta\phi$ is fixed by geometry, so each point on a screen has a definite condition: bright, dark, or somewhere between.

Hidden inside that last sentence is one assumption we have been making without saying so out loud. The assumption is that the two waves we are adding *share a stable phase relationship*. If they did not, $\Delta\phi$ at any fixed point would not be fixed; it would wander; and the bright-dark pattern would slide around faster than the eye could follow. Kipute's two torches are exactly this case. The whole of this section is about why.

Coherence

Two sources are **coherent** if the phase relationship between the waves they emit stays constant in time, or at least changes only on a timescale much longer than the time over which we are observing. They are **incoherent** if the phase relationship wanders unpredictably on a timescale shorter than ours.

To make this picture concrete, look at the figure below. The top panel shows the phase difference between two coherent sources plotted against time: it is flat, because the two sources march in lock-step and the difference between their phases never changes. The bottom panel shows the phase difference between two incoherent sources plotted against the same axes: it is a step function, jumping at random to a new value every short interval of duration τ_c . That short interval has a name; it is called the **coherence time** of the source.

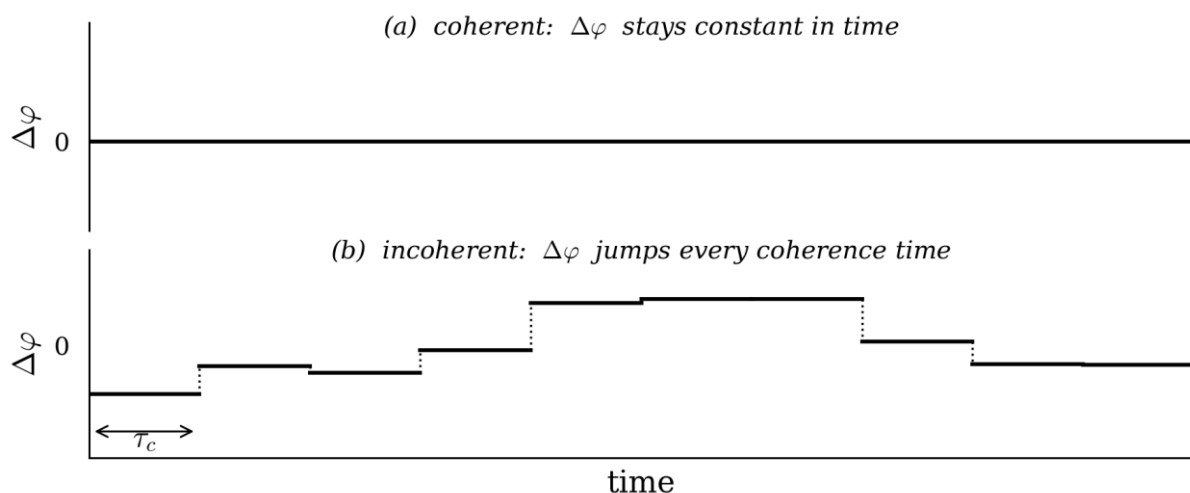


Figure: Phase difference between two sources, plotted against time. Coherent sources keep a constant phase difference; incoherent sources reset their phase difference to a random new value every coherence time τ_c .

How long is the coherence time? It is set by the spread Δf of frequencies the source emits. A source that emits a perfectly single frequency would have an infinite coherence time and a wave that marched in step with itself forever. A real source emits a small range of frequencies; the different frequencies drift out of step with one another, and the coherence time is roughly the time it takes them to drift by one cycle:

$$\tau_c \approx \frac{1}{\Delta f}$$

The narrower the frequency spread, the longer the coherence time. The **coherence length** of the source is the distance the wave travels in one coherence time,

$$\ell_c = c \tau_c \approx \frac{c}{\Delta f} \approx \frac{\lambda^2}{\Delta \lambda}$$

where the last form, equivalent because of $c = f\lambda$, is the version most often quoted in laboratories that measure light by wavelength rather than by frequency. Three real-world numbers fix the orders of magnitude. A sodium discharge lamp at room temperature has a frequency spread of about $1.5 \times 10^9 \text{ Hz}$, giving a coherence length of about 20 cm. A white-light incandescent bulb spans the whole visible spectrum, $\Delta \lambda$ of about 300 nm, giving a coherence length of about $1 \mu\text{m}$. A common helium-neon laser has a frequency spread of a few hundred kHz and a coherence length of several hundred metres. The factor of 30 million between the bulb and the laser is what every interference experiment in this chapter has to negotiate.

Why two independent lamps do not interfere

Now we can resolve Kipute's puzzle. A torch is a hot-filament lamp; the atoms in the filament emit light independently, and any single emission lasts only as long as the coherence time, after which the next emission begins from a fresh, randomly-chosen phase. The coherence time of a torch is comparable with that of a tungsten incandescent bulb, of order 10^{-14} s . Two torches, even if they emit at exactly the same wavelength, randomise their phases independently of each other. Their relative phase, the quantity $\Delta \phi$ the human eye would need to sit still in order to see fringes, jumps to a new random value every femtosecond.

The eye and the brain together integrate the light reaching them over a period of about 50 milliseconds, the so-called persistence of vision. During those 50 milliseconds, the two torches' relative phase has reset itself roughly 5 trillion times. The fringe pattern shifts trillions of times during one moment of seeing. The eye averages over the trillions of shifted patterns, and the average is uniform brightness. Kipute's wall, accordingly, looked uniform. There were fringes there, but they were sliding past too fast for the eye, the camera, the photographic plate, or any benchtop detector to catch.

The condition for fringes to be visible, then, is that *the observation time of the detector should be less than the coherence time of the source*. Either the detector must be unimaginably fast (faster than 10^{-14} s for two torches), or the source's coherence time must be made enormous, or, by far the cleanest workaround, the two waves must be made to share a single source's phase by construction. The third route is the one all the experiments in this chapter take, and it has its own name.

The split-source trick

Take a single source, monochromatic and bright. Split its light into two paths. Recombine the two paths at the screen. Whatever phase the source happens to be in at any one instant, both paths inherit that phase, and the relative phase between them at the screen depends only on the geometric difference of their path lengths. The source's phase wandering, however violent, cancels out: it is added to both waves and subtracted in the difference. The two paths, by construction, are perfectly coherent with each other for all time, regardless of the coherence time of the source itself.

There are exactly two ways to perform the split, and they cover every interference experiment we will meet for the rest of this chapter and every interference experiment that has ever been performed. The following figure shows them side by side.

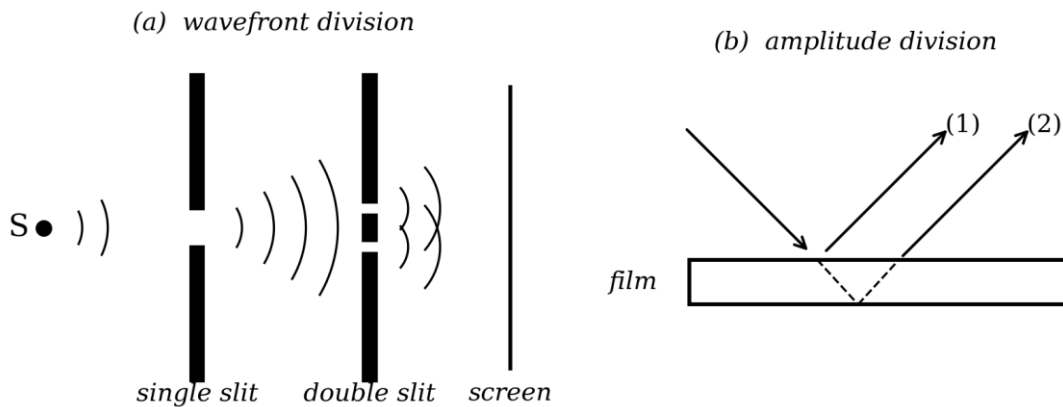


Figure: The split-source trick. (a) Wavefront division: a single wavefront is sliced into two by passing it through two narrow slits. The slits act as new coherent sources. (b) Amplitude division: a single ray is partially reflected at one surface and partially at another, so the same wave returns by two paths.

In **wavefront division**, panel (a) of the figure, the source S sends light through a single narrow slit, which guarantees that the light arriving at the next barrier is in identical phase across its width. That second barrier carries two narrow slits, S_1 and S_2 , which act as two new coherent sources. The geometry of the screen behind them gives different path lengths for waves arriving from S_1 and S_2 , and the path-difference logic produces the bright-dark pattern. Young's double-slit experiment, the Fresnel biprism, the Fresnel double mirror, and Lloyd's single mirror all belong to this family. We will treat Young's apparatus in detail in the next section.

In **amplitude division**, panel (b) of the figure, a single ray is split into two by partial reflection at the surfaces of a thin film. One reflection comes off the top surface; the other reflection comes off the bottom surface, after a round-trip through the film of thickness t and refractive index μ . The optical path-difference between the two reflections is $2\mu t$ plus a small contribution from any phase change at the surfaces (a contribution we will earn properly in a later section). Soap bubbles, oil films on rainwater, anti-reflective coatings on a camera lens, and the Newton's-rings apparatus all belong to this family. We will treat the latter two in their own sections.

The four conditions for sustained interference

We can now state, in clean numbered form, the conditions an experimental setup must satisfy for a stable bright-dark pattern of fringes to appear. There are four of them; the first one is the new physics of this section, the other three follow from common sense once it is on the page.

First condition: *The two waves must be **coherent**.* This means in practice that they must come from the same source by way of one of the two splitting tricks above. Two independent sources of any kind, even of identical wavelength, will not do.

Second condition: *The two waves should be **monochromatic**,* or near enough to it. A wide spread of wavelengths gives each colour its own fringe spacing, and the colours overwrite one another into a uniform glow after only a few orders. White light gives a few coloured fringes very close to the central bright fringe and washes out beyond. Sodium light gives several hundred clean fringes. A laser gives thousands.

Third condition: *The two waves should have **comparable amplitudes**.* The minima of an interference pattern are dark only when destructive interference is complete, and complete cancellation requires equal amplitudes. If one wave is much weaker than the other, the minima are merely dim, the contrast is poor, and the fringes are difficult to see even when they are there.

Fourth condition: *The two waves should have the **same polarisation state**.* Two waves whose electric fields oscillate in mutually perpendicular planes do not interfere with each other; they superpose into a wave with a tilted polarisation, but the intensities at every point on the screen simply add as for incoherent waves. We will return to polarisation in its own section later in this chapter; for now, accept that the experiments we are about to describe all use light whose polarisation is the same on both paths, because the same source feeds both paths and partial reflection does not turn one polarisation into the other.

Two worked examples now follow, because ideas of this kind are understood best, and enjoyed best, when they are put to work.

BINDER Example 3

A sodium discharge lamp emits its yellow line at $\lambda = 589\text{nm}$ with a frequency spread $\Delta f \approx 1.5 \times 10^9\text{Hz}$ arising mostly from Doppler broadening at operating temperature. Calculate (a) the coherence time and (b) the coherence length of this source. Comment on whether the lamp is suitable for a Young's double-slit experiment in which the path-difference between the two slits and the screen never exceeds a few millimetres.

Solution

(a) The coherence time follows directly from the frequency spread:

$$\tau_c \approx \frac{1}{\Delta f} = \frac{1}{1.5 \times 10^9 \text{ Hz}} = 6.7 \times 10^{-10} \text{ s} = 0.67 \text{ ns}$$

(b) The coherence length follows from the coherence time and the speed of light:

$$\ell_c = c \tau_c = (3.0 \times 10^8 \text{ m/s})(6.7 \times 10^{-10} \text{ s}) \approx 0.20 \text{ m} = 20 \text{ cm}$$

The coherence length is 20cm.

Path differences of a few millimetres in a Young's apparatus are smaller by a factor of about a hundred, so the lamp is comfortably suitable. We expect, and we will see, hundreds of clean fringes from a sodium lamp behind a Young's double slit.

Making Sense of the Answer: A coherence length of 20cm is a good benchtop number. It tells us that fringes will survive any path difference up to about 20cm before they wash out, which leaves enormous room for the typical Young's apparatus geometry of slit-separation under one millimetre and screen distance under one metre. It also tells us that the same lamp would not work for a Michelson interferometer with arms of a few metres, or for a holographic exposure that needs path-matching to better than a fraction of a wavelength: those experiments demand laser sources, whose coherence lengths are measured in hundreds of metres rather than centimetres. The single quantity ℓ_c tells us, before we do any geometry, what kind of interference experiment a given source can support.

Think Like a Physicist: When you are handed an interference problem, the first quantity to compute is the coherence length of the source. Compare it with the geometric path differences of the apparatus. If the path difference is much less than ℓ_c , the experiment will work and you can reach for the geometry. If the path difference is comparable with ℓ_c , expect a few faint fringes only. If the path difference exceeds ℓ_c , expect no fringes at all and look for a different source. The coherence length is the licence the source carries; the geometry is what you do with the licence.

REAL Example 4

A soap film of thickness about 500nm and refractive index 1.33, illuminated by ordinary daylight, displays bright iridescent colours that shift as the film thins. A clean window pane of thickness 3mm and refractive index 1.50, illuminated by exactly the same daylight, displays nothing of the kind. Both surfaces have two reflecting interfaces (front and back), so both ought, on the face of it, to give amplitude-division interference. Explain, using the coherence length of daylight, why one shows colours and the other does not.

Solution

Estimate first the coherence length of daylight. Visible light spans the wavelength range from about 400nm to 700nm, so the spread is $\Delta\lambda \approx 300\text{nm}$ centred on $\lambda \approx 550\text{nm}$. The coherence length is:

$$\ell_c \approx \frac{\lambda^2}{\Delta\lambda} = \frac{(550\text{nm})^2}{300\text{nm}} \approx 1.0 \times 10^{-6} \text{ m} = 1\mu\text{m}$$

Daylight has a coherence length of about one micrometre. Now compare it with the optical path differences in the two films.

For the soap film, the optical path between the front-surface reflection and the back-surface reflection is $2\mu\text{t} = 2(1.33)(500\text{nm}) \approx 1.3\mu\text{m}$. This is just barely larger than the coherence length, so the two reflections from the soap film still share a partial phase relationship. The interference is real but weak; only a few orders survive. Those few orders, with the path difference depending on the thickness t of the film and on the wavelength, give different colours at different thicknesses. As the film drains under gravity, the thickness varies smoothly with height and the colours move with it. We see iridescent bands.

For the window pane, the optical path is $2\mu t = 2(1.50)(3\text{mm}) = 9\text{mm} = 9 \times 10^3 \mu\text{m}$. This is nine thousand times larger than the coherence length of daylight. By the time the back-surface reflection emerges from the glass, it has lost any memory of the phase the front-surface reflection had when it began. The two reflections are now mutually incoherent, even though they came from the same source, because they have spent more than a coherence time apart. Their intensities simply add. There is no interference and no colour, only ordinary glass with two reflective surfaces.

Making Sense of the Answer: *This is why thin films display colour and thick plates do not. The threshold thickness is set by the source's coherence length divided by twice the refractive index, which for daylight on glass is about $0.3 \mu\text{m}$. Any film thinner than this gives interference; any film thicker than a few microns does not. Soap bubbles fall right at the limit, which is why they are vivid and ephemeral; the iridescent shimmer on a petrol-rainbow puddle the morning after rain works the same way; the anti-reflection coating on a camera lens, designed to be a quarter of one wavelength thick, sits well inside the limit and exploits it. The same physics, the same threshold, three different applications.*

Think Like a Physicist: *When you see two reflecting surfaces and you are wondering whether they will give interference, do not ask first whether the two waves are coherent at the source. Ask whether the optical path difference between the two reflections is smaller than the coherence length of the source. If yes, expect interference. If no, expect ordinary reflections that add in intensity, not in amplitude. The geometric thickness of the film is the diagnostic; everything else follows.*

We have now installed the new physics this chapter needed: coherence, the split-source trick, and the four conditions a real apparatus must satisfy for a stable fringe pattern to be visible. The next section picks up from here and treats, in geometric detail, the most famous and most heavily examined case of wavefront division: Thomas Young's double-slit experiment of 1801. We will derive the fringe-position formula from Pythagoras, work through five different parameter changes and what they each do to the fringe spacing, and finish with the only formula in the chapter that needs to be memorised verbatim because every paper sits one or two of its variants.

YOUNG'S DOUBLE-SLIT EXPERIMENT

Monday morning. Kipute has brought his two torches as instructed, slightly embarrassed at the thought that they are about to be told off in front of the whole class. The physics laboratory has been darkened with thick black cloth over the windows. A sodium lamp sits on a workbench, glowing the unmistakable yellow of every street lamp on Bagamoyo Road; in front of the lamp Mr. Akilikubwa has placed a vertical card with a single narrow slit cut into it; behind the first card and a metre further back stands a second card with two narrow slits scored a fraction of a millimetre apart; behind the second card stretches the length of the laboratory; behind the laboratory stands the white wall opposite the windows. Kipute's torches sit on the desk, switched off, awaiting their negative-example duty.

Mr. Akilikubwa switches on the sodium lamp, turns off the room lights, and asks the class to look at the wall. After a few seconds for the eyes to adjust, the wall is no longer dark. It carries a pattern of faint, evenly spaced, alternating bright and dark stripes, all yellow, running vertically. Kipute, who had spent half his Saturday looking for those stripes with two torches and finding nothing, leans forward and stares.

What the class is looking at is two hundred and twenty-five years old. Thomas Young, the English physician and physicist, set up exactly this apparatus in 1801, with a candle in place of the sodium lamp and a pair of pinholes in place of slits, and used the pattern on the wall to settle, once and for all, an argument that had run for a century and a half: light is a wave, the same kind of wave we have now spent two chapters deriving, and the bright-dark stripes on the wall are the constructive and destructive maxima of two coherent wavefronts adding their amplitudes at every point on the screen. The remainder of this section earns the formula that says exactly where each stripe falls.

The apparatus

The following figure shows the standard arrangement. A monochromatic source S sits on the left. Its light passes through a single narrow slit S_0 , which acts as the new effective source for the rest of the experiment. After S_0 , the light reaches a second card carrying two narrow slits S_1 and S_2 , separated by a small distance d . By the split-source argument of the previous section, S_1 and S_2 act as two new coherent sources, because both inherit their phase from the same wavefront that passed through S_0 . The two waves then propagate to a screen placed at a perpendicular distance D behind the double slit. The pattern of bright and dark stripes on the screen is the Young's interference pattern, and every formula in this section is about predicting where on the screen the bright stripes fall.

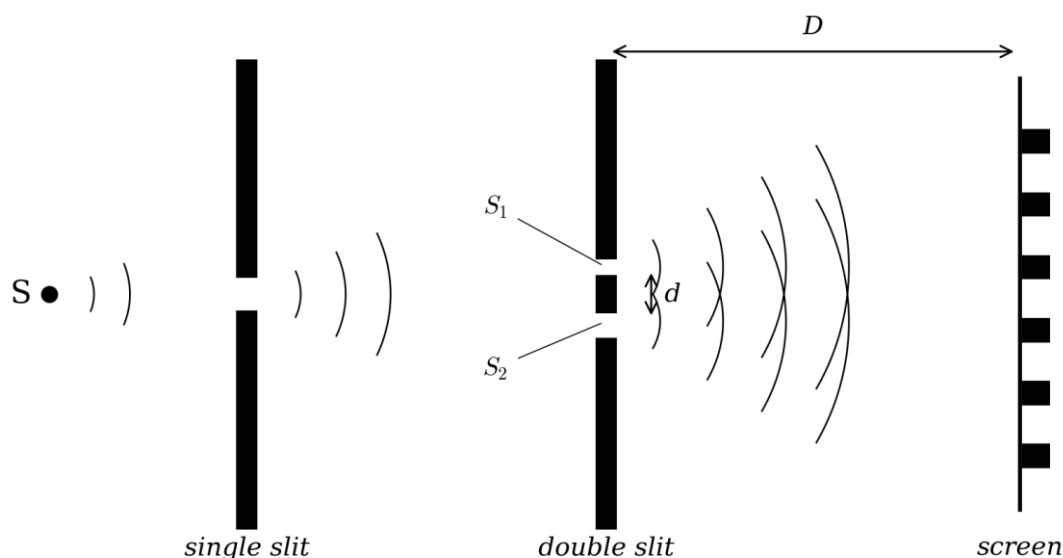


Figure: Young's double-slit apparatus. A monochromatic source S sends light through a single slit S_0 , which then illuminates a second card carrying two slits S_1 and S_2 separated by d . The two slits act as coherent sources; the screen at distance D behind the double slit displays the interference pattern.

Three remarks fix the orders of magnitude in your head before any algebra. **First**, the slit separation d is *small*. In a typical school benchtop apparatus d lies between 0.1mm and 1mm; if d is much smaller the fringes spread out so widely that only one or two fit on a sensible-sized screen, and if d is much larger the fringes crowd so close that the eye cannot resolve them. **Second**, the slit-to-screen distance D is *large*. Typical values run from 50cm to a few metres; D much larger than d is the geometric assumption that allows the small-angle approximation we are about to make. **Third**, the wavelength of light is *tiny*. For the sodium lamp on the bench, $\lambda \approx 589\text{nm}$; for a helium-neon laser, $\lambda \approx 633\text{nm}$; both numbers are tiny compared with even the smallest slit separation. The chapter's central inequality, λ *much less than* d , is what keeps the pattern visible at all.

Derivation of the fringe positions

To find where the bright fringes fall we follow the wave from each slit to a general point P on the screen and ask for the path-difference between the two waves there. The geometry is shown below.

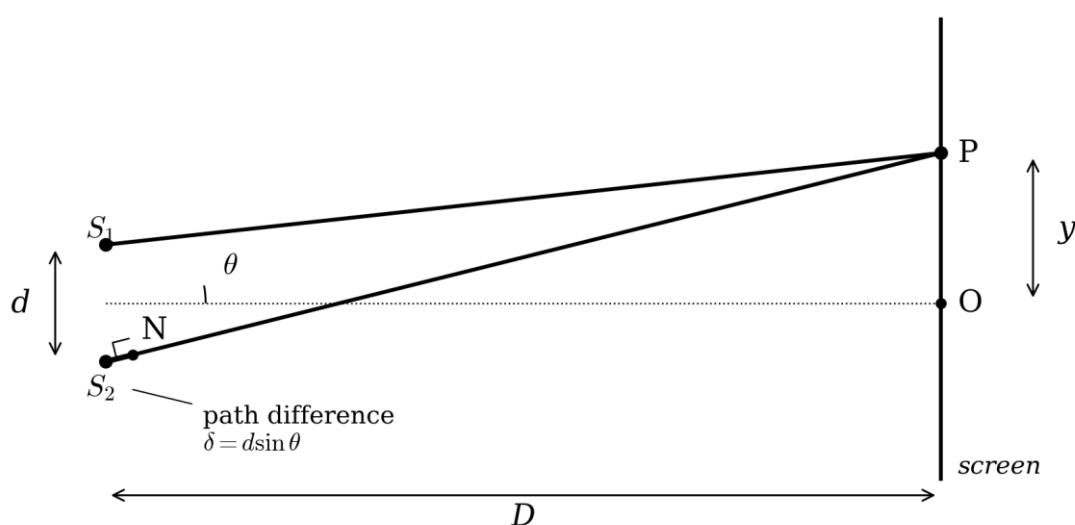


Figure: The path-difference geometry. P is a point on the screen at vertical distance y above the central axis O . The path difference $\delta = S_2P - S_1P$ is the segment from S_2 to the foot of perpendicular N dropped from S_1 onto S_2P ; in the small-angle limit $\delta \approx d \sin \theta \approx dy/D$.

Place S_1 and S_2 symmetrically about the central axis with separation d . Place the screen at distance D behind them. Let O be the foot of the perpendicular from the midpoint of the slits onto the screen, and let P be a general point on the screen at vertical distance y above O . We want the path-difference,

$$\delta = S_2P - S_1P$$

Drop a perpendicular from S_1 onto the line S_2P , and call the foot of that perpendicular N . In the far-field limit, where the screen is much further than the slit separation, the two paths S_1P and NP are very nearly equal in length, so:

$$\delta \approx S_2N$$

That is the path-difference, dressed in geometry. In the small-angle limit, where D is much greater than d and the angle θ between the central axis and the line from the midpoint of the slits to P is small, the segment S_2N equals $d \sin \theta$ by the right triangle S_1S_2N . Therefore:

$$\delta = d \sin \theta$$

In the same small-angle limit, $\sin \theta \approx \tan \theta = \frac{y}{D}$, so the path-difference at the point y on the screen is:

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

This is the master formula of Young's experiment. Every fringe condition follows from it by inserting the right value of δ . Constructive interference requires path-difference equal to an integer number of wavelengths,

$$\delta = n\lambda \quad (n = 0, \pm 1, \pm 2, \dots)$$

Combining the two equations and solving for y , the position of the n -th bright fringe is:

$$y_n = \frac{n \lambda D}{d}$$

The dark fringes lie halfway between the bright ones, with path-difference equal to a half-integer number of wavelengths,

$$\delta = (n + \frac{1}{2}) \lambda \quad (n = 0, \pm 1, \pm 2, \dots)$$

which gives the position of the n -th dark fringe as:

$$y_n = \frac{(n + \frac{1}{2}) \lambda D}{d}$$

These three formulas are the entire output of the section so far. Every numerical Young's-slit problem you will ever solve plugs into one of them. The next subsection notices a quantity that is even more useful in practice than the absolute fringe position, and gives it a name.

The fringe width

The **fringe width**, written β , is the perpendicular distance between any two consecutive bright fringes (or between any two consecutive dark fringes). From the position formula,

$$\beta = y_{n+1} - y_n = \frac{(n+1) \lambda D}{d} - \frac{n \lambda D}{d} = \frac{\lambda D}{d}$$

Three features of this single line are worth pinning down.

First, the fringe width depends only on the wavelength, the slit-to-screen distance, and the slit separation. It does not depend on which fringe you happen to be measuring; the bright fringes are evenly spaced and the dark fringes are evenly spaced. The pattern is a clean periodic stripe pattern with one period of width β .

Second, β scales linearly with λ and with D , and inversely with d . Doubling the wavelength doubles the fringe width. Doubling the screen distance doubles the fringe width. Doubling the slit separation halves it. These three scaling rules are the most useful single piece of arithmetic in the chapter.

Third, every fringe width formula in this chapter, no matter how unfamiliar the apparatus looks, comes back to $\beta = \lambda D/d$ in the limit where the screen is far compared with the slit separation. Newton's rings, single-slit diffraction patterns, diffraction-grating spectra, and the Fresnel biprism all reduce to scaled-up or scaled-down versions of this same formula in the small-angle limit. We are unlocking, in this section, a single physical idea that is going to recur three more times before the chapter is finished.

What happens when you change a parameter

Examiners love the same question dressed up five ways. The question is always: *what happens to the fringe pattern when one parameter of the apparatus is altered?* Five common alterations and their consequences follow.

Change one: the entire apparatus is **immersed in water** of refractive index $n = 1.33$. The wavelength of light in water is λ/n , so the fringe width becomes:

$$\beta' = \frac{(\lambda/n) D}{d} = \frac{\beta}{n}$$

The fringes squeeze together by a factor of n . For water this means the new spacing is $1/1.33 \approx 0.75$ of the old spacing. The pattern survives, only tighter.

Change two: the **slit separation d is reduced**. Since β is inversely proportional to d , halving d doubles the fringe width. Closer slits give wider fringes.

Change three: the **source slit S_0 is moved closer to the double slit S_1S_2** . The fringe width β itself is unchanged, because none of λ , D , or d has changed. What changes is the spatial coherence of the wavefront illuminating S_1S_2 : as S_0 approaches the double slit, the wavefront curvature across the gap becomes detectable, and the fringes become less crisp. The pattern is the same width but the contrast falls.

Change four: the **screen is moved closer to the double slit, halving D** . This is the change that confused Kipanga and Kipute when Mr. Akilikubwa tested them on it during the demonstration.

Mr. Akilikubwa: "I am about to bring the screen halfway closer. Predict what happens to the fringe width before I do."

Kipanga: "Halving the distance to the screen doubles the fringe width. Closer means the fringes spread out."

Kipute: "Halving the distance halves the fringe width. The closer the screen, the smaller the spacing."

Mr. Akilikubwa: "Two opposite predictions, two confident faces, one formula on the board. Read it. $\beta = \lambda D/d$. The fringe width is proportional to D . Half D , half β . Kipute is right and Kipanga is not. The intuition that betrayed Kipanga is the one that says "closer always means more spread out"; that intuition would be correct if the angular spread of the rays was fixed and we were tracking the size of a divergent beam. But the angular spread here is set by $\sin \theta \approx \lambda/d$, which does not depend on D at all. The spacing of the fringes on the screen is the angular spread times the distance from the source to the screen, so spacing scales with D . Move the screen closer, the angular fan stays the same, the screen samples a smaller piece of it, and the spacing between fringes is correspondingly smaller. Move it further away and the spacing grows."

He moved the screen halfway closer. The fringes halved in spacing, exactly as the formula said they would.

Change five, and the most important one for examination purposes, is the introduction of a **thin transparent plate** of refractive index μ and thickness t in front of one of the two slits. This is the topic of the next subsection because it has its own clean formula and its own family of problems.

The fringe-shift formula

Place a thin transparent plate of refractive index μ and thickness t directly in front of slit S_1 , with all other geometry unchanged. The light passing through S_1 now traverses the plate and emerges with an extra optical path of μt instead of the simple geometric path t it would have taken in air. The increase in optical path length, compared with the no-plate case, is:

$$\Delta p = \mu t - t = (\mu - 1) t$$

This extra path is added to the wave from S_1 before recombination. The whole fringe pattern therefore shifts on the screen so that the central bright fringe (formerly at $y = 0$) moves to a new position where the geometric path-difference exactly cancels the extra optical path. From $\delta = dy/D = (\mu - 1)t$, the displacement y :

$$y = \frac{D (\mu - 1) t}{d}$$

This displacement is called the **fringe shift**, so we rename it:

$$\Delta x = \frac{D (\mu - 1) t}{d}$$

Three consequences are worth noting. **First**, the entire pattern shifts as a rigid block; the fringe spacing β does not change, because the plate adds the same extra optical path to every wave from S_1 regardless of the angle. **Second**, the shift is towards the side where the plate was placed (because the central fringe now occurs where the plate-side wave is delayed enough to match the unplated wave). **Third**, the number of fringes that

have crossed any fixed point on the screen during the placement is $\frac{\Delta x}{\beta} = \frac{(\mu - 1)t}{\lambda}$, which gives an experimental method for measuring μ or t by counting the fringes that pass.

With the ideas now simmering nicely, next five worked examples are the moment we lift the lid and finally taste them, where the flavour of the physics rewards the patience.

BINDER Example 5

In a Young's experiment, a helium-neon laser of wavelength 632.8nm illuminates a card scored with two narrow parallel slits 0.20mm apart. The resulting interference pattern is observed on a white screen placed 1.50m behind the slits. Calculate (a) the perpendicular distance from the central bright fringe to the third bright fringe above it on the screen, and (b) the spacing between consecutive bright fringes in the pattern.

Solution

Begin by naming the quantities. Let $\lambda = 632.8\text{nm}$ be the wavelength of the light, $d = 0.20\text{mm}$ the separation between the two slits, and $D = 1.50\text{m}$ the perpendicular distance from the slits to the screen.

(a) The position of the n -th bright fringe is given by:

$$y_n = \frac{n \lambda D}{d}$$

Substitute $n = 3$, with all quantities in SI units:

$$y_3 = \frac{3 \times (632.8 \times 10^{-9}\text{m}) \times 1.50\text{m}}{0.20 \times 10^{-3}\text{m}} = 1.42 \times 10^{-2}\text{m} = 14.2\text{mm}$$

The third bright fringe lies 14.2mm above the central maximum.

(b) The fringe width is

$$\beta = \frac{\lambda D}{d} = \frac{(632.8 \times 10^{-9}\text{m}) \times 1.50\text{m}}{0.20 \times 10^{-3}\text{m}} \approx 4.7 \times 10^{-3}\text{m} = 4.7\text{mm}$$

The fringes are 4.7mm apart on the screen.

Making Sense of the Answer: Two checks confirm the answer. The fringe width times three equals the position of the third bright fringe: $3 \times 4.7\text{mm} = 14.1\text{mm}$, which matches $y_3 = 14.2\text{mm}$ to round-off. The size is also reasonable: a few millimetres is exactly what the eye can comfortably see at arm's length, which is why every Young's apparatus that has ever been built ends up with d in the tenths-of-a-millimetre range and D somewhere between half a metre and a couple of metres. Smaller d gives wider fringes; bigger D gives wider fringes; both choices are made to push the fringes into the comfortable arm's-length range. The apparatus is a piece of pre-tuned arithmetic that someone solved for you, decades before you arrived in the laboratory; it would be ungrateful not to use it. If your computed y_3 comes out in metres rather than millimetres, the school's screen has somehow grown to the size of a football pitch and the calculation needs a unit conversion before anything else.

Think Like a Physicist: Always check that the position of the n -th fringe equals n times the fringe width before you trust your numerical answer. The two formulas are derived from the same logic and any rounding error that breaks one will break the other in the same direction. If they disagree by more than the rounding tolerance you have, with overwhelming probability, made an arithmetic slip somewhere in y_3 ; redo the substitution.

BINDER Example 6

In a Young's apparatus, a green light source of wavelength 500nm illuminates a card carrying two narrow parallel slits. A travelling microscope is focused on the screen, which sits 1.0m behind the double slit, and counts 21 consecutive bright fringes occupying a horizontal distance of 10mm. Calculate the separation between the two slits.

Solution

Let $\lambda = 500\text{nm}$ denote the wavelength of the light and $D = 1.0\text{m}$ the slit-to-screen distance. The slit separation, which we wish to find, is d .

Twenty-one bright fringes have **twenty** intervals between them (the first and the last fringes bracket twenty fringe-widths). The fringe width is therefore:

$$\beta = \frac{10.0\text{mm}}{20} = 0.50\text{mm}$$

Rearrange the fringe-width formula for the slit separation:

$$d = \frac{\lambda D}{\beta} = \frac{(500 \times 10^{-9} \text{m}) \times 1.0 \text{m}}{0.50 \times 10^{-3} \text{m}} = 1.0 \times 10^{-3} \text{m} = 1.0 \text{mm}$$

The slits are 1.0mm apart.

Making Sense of the Answer: *There is exactly one trap in this question, and the trap has caught more students over the years than any other detail in Young's slits arithmetic. Twenty-one bright fringes are not twenty-one intervals; they are twenty intervals, the same way twenty-one fence posts make a fence with only twenty panels. Off-by-one in fringe counting halves the fringe width and doubles the slit separation, neatly converting an honest 1.0mm slit separation into a non-existent 0.50mm one. Whenever you read "N bright fringes spanning x millimetres", divide by N - 1, not by N. The fence post is the metaphor that has saved a thousand examination scripts.*

Think Like a Physicist: *Whenever a Young's-slits problem gives you a fringe count instead of a fringe width, your first arithmetic step is to translate count into spacing using "intervals = fringes - 1". Only after this conversion does the formula $\beta = \lambda D/d$ earn its keep. The translation is the kind of small, quiet step that separates a student who has merely memorised the formula from a student who has understood what the formula is counting.*

REAL Example 7

A Young's apparatus illuminated by a sodium lamp produces bright fringes spaced exactly 1.0mm apart on a screen placed 1.0m behind the double slit. The light from the sodium lamp has a wavelength of 589nm. Without disturbing any of the geometry, the entire apparatus is then submerged in a tank of clean water whose refractive index is 1.33. Predict the new fringe spacing, and explain in physical terms why the change occurs.

Solution.

Let $\lambda = 589 \text{nm}$ be the wavelength of the light in air, $\beta = 1.0 \text{mm}$ the original fringe spacing, $D = 1.0 \text{m}$ the slit-to-screen distance, and $n = 1.33$ the refractive index of the water.

In water of refractive index n , the wavelength of light is shorter:

$$\lambda' = \frac{\lambda}{n}$$

The fringe spacing depends on the wavelength of light in the medium where the interference takes place, not on the vacuum wavelength. With d and D both unchanged (these are geometric distances measured in physical space and are not altered by immersion), the new fringe width is:

$$\beta' = \frac{\lambda' D}{d} = \frac{\lambda D}{nd} = \frac{\beta}{n} = \frac{1.0 \text{mm}}{1.33} \approx 0.75 \text{mm}$$

The new fringe spacing is about 0.75mm. The fringes squeeze together by a factor of $n = 1.33$.

Making Sense of the Answer: *The picture is straightforward once you see what is changing. The geometric path-difference between the two slits and a point at height y on the screen is still dy/D , exactly as it was in air, because none of d , D , or y is altered by the surrounding medium. What is altered is the wavelength used to convert that geometric path-difference into a phase-difference. In air, a path-difference of one wavelength gave the first bright fringe at $y = \lambda D/d$. In water, the wavelength is shorter, so the same first-bright-fringe condition is met at a smaller y , namely $\lambda D/(nd) = \beta/n$. The fringes are tighter because the wave has been compressed by the medium, not because the geometry has changed. The same physics, the same screen, just water in the way of the light.*

Think Like a Physicist: *Whenever a Young's apparatus is altered, ask three questions in this order. **First, has the wavelength of light in the medium changed?** If yes, β scales as the new wavelength. **Second, has the slit separation d changed?** If yes, β scales inversely with d . **Third, has the screen distance D changed?** If yes, β scales linearly with D . The fringe width formula $\beta = \lambda D/d$ is a permission slip with three filled-in fields; whenever any field is altered, β changes in the obvious proportionate way and nothing else is needed. Most exam questions on Young's slits are exactly this template dressed in different clothing.*

HOT Example 8

In a Young's experiment, monochromatic light of wavelength 600nm produces bright fringes that are 1.2mm apart on a screen at some perpendicular distance behind the double slit. Without disturbing the slit separation,

the screen is then moved halfway closer to the slits, and the original light source is replaced with a different monochromatic source of another colour. The new bright fringes are observed to be 0.40mm apart on the screen. Calculate the wavelength of the new light.

Solution

Let $\lambda_1 = 600\text{nm}$ be the wavelength of the original source, $\beta_1 = 1.2\text{mm}$ the original fringe width, λ_2 the wavelength of the new source (the unknown we want), and $\beta_2 = 0.40\text{mm}$ the new fringe width. Let D denote the original slit-to-screen distance, so that the new distance is $D/2$. The slit separation d is the same in both situations, so the ratio $\beta/(\lambda D) = 1/d$ is constant. Apply this twice and divide:

$$\frac{\beta_1}{\lambda_1 D} = \frac{\beta_2}{\lambda_2 (D/2)}$$

Solve for λ_2 :

$$\lambda_2 = \lambda_1 \times \frac{\beta_2}{\beta_1} \times 2 = 600\text{nm} \times \frac{0.40}{1.2} \times 2 = 400\text{nm}$$

The new wavelength is 400nm, which is the deep-violet end of the visible spectrum, very nearly ultraviolet.

Making Sense of the Answer: Two changes acted on the fringe width here, and both shrank it: D was halved (which on its own would halve β), and the wavelength was shortened (which on its own would shrink β in proportion). The combined shrinkage is exactly 1.2mm to 0.40mm, a factor of three. Halving D gave one factor of two; the rest, a factor of 2/3 in the same direction, came from the colour change. The arithmetic is two steps stacked on top of each other, and the trick is to keep them straight in the head: each parameter pulls β in its own direction, and the final β is the product of the pulls.

Think Like a Physicist: When a problem changes more than one parameter at once, never substitute numbers into $\beta = \lambda D/d$ on each side and divide. Instead, write the formula symbolically for both situations, divide one by the other, and watch the unchanged quantities (here, the slit separation d) cancel out. The cancellation is what kept this problem solvable without knowing d at all. Whenever you find yourself needing a quantity the problem refused to give you, suspect that it is meant to cancel.

HOT Example 9

A thin glass plate, of refractive index 1.50 and thickness $6.0\mu\text{m}$, is placed directly in front of one of the two slits of a Young's apparatus. The two slits are 0.50mm apart, and the screen sits 1.0m behind the double slit. The apparatus is illuminated by yellow light of wavelength 500nm. Calculate (a) the lateral shift of the central bright fringe on the screen, and (b) the number of bright fringes that cross any fixed point on the screen as the plate is being slid into position.

Solution

Let $\mu = 1.50$ denote the refractive index of the plate, $t = 6.0\mu\text{m}$ its thickness, $d = 0.50\text{mm}$ the slit separation, $D = 1.0\text{m}$ the slit-to-screen distance, and $\lambda = 500\text{nm}$ the wavelength of the light.

(a) The fringe-shift formula gives:

$$\Delta x = \frac{D(\mu - 1)t}{d} = \frac{(1.0\text{m})(0.50)(6.0 \times 10^{-6}\text{m})}{0.50 \times 10^{-3}\text{m}} = 6.0 \times 10^{-3}\text{m} = 6.0\text{mm}$$

The central bright fringe moves 6.0mm to the side on which the plate was placed.

(b) The fringe width of the apparatus, before any shift, is:

$$\beta = \frac{\lambda D}{d} = \frac{(500 \times 10^{-9}\text{m})(1.0\text{m})}{0.50 \times 10^{-3}\text{m}} = 1.0\text{mm}$$

The number of fringes that pass any fixed point on the screen as the plate is inserted is therefore:

$$N = \frac{\Delta x}{\beta} = \frac{(\mu - 1)t}{\lambda} = \frac{(0.50)(6.0 \times 10^{-6}\text{m})}{500 \times 10^{-9}\text{m}} = 6$$

Six bright fringes cross any fixed point on the screen during the insertion.

Making Sense of the Answer: Notice that the answer to part (b) does not contain D or d at all; the number of fringes is set entirely by the extra optical path $(\mu - 1)t$ and the wavelength λ , both of which are properties of the light and the plate. This is one of the prettiest cancellations in the whole chapter. The geometric details of the apparatus, the very details examiners use to dress up the problem in different clothing, drop out the

moment you ask for the number of fringes that have passed. The fringe count is, in this sense, more fundamental than the fringe shift: it depends only on physics, not on benchtop layout. The same plate inserted into a Newton's rings apparatus, or any other amplitude-division setup, would shift exactly the same number of fringes past any fixed observation point.

Think Like a Physicist: Inverting the relation $N = (\mu - 1)t/\lambda$ gives the experimental method that several Nobel-prize-quality measurements of refractive index and small thicknesses have rested on for over a century: count the fringes, know the wavelength, deduce the optical thickness. Whenever a question gives you fringe shift in fringe-counts rather than millimetres, this is the formula to reach for. Whenever you yourself want to measure the refractive index of an unknown thin film and you have a Young's apparatus on the bench, this is the experiment to perform.

We have now completed the most important section of the chapter. Young's double-slit experiment, with its tightly packed bright and dark fringes a few millimetres apart on the wall, is the cleanest example of the split-source trick the previous section installed. The next section turns to its sibling, amplitude division, and treats the family of interference patterns produced by thin films and air wedges. The geometry there is different, the formula $2\mu t$ for the optical path-difference replaces dy/D , but the physics is identical to what we have just developed: two coherent waves arrive at a single point, their phase-difference is set by a path-difference, and the bright-dark pattern follows by the same constructive-and-destructive logic that has now served us through two chapters of wave physics.

THIN FILMS AND AIR WEDGES

After Young's demonstration, with the sodium lamp still glowing and the blackboard still bearing the wave-counting arithmetic from the previous lesson, Mr. Akilikubwa walked to the laboratory tap, dipped a wire loop into the beaker of soap solution he had left out overnight, and lifted out a thin elastic disc of soap film stretched flat across the loop. He held it up under the desk lamp and waited.

The class watched. For two or three seconds, while the film was still everywhere thick, it appeared a vague pale grey. Then, as the soap drained downward under gravity, the film began to draw colour. Bands of colour started running across the disc, narrowest near the top, broadest near the bottom, slowly creeping downward as the film thinned: a violet just below the black band, then blue, then green, then yellow, then orange, then red, in unbroken sequence from the top of the rainbow to its bottom. At the very top of the loop, where the film had thinned the most, a band of pure black appeared, grew, and rose, until it dropped off the top edge and a fresh wave of colour took its place behind it.

Mr. Akilikubwa, still holding the loop: *"A soap film, lit by ordinary white light, behaves like a rainbow ironed flat and pinned to a wire frame: violet at the top of the colours, just below the black band; blue and green next; yellow, orange, and red at the bottom. Two questions, both of which this section will answer. First, why colours at all? Daylight is white. The film is colourless. The two together produce, in our window, every wavelength of the visible spectrum, neatly stratified by height like volcanoes by altitude. Where did the colours come from? Second, why is the very top of the film black? The black band sits exactly where the film is thinnest, just before it bursts. The thinner the film, the less light it should reflect at all; why does the limit of vanishing thickness produce a perfectly dark reflection rather than a slightly faint one?"*

This section is about the answer. The answer is interference, but of the second kind installed in the coherence section: not the wavefront-division kind of Young's slits but the amplitude-division kind that splits a single wave into two by partial reflection at two parallel surfaces, lets the two parts travel along nearly identical paths through a thin layer, and recombines them at the eye of the observer. We have already seen the trick in outline. We now derive its formulas, apply them to soap films, to the air wedge between two clamped glass plates, and to the anti-reflective coatings that camera lenses depend on, and answer Mr. Akilikubwa's three questions exactly.

Amplitude division revisited

Picture the soap film of the demonstration, or any thin transparent film: a sheet of uniform thickness t and refractive index μ , with air of refractive index 1 above and below. Imagine a single ray of monochromatic light, of wavelength λ in vacuum, descending onto the top surface of the film from above. At that top surface two things happen at once, with the partial-reflection coefficient set by the refractive-index contrast across the boundary. Part of the wave reflects directly off the top surface and travels back up; call this reflection R_1 . The remaining part of the wave refracts into the film, travels down through the film at the refraction angle, reaches the bottom surface, and there partially reflects again. The bottom-surface reflection travels

back up through the film, refracts out at the top surface, and emerges into the air above moving parallel to R_1 but laterally displaced by a small amount; call this twice-bounced wave R_2 .

The two waves R_1 and R_2 are coherent with each other (because they came from the same incident wave by amplitude division at the top surface) and they meet again at the eye or at the screen above the film. Whether the meeting is constructive (a bright fringe in the reflected light) or destructive (a dark fringe in the reflected light) is set, exactly as in Young's slits, by the phase difference between the two waves at the observation point. The phase difference here comes from two contributions: the geometric round-trip through the film, which depends on the film's thickness and its refractive index, and a phase shift at the top-surface reflection that wavefront division never had to worry about.

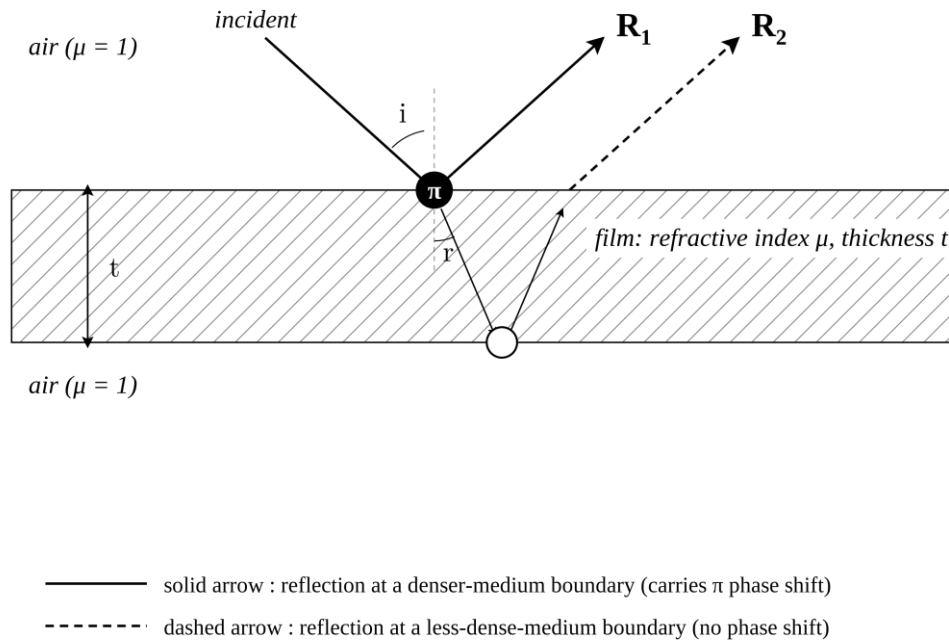


Figure: Amplitude division at a thin film. The incident ray reflects partially at the top surface as R_1 (which suffers a π phase change because the film is denser than the air above it) and partially refracts into the film at angle r . The refracted ray reflects off the bottom surface without a phase change (the air beneath is less dense), returns to the top surface, refracts back out, and emerges as R_2 parallel to R_1 but laterally displaced.

The phase change of π at the denser surface

One physical fact is owed before any formula. When a wave reflects off the boundary between two media, the reflected wave sometimes flips sign and sometimes does not, depending on which medium is denser. The rule, stated cleanly so it can be memorised in one sitting:

A wave reflecting off the surface of a medium that is optically denser than the one it came from suffers a phase change of π radians; equivalently, the reflected wave picks up an extra optical path of $\frac{\lambda}{2}$ relative to the incident wave. A wave reflecting off the surface of a medium that is optically less dense than the one it came from suffers no phase change at all; the reflected wave returns in phase with the incident wave.

The rule is not new. The same rule governed the reflection of a transverse wave on a stretched string from the previous chapter, where the wave was inverted at a tied-down end (a hard end) and was not inverted at a free end (a soft end). Light treats the denser medium as the hard end and the less-dense medium as the soft end, and the analogy is exact. Maxwell's equations supply the proof from the boundary conditions on the electric and magnetic fields, but the qualitative analogy with the string is enough for everything in this chapter.

Apply the rule now to our thin film. The wave R_1 reflects off the top surface of the film: it goes from air (less dense, $\mu = 1$) to film (denser, $\mu > 1$), and therefore picks up a phase shift of π , equivalently an extra $\frac{\lambda}{2}$ of optical path. The wave R_2 reflects off the bottom surface of the film: it is now inside the film going down, and reflects against the air beneath; light is going from the denser medium to the less dense, so no phase shift. Exactly one of the two reflections is hard-ended, and exactly one of the two waves carries an

unavoidable extra $\frac{\lambda}{2}$ with it back to the eye. The asymmetry between the two reflections is what makes thin-film interference different from Young's interference, and it is responsible for every “surprise” the rest of this section will produce, starting with the dark cap on Mr. Akilikubwa's soap film.

The optical path difference through the film

Now the geometry. At near-normal incidence (which is the case we shall treat throughout, because it covers the soap film, the air wedge, and the anti-reflection coating all at once) the wave inside the film travels down through the film and back up, a geometric round-trip of total length $2t$. But a metre of path inside a denser medium is not equivalent to a metre of path in air; the wavelength inside the film is shorter than the wavelength in vacuum by a factor of μ , and the optical path (measured in vacuum wavelengths) is therefore μ times the geometric path. The optical round-trip through the film is thus:

$$\text{Optical round – trip} = 2\mu t$$

(For a general angle of refraction r inside the film, the round-trip becomes $2\mu t \cos r$; at near-normal incidence $\cos r \approx 1$ and the expression collapses to $2\mu t$.)

The total optical path difference between R_1 and R_2 at the observer is the optical round-trip through the film, plus the half-wavelength length-equivalent of the π phase shift at the top surface. We give the optical path difference the symbol Δx . In length units it reads:

$$\Delta x = 2\mu t + \frac{\lambda}{2}$$

A word on terminology, because two closely related quantities live in every interference calculation and they are not interchangeable. The optical path difference Δx just written down has units of length (metres). The phase difference $\Delta\phi$ at the observer has units of radians. The two are related by the master conversion from Chapter 7, $\Delta\phi = \frac{2\pi}{\lambda} \Delta x$, so that any constructive-or-destructive condition can be written in either form. Throughout this section we work in length units, because the bright-fringe condition $\Delta x = n\lambda$ reads more cleanly than its phase-difference cousin $\Delta\phi = 2n\pi$, but the two are saying the same thing.

The extra $\frac{\lambda}{2}$ in Δx is the bookkeeping signature of the asymmetric reflection. Without it, the bright and dark conditions on a thin film would be the constructive and destructive cases of a simple geometric path difference $2\mu t$, with no surprises. With it, the conditions are systematically interchanged: what would have been the bright case is now the dark case, and vice versa. We turn the algebra next.

Conditions for bright and dark fringes

Constructive interference (bright fringe in reflected light) requires the total optical path difference to be an integer number of wavelengths:

$$\Delta x = 2\mu t + \frac{\lambda}{2} = n\lambda$$

Solving for the geometric part of the optical path:

$$2\mu t = \left(n - \frac{1}{2}\right)\lambda, \quad n = 1, 2, 3, \dots$$

Destructive interference (dark fringe in reflected light) requires the total optical path difference to be a half-integer number of wavelengths. Repeating the algebra with a half-integer on the right-hand side:

$$2\mu t + \frac{\lambda}{2} = \left(n + \frac{1}{2}\right)\lambda$$

$$2\mu t = n\lambda, \quad n = 0, 1, 2, \dots$$

Notice the value of n that the dark condition allows: $n = 0$, which corresponds to a film of vanishing thickness $t \rightarrow 0$. A film thinned all the way down to nothing exactly satisfies the dark-fringe condition. It reflects no light at any wavelength. The two reflections R_1 and R_2 from a film of zero thickness arrive at the eye almost coincident in space (because the bottom and the top surfaces have merged), almost coincident in time (because the round-trip through nothing takes no time), and exactly out of phase (because the top-surface reflection has flipped sign while the bottom-surface reflection has not). They cancel each other exactly, at every wavelength. That is the black band at the top of Mr. Akilikubwa's loop. The film has drained itself locally to a thickness much smaller than any visible wavelength, and the interference is destructive for every colour the eye can see. The film has become its own perfect anti-reflection coating, which is also the name we are going to give to the deliberate engineering version of this same physics three subsections from now.

Why the colour bands stack in the order they do

Below the black cap, the film has not yet drained to nothing. At each height up the loop, the film has some local thickness t set by gravity, surface tension, and the time the bubble has been hanging there. The bright-fringe condition can be solved at each height for the wavelength that is reinforced in the reflected light:

$$\lambda = \frac{2\mu t}{\left(n - \frac{1}{2}\right)}$$

With $n = 1$, the lowest non-trivial order, this gives $\lambda = 4\mu t$. For a soap film of refractive index roughly 1.33, the visible wavelengths between 400nm and 700nm are reinforced in the reflected light at thicknesses between 75nm and 132nm. As the film drains, the local thickness at any given height passes downward through this range, and the colour reinforced moves through the spectrum from red (700nm at 132nm thickness) to violet (400nm at 75nm thickness). The colours we see on the bubble are not the source colours but the wavelengths reinforced by the film thickness at each height, and they stack in the order set by the formula above.

The order is the reverse of the prism order, by the way: *a prism orders visible colours by their refractive indices, with red bent the least and violet bent the most. A soap film orders them by film thickness, with red at the thickest film and violet at the thinnest.* The fact that both orderings can be derived from the wave-mechanical account of light, by two completely different physical mechanisms (refraction in the prism, two-surface interference in the soap film), is one of the strongest internal-consistency arguments for the wave theory of light. Two roads, one map.

The Air Wedge

A particularly clean and quantitative version of the same physics is produced not by a soap film, which drains and bursts before any careful measurement can be performed, but by clamping two flat glass plates along one edge and inserting a thin spacer (a wire, a human hair, or a stack of foil) along the opposite edge. The two plates meet at the clamped edge with zero gap between them, and separate gradually as the spacer is approached, with the local thickness of the gap growing linearly along the wedge. The space between the two plates is air, and we therefore call the arrangement an air wedge.

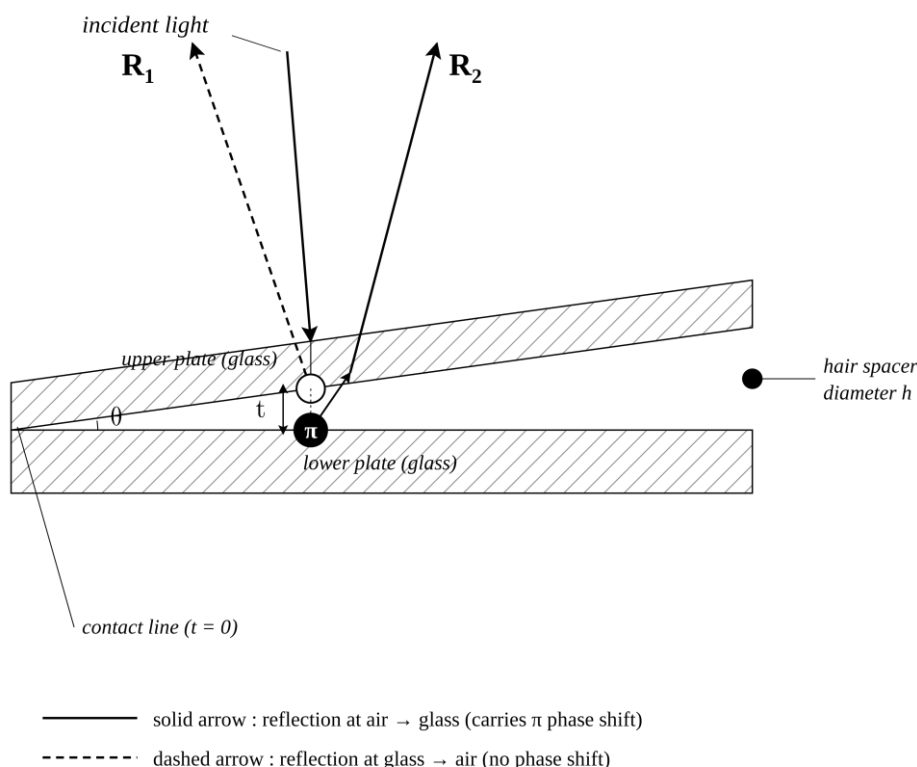


Figure: An air wedge formed by two glass plates clamped at their left edge (the contact line, where $t = 0$) and separated at their right edge by a hair of diameter h . The local thickness t of the air gap grows linearly along the wedge. Light incident from above reflects off the bottom of the upper plate (no phase change, glass → air) and off the top of the lower plate (π phase change, air → glass). The two reflections interfere; the fringe spacing along the wedge encodes the wedge angle θ and the wavelength λ .

The geometry is identical to the soap film with two small differences. **First**, the refractive index of the air between the plates is $\mu = 1$, so the optical path through the wedge reduces from $2\mu t$ to simply $2t$. **Second**, the identification of which reflection picks up the π phase shift is reversed. R_1 , the wave that reflects from the lower surface of the upper glass plate, is going from glass (denser) to air (less dense): no phase shift, soft-end reflection. R_2 , the wave that reflects from the upper surface of the lower glass plate, is going from air (less dense) to glass (denser): π phase shift, hard-end reflection. Exactly one of the two reflections is hard-ended, and the asymmetric phase contribution is the same $\frac{\lambda}{2}$ as before. The bright and dark conditions therefore look the same on the page as the soap-film conditions, with the refractive index dropped out:

$$\text{Bright (reflected): } 2t = \left(n - \frac{1}{2}\right)\lambda, \quad n = 1, 2, 3, \dots$$

$$\text{Dark (reflected): } 2t = n\lambda, \quad n = 0, 1, 2, \dots$$

At the clamped edge, where $t = 0$, the dark condition is satisfied with $n = 0$: the contact line appears as a dark fringe. This is the air-wedge analogue of the soap-film black cap, and it occurs for the same physical reason. As the thickness grows linearly along the wedge, the dark fringes occur at thicknesses $0, \frac{\lambda}{2}, \lambda, \frac{3\lambda}{2}, \dots$, and the bright fringes interleave between them at the half-wavelength thicknesses.

Consecutive dark fringes are therefore separated by a thickness change of:

$$\Delta t = \frac{\lambda}{2}$$

The wedge angle θ (the small angle between the two plates, always small enough that $\tan \theta \approx \theta$ in radians) translates this thickness change into a horizontal distance β between consecutive fringes on the plate:

$$\theta\beta = \frac{\lambda}{2}$$

$$\theta = \frac{\lambda}{2\beta}$$

This is the air-wedge angle formula. It is the easiest experimental method for measuring a very small angle (the angle made by a slightly bent plate, or by a hair trapped between two slides) with a precision the naked eye could not approach.

A second corollary concerns the total number of fringes visible across the entire wedge. If the spacer at the open end has thickness h , the optical path at the open end is $2h$, and the number N of dark fringes spanning the wedge from contact line to spacer is:

$$N = \frac{2h}{\lambda}$$

This formula provides one of the cleanest measurements of the wavelength of light a school laboratory can perform without buying any specialist instrument. A human hair, two glass slides, and a benchtop sodium lamp suffice. Count the dark fringes, measure the hair diameter with a micrometer screw gauge, divide twice the diameter by the count, and you have the wavelength of sodium light to the precision the integers allow.

Application: Anti-reflection coatings

A camera lens or a spectacle lens loses about four per cent of the light incident on it to reflection at the front surface. For a sophisticated zoom lens with twelve or fifteen elements, the cumulative loss adds up to a third or more of the incoming light, and the reflected fraction bounces around between elements until it eventually reaches the camera's sensor as a faint ghost image superimposed on the wanted one. Both effects are undesirable, and a single application of thin-film interference fixes both: deposit on the front surface of the lens a thin transparent coating of carefully chosen thickness, such that the wave reflected off the air-coating boundary interferes destructively with the wave reflected off the coating-glass boundary. The two reflected waves cancel each other; the lens reflects essentially no light at the design wavelength; the lost light is entirely transmitted into the glass as intended.

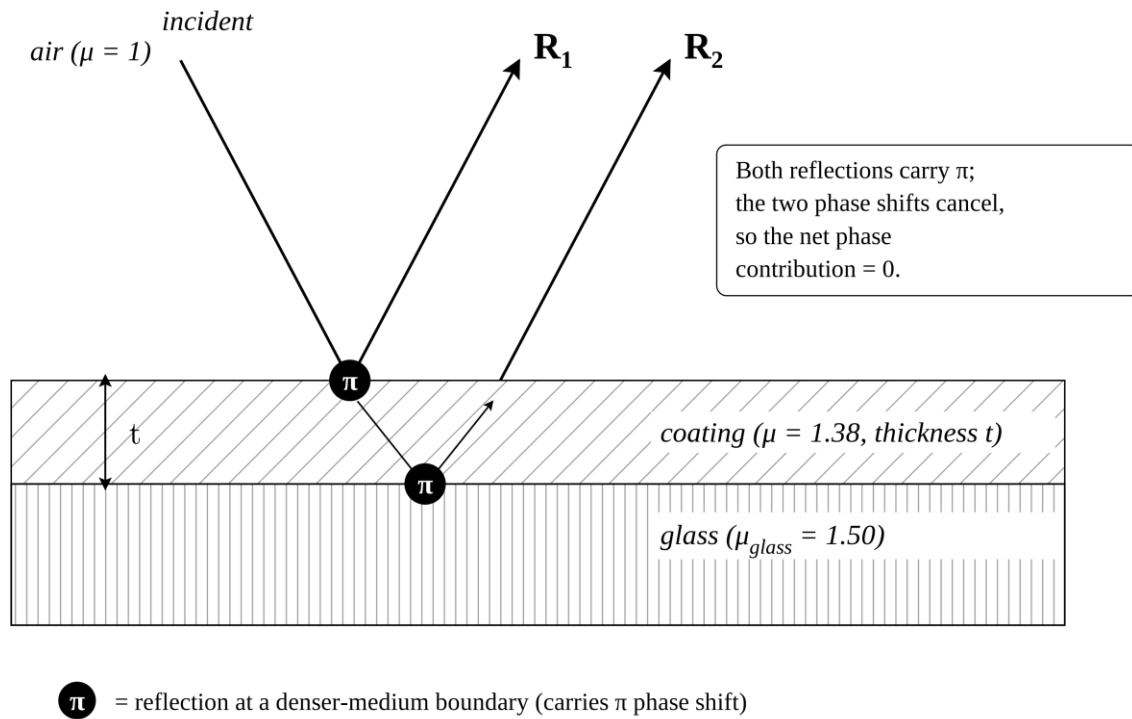


Figure: An anti-reflection coating of refractive index μ and thickness t deposited on a glass surface. Both reflections (R_1 at the air-coating interface and R_2 at the coating-glass interface) suffer a π phase change because the medium beyond each interface is denser. The two phase changes cancel from the optical-path-difference bookkeeping, leaving the geometric round trip $2\mu t$ as the only contributor.

The geometry is a thin film of refractive index μ and thickness t deposited on the front face of glass of refractive index μ_{glass} , with air of refractive index 1 on the outside. The two reflected waves are R_1 (off the air-coating boundary, where coating is denser than air, so a π phase change) and R_2 (off the coating-glass boundary, where glass is denser than coating, so another π phase change). Both reflections carry a π phase shift, and the two contributions cancel each other out of the difference. The condition for destructive interference in the reflected light is therefore the geometric condition alone:

$$2\mu t = \left(n + \frac{1}{2}\right)\lambda, \quad n = 0, 1, 2, \dots$$

The thinnest coating that achieves this corresponds to $n = 0$:

$$2\mu t = \frac{\lambda}{2}$$

$$t = \frac{\lambda}{4\mu}$$

This is the quarter-wave thickness for an anti-reflection coating, and it is the canonical formula for designing a single-layer coating optimised for one wavelength.

Two further refinements are worth knowing without proof, and we shall use one of them in the worked example below.

First, the cancellation between R_1 and R_2 is most complete when the coating has refractive index $\mu = \sqrt{\mu_{\text{air}}\mu_{\text{glass}}}$, because that condition makes the amplitudes of the two reflected waves equal as well as their phases opposite, so they cancel exactly rather than approximately. For air on crown glass at $\mu_{\text{glass}} = 1.50$, the ideal coating refractive index is $\sqrt{1.50} \approx 1.22$; the magnesium fluoride routinely used in commercial lenses has $\mu = 1.38$, which is close enough to give a residual reflected fraction of about 1.5 per cent compared with the uncoated 4 per cent, a substantial improvement but not a perfect cancellation.

Second, the formula $t = \frac{\lambda}{4\mu}$ optimises the coating for one wavelength only; wavelengths above and below it reflect slightly more than the uncoated lens, which is why coated lenses have a characteristic purple sheen in reflection (the green-yellow centre of the spectrum, at the design λ , is suppressed, leaving the blue and the red ends to dominate).

With the ideas now simmering nicely on the page, the next three worked examples are the spoon with which we taste them and feel the flavour of the physics for ourselves. Pen ready.

BINDER Example 10

Two flat glass plates are clamped along one edge to form an air wedge. At the open end, a fine human hair of diameter 0.060mm is inserted between them to act as a spacer. The wedge is illuminated normally with monochromatic light from a sodium lamp, and a travelling microscope counts a total of 240 dark fringes across the wedge from the contact end to the open end. Calculate the wavelength of the sodium light.

Solution

Let h denote the maximum thickness of the air wedge (the diameter of the hair) and N the total number of dark fringes counted from the contact end. The thickness step between consecutive dark fringes is $\frac{\lambda}{2}$, so the relation between h , N , and λ is:

$$N = \frac{2h}{\lambda}$$

Solving for the wavelength:

$$\lambda = \frac{2h}{N}$$

Substituting $h = 0.060\text{mm} = 6.0 \times 10^{-5}\text{ m}$ and $N = 240$:

$$\lambda = \frac{2 \times 6.0 \times 10^{-5}\text{ m}}{240} = 5.0 \times 10^{-7}\text{ m} = 500\text{nm}$$

The wavelength of the sodium light is 500nm.

Making Sense of the Answer: *The answer lies in the green-yellow region of the visible spectrum, which is sensible for a school sodium lamp; the real sodium yellow doublet sits at 589nm in undisturbed atomic vapour, and our 500nm answer is in the right neighbourhood for a school sodium lamp whose lines have been broadened by a hot filament or by pressure-induced shifts. The arithmetic ran on three given numbers (the hair diameter, the fringe count, and the factor of two from the round-trip through the air gap) and produced an answer with five-percent precision without our owning a single specialist instrument.*

Think Like a Physicist: *Air-wedge fringe counts are the cleanest school measurement of the wavelength of light, because the only ingredients are a length (the diameter of the hair, measurable with a micrometer screw gauge to micrometre precision) and a count of integers (the number of fringes, requiring nothing more than patience and a travelling microscope). Neither quantity needs an oscilloscope, a photodetector, or a spectrometer. Notice that the formula has no spare parameters: everything that could carry a measurement uncertainty has been removed by the simple act of using a hair instead of a calibrated reference spacer. This is the cleanest kind of measurement physics knows how to do.*

BINDER Example 11

An air wedge is illuminated normally by monochromatic light of wavelength 600nm. A travelling microscope focused on the top surface of the wedge observes 11 consecutive bright fringes occupying a horizontal distance of 9.0mm along the wedge. Calculate the angle of the wedge in radians.

Solution

Let $\lambda = 600\text{nm}$. Eleven consecutive bright fringes have ten intervals between them (fringes are fenceposts, intervals are panels, and the panel count is always one less than the post count), so the fringe spacing is:

$$\beta = \frac{9.0\text{mm}}{10} = 0.9\text{mm} = 9.0 \times 10^{-4}\text{ m}$$

The air-wedge angle follows directly from the fringe-spacing formula:

$$\theta = \frac{\lambda}{2\beta}$$

Substituting:

$$\theta = \frac{600 \times 10^{-9} \text{ m}}{2 \times 9.0 \times 10^{-4} \text{ m}} = 3.33 \times 10^{-4} \text{ rad}$$

The wedge angle is 3.33×10^{-4} rad, which is about 0.019° .

Making Sense of the Answer: *Three hundred and thirty millionths of a radian is, in any normal sense, a perfectly absurd quantity to be able to measure. With a protractor on a sheet of paper, an angle of one one-thousandth of a degree would not be detected with the most generous reading. The fringe pattern, however, presents itself to the microscope as a series of dark and bright lines just under a millimetre apart, which any Form Five student with a steady hand can resolve to within a few percent. The optics has effectively multiplied the small angle of the wedge into a comfortably visible fringe spacing, and the multiplication factor is the fringe spacing divided by the wavelength of light, $\frac{\beta}{\lambda} \approx 1500$ in our case. Submicroscopic angles are made macroscopic by the wavelength of light; that is the price of admission to optical metrology.*

Think Like a Physicist: *The formula $\theta = \frac{\lambda}{2\beta}$ needed only the wavelength and the fringe spacing; nothing about the length of the wedge or the diameter of the hair appears. Whenever an answer turns out to be independent of a quantity the problem could in principle have asked you to use, that quantity is irrelevant to the physics of the result. Always look for what is missing from the answer as well as what is present; the absences are sometimes more informative than the presences.*

REAL Example 12

A camera-lens manufacturer in Mwanza wishes to apply a non-reflective coating to the front surface of a glass lens, with the goal of minimising reflected light at the green-yellow centre of the visible spectrum (around 550nm). The coating material chosen is magnesium fluoride, which has refractive index 1.38 and forms films of excellent uniformity by vacuum deposition. The glass has refractive index 1.50. Calculate the thinnest coating thickness that achieves destructive interference in the reflected light at the design wavelength, and comment briefly on the choice of magnesium fluoride as the coating material.

Solution

Let $\mu = 1.38$ be the refractive index of the coating, $\mu_{\text{glass}} = 1.50$ that of the glass beneath, and $\lambda = 550\text{nm}$ the design wavelength. Both reflections (air $\rightarrow$ coating and coating $\rightarrow$ glass) pick up a phase shift of π ; the two phase shifts cancel each other out of the difference, and the condition for destructive interference in the reflected light is the purely geometric condition:

$$2\mu t = \left(n + \frac{1}{2}\right)\lambda$$

The thinnest coating corresponds to $n = 0$:

$$t = \frac{\lambda}{4\mu}$$

Substituting:

$$t = \frac{550 \times 10^{-9} \text{ m}}{4 \times 1.38} = 9.96 \times 10^{-8} \text{ m} \approx 100\text{nm}$$

The required coating thickness is about 100nm, or one tenth of a micrometre, a film thinner than a single bacterium and about a fifth the thickness of a soap-bubble band. Modern vacuum-deposition techniques achieve this routinely with thickness uniformity better than one nanometre across the lens.

On the choice of material: the cancellation between R_1 and R_2 is total only when their amplitudes are equal as well as their phases opposite. The phases are made opposite by the quarter-wave thickness condition above. The amplitudes are made equal by the additional condition that the refractive index of the coating should be the geometric mean of the refractive indices on either side, namely $\sqrt{\mu_{\text{air}}\mu_{\text{glass}}} = \sqrt{1.00 \times 1.50} \approx 1.22$. Magnesium fluoride at $\mu = 1.38$ is close to this ideal but not exactly on it; the residual reflected fraction at the design wavelength is about 1.5 per cent compared with the 4 per cent

reflected by the uncoated glass. A reduction by a factor of roughly three, applied to every air-glass interface in a multi-element lens, is what makes modern professional camera lenses optically transparent enough that they appear almost black in reflection.

Making Sense of the Answer: Around 100nm is the canonical thickness of a single-layer anti-reflection coating in the visible. The reason the answer is so robustly set by $\frac{\lambda}{4\mu}$ is that the formula has no free parameters once the design wavelength and the coating material are chosen; the engineering is entirely in the manufacturing precision (depositing exactly the right thickness, uniformly, across the lens), not in the design arithmetic. The difference of a few nanometres between the ideal-index coating and the nearly-ideal magnesium fluoride is absorbed in the residual 1.5 per cent reflection: a small price for the convenience of having a chemically robust and inexpensive material.

Think Like a Physicist: Whenever a problem mentions an anti-reflection coating, a quarter-wave plate, or any single-layer thin-film design, the formula $t = \frac{\lambda}{4\mu}$ is doing the work. Identify which surface gives which phase shift, decide whether you want destructive or constructive interference in the reflected light, and the integer-condition formula collapses to a single thickness for $n = 0$. Multi-layer coatings stack several quarter-wave layers of carefully chosen refractive indices, each designed to suppress the residual reflection of the layer below, and the same physics scales up by repetition: every layer contributes its own $\frac{\lambda}{4\mu}$ thickness condition, and the design problem reduces to choosing the indices.

Amplitude division has now joined wavefront division in our toolkit. The new physical idea, the π phase change at the denser surface, has been spent three times: once on a thin film of uniform thickness, once on an air wedge of linearly varying thickness, and once on a quarter-wave anti-reflection coating of carefully chosen thickness. The next section turns to the most heavily examined apparatus that exploits amplitude division: Isaac Newton's rings, formed between a long-focus plano-convex lens and a flat glass plate. The geometry is a special case of the air wedge in which the gap thickness varies quadratically rather than linearly with horizontal position, producing circular fringes rather than straight ones. Everything else (the phase change, the dark central spot, the brightness ratios, the experimental protocol for measuring wavelengths to four significant figures with two pieces of glass and a microscope) follows by the same logic we have just spent.

NEWTON'S RINGS

After the air wedge of the previous section, one more arrangement of two reflecting surfaces deserves a section of its own. It is older than the soap film by almost two centuries (Isaac Newton himself reported it in 1675, and the 1704 publication of *Opticks* gave it the name it carries today), and it is by far the most heavily examined of all the thin-film phenomena in this chapter. The fringes are circular. They appear in a beautiful concentric ring pattern with a dark spot at the centre. The radii of the rings turn out to depend on the wavelength of light in a way clean enough to measure that wavelength to three significant figures with apparatus that fits on a school desk.

Mr. Akilikubwa, on the Tuesday after the air-wedge demonstration, had wheeled out from the laboratory preparation room a brass mount that held a plano-convex lens in light contact with a flat glass plate. He set a small sodium lamp to one side, clipped a thin glass square at forty-five degrees above the lens to act as a beam splitter, and mounted a travelling microscope vertically above the whole thing. He dimmed the windows, switched on the lamp, and called the class to the eyepiece.

What the students saw, once their eyes had adjusted to the small yellow disc of light reflected up to them, was not a uniform field. They saw a central dark spot the size of a chickpea, surrounded by a series of bright and dark concentric rings that shrank in spacing as one looked outward. The rings were the same yellow as the lamp, alternating with darkness, getting more and more crowded together until they faded into the edge of the field of view.

Kipanga: Sir, what is producing the rings?

Mr. Akilikubwa: An air film. A wedge of air, the same kind we measured in the previous demonstration. But this time the wedge is not between two flat plates clamped at one end. It is between a flat plate and a curved lens resting on top of it. The air gap is zero where the lens touches the plate, and grows steadily as you move outward from the contact point. The fringes are the contours of constant thickness.

That one sentence is the section. We will draw the apparatus, derive the relationship between the radius of a ring and the thickness of the air film beneath it, identify the optical condition for each ring to be dark or bright, read the wavelength of the source from the rings, and end with three small modifications to the experiment that examiners delight in asking about. Have your pen ready.

The apparatus and what the eye sees

The arrangement is shown in the figure below. A plano-convex lens of large radius of curvature sits with its curved side down on a flat glass plate, often called an optical flat. A thin film of air is therefore trapped between the underside of the lens and the upper face of the plate. Directly above the lens, a half-silvered glass is tilted at forty-five degrees so that monochromatic light arriving horizontally from the right is partly reflected downward onto the lens, and the reflected light coming back up from the air film is partly transmitted upward into a low-power microscope mounted vertically above. The microscope magnifies the ring pattern onto your eye.

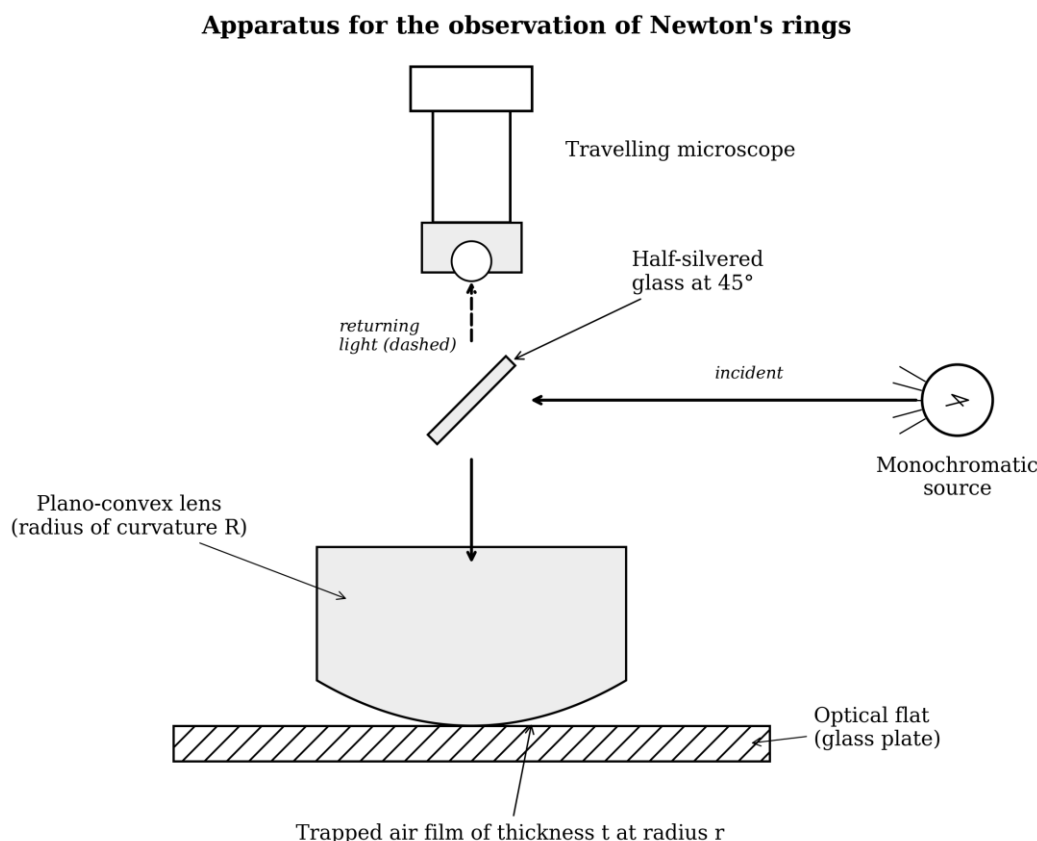


Figure: A plano-convex lens rests on an optical flat. Light from a monochromatic source is bent downward onto the lens by a half-silvered glass at 45° . Two reflections (one from the lower curved surface of the lens, one from the upper surface of the plate) travel back up through the beam splitter and into the travelling microscope. The interference between these two reflections is what produces the ring pattern.

The light striking the lens splits into two reflections of interest. The first, R_1 , reflects off the lower curved surface of the lens (a glass-to-air boundary, with the light moving from glass into air). The second, R_2 , reflects off the upper surface of the flat plate (an air-to-glass boundary, with the light moving from air into glass). The two reflected waves then travel back up to the eyepiece. Whatever pattern of constructive and destructive interference these two reflections produce, the microscope shows it to you.

The radius of curvature of the lower face of the lens, which we shall call R , is large (typically between half a metre and two metres in a school setup). The thickness of the air film at the very centre, where the lens touches the plate, is essentially zero. At a horizontal distance r from the contact point, the thickness t of the air film is determined by the geometry of the spherical lens surface. The next sub-section works out exactly what that thickness is.

A practical word before any algebra: The contact between the lens and the plate must be gentle and the surfaces must be clean. A grain of dust between them will tilt the lens slightly and skew the ring pattern into

ovals. A fingerprint on either glass will paint a haze of additional fringes from the oil. The careful experimenter wipes both surfaces with lens tissue, places the lens on the plate without pressing, and accepts that the central dark spot will sometimes refuse to sit perfectly in the middle of the field of view. The geometry of the rings does not depend on the contact being in the centre of the microscope's field, so all is well.

Why the centre is dark

Before any algebra at all, one observation deserves attention. At the very centre, where the lens touches the plate, the air film has zero thickness. The two reflected waves R_1 and R_2 have travelled the same geometric distance. The naive prediction is that they should arrive in phase, interfere constructively, and produce a bright spot. The naive prediction is wrong. The centre is dark in every Newton's-rings experiment ever performed, and the reason is the same reason a soap film looks black at the top: an asymmetric phase change at reflection.

Kipanga: *Sir, the lens and the plate are touching at the centre. The air gap is zero. Zero path difference. Surely that means constructive interference; the centre should be bright.*

Mr. Akilikubwa: *That is exactly the trap, Kipanga. The geometric path difference at the centre is zero. But you have counted only one of the two ingredients. The other ingredient is the phase change of π that happens at every reflection off a denser medium. Walk through it once carefully.*

Kipanga: *The first reflection is off the bottom of the lens. The wave is inside the glass and reflects off the air gap. Going from denser to less dense. No phase change.*

Mr. Akilikubwa: *Correct.*

Kipanga: *The second reflection is off the top of the glass plate. The wave is inside the air and reflects off the glass plate below. Going from less dense to denser. Phase change of π .*

Mr. Akilikubwa: *Correct. So at the centre, your two reflections have a geometric path difference of zero and a phase difference of π . Reading this as an optical path difference, you have an extra $\lambda/2$ between them. They arrive perfectly out of phase. Destructive interference. The centre is dark.*

Kipanga: *All because one of the two reflections happens at a denser interface and the other does not.*

Mr. Akilikubwa: *All because of that. The asymmetry of the phase change is the whole story. Anyone who remembers only the geometric path difference will predict a bright central spot, and will then have to explain why their prediction disagrees with what the microscope clearly shows. The phase change is not optional.*

The argument generalises. At any point of the air film, the optical path difference between R_1 and R_2 is the geometric round-trip through the air (which is $2t$, because the wave goes down through the air and back up), plus a half-wavelength contribution from the single π phase change at the lower surface. In symbols,

$$\Delta x = 2t + \frac{\lambda}{2}$$

Dark fringes occur where the total optical path difference is a half-integer number of wavelengths (so the two waves arrive π out of phase), and bright fringes occur where it is an integer number of wavelengths (so they arrive in phase). Inserting the half-wavelength contribution and rearranging,

$$\text{dark: } 2t = n\lambda, \quad n = 0, 1, 2, 3, \dots$$

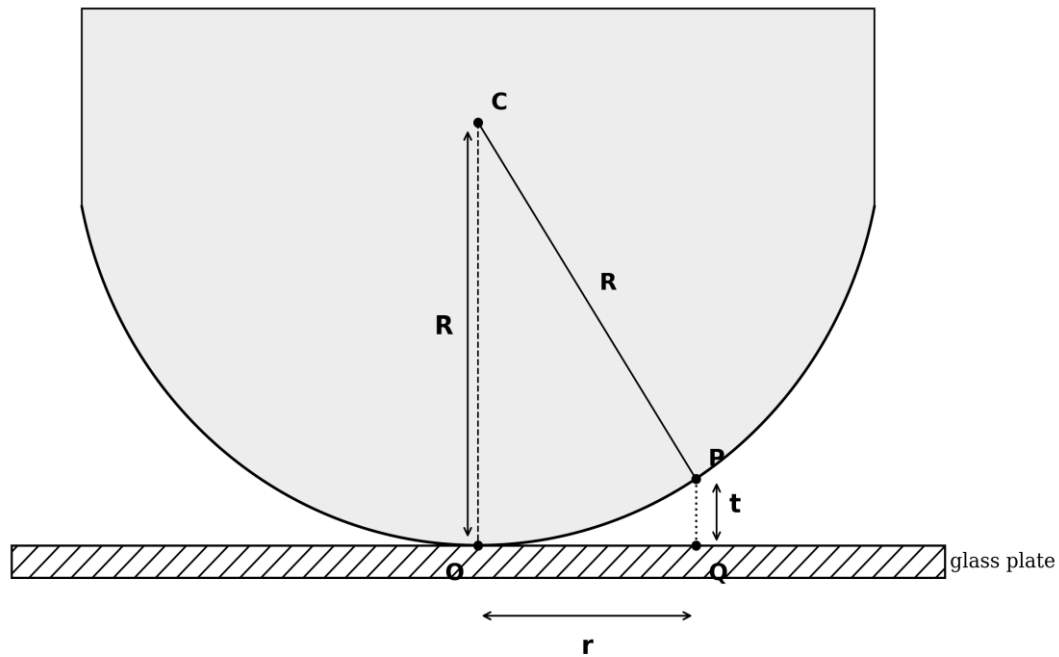
$$\text{bright: } 2t = (n - \frac{1}{2})\lambda, \quad n = 1, 2, 3, \dots$$

The $n = 0$ case of the dark condition is the centre, where $t = 0$ satisfies $2t = 0 \times \lambda$ exactly. The first non-trivial dark ring occurs when $2t = \lambda$, the second when $2t = 2\lambda$, and so on. Each successive dark ring sits on a higher contour of the trapped air film.

The geometry that links radius to thickness

To turn the optical condition into a numerical formula for the radius of a ring, we need a relation between the thickness t of the air film and the horizontal distance r from the contact point. The relation is set entirely by the shape of the lens, and it is one of the cleanest pieces of geometry in the whole chapter.

Geometry of the trapped air film



From the right triangle: $r^2 + (R - t)^2 = R^2$, so $r^2 \approx 2Rt$ when t is much less than R

Figure: Cross-section of the lens on the plate. The lower face of the lens is part of a sphere of radius R centred at C , sitting directly above the point of contact O . A general point P on the lens at horizontal distance r from the axis has the air film of thickness t beneath it (down to Q on the plate). The right triangle PQC immediately yields $r^2 + (R - t)^2 = R^2$, and dropping the t^2 term (because t is much smaller than R) gives $r^2 \approx 2Rt$.

Place the centre of the spherical lens surface at the point C , directly above the contact point O by a distance equal to the radius of curvature R . Take any general point P on the lower face of the lens, at horizontal distance r from the axis. Drop a perpendicular from P down to the plate, meeting it at Q . The length PQ is the thickness t of the air film at that radial position.

The triangle PQC is right-angled at Q . Its horizontal side is r . Its vertical side is $R - t$ (the distance from Q up to C is the full radius minus the thickness of the gap). Its hypotenuse is CP , which equals R because P lies on the sphere. Pythagoras gives:

$$r^2 + (R - t)^2 = R^2$$

Expand the bracket and cancel the common R^2 from both sides:

$$r^2 + R^2 - 2Rt + t^2 = R^2,$$

$$r^2 = 2Rt - t^2$$

The term t^2 is utterly negligible. The thickness t of the air film is at most a few micrometres in a typical Newton's-rings setup, while the radius of curvature R is of order one metre. The ratio t/R is therefore about one in a million; the ratio t^2/R is one in a million times one micrometre, which is utterly invisible. We drop t^2 without losing anything that mattered, and we are left with the master geometric result of this section:

$$r^2 = 2Rt \quad \text{equivalently} \quad 2t = \frac{r^2}{R}$$

Two sentences tell the whole story.

- 1) The thickness of the air film, at horizontal distance r from the contact point, is proportional to the square of r . The constant of proportionality is $1/(2R)$, which depends only on how curved the lens is.
- 2) A flatter lens (larger R) gives a slower-growing air film and therefore wider-spaced rings. A more strongly curved lens (smaller R) gives a faster-growing air film and tightly-packed rings. The rings widen and narrow with the geometry of the lens you happen to have chosen, but their structure follows a single universal formula.

The radii of the dark and bright rings

We now combine the optical condition for a dark fringe ($2t = n\lambda$) with the geometric relation ($2t = r^2/R$). Equating the right-hand sides gives the radius of the n -th dark ring:

$$r_n^2 = n\lambda R, \quad n = 0, 1, 2, 3, \dots$$

Working in diameters rather than radii is often more convenient, because the travelling microscope reads the position of a ring on both sides of centre and the difference is the diameter directly. Doubling r gives $D_n = 2r_n$, and squaring yields:

$$D_n^2 = 4n\lambda R$$

For bright rings, the optical condition is $2t = (n - \frac{1}{2})\lambda$. Combining with the same geometric relation gives the bright-ring radius and diameter as:

$$r_n^2 = (n - \frac{1}{2})\lambda R, \quad D_n^2 = (2n - 1)2\lambda R, \quad n = 1, 2, 3, \dots$$

Three features of these formulas earn a moment of attention before we put them to work.

First, the diameters do not grow linearly with the ring number; they grow with the square root of n . The first ring is one unit out; the fourth ring is two units out; the ninth ring is three units out. The rings crowd together as you move outward, which is what the eye saw at the microscope.

Second, the radii depend on three quantities only: the integer n , the wavelength λ , and the radius of curvature R . Nothing else in the geometry of the apparatus enters.

Third, by measuring one ring and knowing two of the three quantities, you can compute the third. That is the entire experimental method of Newton's rings: pick the unknown you want, build an apparatus that supplies the other two, and read the answer off the microscope.

Reading the wavelength from a graph

Measuring a single ring is risky in practice. The very innermost rings are crowded, broad, and hard to centre under the microscope's cross-hairs. The very outermost rings are dim because the intensity of the source falls off and the contrast washes. The reliable laboratory method is therefore to measure several rings, plot their squared diameters against the ring number, and extract the wavelength from the slope of the resulting straight line.

From the dark-ring formula $D_n^2 = 4n\lambda R$, a plot of D_n^2 against n is a straight line through the origin with slope $4\lambda R$. The wavelength is therefore:

$$\lambda = \frac{\text{slope}}{4R}$$

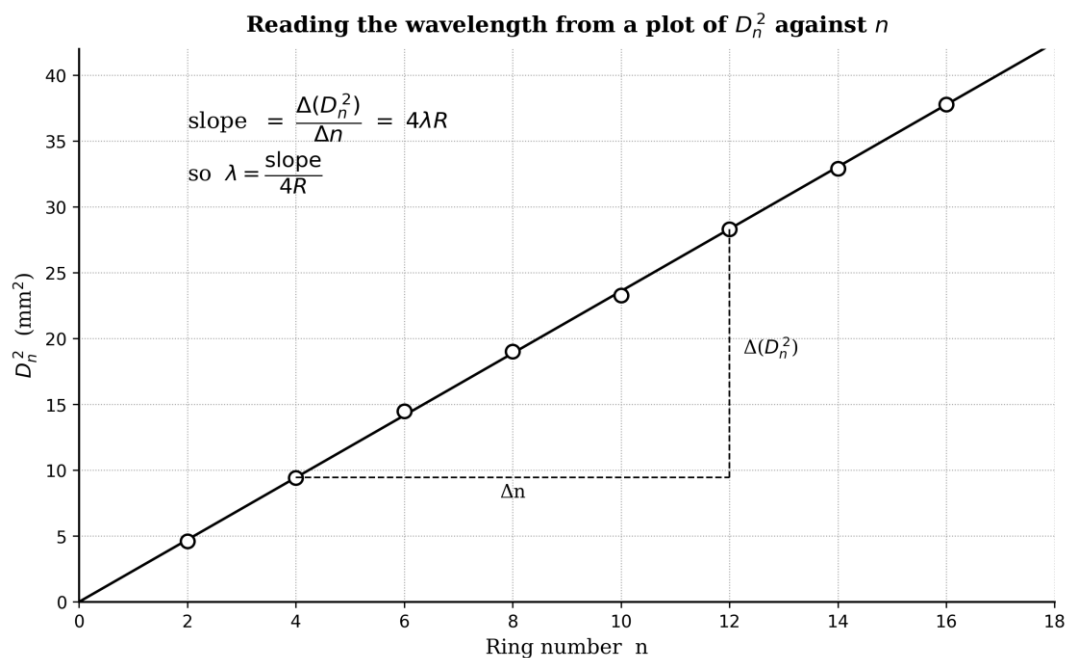


Figure: Squared diameter D_n^2 of the n -th dark ring plotted against the ring number n . Each open circle is one measurement; the solid line is the best fit through the origin. The slope of the line, read off as $\Delta(D_n^2)/\Delta n$ from the dashed right triangle, equals $4\lambda R$, so the wavelength of the source is $\lambda = \text{slope} / (4R)$.

Two advantages of the graphical method are worth pinning down.

First, the slope is the average rate of change over many rings, so any small error in counting which ring is which (off by one at the centre, where the rings are crowded and hard to count) averages out.

Second, the slope does not depend on whether the $n = 0$ ring is at the centre of the microscope's cross-hairs or slightly displaced (the slope is invariant under a constant offset in n).

The graphical method is therefore robust against the very mistakes a careful experimenter is most likely to make. Single-ring measurements are usable but always less trustworthy than a slope reading from five or more rings.

Filling the gap with a liquid

So far the gap between the lens and the plate has been air, with refractive index $\mu = 1$. A favourite extension of the experiment, both in textbooks and in physics laboratories interested in measuring the refractive index of clear liquids, is to pour a small drop of liquid between the lens and the plate and re-record the rings.

With a liquid of refractive index μ filling the gap, the optical round-trip through the gap is no longer $2t$ but $2\mu t$ (each metre of liquid contributes μ metres of optical path, because the wavelength of light shortens to λ/μ inside the liquid). The π phase change at the air-to-glass or liquid-to-glass lower boundary is still there provided the liquid is less dense than the glass plate (which is true for water, glycerine, alcohol, and almost every transparent liquid a school laboratory will ever contain). The optical condition for dark fringes becomes:

$$2\mu t = n\lambda,$$

and combining with the geometric $2t = \frac{r^2}{R}$ gives the new ring radius:

$$r_n^2 = \frac{n\lambda R}{\mu}, \quad D_n^2 = \frac{4n\lambda R}{\mu}$$

The rings shrink by a factor of the square root of μ everywhere. The ratio of the diameter of any given ring in air to the diameter of the same ring in liquid is therefore exactly the square root of μ , and squaring this ratio gives the refractive index directly:

$$\mu = \frac{D_n^2(\text{air})}{D_n^2(\text{liquid})}, \quad \text{at the same ring number } n.$$

Two points of practical good sense.

- 1) The same formula holds if you compare the slopes of the two D_n^2 against n plots, one for air and one for liquid:

$$\mu = \frac{\text{slope}(\text{air})}{\text{slope}(\text{liquid})},$$

which is the graphical version and is more accurate.

- 2) The result is independent of which particular ring you happen to compare, which is a useful self-check on the experiment: if your computed μ comes out different for the fourth ring than for the fourteenth ring, something has gone wrong (probably air has been re-admitted into the gap, or the liquid has not fully replaced the air).

Three small changes that the rings respond to

Three simple modifications of the apparatus produce qualitatively different ring patterns, all of which can be predicted from the physics already in your head. Examiners are fond of these because they reward the student who actually understands the concept and punish the student who has memorised $D_n^2 = 4n\lambda R$ without thinking about where it comes from.

Modification 1: Replace the sodium lamp with a white-light source

Each wavelength now produces its own family of rings, with red rings of any given order at slightly larger radii than violet rings of the same order (because the diameter is proportional to the square root of the wavelength). The central spot remains dark for every wavelength simultaneously, because the half-wavelength contribution from the single π phase change cancels every wavelength's geometric zero at

$t = 0$. A few orders out from the centre, the wavelengths still produce distinguishable coloured rings (violet on the inside, red on the outside, for each order). Further out, the rings of different colours start to overlap, the colours mix, and after perhaps half a dozen orders the pattern washes into a uniform pale grey. The following figure shows both pictures side by side.

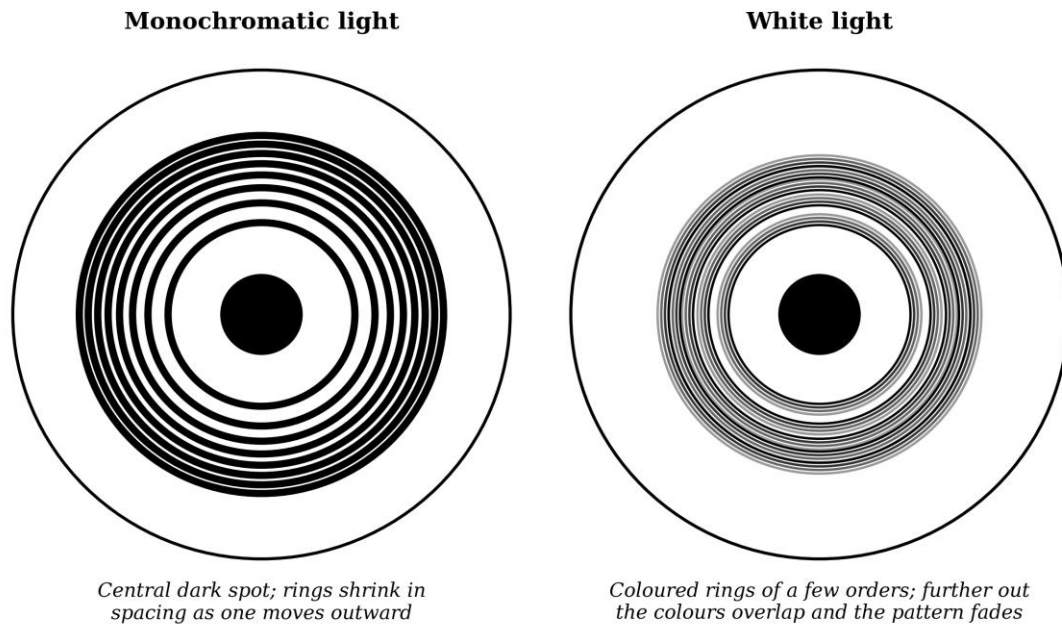


Figure: Left: the ring pattern observed under sodium light, with a central dark spot and outward rings that crowd together as $1/\sqrt{n}$. Right: the same apparatus illuminated by white light, with coloured rings of a few orders before different wavelengths overlap and the pattern fades.

Modification 2: Silver the upper surface of the flat plate, so that the lower reflection becomes a metallic reflection rather than a glass reflection

A metallic reflection also picks up an extra π phase change. Both reflections now suffer the same phase shift, the asymmetry that produced the dark central spot disappears, and the geometric condition for constructive and destructive interference reverses. At the centre, $t = 0$ with no net phase difference between the two waves now gives constructive interference: the central spot becomes **bright**. The pattern outside the centre is the same alternating sequence of rings, but interchanged: dark where it used to be bright, bright where it used to be dark.

Modification 3: Introduce a few drops of a transparent liquid between the lens and the plate

This is the case treated in the previous sub-section. Every ring shrinks in diameter by a factor of $\sqrt{\mu}$, the spacing between consecutive rings contracts by the same factor, and the central spot remains dark (the optical condition $2\mu t = 0$ plus the half-wavelength is still destructive). A side-by-side observation of the same ring before and after the liquid is added gives the refractive index of the liquid as the ratio of squared diameters.

A small warning is necessary for modification 2: The phase change on metallic reflection is not exactly π for all metals and all wavelengths. Silver at sodium yellow comes very close to π , which is why the silvered-plate experiment is described as straightforwardly reversing the pattern. For other metals and other colours, the central spot becomes bright but the contrast of the surrounding rings can be diminished from what the air-gap experiment delivers. Real laboratory work always uses silver.

With this solid theoretical foundation in place, we are now ready to see these ideas come to life through four worked examples.

BINDER Example 13

In a Newton's-rings setup the radius of curvature of the plano-convex lens is 0.80m. The radius of the eighth dark ring is measured under the travelling microscope and found to be 1.94mm. Calculate the wavelength of the light used.

Solution

Let the wavelength be λ , the ring number $n = 8$, the radius of curvature $R = 0.80\text{m}$, and the radius of the eighth dark ring $r_8 = 1.94 \times 10^{-3}\text{m}$. Apply the dark-ring formula:

$$r_n^2 = n\lambda R$$

Rearrange for λ :

$$\lambda = \frac{r_n^2}{nR}$$

Insert numbers with units:

$$\lambda = \frac{(1.94 \times 10^{-3}\text{m})^2}{(8 \times 0.80\text{m})} = 5.88 \times 10^{-7}\text{m}$$

The wavelength of the light used is $5.88 \times 10^{-7}\text{m}$ or 588nm .

Making Sense of the Answer: 588nm sits in the yellow part of the visible spectrum, just a hair below the famous sodium D-doublet at 589.0nm and 589.6nm . Since the apparatus used a sodium lamp, the answer practically tells us we used a sodium lamp, which is exactly the calibration we should hope for from a well-set-up experiment. If our answer had come out at, say, 280nm or 1300nm , we would know without further checking that something was wrong: 280nm is invisible ultraviolet and 1300nm is infrared, and the lamp on the bench was glowing visibly yellow. Use this kind of physical reality-check at the end of every calculation in this chapter; the visible spectrum is roughly **400nm to 700nm**, and any answer outside that range demands an explanation.

Think Like a Physicist: When you are given a single ring radius and asked for the wavelength, the only fact that matters is $r_n^2 = n\lambda R$. Plug, chug, done. The graphical method (slope of D^2 against n) is more accurate when you can measure several rings, but for a school problem where only one ring is given, the single-ring formula is the right tool. Pick the formula that matches the number of data points you have been given.

REAL Example 14

In a Newton's-rings experiment with a lens of radius of curvature 1.00m , the diameters of the sixth and twelfth dark rings are measured to be 3.76mm and 5.32mm respectively. Calculate the wavelength of the light used.

Solution

Let D_6 and D_{12} be the two ring diameters, $R = 1.00\text{m}$ the radius of curvature, and λ the unknown wavelength. From $D_n^2 = 4n\lambda R$, the squared diameter grows linearly with n , so the difference of the squared diameters of two rings depends only on the difference of the ring numbers:

$$D_{12}^2 - D_6^2 = 4(12 - 6)\lambda R = 24\lambda R$$

Rearrange for λ :

$$\lambda = \frac{(D_{12}^2 - D_6^2)}{24R}$$

Insert numbers with units:

$$\lambda = \frac{(5.32 \times 10^{-3}\text{m})^2 - (3.76 \times 10^{-3}\text{m})^2}{24 \times 1.00\text{m}} = 5.90 \times 10^{-7}\text{m} = 590\text{nm}.$$

The wavelength of the light is approximately $5.90 \times 10^{-7}\text{m}$.

Making Sense of the Answer: 590nm is right in the middle of the yellow part of the spectrum and is a sodium lamp again. Notice that the difference-of-squares method has not asked us to know the absolute ring number of either ring; it has only asked us to know that the two rings are separated by six orders. If our count of which ring was the sixth happened to be off by one at the centre (the centre is crowded and easy to miscount), the answer would shift by exactly nothing, because the labels n and $n + 6$ would have shifted together. The method is far more robust than picking a single ring's radius.

Think Like a Physicist: Whenever a problem gives you two ring measurements, do not solve it as two single-ring problems. Subtract the squared diameters, divide by the difference of ring numbers times $4R$, and read the wavelength off. The same arithmetic with three or more ring measurements becomes a slope on a plot, which is even more reliable. Always prefer differences of squared diameters to single radii.

REAL Example 15

In the apparatus of the previous example, the air gap between the lens and the plate is now filled with a clear liquid. The twelfth dark ring, which had a diameter of 5.32mm in air, now has a diameter of 4.60mm with the liquid present. The radius of curvature of the lens is unchanged at 1.00m and the source is the same sodium lamp. Calculate the refractive index of the liquid.

Solution

Let $D_{12}(\text{air})$ and $D_{12}(\text{liq})$ be the two diameters of the same ring before and after the liquid is added, and μ the refractive index of the liquid. The ring formula in the liquid is $D_n^2 = 4n\lambda R/\mu$, and the air formula is $D_n^2 = 4n\lambda R$, so dividing the air diameter squared by the liquid diameter squared at the same ring number gives μ directly:

$$\mu = \frac{D_n^2(\text{air})}{D_n^2(\text{liq})}$$

Insert numbers with units:

$$\mu = \frac{(5.32 \times 10^{-3}\text{m})^2}{(4.60 \times 10^{-3}\text{m})^2} = 1.338$$

The refractive index of the liquid is approximately 1.34.

Making Sense of the Answer: A refractive index of 1.34 is essentially the refractive index of fresh water (which is 1.33 at room temperature for yellow light). If the answer had come out at, say, 0.85 (smaller than 1), we would know without checking that the calculation had gone wrong: a refractive index less than one does occur for some materials at some wavelengths, but it is not on the menu for ordinary liquids in a school laboratory. If the answer had come out at 2.5, we would suspect a transcription error: even diamond, the densest optical solid commonly handled, is at 2.42 in visible light. Trust the answer if it falls between **1.3 and 1.7**, sigh at the answer if it falls outside, and re-do the arithmetic.

Think Like a Physicist: Whenever you compare a ring diameter in two different media at the same ring number, square the ratio of the diameters and you have the refractive index. Whenever you compare ring diameters between two different ring numbers in the same medium, take the difference of the squares (Example 14). The two manoeuvres are the two faces of the same algebra, and a Newton's-rings problem is almost always one or the other.

HOT Example 16

A Newton's-rings apparatus is illuminated by light of wavelength 600nm. When the gap between the plano-convex lens and the flat glass plate contains air, the diameters of the fourth and twelfth dark rings are measured to be 3.80mm and 6.58mm respectively. After the same gap is filled with a transparent liquid, the diameters of the same two rings become 3.10mm and 5.37mm respectively. Calculate (a) the radius of curvature of the lens and (b) the refractive index of the liquid.

Solution

Let R be the unknown radius of curvature of the lens and μ the unknown refractive index of the liquid. We have four pieces of data: two ring diameters in air and two ring diameters in liquid, all at $\lambda = 6.00 \times 10^{-7}\text{m}$. The two unknowns require two independent equations.

Build the first equation from the air measurements. The difference of squared diameters of the twelfth and fourth dark rings in air is $D_{12}^2(\text{air}) - D_4^2(\text{air}) = 4 \times (12 - 4) \times \lambda R = 32 \lambda R$:

$$D_{12}^2(\text{air}) - D_4^2(\text{air}) = 32 \lambda R$$

$$R = \frac{D_{12}^2(\text{air}) - D_4^2(\text{air})}{32 \lambda} = \frac{(6.58 \times 10^{-3}\text{m})^2 - (3.80 \times 10^{-3}\text{m})^2}{32 \times 6.00 \times 10^{-7}\text{m}} = 1.50\text{m}$$

The radius of curvature of the lens is 1.50m.

Build the second equation from the liquid measurements. The same difference, in the liquid is:

$$D_{12}^2(\text{liq}) - D_4^2(\text{liq}) = \frac{32 \lambda R}{\mu}$$

$$\mu = \frac{32 \lambda R}{D_{12}^2(\text{liq}) - D_4^2(\text{liq})} = \frac{32 \times 6.00 \times 10^{-7}\text{m} \times 1.50\text{m}}{(5.37 \times 10^{-3}\text{m})^2 - (3.10 \times 10^{-3}\text{m})^2} = 1.50$$

The refractive index of the liquid is approximately 1.50, which is consistent with a clean cedar oil or a light mineral oil at room temperature.

Making Sense of the Answer: *Both answers carry weight checks. The radius of curvature of 1.50m is in the right range for a school lens (which is typically 0.5m to 2m); a lens with $R = 1.50\text{m}$ has a very shallow curvature, which is exactly what you want for Newton's rings because shallow lenses give wide, well-spaced rings that the microscope can resolve. The refractive index of 1.50 is too high for water (which is 1.33) and too low for diamond (which is 2.42), so it is something organic and oily. Cedar oil at the sodium D-line happens to be 1.516, mineral oil sits between 1.46 and 1.50, and most clear cooking oils are in the same range.*

Think Like a Physicist: *Whenever a problem hands you two unknowns and four measurements, look for two equations. Here the two equations both came from the difference-of-squares manoeuvre, once in air and once in liquid; their ratio cleanly delivered μ without even substituting R , and the absolute value of either delivered R if you wanted it. The general rule is: more measurements than unknowns is a gift, not a complication. Always set up the simultaneous equations first, then solve them. Plugging numbers into a single formula before you have written the system on paper is a common way to drop into algebraic confusion.*

We have spent the air-film geometry three times: on a thin film of uniform thickness, on an air wedge of linearly varying thickness, and now on the thicker-at-the-edge spherical air film of Newton's rings. The same two ingredients, the geometric round-trip $2t$ through the gap and the asymmetric phase change of π at one of the two boundaries, have predicted the dark-fringe-then-coloured-fringes appearance of a soap film, the parallel fringes of an air wedge, and the concentric ring pattern of Newton's apparatus. The wavelength measurement methods are not three different methods either: they are three geometrical dressings of the same idea, namely that a known geometric distance times a known optical condition gives a known wavelength.

Two more interference systems would fit in this section, but the chapter has work to do elsewhere and we shall not stay any longer. The **Michelson interferometer** and the **Fabry-Perot etalon** both repeat the same physics with different geometry; the first is the subject on the speed of light, and the second is the workhorse of laser-quality optical filters. They are flagged here so that the reader who meets them later recognises the family resemblance. For now we close interference and move to the next physical question of the chapter: *when a single wave of light passes a single obstacle whose size is comparable with the wavelength, what does the wave do?*

DIFFRACTION

The first half of the chapter belonged to interference. Two waves, two slits, two reflecting surfaces, and in every case a pattern of bright and dark fringes that the eye could read out as a wavelength. Now, we are diving to a different operation that the same wave model performs with only one of either: a single slit, a single sharp edge, a single hair across a beam. The phenomenon is called **diffraction**, and it is the answer to the puzzle Kipanga handed us in the very first paragraph of this chapter.

Recall the CD that Kipanga had brought to school on that long morning, the sharp white shaft of sunlight falling through the louvre windows, and the band of rainbow colour that walked across the silvery disc. There were no two slits, no thin film, no two reflecting surfaces. There was one shiny disc, one window full of morning light, and an outpouring of rainbow colour that geometric optics had no explanation for. Today we explain it. The grooves of a CD form a regular array of one thousand six hundred slits per millimetre, and each colour of the white sunlight bounces off that array at its own preferred angle. The CD is a diffraction grating, the most elegant single optical instrument the chapter will meet, and by the end of the section you will be able to read the wavelength of the sunlight's red component to two significant figures with nothing more than a CD case and a ruler.

The plan is two derivations and three protocols. The first derivation treats a single slit of width a , illuminated by a plane wave, and asks where the bright and dark bands fall on a distant screen. The second treats a regular array of many slits at common spacing d , the device called a diffraction grating, and asks the same question. Out of the two derivations come three formulas and one observation. The formulas read wavelengths out of angles. The observation tells you why a grating is preferable to a single slit, and why the CD that Kipanga held up to the bulb was, although he could not have known it on the day, doing physics that Newton's first telescope could not have done.

When diffraction shows itself

Diffraction is everywhere. Every time a wave passes an obstacle, the obstacle's edge sets off a chain of secondary wavelets that spread into the geometric shadow. The reason we do not normally see diffraction with light is a matter of scale. Visible light has wavelengths between about four hundred and seven hundred nanometres, which is to say between $4 \times 10^{-7} \text{m}$ and $7 \times 10^{-7} \text{m}$. Everyday obstacles, like fingers, doors, and Mwanza chickens crossing a sunlit road, are between a centimetre and a metre wide. The ratio of obstacle width to wavelength is therefore between ten thousand and ten million, and the diffracted light spills only a few wavelengths into the shadow, which is to say a few micrometres on a length scale where the eye cannot resolve a hundred-micrometre fly. Diffraction is there. We simply cannot see it.

This is the concrete answer to a question that the chapter has been postponing. **Why, despite light being a wave, does the everyday experience of light insist that it travels in straight lines?** Because everyday obstacles are very many wavelengths wider than the wavelengths themselves. The geometric shadow is sharp to a precision set by the obstacle width divided by the wavelength, and that precision is in the millions for ordinary objects. Diffraction is not absent from your living room; it is merely too tightly confined to be noticed.

To see diffraction, we need an obstacle whose smallest dimension is comparable with the wavelength. A slit narrower than a hair, a wire thinner than a hair, a CD's groove pitch, an opal's array of microspheres. Once we get down to those scales, diffraction takes over and dominates. The rest of the section is about two such obstacles.

One more orientation, then to the algebra. A plane wave (a wavefront like a flat sheet of paper) hitting a slit produces a diffraction pattern whose shape depends on the geometry of where the screen sits. *If the screen is far enough away that the secondary wavelets reaching it can be treated as parallel*, we are in the regime called **Fraunhofer diffraction**. *If the screen is so close that the secondary wavelets reach it as diverging spherical waves*, we are in the regime called **Fresnel diffraction**. In school laboratories and in advanced-physics syllabuses we work exclusively in the **Fraunhofer regime**; the rest of the section assumes the screen is comfortably distant or, equivalently, that a converging lens has been used to bring parallel rays to focus.

The single slit

Take a single slit of width a , illuminated by a plane wave of wavelength λ arriving along the optical axis. The geometry is shown in the figure below. Far away on a screen at perpendicular distance L from the slit, we want to find the directions θ along which the diffracted intensity vanishes (the dark fringes) and the directions along which it is maximal (the bright fringes).

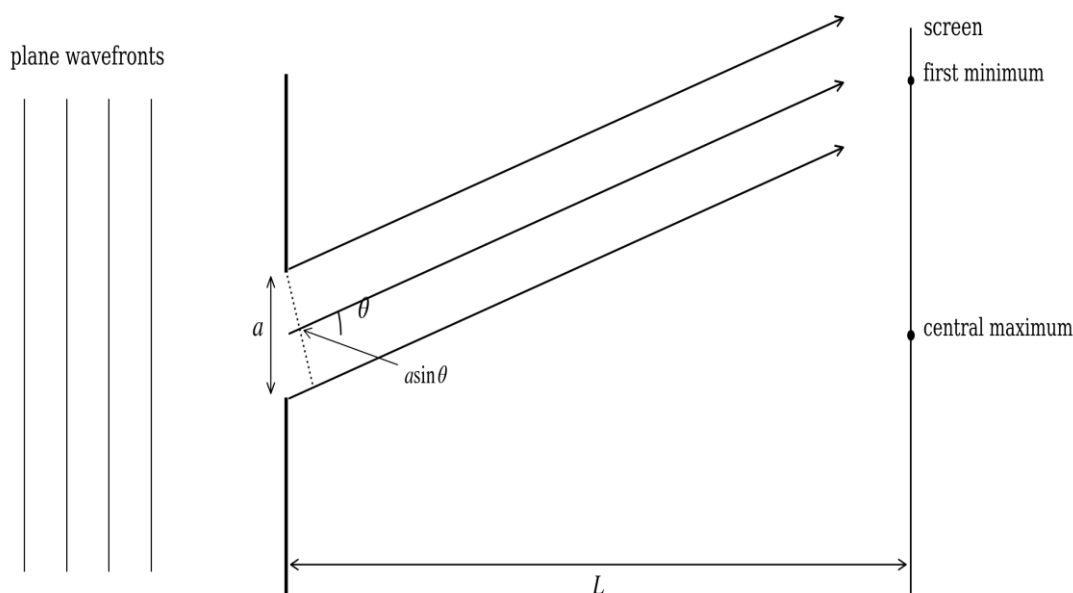


Figure: A plane wave arrives at a slit of width a . Three representative rays leave the slit along the same angle θ to the optical axis. The path difference between the ray from the top of the slit and the ray from the bottom is $a \sin \theta$. The path difference between any strip in the upper half and its partner directly opposite in the lower half is $(a/2) \sin \theta$. The first minimum occurs when this paired difference equals $\lambda/2$.

The slit is not two sources but a continuous infinity of them. Every point on the wavefront crossing the slit acts as a secondary source (Huygens' principle, installed earlier in this chapter). To find the resultant at angle θ , divide the slit mentally into $2N$ equal strips, each of width $\frac{a}{2N}$, and pair the strip at position y in the upper half with the strip directly opposite, at position $y - a/2$, in the lower half. The path difference between any such paired pair, as their light leaves the slit at angle θ , is $(a/2)\sin \theta$, which is the horizontal distance $a/2$ projected on to the direction of propagation.

If this paired path difference equals exactly half a wavelength, the two strips arrive perfectly out of phase and cancel each other on the screen. Pair after pair cancels in the same way, and the whole slit contributes nothing at that angle. The condition for the first minimum is therefore:

$$\frac{a}{2} \sin \theta = \frac{\lambda}{2}$$

or, multiplying both sides by two,

$$a \sin \theta = \lambda$$

The argument that produced the first minimum can be repeated by dividing the slit into four halves, six thirds, eight quarters, and so on. Each subdivision yields one more cancellation condition, and the general condition for a dark fringe in single-slit diffraction is:

$$a \sin \theta = \pm n\lambda, \quad n = 1, 2, 3, \dots$$

The $n = 0$ case is excluded because it corresponds to $\sin \theta = 0$, which is the optical axis itself, and along the optical axis every strip of the slit arrives in phase. This direction is the bright central maximum, and the diffraction pattern is symmetric about it: dark fringes appear in pairs at the same angle to either side, brought to the same position on the screen as $y = \pm n\lambda L / a$ under the small-angle approximation $\sin \theta \approx \tan \theta \approx y / L$.

The intensity envelope and the width of the central maximum

Between the dark fringes lie bright bands, but they are not all the same brightness. A careful calculation of the resultant amplitude as a function of angle gives the intensity envelope shown in the next figure. The mathematical form is:

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

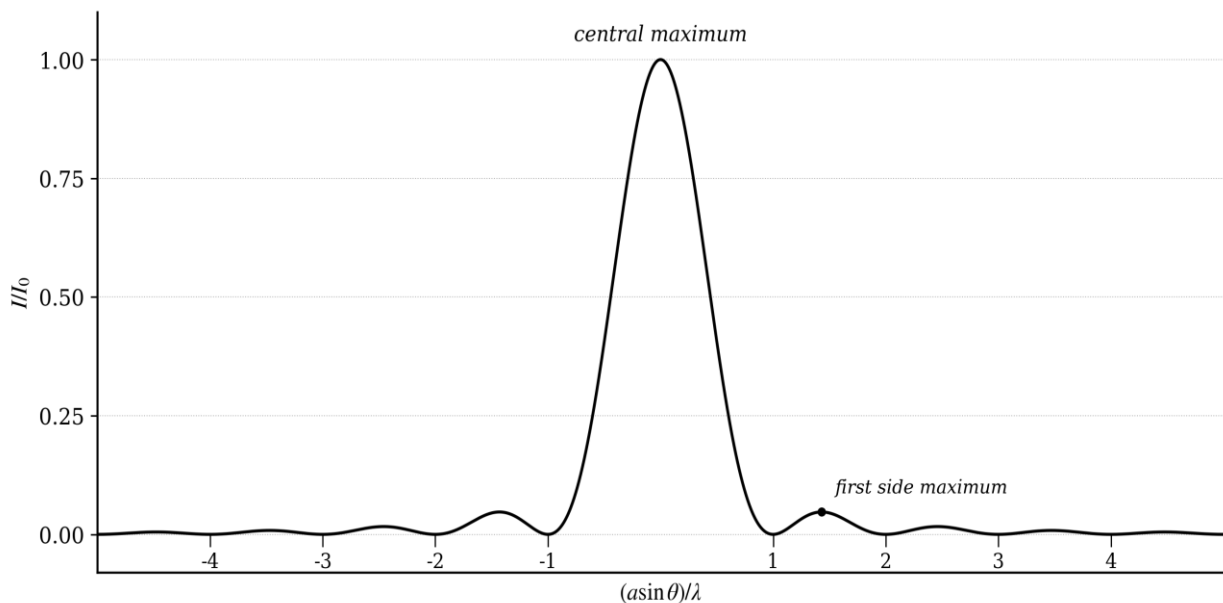


Figure: Intensity envelope $I(\theta) = I_0(\sin \alpha / \alpha)^2$ of single-slit diffraction, plotted as a function of $(a \sin \theta) / \lambda$. The central peak occupies the region from -1 to $+1$, the minima are at integer values, and the first side maximum carries only about 4.7% of the central peak intensity.

Three features of the envelope deserve attention.

First, the central maximum is twice as wide as any of the side maxima, because the first minima occur at $\sin \theta = \pm\lambda/a$ while consecutive side minima are spaced by λ/a only. On a screen at distance L the full width of the central maximum is therefore:

$$W = \frac{2\lambda L}{a}$$

Second, the side maxima are dim. The first side maximum carries only about four-and-three-quarters per cent of the central peak intensity; the second side maximum is dimmer still. The vast majority of the diffracted light sits inside the central peak, and the rest is a faint corona that the eye notices only on a dark screen.

Third, the position of every minimum and every maximum depends on the slit width a and the wavelength λ through the single ratio λ/a . A narrower slit gives a wider pattern; a wider slit gives a narrower pattern. In the limit of a very wide slit the diffraction pattern collapses to a single bright spot on the optical axis, which is the geometric-optics result you have been trusting since O-level. Diffraction recovers geometric optics whenever the slit is many wavelengths wide.

The diffraction grating

Now repeat the experiment with not one slit but many. A **diffraction grating** is a regular array of identical slits at a common spacing d (the slit-to-slit distance, measured between the centres of adjacent slits). A school grating typically carries between three hundred and one thousand slits per millimetre, scratched into a thin glass plate so that the gaps act as the slits and the scratches act as the opaque material between them. Figure below shows the geometry.

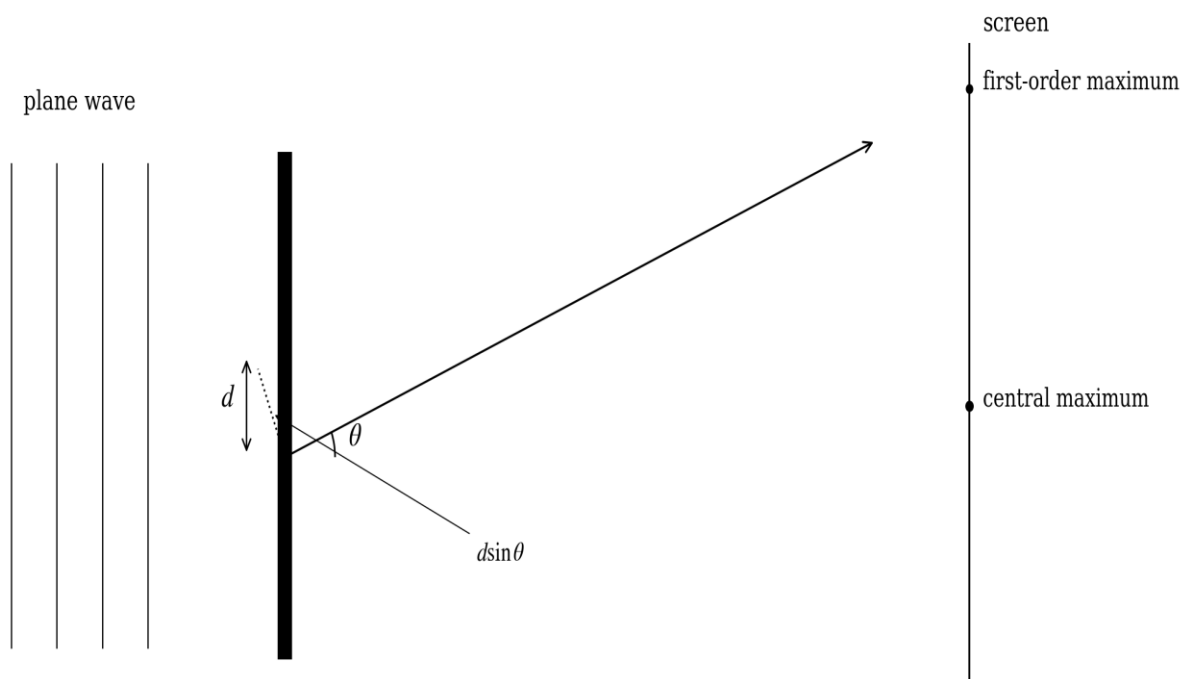


Figure: Plane wave incident on a grating of slit spacing d . Two representative rays leave adjacent slits at the same angle θ to the optical axis. The path difference between them is $d \sin \theta$. The grating produces a bright principal maximum whenever this path difference equals an integer number of wavelengths.

Apply Huygens' principle slit by slit. Each slit acts as a coherent secondary source, and the path difference between waves leaving two adjacent slits, at the common angle θ to the optical axis, is $d \sin \theta$. For all the slits to reinforce one another at that angle, the path difference must equal an integer number of wavelengths,

$$d \sin \theta = n\lambda, \quad n = 0, \pm 1, \pm 2, \pm 3, \dots$$

The integer n is called the order of the principal maximum. The $n = 0$ case is the central maximum at $\theta = 0$; the first-order pair sits at $\sin \theta = \pm\lambda/d$; the second-order pair at $\sin \theta = \pm 2\lambda/d$; and so on.

Two further conveniences of the grating formula are worth pinning down. The slit spacing d is most easily quoted as its reciprocal, the number of lines per unit length,

$$N = \frac{1}{d},$$

so a grating with five hundred lines per millimetre has $d = (1/500)\text{mm} = 2.00 \times 10^{-6}\text{m}$. And the formula $d \sin \theta = n\lambda$ relates four quantities, any three of which determine the fourth. Knowing the wavelength and the spacing, one computes the angle at every order; knowing the angle and the spacing, one computes the wavelength of the source; knowing the angle and the wavelength, one computes the spacing of an unknown grating. The next two subsections exercise each pairing.

How many orders can be seen

Not every value of n produces a physical principal maximum. The grating equation requires $\sin \theta = n\lambda/d$, and the sine of a real angle cannot exceed unity. The highest order that the grating can produce on a screen is therefore the largest integer n satisfying $n\lambda \leq d$, or equivalently:

$$n \leq \frac{d}{\lambda}, \text{ and thus:}$$

$$n_{\max} = \text{floor}\left(\frac{d}{\lambda}\right)$$

The **floor** function (*round-down-to-the-nearest-integer*) is doing essential physical work here, not pedantry. There is no fractional principal maximum sitting on the screen at the angle implied by a non-integer ratio d/λ . The next dialogue makes the point precisely.

Kipute: Sir, for a grating with five hundred lines per millimetre, illuminated by green light of wavelength 526nm, the ratio d divided by λ comes to 3.80. So the maximum order is four, after rounding up.

Mr. Akilikubwa: Stop, Kipute. Read the formula once more. $n_{\max}$ is the floor of d/λ , not the rounded value. Why floor?

Kipute: Because $\sin \theta$ cannot exceed one.

Mr. Akilikubwa: Correct. So whatever $n_{\max}$ is, it is the largest integer for which $\sin \theta$ comes out at less than or at most equal to one. For your grating, n equals three gives $\sin \theta$ equal to 3 times 526 nanometres divided by 2 micrometres, which is 0.789. That fits. n equals four would need $\sin \theta$ equal to 1.052, which does not fit; an angle whose sine is bigger than one does not exist. So the answer is three. There is no 3.8th maximum sitting on the screen. There is the third maximum at one angle, and after that, geometry has no further angles to offer.

Kipute: Always floor.

Mr. Akilikubwa: Always. The floor function in this formula is not a mathematical convention; it is the geometric statement that sines top out at one. The day a sine exceeds one is the day a new physics is being discovered.

Why a grating is sharper than a double slit

Both Young's two-slit setup and the diffraction grating obey the same constructive-interference condition $d \sin \theta = n\lambda$. The angles at which the bright fringes appear are therefore identical. Yet the grating's lines are visibly sharper, narrower, and brighter than Young's. Why?

The answer is in the number of slits, N . With two slits the resultant amplitude oscillates between full reinforcement (every $d \sin \theta = n\lambda$) and zero (every $d \sin \theta = (n + \frac{1}{2})\lambda$) as the angle is swept. With N slits, full reinforcement still occurs at the same angles, but the cancellation between them is far more thorough; between consecutive principal maxima the resultant amplitude oscillates many times and passes through $N - 1$ subsidiary minima and $N - 2$ much weaker subsidiary maxima. The principal maxima themselves become narrower in inverse proportion to N :

$$\Delta\theta_{\text{principal}} \propto \frac{1}{N d \cos \theta}$$

A grating with a thousand slits illuminated across its width therefore gives principal maxima a thousand times narrower than a two-slit pattern at the same d and λ . This is the property that makes the diffraction

grating, and not Young's two slits, the tool of choice for measuring wavelengths precisely. A double-slit pattern can locate a wavelength to perhaps one per cent; a grating can locate the same wavelength to one part in a million. The principle is identical; the resolution differs by six orders of magnitude.

Two wavelengths sharing a direction

A light source rarely emits a single wavelength. A sodium lamp emits two closely spaced yellow lines at 589.0nm and 589.6nm; a mercury vapour lamp emits a spectrum of distinct lines; daylight emits a continuum. When such a source illuminates a grating, each wavelength produces its own family of principal maxima, and the families intersect at angles where two wavelengths happen to share a common diffraction angle. The condition for such a coincidence is simply

$$n_1 \lambda_1 = n_2 \lambda_2 = d \sin \theta,$$

with n_1 and n_2 positive integers and the same product $d \sin \theta$ on both sides. Given the two wavelengths, the smallest positive integers (n_1, n_2) satisfying this equation locate the smallest-angle coincidence; the second coincidence sits at $(2n_1, 2n_2)$, and so on. Several examination problems set the integers and the wavelengths and ask for the angle; several others set the integers and the angle and ask for the missing wavelength. The arithmetic is the same in both directions.

Measuring a wavelength with the spectrometer

All of the above is at its most useful in the school's spectrometer, a compact instrument designed exactly to measure the angles between the principal maxima of a grating, and from them the wavelengths of a source. Figure below shows the layout.

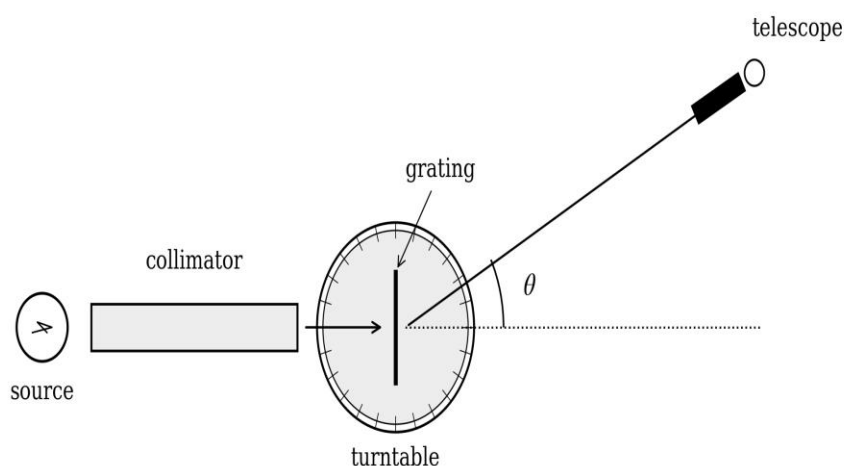


Figure: Bench-top spectrometer arrangement. The source is imaged through a fine slit by the collimator; the parallel beam falls on a grating mounted on the turntable; the diffracted parallel rays are collected by the rotatable telescope, whose pointing angle is read off the turntable scale.

The instrument has three optical components and one calibrated angular scale. The collimator turns the diverging light from the source slit into a parallel beam. The grating, mounted on a rotatable turntable with its plane perpendicular to the incident beam, diffracts the parallel beam into its principal maxima. The telescope, also rotatable on the same turntable, is swung along the angular scale until its cross-wires sit on a chosen principal maximum, and the angle read off the scale is recorded.

Because the turntable's zero is rarely exactly at $\theta = 0$, the standard practice is to swing the telescope to the principal maximum on the right of the central maximum (recording angle θ_R), then swing it across centre to the corresponding principal maximum on the left (recording angle θ_L), and take the half-difference,

$$\theta = \frac{\theta_R - \theta_L}{2}$$

This neatly removes whatever zero-error the turntable carries. With θ so measured and d known from the grating's specification, the wavelength of the source is computed from the grating equation,

$$\lambda = \frac{d \sin \theta}{n}$$

Repeating the measurement at several orders and averaging is the standard refinement. With a thousand-lines-per-millimetre grating, a school spectrometer routinely measures sodium yellow to within one nanometre, which is the precision that decided the discovery of helium on the sun in 1868 before any chemist had isolated it on Earth.

Five worked examples now follow. They are where the formulae stop being decoration and start being instruments.

BINDER Example 17

A laser of wavelength 650nm illuminates a narrow slit. A diffraction pattern is observed on a screen at perpendicular distance 4.0m behind the slit. The two dark fringes that bracket the central bright band are measured to be 52mm apart. Calculate the width of the slit.

Solution

Let a be the slit width, $\lambda = 6.50 \times 10^{-7}\text{m}$ the wavelength, $L = 4.0\text{m}$ the slit-to-screen distance, and $W = 52 \times 10^{-3}\text{m}$ the full width of the central maximum. The full width is the distance between the first minima on either side of the centre, namely:

$$W = \frac{2\lambda L}{a}$$

Rearrange for a :

$$a = \frac{2\lambda L}{W}$$

Insert numerical values:

$$a = \frac{2 \times (6.50 \times 10^{-7}\text{m}) \times (4.0\text{m})}{52 \times 10^{-3}\text{m}} = 1.00 \times 10^{-4}\text{m},$$

which is 0.100mm, or 100 μm . The slit is therefore one tenth of a millimetre wide.

Making Sense of the Answer: A slit width of 100 micrometres is right in the range that produces visible diffraction with red laser light. Anything much wider would give a central maximum thinner than the laser dot itself, and the diffraction would not be visible to the eye; anything much narrower would spread the central maximum so wide that the dark fringes would leave the screen. The fact that the answer comes out at exactly the order of magnitude a school laser slit was designed for is a strong sanity check on the arithmetic.

Think Like a Physicist: Whenever a single-slit problem gives you the full width of the central maximum, write $W = 2\lambda L/a$ and rearrange. Whenever it gives the distance from the centre to the first minimum, write $y_1 = \lambda L/a$ and rearrange. The two formulas are the same formula, differing only by a factor of two for the full-width-versus-half-width convention. Read the problem carefully to know which one you have been handed.

BINDER Example 18

A diffraction grating with 667 lines per millimetre is illuminated normally by monochromatic light of wavelength 530nm. Calculate the angles to the normal at which the first-order and second-order principal maxima are observed.

Solution

Let $N = 667 \text{ lines/mm} = 6.67 \times 10^5 \text{ lines/m}$, so the slit spacing is :

$$d = \frac{1}{N} = \frac{1}{6.67 \times 10^5 \text{ m}^{-1}} = 1.50 \times 10^{-6}\text{m}.$$

The grating equation reads $d \sin \theta = n\lambda$, which rearranges to:

$$\sin \theta = \frac{n\lambda}{d}$$

First-order ($n = 1$):

$$\sin \theta_1 = \frac{1 \times 5.30 \times 10^{-7}\text{m}}{1.50 \times 10^{-6}\text{m}} = 0.3533,$$

$$\theta_1 = 20.7^\circ.$$

Second-order ($n = 2$):

$$\sin \theta_2 = \frac{2 \times 5.30 \times 10^{-7} \text{m}}{1.50 \times 10^{-6} \text{m}} = 0.7067,$$

$$\theta_2 = 45.0^\circ.$$

The first-order principal maximum sits at 20.7° and the second-order at 45.0° to the normal.

Making Sense of the Answer: The second-order angle is a little more than twice the first-order angle, not exactly twice. The doubling of $\sin \theta$ does not produce a doubling of θ because sine is not a linear function. This non-linearity is the reason a grating's higher orders sit closer together as the order increases, and is the reason gratings are easier to read in the first two or three orders than in the eighth or ninth. The arithmetic also exposes how easy it is to misread the grating equation: students who confuse sine and tangent here will get the right answer at small angles and very wrong answers at large ones. Stick to sine.

Think Like a Physicist: In every grating problem, convert lines per millimetre to slit spacing in metres before you do anything else. Carrying the factor of a thousand through the rest of the algebra is the most common source of an answer that is wrong by exactly that thousand-fold.

REAL Example 19

A spectrometer is set up to examine yellow sodium light of wavelength 580nm using a diffraction grating ruled with 500 lines per millimetre. Calculate the highest order of principal maximum that can be observed and, separately, the angle at which it appears.

Solution

The slit spacing is:

$$d = \frac{1}{5.00 \times 10^5 \text{ m}^{-1}} = 2.00 \times 10^{-6} \text{m}$$

Sines have an upper limit of one. The grating equation rearranged for the order is

$$n = \frac{d \sin \theta}{\lambda},$$

so the largest integer n that the grating can accept must satisfy $n\lambda \leq d$, or

$$n_{\text{max}} = \text{floor}\left(\frac{d}{\lambda}\right) = \text{floor}\left(\frac{2.00 \times 10^{-6} \text{m}}{5.80 \times 10^{-7} \text{m}}\right) = \text{floor}(3.448) = 3.$$

The highest visible order is $n = 3$. Its angle follows from the grating equation:

$$\sin \theta_3 = \frac{3 \times 5.80 \times 10^{-7} \text{m}}{2.00 \times 10^{-6} \text{m}} = 0.870,$$

$$\theta_3 = 60.5^\circ.$$

The third-order principal maximum is the highest visible. It appears at an angle of 60.5° to the normal on each side of the central maximum.

Making Sense of the Answer: The floor function is doing the work of keeping the geometry honest. A naive rearrangement of d/λ gives 3.45, and the temptation is to round that up to four. But the fourth order would require $\sin \theta$ to be greater than one, which is geometrically impossible. The grating quietly refuses to deliver a fourth maximum, and the floor function is the mathematical statement of that refusal. The remaining third-order maximum at 60.5 degrees is consistent: $\sin 60.5$ degrees is 0.870, which matches the algebra, and a beam diffracted at 60 degrees is a beam diffracted at a very large angle, so the third order is sitting close to the grating's geometric limit.

Think Like a Physicist: When a problem asks for the highest order, compute d divided by λ , then immediately apply the floor function before doing anything else. Resist the temptation to round in either direction. Only after the floor function gives an integer should you compute the angle, and that angle should always come out at less than ninety degrees if the floor was applied correctly.

HOT Example 20

In a spectrometer experiment, a diffraction grating illuminated normally by blue light of wavelength 480nm produces its second-order principal maximum at an angle of 35.2° to the normal. Calculate the number of lines per millimetre of the grating.

Solution

Let N be the number of lines per metre and $d = 1/N$ the slit spacing. The grating equation rearranges for the spacing as:

$$d = \frac{n\lambda}{\sin \theta}$$

Substituting numerical values (using $n = 2$ and $\sin 35.2^\circ = 0.5764$):

$$d = \frac{2 \times 4.80 \times 10^{-7} \text{m}}{0.5764} = 1.665 \times 10^{-6} \text{m}$$

Therefore:

$$N = \frac{1}{d} = \frac{1}{1.665 \times 10^{-6} \text{m}} = 6.00 \times 10^5 \text{m}^{-1}$$

Converting to lines per millimetre:

$$N = \frac{6.00 \times 10^5}{10^3 \text{mm}} = 600 \text{ lines/mm}$$

The grating has 600 lines per millimetre.

Making Sense of the Answer: 600 lines per millimetre is bang in the middle of the laboratory range for ruled diffraction gratings, which typically run from 100 lines per millimetre at the coarse end (used for the infrared) up to 2400 lines per millimetre at the fine end (used for the ultraviolet). A blue line at 480 nanometres diffracting at 35 degrees in the second order is a sensible setup that anyone with a school spectrometer can reproduce. Had the answer come out at 60000 lines per millimetre or at 60 lines per millimetre, we would suspect a factor-of-a-thousand slip in the arithmetic; coming out at 600 is the right neighbourhood.

Think Like a Physicist: The grating equation always carries four quantities: order, wavelength, angle, and spacing. Any three of the four determine the fourth. Reading the problem to identify which three are given is more than half the work. Once you know the three you have, write the equation with the fourth quantity on its own on the left, and the substitution is mechanical from there.

HOT Example 21

A diffraction grating of slit spacing $2.0\mu\text{m}$ is illuminated normally by light containing two wavelengths. The third-order principal maximum of the longer wavelength is observed to fall on the screen at exactly the same angle as the fourth-order principal maximum of the shorter wavelength. The longer wavelength is 600nm. Calculate the shorter wavelength, and the angle at which the two maxima coincide.

Solution

Let $\lambda_1 = 6.00 \times 10^{-7} \text{m}$ be the longer wavelength and λ_2 the shorter, and let θ be the common angle at which the two principal maxima coincide. The two grating equations are $d \sin \theta = 3\lambda_1$ and $d \sin \theta = 4\lambda_2$. Equating the two right-hand sides,

$$3\lambda_1 = 4\lambda_2, \text{ so } \lambda_2 = \frac{3\lambda_1}{4}$$

Substituting numerical values:

$$\lambda_2 = \frac{3 \times 6.00 \times 10^{-7} \text{m}}{4} = 4.50 \times 10^{-7} \text{m} = 450 \text{nm}$$

The shorter wavelength is 450nm. Its coincidence angle follows from either of the two grating equations:

$$\sin \theta = \frac{3\lambda_1}{d} = \frac{3 \times 6.00 \times 10^{-7} \text{m}}{2.00 \times 10^{-6} \text{m}} = 0.900, \\ \theta = 64.2^\circ.$$

The two maxima coincide at 64.2° to the normal.

Making Sense of the Answer: The two wavelengths span the orange end of the spectrum (600nm) and the blue end (450nm), which together form a fair share of the visible band. The fact that two such different colours can share an angle on the same grating is one of the reasons why high-order spectra are difficult to interpret without prior knowledge of which colours are present; the second-order red can sit on top of the

third-order green, the third-order orange can sit on top of the fourth-order blue, and so on. Spectrometers reserve their highest-precision work for the lower orders where coincidences are rarer.

Think Like a Physicist: Whenever a problem mentions that two orders of two wavelengths share an angle, set the two right-hand sides of the grating equation equal to each other and solve for the missing wavelength. The slit spacing drops out of that step; it returns only when you compute the actual angle. Two parts, both clean, neither difficult.

We have now extracted, from a single slit and from a grating, three formulas that connect the four quantities of any diffraction experiment: the slit width or spacing, the wavelength of the source, the angle of the principal maximum, and the order. With these in hand, the chapter has met its second physical question. A wave passing one obstacle of size comparable with the wavelength produces a pattern of intensity whose geometry is completely determined by the obstacle's size and the wave's wavelength.

Two physical operations now have a complete account: superposition of two beams (interference) and superposition within a single beam passing an obstacle (diffraction). In a real apparatus, both happen at once. A Young's double-slit pattern is not, strictly speaking, made only of two-source interference fringes; it is made of those fringes modulated by the diffraction envelope of each single slit. A grating pattern is not, strictly speaking, only the constructive sum of N sources; the diffraction envelope of each slit weights the principal maxima by an angle-dependent factor. The next short reconciliation makes the relationship between the two operations explicit and answers, in six clean differences.

INTERFERENCE VERSUS DIFFRACTION

The chapter has spent its pages on two physical operations, and the time has come to set them side by side and admit what is the same and what is different. **Interference** asks *what two coherent waves do when they arrive at the same point*. **Diffraction** asks *what one wavefront does when it passes an obstacle whose size is comparable with the wavelength*. The two questions sound different. Their answers, on closer inspection, are made of the same ingredient: **superposition**. Where they part company is in what is superposing, how many of it there are, and whether the superposing pieces were physically separated to begin with.

A six-row comparison brings the differences to the surface in the cleanest available form.

Property	Interference (two-source)	Diffraction (single-source spreading)
Source structure	Two coherent sources of comparable amplitude, physically separated.	One wavefront passing one obstacle of size comparable with λ .
Origin of the pattern	Superposition of two whole waves.	Superposition of infinitely many secondary wavelets from a single obstacle.
Fringe widths	All fringes are equally spaced.	Central maximum is twice as wide as any of the side maxima.
Intensity of fringes	All bright fringes have the same intensity.	Intensity falls off rapidly with order; the envelope is $(\sin\alpha/\alpha)^2$.
Order count	Bounded by $\sin\theta \leq 1$: $n_{\max} \approx d/\lambda$; large for typical d .	Bounded by $\sin\theta \leq 1$: $n_{\max} \approx a/\lambda$; smaller, because a single slit is narrower than a slit spacing.
Visibility	Crisp fringes visible without lenses, simply by projecting the screen far behind the slits.	Often requires a converging lens to focus the diffracted parallel rays onto a finite screen.

The two columns above are the differences. What follows is the unity behind them: in every real apparatus, both operations happen at once, and the pattern the eye finally sees is the product of the two.

Why both phenomena coexist in every real apparatus

Take Young's double-slit experiment, the first apparatus the chapter built. The setup is two narrow slits, separated by a distance d , illuminated by a coherent wavefront. The interference formula tells us that the screen will carry bright fringes at every angle where the path difference between the two slits is an integer number of wavelengths,

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

That is the truth, but not the whole truth. A real slit is not a mathematical line. It is a slit of some finite width a , and the wave passing through it does not emerge as a clean spherical wavelet; it emerges as a full single-slit diffraction pattern with the intensity envelope we met in the previous section,

$$I_{\text{diff}}(\theta) \propto \left(\frac{\sin \alpha}{\alpha} \right)^2, \quad \alpha = \frac{\pi a \sin \theta}{\lambda}$$

The two-slit interference pattern, with peaks at $d \sin \theta = n \lambda$, is therefore multiplied at every angle by the single-slit envelope. What appears on the screen, $I_{\text{obs}}(\theta)$, is the product:

$$I_{\text{obs}}(\theta) \propto \cos^2 \left(\frac{\pi d \sin \theta}{\lambda} \right) \times \left(\frac{\sin \alpha}{\alpha} \right)^2$$

Three observations explain why this product is the right picture of what is actually seen at the school bench.

First, the $\cos^2$ factor oscillates much faster than the $(\sin \alpha / \alpha)^2$ envelope, because the ratio d/a is typically four or five for a Young's-slit card. The envelope wears the $\cos^2$ -fringes as a slow modulation. **Second**, the angle θ at which the diffraction envelope reaches its first zero, $a \sin \theta = \lambda$, is far smaller than the angle θ at which the interference reaches its first zero, $d \sin \theta = \lambda/2$, so most of the visible fringes sit inside the central diffraction lobe and the eye sees a few orders of bright $\cos^2$ fringes before the envelope dims them into invisibility. **Third**, the fringe whose centre would have been at the diffraction minimum is missing from the screen entirely, because the envelope multiplies it by zero. This is the so-called **missing order**, and it gives a quick experimental measure of the ratio d/a : count the number of bright $\cos^2$ peaks inside the central diffraction lobe, divide by two, and the result is d/a almost exactly.

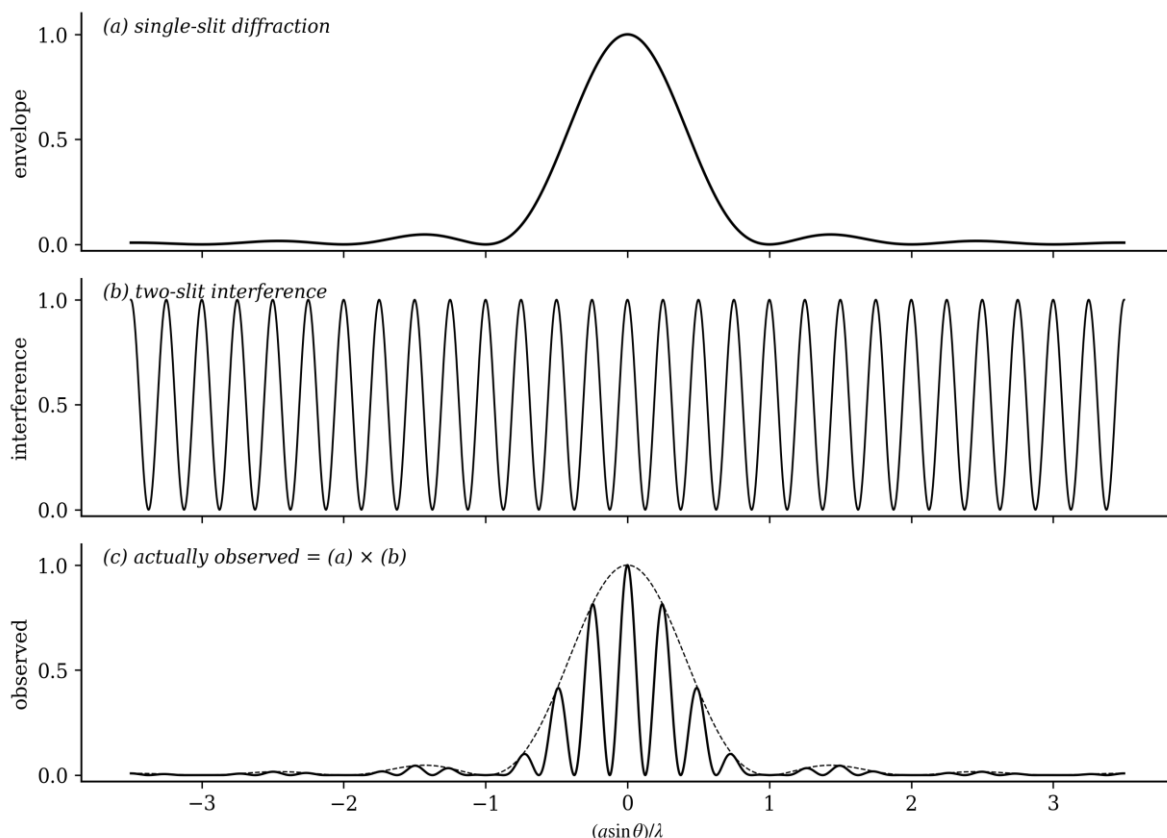


Figure: How the two operations combine on the screen in Young's double-slit experiment. Panel (a) is the diffraction envelope of one slit on its own. Panel (b) is the interference pattern of two infinitesimally narrow slits on their own. Panel (c), the pattern an experimenter actually observes, is the point-by-point product of the two. The fast oscillation in (c) is interference; the slow shape that wraps around it is diffraction; the small fringes that go missing where the envelope passes through zero are the missing orders.

Two further points generalise the picture to a diffraction grating. The grating with N slits replaces the $\cos^2$ interference factor with the much sharper N -slit interference function, whose principal maxima are at the same angles $d \sin \theta = n\lambda$ but whose linewidth shrinks in inverse proportion to N . The diffraction envelope of each individual slit is unchanged. The actually observed pattern is, once again, the product of an interference factor that depends only on d and an envelope that depends only on a . The envelope is the slower; the interference is the faster; and the missing orders of Young's slits become, in the grating, suppressed orders that the manufacturer routinely chooses where to place by tuning d/a during the ruling.

So the difference between interference and diffraction is a difference of vocabulary. Once both operations are admitted to the same apparatus, the difference between them is no longer a contradiction; it is a multiplication. The wave model that has carried the chapter for about seventy pages does not care whether the screen pattern is read out as fringes inside an envelope or as the product of two superposition sums. The two readings are the same thing said twice.

Two questions of the chapter have now received their full account. Where the wave goes when two waves arrive at one point, and where the wave goes when one wave passes one obstacle, are both, in the end, exercises in adding amplitudes and squaring. A third question remains for the chapter to answer. The wave has been described, all the way through, as a transverse disturbance. The direction of the disturbance, the way the wave shakes as it travels, has been treated as a quiet detail in the background. It is time to bring that detail forward. The next section is about polarisation, the way in which a transverse wave commits itself to a single plane of vibration and what happens when you ask it to commit to a different one.

POLARISATION

The chapter has so far asked the wave only two questions. *Where does it go when two of itself meet at a point? Where does it go when one of itself meets a slit?* Both questions take the wave as a scalar quantity, a height above the average that adds, subtracts, and squares. There is a third question that the previous two have left unattended. The wave is transverse: it shakes across the direction in which it travels. Across is not one direction. It is two, and within those two there is an infinite family of intermediate directions. The question of polarisation is the question of which of those directions the wave has chosen, and what happens when an apparatus or a medium asks it to choose differently.

Two reasons make this a topic the syllabus cannot afford to bury at the back. First, polarisation is the simplest experimental signature by which a transverse wave is told apart from a longitudinal wave. Sound, the most familiar wave of school physics, is longitudinal and cannot be polarised; light, the wave the chapter has been studying, is transverse and can. Second, polarisation has wandered out of the laboratory and into the engineering of nearly every optical device the candidate will hold during their working life. Liquid-crystal screens, photographic filters, three-dimensional cinema spectacles, and the strain test of a windscreen all depend on it. The opening pages of this section build the idea from a single experiment, derive the law that governs it, and then take a quick tour of the four ways in which plane-polarised light can be produced.

A test that tells a transverse wave from a longitudinal one

Take two identical pieces of Polaroid film, the kind sold for camera filters or polarising sunglasses. Place one in front of a window through which the morning sun is falling on the louvre slats of the Miono Secondary classroom. Looking through the single film, the world is dimmed but still visible. Now place the second film in front of the first and slowly rotate it about the line of sight. The brightness rises and falls; at one orientation the view is at its brightest, at the orientation a quarter-turn away it has gone almost completely dark, and the cycle repeats every half-turn. Two crossed Polaroids extinguish the light.

This is not what would happen if light were a longitudinal wave. A longitudinal disturbance shakes along the direction of travel, and there is only one such direction; nothing rotated about the line of travel could ever block such a wave. The fact that two films can be arranged to extinguish the light, and then unblock it by a quarter-turn, proves at once that whatever is shaking inside the wave is shaking sideways to its motion, and that the sideways direction can take more than one orientation. The wave is transverse, and the two films are sorting it by the orientation of its shake. The film that emits the sorted wave is the **polariser**; the second film, used to detect the orientation that the first has chosen, is the **analyser**.

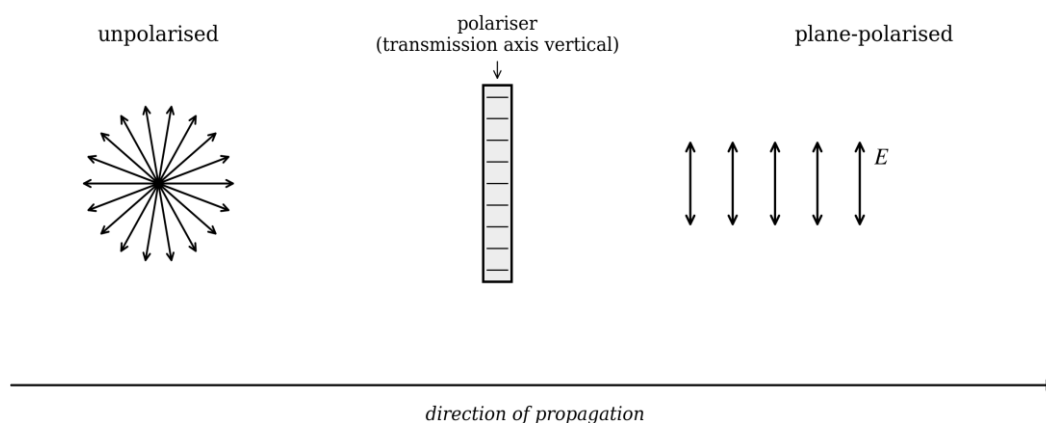


Figure: Unpolarised light has its transverse vibrations spread evenly over every direction perpendicular to the propagation direction; on the left of the diagram, each short double-headed arrow stands for one such oscillation direction, picked at random. A polariser passes only the component along its transmission axis, which here is drawn vertical; what emerges on the right is plane-polarised light, every oscillation now confined to a single direction.

Plane of vibration and plane of polarisation

Two pieces of geometric language need fixing before any law is written down. A light wave is, as the candidate already knows from earlier chapters, a coupled pair of oscillating fields: an electric field, conventionally written $\mathbf{E}$ and pointing along the direction in which the wave shakes its charges, and a magnetic field, conventionally written $\mathbf{B}$ and pointing along a direction perpendicular to $\mathbf{E}$ and also perpendicular to the line of travel. The two fields are tied together: wherever $\mathbf{E}$ is large, $\mathbf{B}$ is also large; wherever $\mathbf{E}$ reverses sign, so does $\mathbf{B}$. The three directions, $\mathbf{E}$, $\mathbf{B}$, and the direction of travel, form a right-angled triple.

With $\mathbf{E}$ identified as the line in which the wave actually shakes, the geometry is straightforward. The plane that contains the line of $\mathbf{E}$ and the direction of travel is the **plane of vibration**. The plane that is perpendicular to it, and which also contains the direction of travel, is the **plane of polarisation**; this is the plane in which $\mathbf{B}$ lies. The two planes intersect along the direction of travel and are at right angles to each other. Older textbooks sometimes use the two names the other way round, so the candidate should be ready to identify which of the two planes is meant by inspecting whether the author is talking about $\mathbf{E}$ or about $\mathbf{B}$. Throughout this book the convention is that the plane of vibration is the plane of $\mathbf{E}$, because $\mathbf{E}$ is the field that polarisers, analysers, photographic plates, and biological detectors actually respond to.

Three states of polarisation will appear in what follows.

- **Unpolarised** light has its electric vector pointing in every direction perpendicular to the line of travel with equal probability; an ordinary lamp, the Sun, and a candle all emit such light.
- **Plane-polarised** (or **linearly polarised**) light has its electric vector confined to one fixed plane; the output of a Polaroid filter is of this kind.
- **Partially polarised** light has a preferred direction of vibration but not an exclusive one; light reflected off a window pane at most angles is of this kind.

The remainder of this section concerns plane-polarised light, because the syllabus and the laboratory both treat it as the case from which everything else follows.

Four ways to make plane-polarised light

Nature does not insist that plane-polarised light be produced by a Polaroid film. Four physically distinct mechanisms are catalogued by the syllabus and will be examined in turn in the next four subsections. They share no apparatus, but they share one outcome: every one of them converts unpolarised light into a wave whose electric vector lies in a single plane. The four mechanisms are:

- 1) **Selective absorption** by an oriented molecular grid (the Polaroid filter);
- 2) **Reflection** at a denser medium at a particular angle (Brewster's law);
- 3) **Double refraction** inside an anisotropic crystal followed by suppression of one of the two refracted beams (the Nicol prism); and
- 4) **Scattering** by particles smaller than the wavelength (the blue of the daytime sky).

The first two are the most often examined and will receive the longest treatment; the third introduces the most beautiful piece of nineteenth-century optical engineering ever placed on a school syllabus; the fourth explains why the sky is the colour it is.

Polaroid film and Malus's law

The film named after the company that first manufactured it consists of long-chain organic molecules, drawn during manufacture into parallel alignment and then doped with iodine atoms that make them electrically conducting along their length. The electric field of a light wave whose vibration is parallel to the molecular chains drives free charge up and down the chain, loses its energy to the chain as heat, and is therefore absorbed. The component whose vibration is perpendicular to the chains finds no charge to push, dissipates no energy, and is transmitted. The transmission axis of the Polaroid is, perhaps counter-intuitively, perpendicular to the direction in which the molecules lie.

Unpolarised light incident on such a film has its electric vector pointing equally in every direction. The component perpendicular to the molecular chains is half of the total energy on average; the component parallel to the chains, also half, is absorbed. A perfect Polaroid film therefore halves the intensity of unpolarised light and turns the transmitted half into plane-polarised light whose plane of vibration is the transmission axis of the film. Write this as a numerical statement to fix it: if I_{unp} is the intensity of unpolarised light, then after one Polaroid film,

$$I = \frac{1}{2} I_{\text{unp}}$$

Now the geometry of two films. Pass plane-polarised light, of intensity I_0 and electric-vector amplitude E_0 , onto a second Polaroid whose transmission axis makes an angle θ with the plane of vibration of the incoming wave. Decompose E_0 into one component along the analyser's transmission axis and one perpendicular to it. The component along the axis has magnitude $E_0 \cos \theta$; the perpendicular component is absorbed. The transmitted amplitude is therefore $E = E_0 \cos \theta$, and the transmitted intensity, being proportional to the square of the amplitude, is:

$$I = I_0 \cos^2 \theta$$

This is **Malus's law**, named for Étienne-Louis Malus, who first wrote it down in 1809 after watching the setting Sun reflected in the windows of the Luxembourg Palace through a piece of Iceland spar. In words:

When plane-polarised light passes through an ideal polarising film, the intensity of the transmitted beam is equal to the intensity of the incident beam multiplied by the square of the cosine of the angle between the plane of vibration of the incident light and the transmission axis of the film.

The law will return in every numerical problem that involves two or more polarising films in series. Notice three details. At $\theta = 0$ the analyser passes the entire incident intensity, because $\cos^2 0 = 1$. At $\theta = \pi/2$, that is, with axes crossed, it passes nothing, because $\cos^2(\pi/2) = 0$; this is the extinction the morning observer saw when crossing the two films. At $\theta = \pi/4$ the transmitted intensity is one half of the incident intensity, because $\cos^2(\pi/4) = 1/2$.

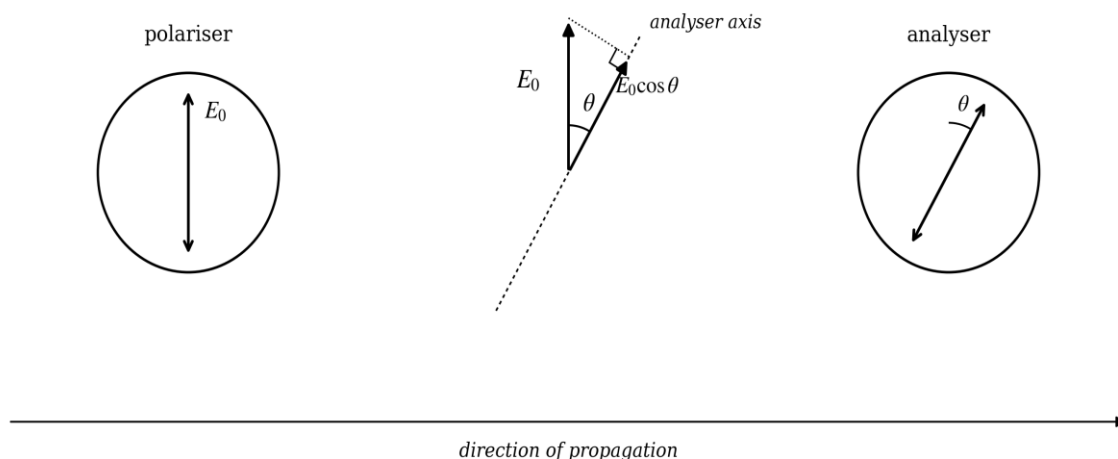


Figure: The geometry from which Malus's law follows. Plane-polarised light leaves the polariser with electric vector along the vertical transmission axis, of magnitude E_0 . The analyser's transmission axis makes an angle θ with the polariser's axis. The component of E_0 along the analyser axis is $E_0 \cos \theta$; the transmitted intensity is therefore proportional to the square of this projected amplitude.

Three polaroids in series

Malus's law has a consequence that is easy to state and unsettling to think about until it is grasped. Cross two polaroids so that no light comes through. Now slide a third polaroid between them at forty-five degrees to either of the outer two. The light that had been extinguished is reborn. A film placed in the path of an already extinguished beam apparently emits light it does not receive. The puzzle is sharp enough to merit a dialogue.

Kipanga: *Mr. Akilikubwa, two polaroids at 0° and 90° give zero light. I just put a third in the middle and the room got brighter behind the third one. The polaroid is making light?*

Mr. Akilikubwa: *The polaroid is not making any light. It is rearranging the light that does pass it. Step through it with me one stage at a time, please.*

Kipute: *First polaroid is vertical. Unpolarised light goes in; half of it comes out, plane-polarised vertical. Intensity $I_0/2$.*

Mr. Akilikubwa: *Correct. Now the middle polaroid is at 45° to the vertical. What does Malus's law say?*

Kipanga: *Transmitted intensity is the incoming intensity multiplied by $\cos^2(45^\circ)$, which is $1/2$. So we have $(I_0/2) \times (1/2) = I_0/4$ coming out of the middle polaroid, with its plane of vibration now along the middle polaroid's axis at 45° .*

Mr. Akilikubwa: *Good. Now the last polaroid is at 90° from the first, which is 45° from the middle one. Apply Malus's law again.*

Kipute: *$(I_0/4) \times \cos^2(45^\circ) = I_0/8$. So we get one eighth of the original intensity, not zero.*

Mr. Akilikubwa: *And now compare with the case where the middle polaroid is taken out. Two polaroids at 0° and 90° : Malus's law gives $I_0/2$ multiplied by $\cos^2(90^\circ)$ which is zero. Removing the middle polaroid extinguishes the light completely; putting it back gives back $I_0/8$.*

Kipanga: *So the middle one is not emitting light from itself; it is changing the plane of vibration so that something is allowed through the third one. A polaroid is not a curtain that blocks light. It is a rotor that turns the plane of vibration to a new direction and discards what cannot fit.*

Mr. Akilikubwa: *Exactly. A polaroid in the path of plane-polarised light absorbs the perpendicular component and keeps the parallel one. Without the middle film, the third film is being asked to pass light whose plane is at 90° to its own axis; nothing to project, nothing to keep. With the middle film, the third film is being asked to pass light whose plane is at 45° to its own axis; half by amplitude, a quarter by intensity, survives. The same physics in both cases.*

The dialogue makes a general point. Adding a polariser to the path of a plane-polarised beam can never increase its intensity; that is obvious from the $\cos^2\theta$ factor, which is at most one. Adding a polariser to the path between two crossed polarisers can however increase the intensity from zero to a positive value, because the second polariser, no longer required to absorb the parallel component of a beam, is being given a beam whose plane has been rotated. Polarisers are not curtains; they are decomposers, keeping the part of the wave that fits and absorbing the rest.

Brewster's law

A second route to plane-polarised light requires no special film. Allow unpolarised light to fall on the surface of glass, water, or any transparent dielectric medium. Some of the light reflects, some refracts. At most angles, both the reflected beam and the refracted beam are partially polarised; the reflected component has a slight excess of vibrations parallel to the surface and the refracted component has a slight excess perpendicular to the surface. At one special angle of incidence, however, the reflected beam is **completely** plane-polarised, with its electric vector lying parallel to the reflecting surface and perpendicular to the plane of incidence. That special angle is the **polarising angle** of the medium and is given the symbol i_p .

Sir David Brewster discovered in 1815 that the polarising angle of a medium is fixed by its refractive index alone, through a relation now called **Brewster's law**. Its derivation is short and instructive. Take a ray of unpolarised light incident at the polarising angle i_p on the boundary between a medium of refractive index unity (air) and one of refractive index n . The reflected ray leaves at angle i_p on the same side of the normal, and the refracted ray bends into the denser medium at some angle r . Brewster's observation, which is the

experimental input to the derivation, is that at the polarising angle the reflected ray and the refracted ray are perpendicular to each other. In symbols,

$$i_p + r = 90^\circ$$

Snell's law for the boundary reads $\sin i_p = n \sin r$. Substitute $r = 90^\circ - i_p$, so that $\sin r = \sin(90^\circ - i_p) = \cos i_p$, and the equation becomes $\sin i_p = n \cos i_p$. Divide both sides by $\cos i_p$:

$$\tan i_p = n$$

Brewster's law in its compact form. In words:

The tangent of the polarising angle of a transparent dielectric medium is equal to the refractive index of that medium relative to the medium of incidence.

Three numerical examples drive home what the law actually says. For water, $n = 1.33$; the polarising angle is $\tan^{-1}(1.33) \approx 53.1^\circ$ measured from the normal, which is 36.9° measured from the water surface. For crown glass, $n = 1.52$; the polarising angle is $\tan^{-1}(1.52) \approx 56.7^\circ$. For diamond, $n = 2.42$; the polarising angle is $\tan^{-1}(2.42) \approx 67.5^\circ$. The angle creeps upward with refractive index but stays well clear of grazing incidence even for the densest natural medium normally encountered.

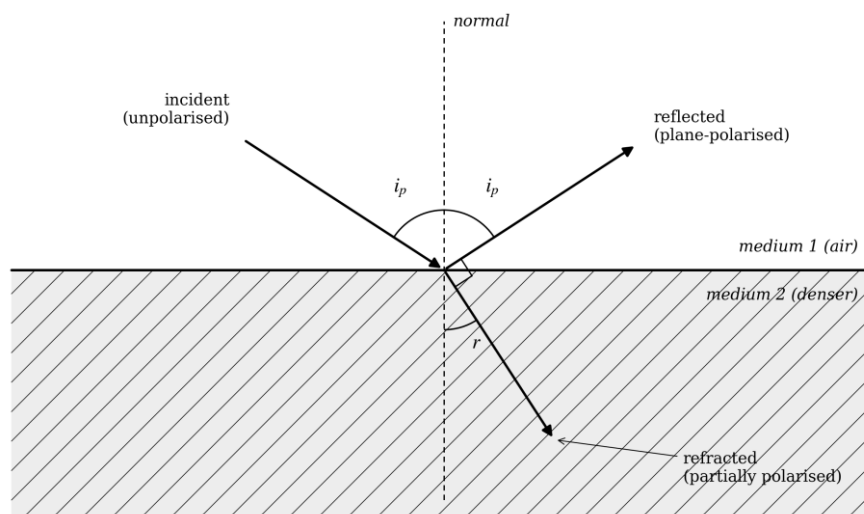


Figure: Reflection and refraction at the polarising angle. The incident ray is unpolarised. The reflected ray, which leaves the surface at angle i_p from the normal, is completely plane-polarised with its electric vector parallel to the reflecting surface. The refracted ray, bent into the denser medium at angle r , is partially polarised. The small square at the origin marks the right angle between the reflected and refracted rays, the geometric condition that defines i_p .

Two practical observations close the topic.

- 1) **Brewster's law explains why glare from a road, the surface of a lake, or a glass shop window is partially or fully horizontal:** it is the horizontal component of the original light that the surface preferentially reflects. Polarising sunglasses are deliberately oriented to absorb horizontal vibrations, and so dim the reflected glare without dimming everything else as much.
- 2) **Brewster's relation gives an easy way to measure the refractive index of an unknown sample:** place it on the bench, illuminate it at a variable angle, and find the angle at which the reflected light vanishes when viewed through an analysing polaroid whose axis is vertical. That angle is i_p ; its tangent is the index.

Double refraction and the Nicol prism

Iceland spar, a clear variety of calcite (CaCO_3), does something unusual when light is sent through it. A single ray entering the crystal emerges as two rays. Pull the crystal across a printed page and the print is seen double. One of the two refracted rays obeys Snell's law, with a refractive index $n_o = 1.658$; the other does not, and its effective refractive index n_e depends on the direction of travel inside the crystal, ranging from 1.486 along one crystal axis to 1.658 along another. The first is called the **ordinary ray** (the o-ray); the

second is called the **extraordinary ray** (the e-ray). The phenomenon as a whole is **double refraction** or **birefringence**, and it is found in any crystal whose lattice has different mechanical stiffness along different axes.

Two further facts make double refraction a useful tool.

First, the o-ray and the e-ray are each individually plane-polarised, and their planes of vibration are at right angles to each other. The crystal does not destroy the polarisation of the original beam; it sorts it into two beams of perpendicular polarisations.

Second, the two beams travel in different directions inside the crystal because of their different indices, and so emerge displaced from each other. If the crystal is thick enough and the user is patient enough, the two beams can be separated by an aperture and one of them used as a source of plane-polarised light.

William Nicol of Edinburgh, in 1828, replaced this awkward aperture trick with a piece of optical engineering of striking economy. He took a parallelepiped of Iceland spar, cut it diagonally along a plane making sixty-eight degrees with the long faces, polished the two pieces flat, and cemented them back together with Canada balsam, a clear resin of refractive index about 1.55. Unpolarised light entering one end of the cemented prism splits into the o-ray and the e-ray as before. The two rays then hit the slanting cement layer at an angle. The o-ray, with its high refractive index of 1.658 inside the calcite, is travelling from a denser medium into a less dense one (the balsam at 1.55); its angle of incidence at the cement is greater than the critical angle for that pair of indices, and it suffers total internal reflection. The e-ray, with its lower effective index of 1.486, is travelling from a less dense medium into a denser one and is transmitted through the cement into the second piece of calcite, eventually emerging at the far end. The o-ray, after total internal reflection, strikes the long side face of the prism, which Nicol blackened with a layer of paint so that the o-ray would be absorbed and not return to confuse the user.

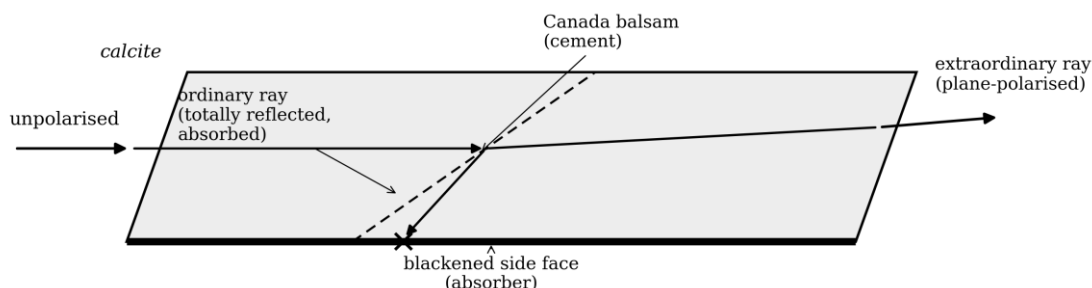


Figure: The Nicol prism. A parallelepiped of calcite is cut diagonally and the two halves cemented back together with Canada balsam (dashed line). Unpolarised light entering the left face is split inside the crystal into an ordinary ray and an extraordinary ray, each plane-polarised. The o-ray is totally internally reflected at the cement and absorbed at the blackened side face; the e-ray passes through the cement and emerges at the right face, plane-polarised.

What emerges from a Nicol prism is therefore a single beam of plane-polarised light, with the unwanted polarisation absorbed by the blackened side face rather than allowed to escape. Two Nicol prisms in series form a polariser-analyser pair as good as any made from Polaroid film, and historically far more accurate. The high cost of optical-quality calcite has now reduced the prism to a museum exhibit in most countries, but its design is examined because it lets the candidate apply the laws of geometrical optics (Snell's law, total internal reflection) to a real device of unmistakable elegance.

Why the sky is blue

The fourth mechanism by which unpolarised light becomes plane-polarised takes place not in a manufactured device but in the air overhead. Sunlight passing through the atmosphere is scattered by the nitrogen and oxygen molecules of the air. The wavelength of visible light is hundreds of nanometres; the size of one of these molecules is less than a nanometre. Scattering by particles much smaller than a wavelength is described by Lord Rayleigh's theory, which gives two predictions that bear directly on this section.

Rayleigh's first prediction is the **inverse-fourth-power scattering law**. Stated in words: *the intensity of light scattered sideways by a particle whose size is much smaller than the wavelength of the light is inversely proportional to the fourth power of the wavelength*. In symbols, with I_{scat} the scattered intensity and λ the wavelength,

$$I_{\text{scat}} \propto \frac{1}{\lambda^4}$$

Blue light has roughly half the wavelength of red light; the factor of two in the wavelength becomes a factor of sixteen in the scattered intensity, and the sky away from the Sun appears blue because it scatters short wavelengths much more efficiently than long ones. The same argument run backwards explains the redness of the Sun at sunrise and sunset: the long path through the atmosphere strips the blue wavelengths out of the direct beam by scattering them sideways, leaving the residual direct beam dominated by the long wavelengths it could not lose.

Rayleigh's second prediction, the one that earns scattering its place in this section, is that the side-scattered light is **plane-polarised**. The mechanism is the same one that drives the polaroid film. Each air molecule is set into oscillation by the electric vector of the incoming sunlight; the oscillating charge then radiates as a tiny dipole, and a dipole radiates most strongly perpendicular to its own line of oscillation and not at all along it. Light scattered through a right angle from the Sun therefore contains only the component whose oscillation is perpendicular to the scattering direction; the parallel component cannot be radiated in that direction at all. The blue patch of sky ninety degrees from the Sun is plane-polarised, and a single polaroid filter rotated against it darkens and brightens by a factor of roughly four as the filter turns through a right angle. The student can do this with a borrowed pair of polarising sunglasses on a bright afternoon.

Optical activity

One last property of plane-polarised light is worth introducing before the section closes. A small number of substances have the curious power of rotating the plane of vibration of plane-polarised light that passes through them. Quartz crystals do it; solutions of cane sugar, tartaric acid, and a number of other organic molecules do it; many crystalline minerals do it; the antibiotic penicillin does it. Substances that have this property are called **optically active**, and the rotation is described as **dextrorotatory** (right-handed, as viewed by an observer looking back into the oncoming beam) or **laevorotatory** (left-handed) according to the sense of the rotation.

The total angle of rotation produced by a given sample is described by **Biot's law of optical activity**, first stated by Jean-Baptiste Biot in 1815. In words:

When plane-polarised light passes through a solution of an optically active substance, the angle through which the plane of vibration is rotated is directly proportional to the path length of the light through the solution and to the concentration of the solution.

Symbolically, for a solution of **concentration c** (in grams per cubic centimetre) and a light path of **length L** (in decimetres), the rotation α in degrees is:

$$\alpha = \alpha_0 L c,$$

where α_0 is the **specific rotation** of the substance, a characteristic constant determined by the **nature of the substance**. For cane sugar in water at twenty degrees Celsius and at the yellow sodium wavelength 589nm, $\alpha_0 = +66.5^\circ \text{dm}^{-1} \text{g}^{-1} \text{cm}^3$. An instrument that exploits the linearity of this relationship to determine the concentration of a sugar solution by measuring its optical rotation is called a **saccharimeter**, and is used in the sugar industry to assay the strength of cane juice and the purity of refined sugar. The student should be able to recognise that optical activity is the rotation of the plane of vibration, and to apply the linear law above to a numerical problem.

Three short problems now turn these ideas into numbers.

BINDER Example 22

Plane-polarised light of intensity 12.0 Wm^{-2} falls on an analysing polaroid whose transmission axis makes an angle of 40° with the plane of vibration of the light. Calculate the intensity of the light that emerges from the analyser.

Solution

Let $I_0 = 12.0 \text{ Wm}^{-2}$ be the incident intensity and $\theta = 40^\circ$ the angle between the analyser axis and the plane of vibration of the incoming light. Malus's law gives the transmitted intensity directly,

$$I = I_0 \cos^2 \theta.$$

$$I = 12.0 \text{ Wm}^{-2} \times \cos^2 40^\circ = 7.04 \text{ Wm}^{-2}$$

The analyser transmits 7.04 Wm^{-2} of light, which is fifty-nine per cent of the incident intensity.

Making Sense of the Answer: The number is sensible: at zero degrees the answer would be the full incident value; at ninety degrees it would be zero; the fortieth-degree value sits closer to the upper end than to the middle of that interval because forty degrees is closer to zero than it is to a right angle. A quick sanity check is that $\cos^2 45^\circ$ would give exactly half the intensity, 6.00 Wm^{-2} , which it does to three significant figures.

Think Like a Physicist: Malus's law is the law of projection. The cosine factor in it is the projection of the incoming electric vector onto the analyser's axis; the square is there because intensity is amplitude squared. Whenever a polariser problem hands you an angle, ask yourself whether that angle is the one between the two transmission axes or between the incoming plane of vibration and one of the axes. Both ways of stating the geometry give the same answer, but only one of them is the angle that goes straight into the formula.

REAL Example 23

Unpolarised light of intensity I_0 passes in succession through three identical polaroids whose transmission axes are oriented at 0° , 30° , and 60° to the vertical. Calculate (a) the intensity of the light that emerges from the third polaroid, and (b) the intensity that would emerge if the middle polaroid were removed. Compare the two.

Solution

Treat the journey one polaroid at a time, applying the unpolarised-light rule at the first polaroid and Malus's law at every subsequent polaroid. The angle in Malus's law is the angle between the transmission axes of consecutive polaroids, not the absolute orientation of either.

(a) The first polaroid converts unpolarised light into plane-polarised light at half the original intensity,

$$I_1 = \frac{1}{2} I_0$$

The second polaroid is at 30° to the first, so by Malus's law:

$$I_2 = I_1 \cos^2(30^\circ) = \frac{1}{2} I_0 \times \frac{3}{4} = \frac{3}{8} I_0$$

The third polaroid is at $60^\circ - 30^\circ = 30^\circ$ to the second, so by Malus's law again:

$$I_3 = I_2 \cos^2(30^\circ) = \frac{3}{8} I_0 \times \frac{3}{4} = \frac{9}{32} I_0 = 0.28125 I_0$$

The intensity of the light that emerges from the third polaroid is $0.28125 I_0$.

(b) With the middle polaroid removed, the third polaroid in the original line-up is now the second in sequence and the angle in Malus's law becomes the full 60° ,

$$I_3' = \frac{1}{2} I_0 \times \cos^2(60^\circ) = \frac{1}{2} I_0 \times \frac{1}{4} = \frac{1}{8} I_0 = 0.125 I_0$$

The intensity that would emerge if the middle polaroid were removed is $0.125 I_0$.

Comparison: Adding the middle polaroid more than doubles the final intensity, from $0.125 I_0$ up to $0.281 I_0$.

Making Sense of the Answer: Inserting a polaroid in the path of a polarised beam adds more light to the beam beyond it. The middle polaroid was not blocking the light; it was rotating its plane of vibration into one that the third polaroid could deal with more efficiently. Two successive thirty-degree projections lose less amplitude than one single sixty-degree projection, because $\cos 30^\circ \times \cos 30^\circ$ is three-quarters while $\cos 60^\circ$ is one-half.

Think Like a Physicist: A polariser is a decomposer of the wave, not a curtain that lets some photons through and stops the others. Each polaroid in a series projects the incoming amplitude onto its own axis

and discards the perpendicular part. Whenever a problem asks for the intensity through a chain of polaroids, walk through them one at a time, multiplying by $\cos^2\theta$ at each step, where θ is the angle between the current polaroid's axis and the axis of the polaroid immediately before it.

HOT Example 24

Light falls from air onto the calm surface of a freshwater lake whose refractive index is 1.33. Calculate (a) the polarising angle for this surface, measured both from the normal and from the surface itself, (b) the angle of refraction of the light that enters the water at this incidence, and (c) verify by direct calculation that the reflected and refracted rays are perpendicular to each other. (d) State the orientation of the electric vector in the reflected ray.

Solution

Apply Brewster's law to the air-water boundary, then Snell's law to find the refraction angle, then check the geometry.

(a) Brewster's law gives the polarising angle directly,

$$\tan i_p = n = 1.33, \quad i_p = \tan^{-1}(1.33) = 53.06^\circ (\text{from normal})$$

Measured from the surface itself, the polarising angle is the complement, $90^\circ - 53.06^\circ = 36.94^\circ$.

(b) Snell's law for the same ray, with the air-side angle equal to i_p , reads:

$$\sin i_p = n \sin r, \quad \sin r = \frac{\sin 53.06^\circ}{1.33} = 0.6007,$$

$$r = \sin^{-1}(0.6007) = 36.94^\circ$$

(c) The angle between the reflected ray and the refracted ray is:

$$i_p + r = 53.06^\circ + 36.94^\circ = 90.00^\circ,$$

confirming Brewster's geometric condition: at the polarising angle the reflected and refracted rays are mutually perpendicular.

(d) At Brewster incidence the reflected ray is completely plane-polarised, with its electric vector perpendicular to the plane of incidence and therefore parallel to the lake surface. The component of the original light whose electric vector lay in the plane of incidence has been fully transmitted into the water as part of the refracted ray; only the component perpendicular to that plane survives in the reflected beam.

Making Sense of the Answer: The polarising angle for water, 53.06° , is one of the three numerical anchors quoted earlier in the section and reappears here as the experimentally testable consequence of $n = 1.33$. The perpendicularity of reflected and refracted rays is the geometric heart of Brewster's law; it is what makes the law work, not a separate fact. The horizontal orientation of the reflected electric vector is the reason polaroid sunglasses with their absorption axis horizontal can dim the glare off the water: the very component the sunglasses are built to remove is the one that Brewster's law has just selected for the reflection.

Think Like a Physicist: Whenever a problem gives you a refractive index and asks for the polarising angle, write $\tan i_p = n$ and rearrange. The complement of i_p is automatically the refraction angle r , because Brewster's condition forces $i_p + r = 90^\circ$. Two parts of the problem can therefore be answered without Snell's law at all; Snell's law is then a check on the arithmetic, not an independent step.

Three sections of the chapter have now studied what the wave does. It interferes when two of itself meet; it diffracts when one of itself passes an obstacle; it polarises when the direction of its transverse vibration is filtered. The pages that follow take the same physics out of the laboratory and into the world. The thin films that colour a soap bubble and the anti-reflection coating on a camera lens are interference. The compact disc that throws a spectrum across the table when held under a lamp is diffraction. The screen of a calculator and the lens of a pair of sunglasses are polarisation. Each is, in its own way, an answer to a question of engineering that this chapter has now equipped the candidate to ask.

APPLICATIONS OF PHYSICAL OPTICS

So far, the chapter has built the theory of physical optics through definitions, derivations, and worked examples. This section now turns from the page to the world beyond it. It asks what these results actually do once they leave the classroom and enter everyday life. The purpose of the final section is therefore not to introduce new physics, but to reveal how remarkably far the physics already in hand can reach.

1. Holography

A hologram is a photographic record of the interference pattern between two coherent beams of light. The first beam is the object beam, scattered from the surface of the object that is being recorded; the second is the reference beam, taken straight from the source. Both come from a single laser, so their phase relationship is fixed at every point on the photographic plate. The plate records, at every pixel, the fringe contrast that the two beams produce locally. The result is not an image of the object but an interference pattern that encodes every wavefront the object would have sent toward the plate.

To replay the hologram, the developed plate is illuminated by a reproduction of the original reference beam. The plate now acts as a diffraction grating whose pitch varies from point to point exactly as the recorded fringes vary; the diffracted light reproduces the object's wavefront, and an observer on the far side sees a three-dimensional image floating in the position the original object occupied. The technique exploits **coherence** (to record clean fringes), **interference** (to make the recording), and **diffraction** (to read it out), and so brings together three sections of this chapter into one device. The same principle, scaled down to a few millimetres, makes the security holograms printed on bank notes and credit cards: they cannot be photocopied, because a photocopier captures an image rather than the wavefront that produced the image.

2. Anti-reflection coatings

A bare glass surface reflects about four per cent of the light incident on it. A modern camera zoom lens has fifteen or more such surfaces; without intervention the light reaching the sensor would be reduced to less than half of what entered the front element, with much of the lost light bouncing around inside the lens barrel and producing flare. The intervention is a thin coating of a transparent material, magnesium fluoride being the standard, whose refractive index 1.38 sits between the index of air and the index of crown glass at 1.52. A film of this material of thickness one quarter of the wavelength inside the film,

$$t = \frac{\lambda}{4n},$$

makes the two reflections, one from the top surface and one from the bottom, arrive in antiphase at every point above the lens. They cancel by destructive interference; the suppressed reflection is added to the transmitted beam, and the lens transmits ninety-five per cent of the light it receives instead of seventy. The same coating is used on spectacle lenses, on the cover glass of solar panels, and on the front surface of every modern smartphone screen, with the design wavelength chosen at the centre of the visible band so that the residual reflection is faint and purplish rather than at any particular spectral peak.

3. Glare-reducing polaroid sunglasses

Light reflected from a horizontal surface at large angles of incidence (a wet road, the bonnet of a car, the surface of a lake, a snowfield) is partially polarised, with its electric vector preferentially horizontal. At the Brewster angle the polarisation is complete; at angles either side of it the polarisation is substantial. A pair of Polaroid sunglasses whose transmission axes are oriented vertically absorbs the horizontal component and lets through only the vertical one. The reflected glare is dimmed by a factor of three or four; the rest of the scene, whose light is on average unpolarised, is dimmed only by the constant factor 1/2 that any single polaroid imposes on unpolarised light.

The behaviour of the lenses when tilted is the test by which a wearer can confirm that the glasses really are polaroid rather than tinted. Tilt the head sideways by ninety degrees, and the absorption axis is now vertical; the previously dim glare returns, and the rest of the scene is roughly unchanged. Fishermen and motorists use the effect routinely; landscape painters and photographers use the same effect, mounted on a camera lens instead of on a pair of spectacles, to darken the blue sky against the clouds.

4. Photoelastic stress analysis

Many transparent plastics, when bent or compressed, become temporarily birefringent. Their refractive index for light polarised parallel to the local stress direction differs from that for light polarised perpendicular to it. A stressed plastic part viewed between two crossed Polaroid filters develops a pattern of bright and dark

coloured fringes, in which the colour at each point is the wavelength for which the two polarisations have accumulated a relative phase difference of an odd integer number of half-wavelengths across the thickness of the part. The fringe pattern maps directly onto the local stress field.

Aircraft-window manufacturers, dam designers, and the makers of large plastic optical components used the technique to find stress concentrations in scale models of their parts long before the development of finite-element computer modelling. The technique survives today as the cheapest and most direct way to demonstrate to a student that stress in a plastic is a genuine physical quantity. A transparent plastic ruler bent gently between crossed Polaroid filters held up to the light shows the fringes immediately; the heaviest fringe concentration sits at the corners of any notches or holes in the ruler, exactly where an engineer would expect stress to concentrate.

5. Three-dimensional cinema and stereo viewers

A three-dimensional cinema projects two slightly different images of the same scene onto the screen at the same time, one for each eye. The two images are polarised in perpendicular linear planes, or, in the modern variant, in opposite circular polarisations. The viewer wears spectacles whose left and right lenses are Polaroid filters of perpendicular orientations (or circular polarisers of opposite handedness). Each lens passes only the image meant for its eye and blocks the other; the brain interprets the resulting pair of slightly different images as a single three-dimensional scene, exactly as it does with the two slightly different images delivered by the eyes of an ordinary observer.

The trick relies on the two perpendicular polarisations surviving the trip from projector to screen to filter. An ordinary white screen depolarises the light it scatters, because every microscopic facet of paint scatters in a random orientation; modern three-dimensional cinemas therefore use silvered screens, whose flat metallic surfaces preserve the polarisation of the light reflected from them. The same physics drives small hand-held stereo viewers that combine a left-eye and a right-eye image on a single transparency between two polarising filters.

6. X-ray diffraction

The grating equation continues to hold at much shorter wavelengths than the visible band, provided a much smaller slit spacing is available. The lattice planes of a crystal, with spacings of about 0.1nm, are such a grating, and the wavelengths that can be diffracted by them are X-rays, of comparable order of magnitude. The grating equation, written for reflection off adjacent lattice planes a distance d apart, becomes **Bragg's law**, $2d\sin\theta = n\lambda$, where θ is the glancing angle of the X-ray to the lattice planes and n is the order of the maximum. The diffraction pattern that comes off a crystal is therefore a fingerprint of the crystal's lattice spacings, and from a sufficiently complete pattern, in three dimensions, the positions of every atom in the unit cell can be deduced.

The technique determined the structure of common salt in 1913, the structure of penicillin in 1945, and, most famously of all, the double-helix structure of DNA. Photograph 51, taken by Rosalind Franklin at King's College London in May 1952, is an X-ray diffraction pattern of a fibre of DNA; the characteristic cross of streaks at the centre of the pattern is the diffraction signature of a helix, and the spacing of the streaks gave Crick and Watson the helix's pitch. X-ray crystallography is, with some justification, the most consequential single application of diffraction physics in the history of science.

7. Optical-fibre interference filters

Modern telecommunications send data along optical fibres by encoding several independent streams on slightly different wavelengths of infrared light and combining them onto the same fibre. At the receiving end of the fibre the streams have to be separated. The standard device for the separation is a stack of dielectric thin films deposited on a glass substrate; the films are designed so that, by interference between their many internal reflections, the stack transmits a narrow band of wavelengths and reflects everything else. A different filter, with a different stack design, is placed in the path for each independent channel.

The technique is called wavelength-division multiplexing. A single fibre crossing the Indian Ocean can carry a hundred or more independent channels in this way, each carrying tens of gigabits of data per second; without the technique, every channel would require its own fibre, and the cost of a transatlantic cable would be a hundred times what it is. The thin films at each end of the cable are designed using exactly the quarter-wave logic that earlier in this section produced the anti-reflection coating, scaled and stacked to make passbands rather than zero-reflection points.

8. Liquid-crystal displays

The dominant electronic display technology of the modern world is built around a sandwich of two crossed Polaroid films with a liquid-crystal layer between them. The liquid-crystal molecules, when no voltage is applied across the cell, twist the plane of polarisation by ninety degrees as the light passes through them. This rotation matches the angle between the two Polaroid axes, and the cell looks bright. When a voltage is applied across the cell, the molecules untwist; the plane of polarisation passes through unchanged, the second Polaroid blocks it, and the cell looks dark.

Every pixel of every laptop screen, smartphone display, calculator, and digital watch is one such cell with its voltage independently controllable. A colour display adds three coloured filters, red, green, and blue, behind the pixel array; varying the voltages independently in three sub-pixels per pixel produces a full-colour image. The whole technology is, in effect, a programmable curtain made of Polaroid, in which the rotating part of the curtain is a fluid whose molecules respond to electric fields. Every other physical-optics phenomenon studied in this chapter is also present somewhere in the device, but the polarisation rotation is the piece that does the central work.

Now that each section has been understood individually, it is time to see the ideas interact. The following miscellaneous worked examples combine concepts from across the chapter into single, integrated problems.

MISCELLANEOUS WORKED EXAMPLES ON PHYSICAL OPTICS

Example 25

- Explain in physical terms why a true double-slit interference pattern is obtained when the two slits are illuminated by light from a single source through a narrow slit, but not when the same two slits are illuminated by two physically distinct lamps of the same nominal colour.
- Light of coherence length 0.6mm illuminates a thin oil film of refractive index 1.50. Determine the maximum film thickness above which interference colours can no longer be observed in reflected light.

Solution

(a) A stable interference pattern requires that the two slit-illuminating sources stay in a fixed phase relationship over a time long compared with the eye's or detector's integration time. Light from a single primary source split into two paths by a pair of slits originates from the same emission events; the phase relationship between the two paths is therefore the constant zero (apart from path-length differences), and the pattern is stable. Two physically separate lamps emit independent sequences of wavetrains, each lasting only a few nanoseconds. The phase relationship between the two lamps shifts randomly on this nanosecond timescale, while the detector integrates over much longer times. The resulting pattern is averaged to a uniform illumination, as the lack of coherence between the two lamps washes the fringes out.

(b) For interference to be visible from a thin film in reflected light, the optical path difference between the front-reflected and back-reflected beams must lie within the coherence length of the source. The path difference for normal incidence is $2nt$, where n is the refractive index of the film and t its thickness. The maximum thickness for visible fringes is therefore set by:

$$2nt_{\max} = L_c$$

$$t_{\max} = \frac{L_c}{2n} = \frac{0.6\text{mm}}{2(1.50)} = 0.2\text{mm}$$

Above 0.2mm the front- and back-reflected beams are too far separated in time for interference colours to be observed.

Example 26

- Account in physical terms for the experimental observation that the central fringe of a Young's double-slit pattern remains bright when the entire apparatus is immersed in water, while every other fringe moves closer to the centre.
- In a Young's double-slit experiment, light of wavelength 600nm illuminates two slits 0.40mm apart, and fringes are observed on a screen 1.50m away. Find the fringe width on the screen, and the distance between the third bright fringe and the third dark fringe on the same side of the central fringe.

Solution

(a) The central fringe corresponds to zero path difference between light reaching the screen from the two slits. Submerging the apparatus in water does not alter the geometry of the two paths; both paths run through the same medium and remain equal in length, so the central fringe stays at the same position on the screen and remains bright.

Higher-order fringes occur at path differences equal to integer multiples of the wavelength. The wavelength of light inside water is smaller than the in-air value by a factor equal to the refractive index of water. The fringe spacing $\beta = \lambda D/d$ shrinks by the same factor. All non-central fringes therefore move closer to the centre, with the central fringe anchored as the only fixed reference.

(b) Fringe width:

$$\beta = \frac{\lambda D}{d} = \frac{(600 \times 10^{-9} \text{m})(1.50 \text{m})}{0.40 \times 10^{-3} \text{m}} = 2.25 \times 10^{-3} \text{m} = 2.25 \text{mm}$$

Position of the third bright fringe: $y_{3,\text{br}} = 3\beta = 6.75 \text{mm}$.

Position of the third dark fringe (third dark counted from the central bright, at half-integer fringe positions): $y_{3,\text{dk}} = (3 - 1/2)\beta = 2.5\beta = 5.625 \text{mm}$.

Distance between them:

$$\Delta y = y_{3,\text{br}} - y_{3,\text{dk}} = 6.75 \text{mm} - 5.625 \text{mm} = 1.125 \text{mm} = \frac{\beta}{2}$$

Example 27

- (a) Explain in physical terms why, when a Young's double-slit experiment is performed with white light, the central fringe is white but higher-order fringes show coloured fringing.
- (b) Two wavelengths, 650nm and 520nm, simultaneously illuminate a Young's double-slit whose slit separation is 0.50mm and screen distance is 1.20m. Find the distance from the central fringe to the first position at which a bright fringe of each wavelength coincides.

Solution

(a) The central fringe is at zero path difference. At zero path difference every wavelength interferes constructively, so all the wavelengths in white light arrive at the central fringe in phase and add. The combined fringe is therefore bright and white.

Away from the centre, the constructive condition becomes path difference equal to an integer number of wavelengths. The wavelength entering this condition is different for each colour, so the first bright fringe of red light is farther from the centre than that of violet light. Consecutive first-order fringes from different wavelengths therefore fan out into a band of colour, with violet inside and red outside the central white fringe. Higher orders overlap and lose distinct colour identity.

(b) Coincidence of bright fringes requires $n_1\lambda_1 = n_2\lambda_2$ for integers n_1 and n_2 .

$$\frac{n_1}{n_2} = \frac{\lambda_2}{\lambda_1} = \frac{520}{650} = \frac{4}{5}$$

The smallest such integers are $n_1 = 4$, $n_2 = 5$ (and $4 \times 650 \text{nm} = 5 \times 520 \text{nm} = 2600 \text{nm}$).

Position on the screen:

$$y = \frac{n_1\lambda_1 D}{d} = \frac{4(650 \times 10^{-9} \text{m})(1.20 \text{m})}{0.50 \times 10^{-3} \text{m}} = 6.24 \times 10^{-3} \text{m} = 6.24 \text{mm}.$$

Example 28

- (a) Distinguish between the appearance of the pattern observed on the screen in a Young's experiment when one of the two slits is covered with an opaque card and when both slits are open.
- (b) In a Young's double-slit experiment, the slit separation is 0.30mm and the screen is at a distance of 1.20m from the slits. The wavelength of the illuminating laser is changed from 632.8nm (red) to 532nm (green). Find the change in fringe width.

Solution

(a) When one slit is covered, only the uncovered slit acts as a source. The screen shows a single-slit diffraction pattern: a wide central bright maximum, flanked by progressively weaker secondary maxima separated by dark minima. No interference fringes are present, because there is no second source whose light can superpose with that of the first.

When both slits are open, light from each slit reaches every point on the screen and superposes. The screen now shows the Young's interference fringes of equal spacing $\beta = \lambda D/d$, modulated by the single-slit diffraction envelope from each slit. The interference fringes are evenly spaced, the diffraction envelope is broad and bell-shaped, and the visible pattern is a series of evenly spaced bright bars whose intensity decreases away from the centre according to the diffraction envelope.

(b) Fringe width depends on wavelength as $\beta = \lambda D/d$, so the change in fringe width when the wavelength changes from λ_1 to λ_2 is:

$$\Delta\beta = \frac{D(\lambda_1 - \lambda_2)}{d} = \frac{(1.20\text{m})(632.8 \times 10^{-9}\text{m} - 532 \times 10^{-9}\text{m})}{0.30 \times 10^{-3}\text{m}} = 4.03 \times 10^{-4}\text{m} \approx 0.40\text{mm}.$$

The fringe width decreases by approximately 0.40mm when the laser changes from red to green.

Example 29

- (a) Explain in physical terms why placing a thin transparent sheet over one of the two slits in a Young's double-slit experiment shifts the entire fringe pattern toward the side covered by the sheet.
- (b) In a Young's double-slit experiment with slit separation 0.40mm and screen distance 1.50m, illuminated by light of wavelength 550nm, a thin mica sheet of refractive index 1.58 is placed over one of the slits. The central fringe shifts on the screen by 0.20cm. Determine the thickness of the mica sheet.

Solution

(a) Light passing through a thin transparent sheet of refractive index μ and thickness t covers an optical path length μt inside the sheet, whereas in the absence of the sheet the geometric path length over the same distance would be t . The light traversing the covered slit therefore accumulates an extra optical path of $(\mu - 1)t$ compared with the light through the uncovered slit. To re-establish zero net path difference between the two beams (which defines the central bright fringe), the geometric path on the covered side must shorten by the same amount. The central fringe therefore moves toward the covered side, and the entire pattern shifts with it.

(b) Fringe shift formula:

$$\Delta y = \frac{D(\mu - 1)t}{d}$$

Solving for t :

$$t = \frac{\Delta y \times d}{D(\mu - 1)} = \frac{(0.20 \times 10^{-2}\text{m})(0.40 \times 10^{-3}\text{m})}{(1.50\text{m})(1.58 - 1)} = 9.2 \times 10^{-7}\text{m} = 0.92 \mu\text{m}$$

Example 30

- (a) Account in physical terms for the changing pattern of colours that moves across the surface of a soap film stretched in a wire ring as the film drains under gravity.
- (b) A soap film of refractive index 1.33 is illuminated normally by white light. Find the smallest film thickness that produces a bright reflected fringe at a wavelength of 500nm.

Solution

(a) A soap film stretched in a wire ring drains under gravity, becoming thicker at the bottom and thinner at the top. The thickness varies smoothly and continuously with height. Light incident on the film reflects from both the front surface and the back surface; the back-reflected beam travels an extra optical path of $2nt$ inside the film (with an additional half-wavelength shift due to the π phase change on reflection from the front surface). The condition for bright reflection therefore selects, at each thickness, a particular wavelength satisfying $2nt = (m - 1/2)\lambda$. Because t varies along the height of the film, the wavelength satisfying the bright condition also varies smoothly, producing the bands of colour that flow visibly down the film as it drains.

(b) For bright reflection from a thin film of refractive index n and thickness t at wavelength λ , with the half-wavelength shift on reflection accounted for:

$$2nt = (m + 1/2)\lambda, \quad m = 0, 1, 2, \dots$$

Smallest thickness corresponds to $m = 0$:

$$t_{\min} = \frac{\lambda}{4n} = \frac{500 \times 10^{-9} \text{m}}{4(1.33)} = 9.40 \times 10^{-8} \text{m} = 94.0 \text{nm}$$

Example 31

- (a) A student wishes to coat a glass lens of refractive index 1.50 with magnesium fluoride of refractive index 1.38 so that the intensity of light reflected at a wavelength of 550nm is reduced. Account in physical terms for why magnesium fluoride works as an anti-reflection coating on glass.
- (b) A piece of thin copper wire is laid between two flat glass plates at one end to form a small air wedge. When the wedge is illuminated normally with sodium light of wavelength 589nm, a total of 102 dark fringes are counted across the entire length of the wedge, from the closed end to the open end where the wire sits. Calculate the diameter of the wire.

Solution

(a) Light incident on the coated lens reflects partially at the air–coating interface and partially at the coating–glass interface. Both reflections are from a lower-index medium into a higher-index medium, so both undergo a π phase change on reflection. The two π shifts cancel, leaving the relative phase of the two reflected beams to be set entirely by the optical path difference $2nt$. Choosing the coating thickness so that $2nt$ equals an odd half-wavelength produces destructive interference between the two reflections, and the net reflected intensity is minimised. The refractive index of magnesium fluoride (1.38) lies between that of air (1.00) and that of glass (1.50), so the two interface reflections have comparable amplitude, which is a requirement for nearly complete cancellation.

(b) The wire's diameter equals the thickness of the air wedge at the open end. The total number N of dark fringes spanning the wedge from $t = 0$ at the closed end to $t = h$ at the open end is given by:

$$N = 2h / \lambda$$

Solving for h :

$$h = N\lambda / 2$$

Substituting $N = 102$ and $\lambda = 589\text{nm}$:

$$h = (102 \times 589 \times 10^{-9} \text{m}) / 2 = 3.00 \times 10^{-5} \text{m} = 30.0\mu\text{m} = 0.030\text{mm}$$

The diameter of the copper wire is therefore 0.030mm, or 30 μm .

Example 32

- (a) Account for the experimental fact that the centre of the Newton's-rings pattern observed in reflected light is dark, not bright.
- (b) In a Newton's-rings experiment, the diameter of the eighth dark ring (measured from the centre) is 4.80mm and the radius of curvature of the lens is 1.20m. Calculate the wavelength of the light used.

Solution

(a) At the centre of the Newton's-rings arrangement the lens makes contact with the flat glass plate, and the air gap between them is essentially zero. Light reflecting from the lower surface of the lens (going from glass into air) undergoes no phase change, because it is reflecting from a less-dense medium. Light reflecting from the upper surface of the glass plate (going from air into glass) undergoes a π phase change, because it is reflecting from a denser medium. Even though the geometric path difference between the two reflections is zero at the centre, the phase difference is π . The two beams therefore interfere destructively, and the centre is dark.

(b) For dark rings in Newton's rings (reflected light), the diameter of the n th dark ring satisfies

$$D_n^2 = 4nR\lambda,$$

where R is the radius of curvature of the lens. Solving for λ :

$$\lambda = \frac{D_n^2}{4nR} = \frac{(4.80 \times 10^{-3} \text{m})^2}{4(8)(1.20 \text{m})} = 6.00 \times 10^{-7} \text{m} = 600 \text{nm}$$

Example 33

- (a) Account in physical terms for the fact that, in a Newton's-rings experiment, a plot of the squared diameter of successive dark rings against the ring number gives a straight line through the origin.
- (b) In a Newton's-rings experiment, the squared diameter of the twelfth dark ring exceeds that of the sixth dark ring by 12.0mm^2 . The radius of curvature of the lens is 0.80m . Calculate the wavelength of the light used.

Solution

(a) The condition for the n th dark ring in reflected light is that the air-gap thickness t at that radius gives a path difference satisfying $2t = n\lambda$. The geometry of a spherical lens resting on a flat plate gives, for small t , $t \approx r^2/(2R)$, where r is the radius of the ring and R is the radius of curvature of the lens. Substituting:

$$\frac{r^2}{R} = n\lambda \quad \Rightarrow \quad D_n^2 = 4r^2 = 4nR\lambda$$

This is a linear relation between D_n^2 and n , with slope $4R\lambda$ and zero intercept. A plot of D_n^2 versus n is therefore a straight line through the origin (and either of R and λ may be determined graphically).

(b) From $D_n^2 = 4nR\lambda$, the difference of squared diameters of two rings at orders $n_2 > n_1$ is:

$$D_{n_2}^2 - D_{n_1}^2 = 4R\lambda(n_2 - n_1).$$

For $n_2 = 12$, $n_1 = 6$:

$$\lambda = \frac{D_{12}^2 - D_6^2}{4R(12 - 6)} = \frac{12.0 \times 10^{-6}\text{m}^2}{24(0.80\text{m})} = 6.25 \times 10^{-7}\text{m} = 625\text{nm}$$

Example 34

- (a) Distinguish between the diffraction pattern produced by a single slit and the interference pattern produced by a Young's double-slit, with reference to the spacing and intensity of consecutive bright fringes.
- (b) A monochromatic beam of wavelength 600nm passes through a single slit of width 0.10mm and forms a pattern on a screen 2.0m away. Find the width of the central maximum, and the angular position of the second-order minimum.

Solution

(a) The single-slit diffraction pattern consists of a wide central bright maximum flanked by progressively narrower secondary maxima separated by dark minima. The central maximum is roughly twice the width of the secondary maxima, and the intensity of the secondary maxima falls off rapidly (the first secondary maximum is only about 4.5% of the central maximum).

The Young's double-slit interference pattern consists of equally spaced bright fringes of nearly equal intensity, modulated by the single-slit diffraction envelope of each individual slit. Successive bright fringes are evenly spaced (in contrast to the single-slit pattern's narrowing secondary maxima), and the intensity of consecutive bright fringes is nearly equal across the central region of the envelope.

(b) The central maximum extends between the first dark minima on either side. The first minimum lies at $a\sin\theta = \lambda$, so $\sin\theta \approx \lambda/a$ and the half-width on the screen is $y_1 \approx D\lambda/a$. The full width is:

$$W = 2y_1 = \frac{2D\lambda}{a} = \frac{2(2.0\text{m})(600 \times 10^{-9}\text{m})}{0.10 \times 10^{-3}\text{m}} = 2.4 \times 10^{-2}\text{m} = 24\text{mm}$$

Angular position of the second-order minimum:

$$\sin\theta_2 = \frac{2\lambda}{a} = \frac{2(600 \times 10^{-9})}{0.10 \times 10^{-3}} = 1.2 \times 10^{-2} \quad \Rightarrow \quad \theta_2 = \sin^{-1}(1.2 \times 10^{-2}) = 0.69^\circ.$$

Example 35

- (a) Account for the experimental observation that, no matter how much the line density of a diffraction grating is increased, only a finite number of principal maxima can ever be observed for a given wavelength of light.
- (b) A diffraction grating has 500 lines per millimetre. Light of wavelength 580nm is incident normally on the grating. Find all the diffraction orders that can be observed, and the diffraction angle of the highest visible order.

Solution

(a) The grating equation $d \sin \theta = n\lambda$ requires $\sin \theta = n\lambda/d$ for the n th principal maximum, where d is the grating spacing. The function $\sin \theta$ cannot exceed unity, so the order number is bounded by

$$n \leq \frac{d}{\lambda}.$$

A grating with a higher line density has a smaller d , and a smaller d means a smaller value of d/λ and therefore a smaller maximum order. Increasing the line density indefinitely makes d approach λ from above; below that, no orders beyond the central maximum exist at all. For any given wavelength, the number of observable principal maxima is therefore finite, and the upper bound is set by the ratio d/λ .

(b) Grating spacing:

$$d = \frac{1}{500 \times 10^3 \text{ m}^{-1}} = 2.0 \times 10^{-6} \text{ m}.$$

Maximum order:

$$n_{\text{max}} = \left\lfloor \frac{d}{\lambda} \right\rfloor = \left\lfloor \frac{2.0 \times 10^{-6}}{5.80 \times 10^{-7}} \right\rfloor = [3.45] = 3.$$

Observable orders: $n = 0, \pm 1, \pm 2, \pm 3$. The highest visible order is the third.

Diffraction angle of the third order:

$$\sin \theta_3 = \frac{3\lambda}{d} = \frac{3(5.80 \times 10^{-7})}{2.0 \times 10^{-6}} = 0.870 \quad \Rightarrow \quad \theta_3 = \sin^{-1}(0.870) = 60.5^\circ.$$

Example 36

- (a) Explain in physical terms why a grating made of equally spaced opaque wires separated by transparent gaps produces a diffraction pattern whose principal maxima lie at the same angles as those produced by a grating made of equally spaced transparent rulings on glass.
- (b) Infrared light of wavelength $1.0 \times 10^{-5} \text{ m}$ is incident normally on a grating made of fine wires of diameter $1.0 \times 10^{-5} \text{ m}$ with transparent gaps of width $1.5 \times 10^{-5} \text{ m}$ between consecutive wires. Find the angle of the first-order maximum.

Solution

(a) The grating equation $d \sin \theta = n\lambda$ relates the diffraction angles of the principal maxima to the period d of the grating, defined as the distance from one feature to the next identical feature. Whether each period consists of one opaque wire and one transparent gap, or one transparent ruling and one opaque gap, the period is the same: the centre-to-centre spacing of the wires equals the centre-to-centre spacing of the corresponding transparent rulings on the complementary grating. The principal maxima therefore occur at the same angles in the two cases. The two gratings differ in the relative intensities of the orders (because the diffraction envelope from a single opaque wire is not the same as from a single transparent ruling), but not in the angular positions of the principal maxima.

(b) Grating period:

$$d = d_{\text{wire}} + d_{\text{gap}} = 1.0 \times 10^{-5} \text{ m} + 1.5 \times 10^{-5} \text{ m} = 2.5 \times 10^{-5} \text{ m}$$

First-order angle:

$$\sin \theta_1 = \frac{\lambda}{d} = \frac{1.0 \times 10^{-5}}{2.5 \times 10^{-5}} = 0.40 \quad \Rightarrow \quad \theta_1 = 23.6^\circ$$

Example 37

- (a) Account in physical terms for the fact that sound waves in air cannot be polarised, while light waves can.
- (b) Unpolarised light of intensity 16W/m^2 passes through a Polaroid sheet, then through a second Polaroid sheet whose transmission axis is at 30° to that of the first. Find the intensity of the light emerging from the second sheet.

Solution

(a) Polarisation is the restriction of a wave's vibration to a specified plane perpendicular to the direction of propagation. Sound waves in air are longitudinal: the air molecules oscillate along the direction of propagation, so there is no plane transverse to propagation in which the vibration could be restricted. Polarisation is meaningless for a longitudinal wave. Light waves are transverse: the electric (and magnetic) field oscillates in a plane perpendicular to the direction of propagation, and that plane has two independent directions of oscillation. Restricting the electric vector to a single direction in that plane (by a Polaroid filter, by Brewster reflection, by selective scattering, and so on) defines a polarised state. The contrast between sound and light is a direct consequence of their longitudinal versus transverse character.

- (b) Unpolarised light passing through the first Polaroid:

$$I_1 = \frac{I_0}{2} = \frac{16\text{W/m}^2}{2} = 8\text{W/m}^2$$

The transmitted light is linearly polarised along the first sheet's axis. Through the second Polaroid (Malus's law):

$$I_2 = I_1 \cos^2 \theta = (8\text{W/m}^2) \cos^2(30^\circ) = 6\text{W/m}^2$$

Example 38

- (a) Account in physical terms for the observation that inserting a third Polaroid sheet between two crossed Polaroids causes light to be transmitted through the combination, even though no light passes through the two crossed Polaroids alone.
- (b) Unpolarised light of intensity 24W/m^2 is incident on three Polaroid sheets stacked in series. The first and third are crossed (their transmission axes are at 90° to each other); the second is placed midway between them with its transmission axis at 45° to each of the others. Find the intensity of the light emerging from the third sheet.

Solution

(a) Two crossed Polaroids transmit no light because the second sheet is oriented to block exactly the polarisation that the first sheet transmits. When a third Polaroid is inserted between them at, say, 45° to the axis of the first, the situation changes. After the first Polaroid, the light is polarised along axis-1. The third (middle) Polaroid, at 45° to axis-1, transmits the component along its own axis (the cosine-squared factor at 45° , i.e., one half), and re-polarises the light along axis-3. Light now reaches the last Polaroid polarised along axis-3, which is at 45° to the last polaroid's axis. The last Polaroid transmits another factor of one half. Net: the previously blocked light now reaches the screen because the inserted Polaroid has rotated the polarisation in steps of 45° rather than blocking it outright.

- (b) After the first Polaroid:

$$I_1 = \frac{I_0}{2} = \frac{24\text{W/m}^2}{2} = 12\text{W/m}^2$$

After the middle Polaroid (at 45° to the first):

$$I_2 = I_1 \cos^2(45^\circ) = (12\text{W/m}^2)(0.5) = 6\text{W/m}^2$$

After the third Polaroid (at 90° to the first, i.e., 45° to the middle):

$$I_3 = I_2 \cos^2(45^\circ) = (6\text{W/m}^2)(0.5) = 3\text{W/m}^2$$

Example 39

- (a) Account in physical terms for the experimental observation that Polaroid sunglasses with a vertical transmission axis reduce the glare of sunlight reflected from a wet road but not the glare of sunlight reflected from a polished mirror at normal incidence.

- (b) Sunlight reflects from the surface of a calm lake of refractive index 1.33 at the Brewster angle. Find the angle of the reflected ray above the horizontal, the angle of refraction in the lake, and the orientation in space of the electric vector of the reflected light.

Solution

(a) Light reflected from a horizontal surface at a large angle of incidence is partially polarised in the horizontal direction (the component of the electric vector parallel to the surface is preferentially reflected). Polaroid sunglasses with a vertical transmission axis block most of this horizontally polarised glare from the wet road, sharply reducing the perceived brightness of the reflection.

Light reflected from a polished mirror at normal incidence is, however, unpolarised. At normal incidence, the two orthogonal components of the electric field are reflected with equal intensity, and the reflected light shows no preferred direction of vibration. Polaroid sunglasses transmit only one component (half the intensity) regardless of orientation, so they do not selectively suppress the mirror reflection; the sunglasses simply dim the reflection by half.

- (b) Brewster angle:

$$\tan i_p = n_{\text{water}} = 1.33 \quad \Rightarrow \quad i_p = \tan^{-1}(1.33) = 53.06^\circ$$

Angle of reflection above the horizontal: $90^\circ - i_p = 36.9^\circ$

At the Brewster angle, the refracted ray is exactly perpendicular to the reflected ray, so:

$$r = 90^\circ - i_p = 36.9^\circ$$

The electric vector of the reflected light is perpendicular to the plane of incidence, i.e., parallel to the surface of the lake (horizontal and along the line of the shore).

Example 40

- (a) A clear plastic ruler placed between two crossed Polaroid sheets shows a delicate, multi-coloured pattern when bent slightly. Account in physical terms for this observation.
- (b) X-rays of unknown wavelength are reflected by a crystal of sodium chloride whose interplanar spacing is 0.282nm. The first-order Bragg reflection is observed at a glancing angle of 7.5° . Find the wavelength of the X-rays.

Solution

(a) Most transparent solids are optically isotropic, so light propagates through them at a single refractive index regardless of polarisation. When such a solid is mechanically stressed, the strain field breaks the isotropy and makes the material birefringent: the refractive index for light polarised along the stress direction is now slightly different from that for light polarised at right angles. Plane-polarised light passing through such a stressed region therefore emerges elliptically polarised, with the ellipse's shape and orientation depending on the local stress. When this elliptical light is viewed through the analyser (the second Polaroid, crossed with the first), the transmitted intensity depends on the elliptical state, and therefore on the local stress. White-light illumination produces a different colour for each combination of strain magnitude and direction, mapping the stress field across the ruler into a coloured pattern of fringes. This is the principle of photoelastic stress analysis.

- (b) Bragg's law for the n th-order reflection:

$$2d \sin \theta = n\lambda$$

For $n = 1$, $d = 0.282\text{nm}$, $\theta = 7.5^\circ$:

$$\lambda = 2d \sin \theta = 2(0.282 \times 10^{-9}\text{m}) \sin 7.5^\circ = 7.36 \times 10^{-11}\text{m}$$

The wavelength of the X-rays is approximately $7.36 \times 10^{-11}\text{m} = 0.0736\text{nm}$.

Example 41

- (a) Account in physical terms for the experimental observation that a hologram of an object recorded with red laser light reproduces a sharp three-dimensional virtual image only when the hologram is illuminated by light of the same wavelength as that used to record it.
- (b) A laser used to make a hologram has a wavelength of 632.8nm in air. The hologram is then submerged in water of refractive index 1.33. Find the wavelength of the laser light inside the water.

Solution

(a) A hologram is a finely detailed interference pattern recorded in a photosensitive emulsion. The pattern is the result of interference between a reference beam and the light scattered from the object during recording. The recorded fringe structure encodes both the amplitude and the phase of the object wavefronts. When the hologram is later illuminated with the original reference beam (same wavelength), the recorded fringe pattern diffracts the reference beam in the geometric pattern of a grating, reconstructing the original object wavefronts and producing a sharp three-dimensional virtual image.

Diffraction from any periodic structure follows $d \sin \theta = n\lambda$, in which the angles depend on λ . Illuminating the same hologram with a different wavelength changes all the diffraction angles, so the reconstructed wavefronts no longer match the original object geometry. The image is correspondingly distorted or blurred. Only the original wavelength produces a sharp reconstruction.

(b) The wavelength of light inside a medium of refractive index n is:

$$\lambda_{\text{water}} = \frac{\lambda_{\text{air}}}{n_{\text{water}}} = \frac{632.8\text{nm}}{1.33} = 475.8\text{nm}$$

Example 42

- (a) Explain in physical terms why a beam of plane-polarised light, after passing through a column of sugar solution, has its plane of polarisation rotated.
- (b) A column of sugar solution of length 10cm and concentration 0.20g/cm^3 rotates the plane of polarisation of yellow light by 13° . Calculate the specific rotation of the sugar at this wavelength.

Solution

(a) Sugar molecules are chiral: their three-dimensional structure has a definite handedness (sugar is right-handed). A plane-polarised wave entering the solution can be decomposed into two circularly polarised components of equal amplitude, namely one right-circular and one left-circular. The chiral molecules interact differently with the two circular components, giving them slightly different refractive indices and therefore slightly different propagation speeds inside the solution. After traversing the column, the two circular components are no longer at the same phase. When they recombine, the resultant is again plane-polarised, but the plane of polarisation has been rotated through an angle proportional to the path length and to the concentration of the chiral molecules.

(b) Specific rotation of a chiral solution is defined as the rotation per unit path length per unit concentration:

$$\alpha_0 = \frac{\alpha}{Lc} \quad (\text{from } \alpha = \alpha_0 Lc)$$

For convenience, L is conventionally expressed in decimetres ($1\text{dm} = 10\text{cm}$) and c in grams per cubic centimetre:

$$\alpha_0 = \frac{13^\circ}{(1.0\text{dm})(0.20\text{g/cm}^3)} = 65^\circ \text{ dm}^{-1} (\text{g/cm}^3)^{-1}.$$

With these miscellaneous worked examples now complete, the threads of Physical Optics: wavefronts, coherence, interference, diffraction, and polarisation, should feel woven into a single picture. It is time to sharpen our thinking. Welcome to the Digging Deeper Exercise!

DIGGING DEEPER EXERCISE 8

EXERCISE 8A: BINDER QUESTIONS

Question 1

Two ordinary tungsten-filament lamps placed close together fail to produce a stable interference pattern on a nearby screen, no matter how carefully the geometry is set up. Account in physical terms for this observation.

Question 2

Account for the experimental observation that the spacing between consecutive dark fringes formed by an air wedge is the same whether the wedge is viewed close to its apex or far from it.

Question 3

A compact disc held under a fluorescent lamp shows iridescent colours which shift continuously as the disc is tilted relative to the lamp. Account in physical terms for the shifting of the colours.

Question 4

A student claims that, in a Young's double-slit experiment, the dark fringes are arbitrary regions of low intensity lying at no particular position relative to the bright fringes. Indicate, with reasons, whether the dark fringes occur at any specific positions relative to the bright fringes.

Question 5

Kipanga and Kipute argue about a Newton's-rings experiment. Kipanga claims that the centre of the ring pattern viewed in transmitted light is dark, just as it is in reflected light. Kipute claims that the centre viewed in transmitted light is bright. Resolve the disagreement.

Question 6

A student claims that diffraction and interference are two completely different physical phenomena, with diffraction occurring at slits and interference occurring between separate wave trains. Indicate, with reasons, whether the student's distinction is physically meaningful.

Question 7

A student watching a single-slit diffraction pattern claims that the central maximum is twice as wide as each of the secondary maxima, while a colleague claims it is the same width. Resolve the disagreement, giving the physical reason for the correct answer.

Question 8

Account in physical terms for the experimental fact that, when unpolarised light passes through a single Polaroid sheet, exactly half of the incident intensity is transmitted, regardless of the orientation of the Polaroid's transmission axis.

Question 9

Kipanga claims that Brewster's angle for a transparent medium is the same for every colour of light, since it depends only on the refractive index of the medium. Identify the error in this claim and account for the small dependence of Brewster's angle on wavelength.

Question 10

Show that the dark fringes in a Young's double-slit pattern occur exactly midway between consecutive bright fringes.

Question 11

A student claims that immersing a Young's double-slit apparatus in water should shift the entire interference pattern toward one side of the slits, because the wavelength has changed. Identify the error in this claim and account in physical terms for what actually happens to the pattern.

Question 12

Account in physical terms for the experimental observation that the radii of the dark rings in a Newton's-rings pattern increase as the square root of the ring number, rather than in direct proportion to the ring number.

Question 13

A thin soap film viewed in white light shows bright bands of colour, but a glass window-pane several millimetres thick viewed in the same white light shows no such colours at all. Account in physical terms for the difference.

EXERCISE 8B: REAL QUESTIONS**Question 14**

In a Newton's-rings demonstration at a science fair in Morogoro, the plano-convex lens has been silvered on its lower surface, so that all the incident light is reflected at the lens itself and none reaches the flat glass plate below. Account in physical terms for the change, if any, that the silvering produces at the centre of the ring pattern observed in reflected light.

Question 15

A student wearing polarised sunglasses with a vertical transmission axis observes that the screen of her smartphone, viewed in landscape orientation under bright daylight, appears bright and clear when the phone is held one way, but turns almost completely black when the phone is rotated through ninety degrees about the viewing axis. Account in physical terms for this observation.

Question 16

A child blowing soap bubbles in a courtyard in Mtwara observes that each bubble shows a shifting pattern of rainbow colour bands that flow downward across the bubble's surface before the bubble bursts. Account in physical terms for the downward flow of the colour bands.

Question 17

A motorcyclist parking near a workshop in Mwanza notices, after a brief shower of rain, that a thin layer of petrol on a puddle shows brilliant colours that change as he tilts his head and changes his viewing angle. Account in physical terms for the change of colour with viewing angle.

Question 18

A fisherman on the northern shore of Lake Tanganyika wears polarised sunglasses with a vertical transmission axis to reduce the glare of midday sunlight reflected from the surface of the lake. Account in physical terms for the orientation of his line of sight relative to the lake surface at which the sunglasses are most effective.

Question 19

A driver travelling along the Morogoro–Iringa highway at midday observes a shimmering pool of “water” stretching across the tarmac far ahead of the vehicle, which appears to recede as he approaches. Account in physical terms for this observation.

Question 20

A spectacle-maker in Arusha coats a glass lens with a thin transparent layer of magnesium fluoride to reduce the intensity of light reflected at a wavelength of 550nm. Account in physical terms for the reduction in reflected intensity at that wavelength.

Question 21

A photographer in Iringa observes the clear midday sky through a polarising filter rotated in front of her camera lens. She notices that the sky directly overhead darkens visibly when the filter is rotated to one particular orientation, while the sky near the sun barely changes its brightness. Account in physical terms for the orientation-dependent darkening of the overhead sky.

Question 22

A holographic image embedded in the security label of an identification card has been recorded with red laser light. In a shop in Dar es Salaam, the image appears sharp and clearly three-dimensional under a red LED inspection lamp, but appears blurred and washed-out under the yellow sodium street-lamps outside. Account in physical terms for the difference.

Question 23

A student in Tanga places a clear plastic protractor between two crossed Polaroid sheets and observes a faint multicoloured pattern. When she squeezes the protractor gently between her fingers, the pattern brightens and becomes vivid; when she releases the squeeze, the pattern fades back to its faint state. Account in physical terms for the brightening of the pattern under squeezing.

Question 24

A trader at Kariakoo market in Dar es Salaam holds up a compact disc to a customer to demonstrate the spectrum of colours that appears on its surface in bright sunlight. The trader claims that the CD is “*splitting white light like a prism*”. Account in physical terms for the distinction between the colour-spreading mechanism of the CD and that of a prism.

Question 25

A radiology technician at the Muhimbili National Hospital in Dar es Salaam is calibrating an ultrasound scanner. He notes that a 5MHz pulse maintains its frequency when crossing from a layer of subcutaneous fat into the underlying muscle, but that the wavelength inside the muscle is longer than inside the fat. Account in physical terms for this observation.

EXERCISE 8C: HOT QUESTIONS**Question 26**

Light of wavelength 600nm in vacuum is incident at 50.0° on the surface of a glass slab of refractive index 1.50. Calculate the angle of refraction inside the glass, the wavelength of the light inside the glass, and the speed of the light inside the glass.

Question 27

A sodium lamp emits two close spectral lines whose mean wavelength is 589nm and whose frequency spread is $5.0 \times 10^9 \text{ Hz}$. Calculate the coherence time and the coherence length of the emitted light.

Question 28

In a Young’s double-slit experiment, light of wavelength 550nm illuminates two slits separated by 0.25mm, and the fringes are observed on a screen 1.20m from the slits. Find the fringe width on the screen and the angular separation of consecutive bright fringes as seen from the slits.

Question 29

In a Young’s double-slit experiment, light of wavelength 620nm produces fringes of spacing 1.55mm on a screen 1.50m from the slits. Determine the slit separation, and find the new fringe width if the slit separation is doubled while the wavelength and the screen distance are unchanged.

Question 30

In a Young’s double-slit experiment with slit separation 0.45mm, screen distance 1.20m, and illumination of wavelength 589nm, a thin sheet of glass of refractive index 1.52 is placed in front of one of the slits and the central fringe shifts by 4.50mm from its original position. Calculate the thickness of the glass sheet.

Question 31

An air wedge is formed between two glass plates separated at one end by a strip of paper of thickness 0.040mm and length 8.0cm. Light of wavelength 589nm is incident normally on the wedge. Find the spacing between consecutive dark fringes, and the total number of dark fringes observed along the wedge.

Question 32

A thin film of refractive index 1.40 is illuminated normally by white light. Find the smallest film thickness that gives a bright reflected fringe at a wavelength of 480nm, and the next-smallest film thickness that gives a bright reflected fringe at the same wavelength.

Question 33

In a Newton's-rings experiment, light of wavelength 600nm is used and the plano-convex lens has a radius of curvature of 0.90m. Calculate the radius of the tenth dark ring measured from the centre, and find the radius of curvature of a lens that would give the same tenth-ring radius if the wavelength were changed to 500nm.

Question 34

In a Newton's-rings experiment, the diameter of the tenth dark ring is 5.00mm when there is air between the lens and the flat plate. When the gap is filled with a transparent liquid, the diameter of the same tenth ring shrinks to 4.30mm. Calculate the refractive index of the liquid.

Question 35

A monochromatic beam of wavelength 589nm passes through a single slit and forms a diffraction pattern on a screen 2.50m from the slit. The full width of the central maximum on the screen is 1.20cm. Find the width of the slit and the angular position of the third-order dark fringe.

Question 36

A compact disc has reflective tracks spaced $1.6\text{ }\mu\text{m}$ apart and is illuminated normally by white light. Find the diffraction angle of the first-order principal maximum for red light of wavelength 700nm and for blue light of wavelength 450nm.

Chapter 9

SOUND, ELECTROMAGNETIC WAVES, AND THE DOPPLER EFFECT**INTRODUCTION**

Jangwani Secondary School in Dar es Salaam, mid-morning on a Wednesday. The Inter-School Physics Club is in session. Mr. Akilikubwa has brought Kipute and Kipanga from Miono Secondary School on a three-day study tour, and the Jangwani host teacher has loaned them a Form Five classroom that faces United Nations Road, the busy corridor between Kariakoo and Muhimbili National Hospital. The windows are open because the fan is broken and the morning is already 32 degrees. Outside, traffic crawls. Inside, the room is quiet.

Monalisa, a Form Five girl from Jangwani who lives in Upanga and has heard ten thousand sirens in her life, is at the blackboard finishing a diagram of a stationary wave. **Khadija**, a Form Six girl from Jangwani who reads physics textbooks for entertainment, is correcting a sign in one of her equations with the calm precision of a someone who has waited four years for this exact moment. **Kipanga** is staring at the diagram and pretending to understand. **Kipute** is checking Khadija's correction and quietly noting that she is, on this occasion, correct. **Mr. Akilikubwa** is at the back of the room, watching the four of them with the expression of a fisherman who has just seen his float twitch.

From the road outside, a sound begins.

NEEEEEEEEEEEEE—

An ambulance, racing from Kariakoo north to Muhimbili. It is coming fast. The class falls completely silent. Khadija's chalk freezes against the board. Kipanga stops pretending to understand. The siren reaches the school gate at maximum loudness, the most arresting noise a human throat or an electronic horn can make.

NEEEEEEEEEEEEE—

And then it passes, accelerating north towards the hospital, and the sound changes.

nooooooooo—

Same siren. Same machine. Same driver. Same morning. A completely different note.

Kipanga, who has been silent since arriving at Jangwani because the city classroom intimidates him slightly, finally finds a subject he has authority on, namely sirens.

Kipanga: *"Sir, this one is easy. The driver pressed the siren button hard while coming, and soft while going. At Miono we have the same thing with the thief-alarm whistles at the school shop. A hard press is louder and a soft press is quieter."*

Monalisa turns from the blackboard. Her eyebrow rises with the slow patience of someone who has explained Dar to visitors many times before.

Monalisa: *"Kipanga, that is an electronic siren on a Land Cruiser ambulance. It does not have a press-button. It has one switch, on or off, and the people inside are not playing it like a piano. They are too busy keeping the patient alive. Also, in the four years I have lived next to this road, I have not once heard an ambulance pass slowly. They come fast, they leave fast, and the note still changes."*

Kipanga is not yet ready to surrender.

Kipanga: *"Then perhaps the driver opened the window while coming and closed it while going?"*

Khadija puts down her chalk and joins in, with the composed confidence of a Form Six student who has read about this before and now intends to demonstrate it.

Khadija: *"Kipanga, it is the air itself. The sound waves bunch together in front of the ambulance, the way daladadas pile up at the Mwenge roundabout in the evening rush, and they stretch apart behind it, the way the same daladadas look on the Bagamoyo Road at midnight. Bunched-up waves are short. Short waves are high in pitch. Stretched-out waves are long. Long waves are low in pitch. There. The driver and the windows have nothing to do with it."*

Kipute, who has been resting her chin in her hand while thinking, now drops her hand and taps Khadija on the elbow.

Kipute: “Khadija, your *daladalas-at-Mwenge* picture is mostly right. But why does the ambulance bunch the waves up? The siren is producing the same number of beeps per second whether the ambulance is parked, or moving slowly, or racing to Muhimbili. The frequency at the source is fixed. So what changes?”

Khadija opens his mouth, closes it, opens it again.

Khadija: “Because the source itself moves between one beep and the next?”

Kipute: “Yes. But say more.”

Khadija: “I... will think about it.”

Mr. Akilikubwa, who has been watching this with the expression of a man whose float has now disappeared under the surface, walks slowly to the blackboard. He erases Monalisa's stationary wave with one sweep of the duster. Monalisa winces. He writes, in chalk, a single equation.

$$f' = f \times \frac{v \pm u_o}{v \mp u_s}$$

He turns to face the class.

Mr. Akilikubwa: “This equation is the entire ambulance. Everything Kipanga said is wrong, and we love him for trying. Everything Khadija said is half right, which is the most dangerous kind of right because the other half is missing. Everything Kipute is about to say is correct, and that is why we sent her on this study tour. The siren did nothing. The driver did nothing. The universe did one very specific thing. Motion bends perception before it bends sound. This chapter is the unpacking of that sentence.”

Monalisa, recovered from the loss of her stationary wave.

Monalisa: “Sir, but why is there a plus on top and a minus on the bottom? And why are the signs paired like this instead of being the other way round?”

Mr. Akilikubwa: “That, is what the next seventy pages are for. By the time we finish this chapter, you will know not only why the formula looks the way it does, but also why the same kind of formula explains the colour of a distant galaxy, the speed of a car caught by a speed camera on the Morogoro Road, the way a bat finds a moth in the dark inside Kondoa cave, and why your TBC radio sometimes picks up broadcasts from Mozambican radio stations at night but never at noon. All of it, every one of those phenomena, follows from one ambulance, one road, and one equation.”

Kipanga, in a whisper to Kipute.

Kipanga: “Kipute, I think I should pay much more attention in Form Five than I did in Form Four.”

Kipute: “Kipanga, you are in Form Five.”

Welcome to Chapter 9. The previous three chapters built the toolkit that this chapter spends.

This chapter takes those three toolkits and applies them to two specific media that students hear and see every day, namely sound waves in air and electromagnetic waves in vacuum. It then introduces one phenomenon, the Doppler effect, which acts on both media in essentially the same way and which turns out to be the single most useful idea in classical physics for measuring motion at a distance.

Two debts from Chapter 7 are paid in this chapter, and you should know which ones before we begin.

The first debt is the speed of a wave in a gas. Chapter 7 derived the wave speed for a stretched string, a solid rod, and a liquid, all of them from the same square-root rule that the speed equals the root of an elastic property over an inertial property of the medium. Gases were postponed at the time, because the elastic property of a gas depends on whether the compression at each cycle of the wave is isothermal (slow enough for heat to flow) or adiabatic (too fast for heat to flow), and Chapter 7 had not yet earned the right to bring in the adiabatic process from Chapter 5. We have done the homework since. The section that follows on the speed of sound applies the square-root rule to a gas, identifies the adiabatic bulk modulus as the elastic property that fits, and pays the debt in full.

The second debt is the harmonic spectrum of an air column. Chapter 7 derived the harmonic spectrum of a stretched string fixed at both ends, and then promised that the same boundary-condition reasoning would deliver the harmonic spectrum of an air column in a pipe. The pipe argument was postponed on the grounds that pipes are really about sound, and that sound deserved its own chapter. The section on resonance in pipes that comes immediately after, honours that promise.

Let us begin.

THE NATURE OF SOUND

The introduction has done the dramatic work for us. We have an ambulance, a road, a classroom at Jangwani with the windows open because the fan is broken, and Mr. Akilikubwa's equation sitting on the blackboard like a man waiting for a daladala. Before we can ask why the siren rose in pitch coming towards the school and dropped in pitch going away, we must first agree, calmly and without sirens, on what a sound actually is. This first section is the calm one. By its end you will know three things, and you will know them firmly enough that the rest of the chapter can rest its weight on them without trembling. The first is what kind of wave a sound is. The second is what the ear hears when a sound arrives, and what each piece of that hearing corresponds to in the wave itself. The third is why a diver paddling five metres below the surface at Kunduchi Beach cannot hear a friend in the boat above him shouting his name. No equation in this section will frighten you. The serious calculations begin in the next section, when the speed of sound enters the chapter and refuses to leave.

What kind of wave is a sound, really?

The biggest fact about sound is the one we already know from Chapter 7, but we will state it once more because the rest of the chapter swings off it. A sound is a **longitudinal mechanical wave**. **Longitudinal** because the particles of the medium oscillate in the same direction in which the wave is travelling, not perpendicular to it. **Mechanical** because it needs a medium: there is no sound in a vacuum, a point that Mr. Akilikubwa will return to several times before the chapter is finished, often at exam time.

What Chapter 9 adds to Chapter 7 is the medium-specific detail. The medium that carries sound in everyday life is a **fluid**, almost always a **gas** (the air around us), sometimes a **liquid** (the seawater the diver is swimming through), and on **rare** interesting occasions a **solid** (the steel rail along the TAZARA line that the trackside worker presses his ear to). In every one of those media the disturbance is the same: a small local fluctuation in pressure and density. A vibrating source pushes the air immediately around it slightly inwards; the air recoils outwards when the push is released; the disturbance propagates outwards through the medium as alternating regions of slightly increased pressure (called **compressions**) and slightly decreased pressure (called **rarefactions**); each particle of the medium, on average, ends up exactly where it started. Energy travels outwards. Air does not. The diagram below shows exactly this picture.

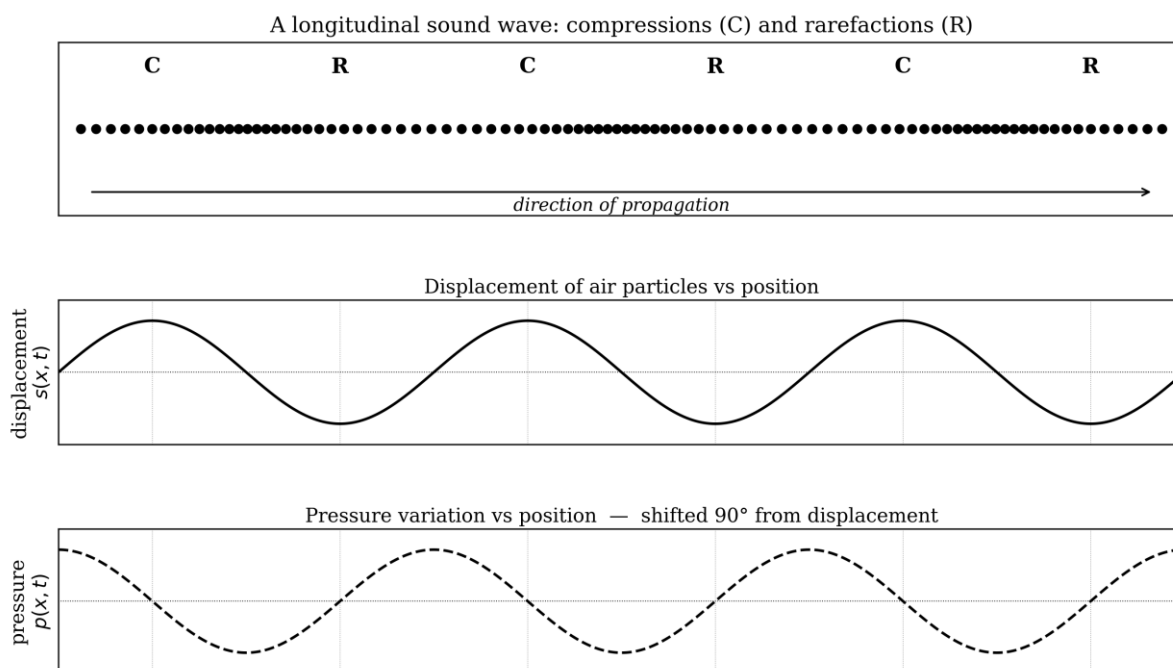


Figure: A longitudinal sound wave. Top: the air molecules cluster (compressions, marked C) and spread out (rarefactions, marked R) as the wave moves along. Middle: displacement of each particle from its rest position. Bottom: pressure variation. The two curves are sinusoids of the same wavelength, but pressure peaks fall exactly where displacement is zero.

Two graphs sit next to every textbook picture of a sound wave, and the relationship between them is the source of more confusion in this chapter than any other piece of geometry. The first graph shows the **displacement** of each particle from its rest position. The second graph shows the **pressure variation** (the

small departure of the pressure from atmospheric) at the same instant, as a function of the same position. Both graphs are sinusoidal. Both have the same wavelength. Both travel at the same speed. But they are **ninety degrees out of phase with each other**, and they are out of phase for a reason that becomes obvious the moment you stop to think about it.

Consider the place along the wave where two groups of particles are converging from opposite sides. The particles on the left are moving rightwards; the particles on the right are moving leftwards. Right at the spot where the two groups meet, the air is being squeezed from both sides at once. The pressure there is at its **maximum**. But the particles at that exact spot have zero net displacement: equal amounts of particle have come in from each side, the local mean position has not moved. So at the point of maximum pressure, the displacement is zero (a **node** of displacement). Where the air has bulged most strongly to one side, the pressure is back to atmospheric (a node of pressure), but the displacement is at its maximum there (an **antinode** of displacement). The two patterns interlock like the teeth of a zipper. Pressure peaks fall exactly where displacement is zero, and the other way around. This is what “ninety degrees out of phase” means in pictures, and it is the geometry that the resonance-tube experiment in the next-but-one section lives on. If you can hold this picture clearly in your head, you will save yourself a lot of suffering when we come to the closed-pipe boundary condition.

Mathematically, the displacement of a particle at position x at time t can be written, exactly as in Chapter 7:

$$s(x, t) = s_{\max} \sin(\omega t - kx)$$

The pressure variation at the same point at the same time is then:

$$p(x, t) = p_{\max} \cos(\omega t - kx)$$

A cosine is a sine shifted by a quarter of a cycle, and that quarter-cycle shift is exactly the ninety-degree offset we have just argued for. The two amplitudes are not independent. They are linked by the elastic property of the medium, and the link is worth deriving once because it shows the wave-speed formula from Chapter 7 in action.

Consider a slab of air whose rest position stretches from x to $x + \Delta x$, with cross-sectional area A . When the wave passes through, the left face moves to $x + s(x, t)$ and the right face moves to $x + \Delta x + s(x + \Delta x, t)$. The slab's new length is $\Delta x + s(x + \Delta x, t) - s(x, t)$. The fractional change in length, which (because the slab is thin) equals the fractional change in volume, is obtained by dividing by Δx and letting $\Delta x \rightarrow 0$:

$$\frac{\Delta V}{V} = \frac{\partial s}{\partial x}$$

Differentiate the wave. From the wave expression for $s(x, t)$ above:

$$\frac{\partial s}{\partial x} = -k s_{\max} \cos(\omega t - kx)$$

Apply the bulk modulus:

The bulk modulus K of the medium, defined in Chapter 5, says that a fractional volume change produces a pressure change in the opposite sense (a compression raises pressure, a rarefaction lowers it):

$$p = -K \frac{\Delta V}{V} = -K \frac{\partial s}{\partial x} = K k s_{\max} \cos(\omega t - kx)$$

Comparing with the form $p = p_{\max} \cos(\omega t - kx)$ we wrote earlier, the pressure amplitude is therefore:

$$p_{\max} = K k s_{\max}$$

Use the wave-speed formula:

The unified formula from Chapter 7 gives the wave speed in a fluid as $v = \sqrt{K/\rho}$, from which $K = \rho v^2$. The wave number is $k = \omega/v$. Substituting both into the line above:

$$p_{\max} = (\rho v^2) \left(\frac{\omega}{v} \right) s_{\max}$$

The factor of v in the numerator and denominator partially cancels; the surviving v carries the whole imprint of the elastic property of the medium. The result is:

$$p_{\max} = \rho v \omega s_{\max}$$

where ρ is the density of the medium, v is the speed of sound in the medium, and $\omega = 2\pi f$ is the angular frequency.

Pitch, loudness, and quality: the three things the ear hears

A musician sitting at the front of a hall on Sokoine Avenue can listen to a single note from a single instrument and describe three completely different things about it.

The first is how high or low the note is. A piccolo is high; a tuba is low. This perceptual property is called the **pitch** of the note.

The second is how loud the note is. A whisper is quiet; a market trader at Kariakoo announcing the price of mangoes is loud. This perceptual property is called the **loudness**.

The third is what kind of instrument is making the note, even when the pitch and loudness of two notes are matched. A clarinet and a guitar playing the same note at the same loudness do not sound the same; the listener can tell them apart instantly. This perceptual property is called the **quality** of the note, also called the **timbre**.

Each of these three perceptual properties corresponds to one specific feature of the wave that the ear receives. The correspondence is not vague; it is one-to-one, and it is worth memorising once and then using freely for the rest of the chapter.

Pitch corresponds to **frequency**. A high-pitched note is a wave of high frequency; a low-pitched note is a wave of low frequency. Doubling the frequency raises the pitch by exactly one **octave**, which is the musical interval between middle C and the next C above it. The piano A above middle C has a frequency of 440Hz; A two octaves above that is 1760Hz; A two octaves below is 110Hz. Two notes sound musically equivalent (same letter name, just at different heights) when their frequencies are in a simple ratio.

Loudness corresponds to **intensity**. The intensity of a wave was defined in Chapter 7 as the power carried across a unit area per second, and intensity scales as the square of the amplitude. A louder note is a wave of larger pressure amplitude (or equivalently larger displacement amplitude). The detailed mathematics of the loudness scale, including the logarithmic decibel scale that the rest of the world uses to write about loudness, lives a few sections from now in the section on the decibel.

Quality corresponds to the **harmonic spectrum**. A real musical note from a real instrument is almost never a single sinusoid at a single frequency. It is a sum of many sinusoids: a strongest one at the fundamental frequency f_1 , and weaker ones at integer multiples of f_1 , called the second harmonic at $2f_1$, the third harmonic at $3f_1$, and so on. Chapter 7 met this idea when it discussed the vibrating string. The exact relative amplitudes of the higher harmonics depend on how the source is excited: a string plucked at its middle gives a different harmonic mix from the same string plucked near the bridge, a clarinet's column of air gives a different mix from a guitar's string, and a human voice singing the vowel AAH gives a different mix from the same voice singing EEH. The ear takes the full mix in and presents it to the brain as a single percept called timbre. Same fundamental, different harmonic recipe, different instrument: that is the whole of quality.

A small but important consequence of this. When you hear your own voice from a recording for the first time, you flinch, because the recording does not sound like the voice in your head. The voice in your head is the part your skull bones conduct directly to your inner ear, which favours the lower harmonics. The voice in the recording is what a microphone in front of your mouth heard, which favours the higher harmonics carried through the air. Same fundamental, different harmonic mix, different timbre, and so a different perceived voice. This is not your phone playing tricks on you. It is the harmonic spectrum doing its accounting.

The audible range, and the worlds on either side of it

The human ear is sensitive to wave frequencies between roughly **20Hz at the low end and 20000Hz (20kHz) at the high end**, in a healthy young person. The exact bounds shift with age: the upper bound retreats first, so that an adult of fifty rarely hears above 12kHz, a finding that hi-fi shop assistants in Kariakoo have learned to use to their advantage. The full span of two-and-a-half orders of magnitude is called the **audible range**. Below the lower bound and above the upper bound, the disturbance is still a perfectly good sound wave; the human ear simply does not register it.

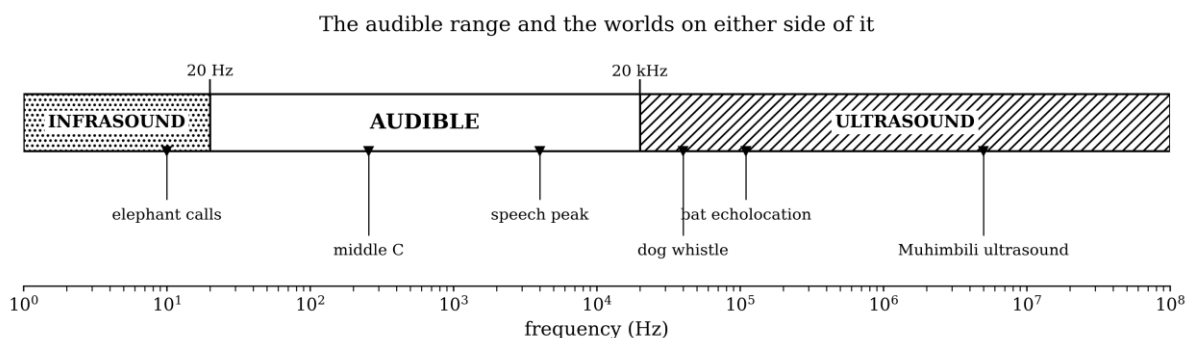


Figure: The audible range on a logarithmic frequency axis, with neighbouring infrasound and ultrasound. The 50Hz tick is the hum a Tanzanian mains transformer makes when the wiring is sloppy.

Below 20Hz the vibrations are called **infrasound**. Earthquakes produce them; large explosions produce them; elephants use them to communicate over kilometres of savannah, which is one of the more surprising practical facts of biology. The vibrations are inaudible but felt: a powerful infrasonic pulse can make a window rattle without making any audible noise.

Above 20kHz the vibrations are called **ultrasound**. Dogs can hear up to about 45kHz, which is why a dog whistle blown at 35kHz drives a dog mad and a human does not even know the whistle has been blown. Bats hunt at frequencies up to 110kHz and read the echoes of their own ultrasonic pulses to navigate; this is the topic of a Doppler example a few sections from now. Medical ultrasound scanners at Muhimbili operate between 2MHz and 15MHz, three orders of magnitude above audibility, because the wavelength at those frequencies is short enough to resolve organs of a few millimetres. None of these waves are physically different from the sound of a school bell. They are simply at frequencies the human ear cannot follow.

Why the diver cannot hear the friend in the boat

A swimmer at Kunduchi Beach, paddling at the surface, hears the boat engine perfectly. A diver swimming five metres below the surface hears nothing, even when a friend in the boat is shouting at the top of his lungs. The diver's ear works perfectly. The friend's lungs are at full pressure. *So why does no sound reach the diver?*

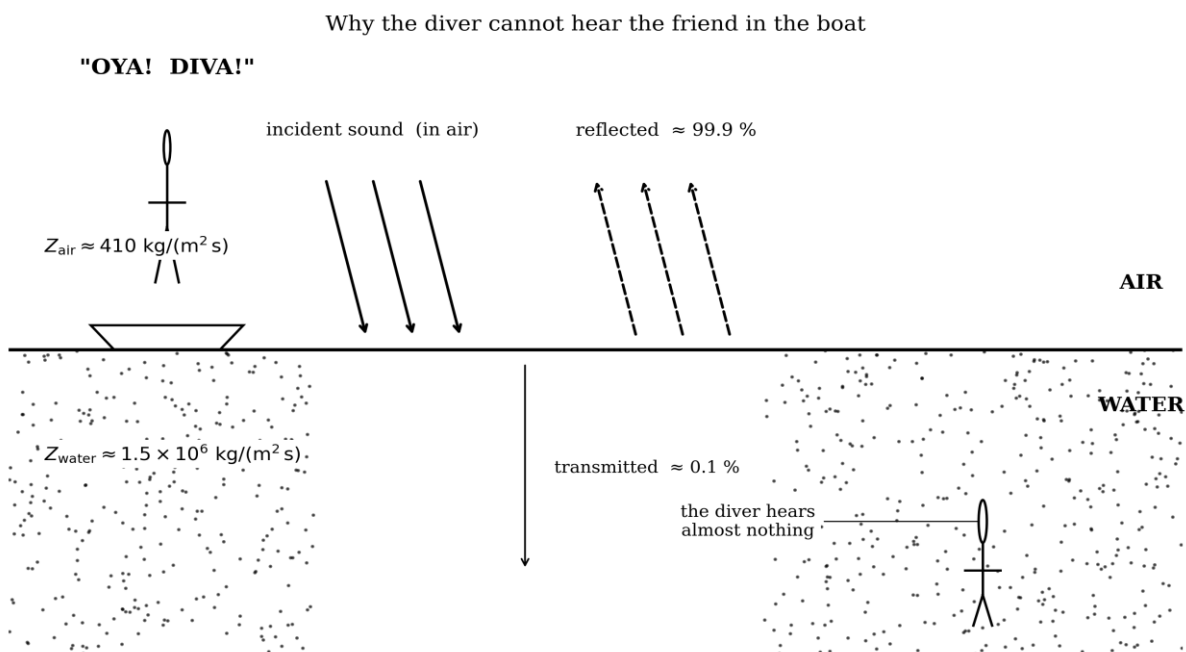


Figure: A loud shout in air carries less than 0.1% of its power into water. The other 99.9% is reflected back at the surface. The diver swims, in practice, in silence.

The answer is one of the most useful pieces of physics in the chapter, and it is the first place where Chapter 9 pays back its debt to Chapter 7's discussion of the wave-speed formula. The key idea is **acoustic impedance**. When a sound wave travelling in one medium meets the boundary of a second medium, two things can happen. Part of the wave *passes into the second medium (transmission)*; part of the wave *bounces*

back into the first medium (**reflection**). The fraction that gets through depends on how different the two media are, in a very precise sense. Each medium has an **acoustic impedance Z** , defined by:

$$Z = \rho v$$

the product of the medium's density and its speed of sound. For air at room temperature, $\rho \approx 1.2 \text{ kg/m}^3$ and $v \approx 340 \text{ m/s}$, giving $Z_{\text{air}} \approx 410 \text{ kg/(m}^2\text{s)}$. For seawater at twenty degrees Celsius, $\rho \approx 1030 \text{ kg/m}^3$ and $v \approx 1500 \text{ m/s}$, giving $Z_{\text{water}} \approx 1.5 \times 10^6 \text{ kg/(m}^2\text{s)}$. The two impedances differ by a factor of about 3700, which is the kind of ratio that sets off alarm bells in every wave physicist who has ever stared at a boundary problem.

The fraction of sound power transmitted across the boundary at normal incidence is given by:

$$T = \frac{4 Z_{\text{air}} Z_{\text{water}}}{(Z_{\text{air}} + Z_{\text{water}})^2}$$

Substituting our two numbers gives $T \approx (4 \times 410 \times 1.5 \times 10^6) / (1.5 \times 10^6)^2 \approx 1.1 \times 10^{-3}$. About 0.1% of the power in the shout makes it through the water surface; the other 99.9% is reflected back into the air. By the time the shout has spread spherically through five metres of water as well, the intensity at the diver's ear is well below the threshold of hearing. The shout is, for the diver, silent.

This is called the **impedance-mismatch argument**, and it is everywhere in physics whenever a wave crosses a boundary between two media of very different stiffness or density. We will meet it again in the applications section when we discuss why an ultrasound scanner needs a gel layer between the transducer and the patient's skin: the gel matches the impedance between the transducer and the body well enough that almost all the ultrasound energy enters the body, instead of bouncing back off the skin like an air-shout off the surface of the sea.

With the picture now firmly in place, the worked examples can begin.

BINDER Example 1

A loudspeaker in the studio of Clouds FM in Kinondoni emits a steady note at 256Hz. The pressure amplitude of the wave one metre in front of the speaker is 0.20Pa. Take the speed of sound in air as 343m/s and the density of air as 1.2 kg/m^3 .

- Find the wavelength of the wave.
- Find the displacement amplitude of an air particle one metre in front of the speaker.

Solution

- The wavelength is the speed of sound divided by the frequency:

$$\lambda = \frac{v}{f} = \frac{343 \text{ m/s}}{256 \text{ Hz}} = 1.34 \text{ m}$$

- The displacement amplitude is connected to the pressure amplitude by:

$$s_{\text{max}} = \frac{p_{\text{max}}}{\rho v \times 2\pi f}$$

Substituting the numerical values:

$$s_{\text{max}} = \frac{0.20 \text{ Pa}}{(1.2 \text{ kg/m}^3)(343 \text{ m/s})(2\pi)(256 \text{ Hz})} \approx 3.0 \times 10^{-7} \text{ m} \approx 0.30 \mu\text{m}$$

Making sense of the answer: The wavelength of 1.34m is comparable to a child's height, which is the reason a 256Hz note bends easily round door frames and corners of buildings: a wavelength of about a metre diffracts strongly around obstacles a metre or two in size. The displacement amplitude of 0.30 micrometres is far smaller than the diameter of a human hair (which is about 80 micrometres), and it is one of the central facts about hearing: the ear can detect particle displacements that are smaller than a wavelength of visible light.

Think like a physicist: Two amplitudes describe the same wave, and they are connected by a single equation. The pressure amplitude p_{max} is what an engineer designing a loudspeaker measures with a microphone. The displacement amplitude s_{max} is what a physicist analysing the air molecules computes from the equation of motion. They are the same physical wave seen through two different mathematical windows. Whenever a problem gives you one and demands the other, the bridge between them is always $p_{\text{max}} = \rho v \omega s_{\text{max}}$.

REAL Example 2

A dog whistle in the hand of a frustrated dog-owner in Wete emits a steady note at 40kHz. A human a few metres away hears absolutely nothing; the family dog, sitting equally near, looks up sharply and starts whining about an injustice he cannot quite name.

- Explain why the dog hears the whistle and the human does not.
- Find the wavelength of the whistle's note in air at 343m/s.
- A fruit bat flying near the house uses echolocation at frequencies up to 110kHz. Would the bat hear the same whistle as a low note or a high note compared with its own calls?
- Why would the same whistle, applied a few centimetres from the dog's ear, hurt the dog rather than charm him?

Solution

(a) The human auditory range tops out at about 20kHz in a young adult and lower in an older one. 40kHz lies well beyond the human upper cut-off; the human eardrum cannot respond fast enough, so the wave passes through the room undetected by the human ear. The dog auditory range extends up to about 45kHz, so 40kHz lies inside what the dog can hear. The whistle is loud, but the loudness is delivered in a frequency band that only one of the two species has equipment for.

(b) The wavelength in air:

$$\lambda = \frac{v}{f} = \frac{343\text{m/s}}{40000\text{Hz}} = 8.6 \times 10^{-3}\text{m} = 8.6\text{mm}$$

(c) The bat's echolocation calls reach up to 110kHz. The whistle at 40kHz lies near the bottom of the bat's hearing range. For the bat, the whistle would sound like a low note, just as a 200Hz human voice sounds low compared with a 1000Hz whistle to a human listener. Frequencies are perceived as high or low relative to the upper end of one's own audible range, and the bat's upper end is much higher than the human's.

(d) Loudness has nothing directly to do with frequency. A 40kHz ultrasonic note can carry enormous pressure amplitude and intensity. The dog's eardrum, although it can follow the frequency, cannot withstand arbitrary intensity. A close-range blast delivers a large pressure amplitude through a wave whose frequency happens to be inside the dog's hearing band, and the eardrum responds painfully to the intensity. The same intensity at 1000Hz, well inside both species' hearing, would hurt a human ear just as much. Frequency makes a sound audible to a given species; intensity makes it painful.

Making sense of the answer: An 8.6mm wavelength is about the width of a small fingernail. Such short wavelengths reflect efficiently off small surfaces, which is exactly why bats can resolve insects a few millimetres in size by echolocation; the same short wavelengths diffract poorly round large obstacles, which is why the ultrasonic part of the whistle does not travel well behind furniture or round corners, while the lower-frequency hisses produced as side-effects of the blowing do. This same wavelength argument is why an ultrasound transducer at Muhimbili operating at 5MHz, with a wavelength in tissue of about 0.3mm, can resolve features one or two millimetres across; an audible-frequency probe with a wavelength of one full metre would resolve nothing finer than the patient's leg.

Think like a physicist: Audibility is a species-specific window onto a universal physics. There is no special category of “sound that humans cannot hear”, and no special category of “sound that dogs can hear”; there is only the physical wave, with its frequency, and there are different ears with different upper cut-offs. The same physical wave that is silent to a human and audible to a dog is audible-but-low to a bat and audible-but-high to a baby (whose audible range typically extends slightly above an adult's). When a question talks about hearing, always separate the physics of the wave from the biology of the listener; they live in different sentences.

With this section behind us, three perceptual properties (pitch, loudness, quality) have been mapped cleanly onto three wave properties (frequency, intensity, harmonic spectrum). One useful boundary trick (acoustic impedance mismatch) has explained why most sound waves bounce back at most boundaries; we will use the same trick again at the resonance tube and again at the medical-ultrasound transducer. We have not yet, however, addressed the single most-asked question of any chapter on sound: *how fast does it travel?* Chapter 7 promised the answer would arrive in this chapter, and the moment has come to deliver. The next section unwraps the speed of sound: what Newton calculated, why his answer was wrong by sixteen per cent, and what Laplace did about it.

SPEED OF SOUND: NEWTON, LAPLACE, AND TEMPERATURE

The previous section settled the perceptual side of sound. We know what a sound is, what the ear does with it, and why one species' silence is another species' alarm. But every one of those statements quietly assumed something we have not yet computed: the speed at which a sound wave actually travels through the medium. Chapter 7 promised us this speed and then walked away whistling, on the grounds that gases required an adiabatic argument that Chapter 7 had not yet earned the right to make. The argument has since been earned in Chapter 5, where the adiabatic process was derived in full. The promise comes due here.

The story has three protagonists. Sir Isaac Newton, who in 1687 wrote down a formula for the speed of sound in a gas and was wrong by sixteen per cent. Pierre-Simon Laplace, who in 1816 looked at the same formula, identified one missing piece, slotted it in, and made the formula right to the third significant figure. And the speed of sound itself, which has the inconvenient habit of caring strongly about temperature but not at all about pressure, an asymmetry that catches the unwary on examination day.

The unified wave-speed formula meets a gas

From Chapter 7 we already know that, in every elastic medium, the speed of a mechanical wave is given by the square root of an elastic property of the medium divided by an inertial property of the medium:

$$v = \sqrt{\frac{\text{elastic property}}{\text{inertial property}}}$$

For a gas the inertial property is straightforward: it is the volume mass density ρ . That part needs no debate. The elastic property is the part where Newton and Laplace will disagree, and it is the **bulk modulus** of the gas. Recall from Chapter 5 that the bulk modulus is the pressure required to produce a unit fractional decrease in volume, $K = -V(dP/dV)$. The trouble is that a gas can have two different bulk moduli depending on whether the compression that defines dV is carried out slowly (with enough time for heat to flow in or out and keep the temperature equal to the surroundings) or quickly (so fast that no appreciable heat flows during the compression). One number does not fit both situations, and the choice between them is our next task.

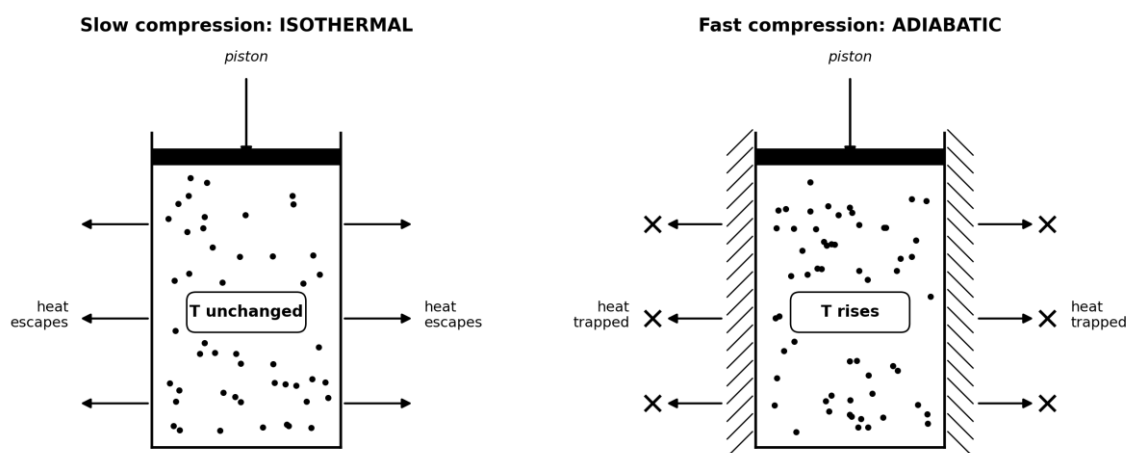


Figure: A gas in a cylinder being pushed by a piston in two limits. On the left, the piston moves slowly enough that heat flows freely through the walls; the temperature of the gas stays equal to the surroundings; the compression is isothermal. On the right, the piston moves fast enough that no heat escapes during the compression (the arrows with crosses indicate that heat transfer is blocked or absent); the temperature of the gas rises; the compression is adiabatic.

Newton's first attempt: isothermal compression

Newton's instinct in 1687 was simple. As the wave passes through the air, the air at each location alternately compresses and rarefies. Newton assumed the compression was slow enough that the air at each location stayed at the same temperature as the air around it: an **isothermal** compression. For an ideal gas held at constant temperature, the equation of state $PV = nRT$ reduces to $PV = \text{constant}$, from which a short calculation gives the isothermal bulk modulus:

$$K_T = -V \frac{dP}{dV} = P$$

So Newton substituted P for the elastic property:

$$v_{\text{Newton}} = \sqrt{\frac{P}{\rho}}$$

Plug in the values for air at standard conditions: pressure $P = 1.013 \times 10^5 \text{Pa}$, density $\rho = 1.29 \text{kg/m}^3$. The arithmetic gives:

$$v_{\text{Newton}} = \sqrt{\frac{1.013 \times 10^5 \text{Pa}}{1.29 \text{kg/m}^3}} \approx 280 \text{m/s}$$

Newton then went to find experimentalists to confirm the figure.

The sixteen-per-cent embarrassment

The experimentalists came back with a different number. Careful measurements of the time taken for a gun's report to reach an observer at a known distance gave the speed of sound in air at 0°C as roughly 332m/s , not 280m/s . The discrepancy is too large to be experimental error. Newton's calculation is off by:

$$\frac{332 - 280}{332} \times 100\% \approx 16\%$$

Sixteen per cent is the gap between physics and almost-physics. Something in the model is physically wrong, and the model has to be repaired. For more than a hundred years no one quite knew what to repair. The repair, when it arrived, did not change a single number in the formula; it changed the interpretation of one of the symbols. That is the kind of fix every physicist envies.

Laplace's correction: adiabatic compression

Laplace's 1816 insight was this. Sound is not a slow phenomenon. The audible frequencies start at 20Hz , which means each compression-and-rarefaction cycle takes at most 50ms ; at 1kHz each cycle takes one millisecond; at 20kHz each cycle takes fifty microseconds. Heat conduction through air at this speed is far too slow to do anything useful. During a single cycle of the wave, the gas at any given location is compressed and then rarefied without any appreciable heat being able to leave or enter. The compression is therefore **adiabatic**, not isothermal. The right elastic property is not the isothermal bulk modulus but the **adiabatic bulk modulus**, derived in Chapter 5:

$$K_S = \gamma P$$

where $\gamma = C_p/C_v$ is the ratio of the specific heat capacities of the gas at constant pressure and constant volume. For diatomic gases such as air (mostly nitrogen and oxygen, both diatomic), $\gamma = 1.40$:

$$v = \sqrt{\frac{\gamma P}{\rho}}$$

This is the **Newton–Laplace formula**. Substituting $\gamma = 1.40$, $P = 1.013 \times 10^5 \text{Pa}$, $\rho = 1.29 \text{kg/m}^3$:

$$v = \sqrt{\frac{(1.40)(1.013 \times 10^5 \text{Pa})}{1.29 \text{kg/m}^3}} \approx 331 \text{m/s}$$

This agrees with the measured value to better than half a per cent. Laplace's correction is the difference between physics and almost-physics, and the whole of it lives in one symbol: the factor of γ inside the square root.

Temperature: the dependence the formula does have

Examiners love a question that asks for the speed of sound at a temperature that is not standard. The formula above is in terms of pressure and density, and both of those change with temperature. There is a cleaner form in which only the temperature appears. Substitute the ideal-gas law in the convenient form $P = \rho RT/M$

(where $R = 8.314\text{J}/(\text{molK})$ is the universal gas constant and M is the molar mass of the gas) into the Newton–Laplace formula:

$$v = \sqrt{\frac{\gamma P}{\rho}} = \sqrt{\frac{\gamma \rho R T / M}{\rho}} = \sqrt{\frac{\gamma R T}{M}}$$

The density cancels exactly. The speed of sound in an ideal gas depends only on the absolute temperature and on the molar mass of the gas. In particular, for any fixed gas (so γ and M are fixed):

$$v \propto \sqrt{T}$$

This is the most important single line in this section. Two important consequences:

- 1) Air at 0°C (273K) gives $v = 331\text{m/s}$; air at 25°C (298K) gives $v = 331 \times \sqrt{(298/273)} \approx 346\text{m/s}$. An engineering rule of thumb worth memorising is the linear approximation $v \approx 331 + 0.6T_c$, with T_c in Celsius; the rule is good to better than one per cent over the everyday range **0–40°C**.
- 2) The temperature at which v doubles compared with its value at 0°C satisfies $\sqrt{(T/273)} = 2$, giving $T = 4 \times 273\text{K} = 1092\text{K} = 819^\circ\text{C}$. You will not encounter that temperature when whistling on the school field, but a furnace at a cement plant at Tanga reaches it without difficulty.

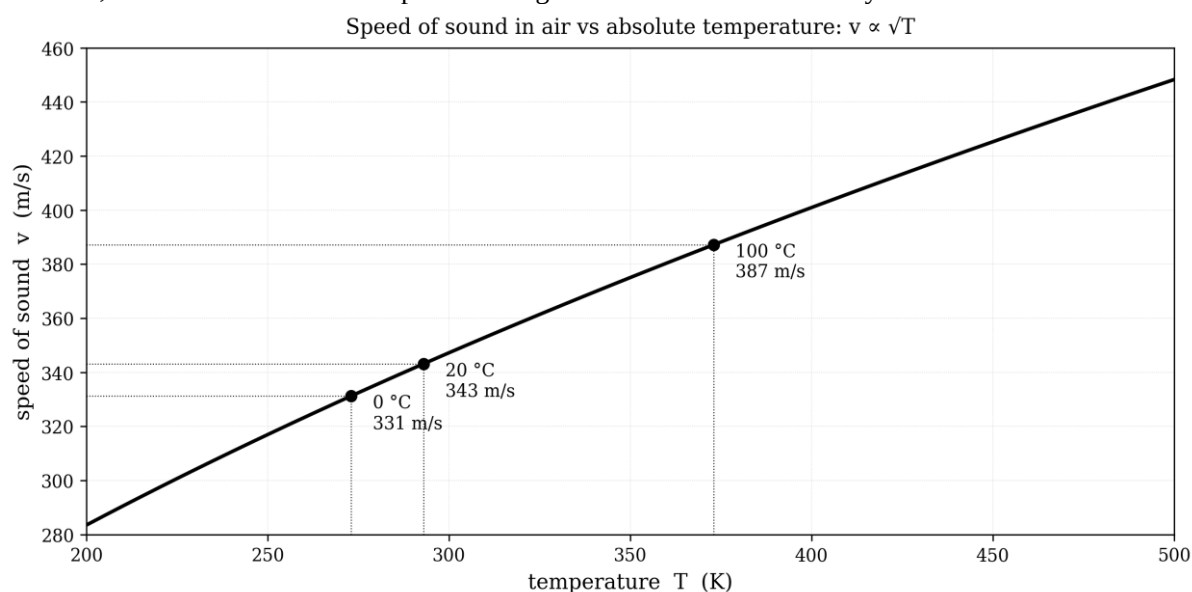


Figure: Speed of sound in air vs absolute temperature. The curve is a square root of T , anchored at $273\text{K} \rightarrow 331\text{m/s}$. Doubling v means quadrupling T .

Pressure: the dependence the formula does NOT have

From $v = \sqrt{(\gamma P / \rho)}$ it certainly looks as though v should depend on P . At first glance, halving the pressure should multiply the speed by $1/\sqrt{2}$, and doubling the pressure should multiply it by $\sqrt{2}$. But this is a trick the formula plays on the careless reader. At a **fixed temperature**, the ideal-gas law gives $\rho \propto P$: *doubling the pressure* of a fixed mass of gas at fixed temperature also *doubles its density*. The ratio P/ρ therefore stays exactly constant. The speed does not change.

This is why the speed of sound at the summit of Mount Kilimanjaro, where the air pressure is roughly half its sea-level value, is the same as the speed of sound on the beach at Dar es Salaam at the same temperature. The **lower** pressure on Kilimanjaro is exactly compensated by the **lower** air density there. Only the temperature, which is about 30°C lower on Kilimanjaro, gives the climber a slower sound; and that slowing is about 5%, much less than a naive use of the pressure ratio would predict. The asymmetry, in which temperature matters while pressure does not, is one of those facts the student must internalise and never write down in reverse on an examination paper.

Humidity: a small but real correction

Humid air is slightly less dense than dry air at the same temperature and pressure, because water vapour ($M = 18\text{g/mol}$) replaces nitrogen ($M = 28\text{g/mol}$) and oxygen ($M = 32\text{g/mol}$), both of which are heavier.

From $v = \sqrt{(\gamma RT/M)}$, smaller M means larger v . In Dar es Salaam at 30°C and 80% relative humidity, the speed of sound is about 0.6% higher than in completely dry air at the same temperature. This is the reason a singer's voice carries a fraction further on a humid coastal morning than in dry Dodoma highland air at the same thermometer reading. The correction is usually too small to matter in calculations, but it deserves to appear at least once in your notes so that it never catches you by surprise later.

Before the ideas have a chance to drift away, let us anchor them with three worked examples.

BINDER Example 3

Calculate the speed of sound in dry air at 20°C using the Newton–Laplace formula, taking $\gamma = 1.40$, molar mass of dry air 28.96g/mol and $R = 8.314\text{J}/(\text{mol} \cdot \text{K})$. Cross-check the answer using the engineering rule $v \approx 331 + 0.6T_C$.

Solution

Convert the temperature: $T = 20 + 273 = 293\text{K}$.

Substitute into the formula:

$$v = \sqrt{\frac{\gamma RT}{M}} = \sqrt{\frac{(1.40)(8.314\text{J}/(\text{molK}))(293\text{K})}{28.96 \times 10^{-3}\text{kg/mol}}} \approx 343\text{m/s}$$

Engineering check: $v \approx 331\text{m/s} + 0.6(20\text{m/s}) = 331 + 12 = 343\text{m/s}$. The two answers agree to the nearest metre per second.

Making sense of the answer: The figure 343m/s is the value engineers, musicians and air-traffic controllers in every city in Tanzania use when they don't know the temperature precisely. At room temperature it is correct to better than half a per cent. Notice the two completely different lines of reasoning, namely a fundamental formula from kinetic theory and a hand-fitted linear rule from data, converging on the same answer. When two unrelated paths agree, a working physicist can carry on with confidence.

Think like a physicist: Every gas-temperature problem becomes a one-line substitution once $v = \sqrt{(\gamma RT/M)}$ is in your head. The two boxes you must fill before computing are the absolute temperature (always in kelvin, never in celsius) and the molar mass (always in kg per mole, never in g per mole). A wrong unit on M is the single most popular error on this topic on every examination paper that touches it.

BINDER Example 4

Find the speed of sound in helium at 20°C. Helium is monatomic with $\gamma = 5/3$ and $M = 4.00\text{g/mol}$. Then explain, in two sentences, why breathing helium makes the human voice sound much higher than usual.

Solution

At $T = 293\text{K}$ (20°C):

$$v_{\text{He}} = \sqrt{\frac{(5/3)(8.314\text{J}/(\text{molK}))(293\text{K})}{4.00 \times 10^{-3}\text{kg/mol}}} \approx 1007\text{m/s}$$

That is roughly $1007/343 \approx 2.94$ times the speed of sound in air at the same temperature, or about three times. When the air inside a person's vocal tract is replaced by helium, the resonant frequencies of the tract scale with the speed of sound inside it: each resonance now sits at roughly three times its normal frequency. The vocal cords themselves vibrate at the same fundamental frequency as before (the fold tension has not changed), but the higher resonances of the tract amplify the harmonics far more than usual, shifting the perceived timbre drastically upwards. The voice does not change its pitch in the strict musical sense, because the fundamental is unchanged, but the listener hears it as much higher because the harmonic mix has been radically reshaped.

Making sense of the answer: Notice what changed and what did not. The helium has $M = 4$ instead of 28.96, a factor of about seven smaller; that pulls v up by a factor of about $\sqrt{7} \approx 2.65$. Helium also has $\gamma = 5/3$ instead of $7/5$ (a factor of $25/21 \approx 1.19$ inside the square root, i.e. $\sqrt{1.19} \approx 1.09$ on the speed). Multiplying gives $2.65 \times 1.09 \approx 2.89$, which lands inside one per cent of the full calculation. Two competing factors, lighter molecules and stiffer adiabatic behaviour, each push the speed up, and helium wins on both.

Think like a physicist: Always separate what depends on the gas (its γ and M) from what depends on the conditions (temperature). The trick of comparing speeds in two different gases is to take their ratio under the square root and let identical factors cancel. Whenever a problem asks for a ratio rather than a number, take the ratio early.

HOT Example 5

A naval observer is positioned due north of a heavy gun, on flat ground, on a morning when the temperature is 20°C. The observer sees the flash of the gun and starts a stopwatch; the report of the gun reaches her ears 4.0s later. A steady wind is blowing from east to west at 48km/h. Treating the speed of light as effectively infinite over the relevant distance, find the distance from the gun to the observer.

Solution

Speed of sound in still air at this temperature:

Using the engineering rule $v \approx 331 + 0.6T_C$:

$$v_{\text{still}} = 331\text{m/s} + 0.6 \times 20\text{m/s} = 343\text{m/s}$$

Decompose the wind:

The wind is from east to west; the observer is due north of the gun, so the sound is travelling due north on its way to her ear. The wind's velocity vector is along east–west; its component along the sound's direction of travel (north) is zero. Convert the wind speed for completeness:

$$w = 48\text{km/h} = \frac{48000\text{m}}{3600\text{s}} = 13.3\text{m/s}$$

The component of the wind in the direction the sound is travelling is the wind's full magnitude multiplied by the cosine of the angle between the wind vector and the line gun-to-observer. That angle is 90°, so the component is $13.3\text{m/s} \times \cos 90^\circ = 0$. A perpendicular wind carries no energy along the sound's path; it advects the sound sideways but does not speed it up or slow it down along the gun-to-observer line.

Effective speed of sound towards the observer

Since the wind contributes nothing along the line of travel, the effective speed of sound from gun to observer is just the still-air value:

$$v_{\text{eff}} = v_{\text{still}} + 0 = 343\text{m/s}$$

Distance from the gun:

Distance equals effective speed times time of flight:

$$d = v_{\text{eff}} \times t = (343\text{m/s})(4.0\text{s}) = 1372\text{m} \approx 1.4\text{km}$$

Making sense of the answer: Many students panic the moment a wind appears in a sound problem and automatically add or subtract the full wind speed to the speed of sound. The correct rule is: only the component of the wind along the line joining source and observer modifies the effective speed of sound between them. A wind blowing perpendicular to the line of travel contributes nothing to the time of flight, however violent it may be. The wind in this problem is doing two things: it is bending the sound's wavefronts sideways and it would shift the apparent direction the sound seems to come from, but it is not changing the time of flight from gun to ear.

Think like a physicist: Vector components are not a piece of mechanics nostalgia tacked onto sound for nostalgia's sake. They are the rule that decides which part of any vector influences any scalar quantity defined along a line. A wind is a vector. Time of flight is a scalar along the source-to-observer line. The rule that connects them is the dot product, and the dot product of a perpendicular wind with the line of travel is zero. Resolving the wind into components, along and across, is the safe move for every Doppler-and-wind problem in the rest of the chapter as well.

With the speed of sound now in hand, sound has acquired its first hard number. The next section asks what happens when a sound wave is confined to a pipe rather than free to spread through open air. The pipe imposes a boundary condition at each end; the boundary condition selects a discrete set of allowed frequencies; and out of that selection falls the whole physics of organ pipes, flutes, clarinets, marimbas, and the resonance-tube experiment that you perform in the laboratory. Chapter 7's promise about pipes is the next one to be paid.

RESONANCE IN PIPES (OPEN AND CLOSED)

Sound that travels through open air spreads, weakens, and dies away within seconds. Sound trapped inside a pipe behaves very differently. The two ends of the pipe send the wave back the way it came, and the wave returning meets the wave still arriving. Most of the time the two cancel each other within a few cycles and the pipe falls silent. At certain special frequencies, however, the two reinforce each other perfectly and the pipe sings out a loud sustained note, picking that one frequency out of every other frequency the source might be supplying. These special frequencies are called the **resonant frequencies** of the pipe, and the phenomenon is called **acoustic resonance**. An organ pipe in a church sings on its own particular notes and no others, a flute plays one fingered pitch and refuses to play its neighbour, and the resonance tube experiment in your school laboratory is loud at one column length and almost silent at every neighbouring length. The pipe, in every case, is acting as a frequency filter, and it filters by the very same boundary-condition reasoning that Chapter 7 used for stationary waves on a stretched string.

Chapter 7 left us with two equivalent forms for the stationary-wave equation, the form with the spatial factor $\sin kx$ and the form with the spatial factor $\cos kx$, and explained that the boundary chooses the form. The sine form has a node at the position labelled $x = 0$ and is selected by a clamped boundary such as a string fixed to a wall. The cosine form has an antinode at $x = 0$ and is selected by a free boundary, where the medium is unconstrained. We promised then that the same boundary-condition reasoning would deliver the harmonic spectrum of an air column. The promise comes due here. The new ingredient is that a pipe can be sealed at one end and open at the other, or open at both ends, and the two boundary situations are not the same. A closed end traps the air and forbids displacement at that end. An open end leaves the air free to move but forbids any excess pressure there, because the air just outside the pipe is sitting at atmospheric pressure and the wave must match it. Out of those two simple rules falls the entire physics of pipe modes.

The closed pipe: one end sealed, one end open

Imagine a length of brass tubing held shut at the left end by a flat lid and left open at the right end. A clarinettist blows across the open end and pushes a sound wave into the tube. The wave runs to the lid, bounces back, and meets the next wave coming in. Some pairs of incoming and reflected waves will cancel each other within a fraction of a second and produce no sound. Others will reinforce each other and sustain a loud note. We want to find which wavelengths qualify.

Set up a coordinate system. Place the origin at the closed end and let $x = L$ mark the open end. The standing wave inside the pipe must satisfy two boundary conditions, one at each end, and the conditions are not the same.

At the closed end the air has nowhere to go: the lid is solid, the molecules cannot pass through it, and the displacement of the air there is forced to zero at every instant. That is a node of displacement at $x = 0$. The sine spatial form is the one that puts a displacement node at $x = 0$ without any further work, so we take it as our starting point:

$$s(x, t) = 2A \sin(kx) \cos(\omega t)$$

At the open end the air is free to move; the displacement is unconstrained. The pressure, however, is constrained. Just outside the open end the air sits at atmospheric pressure, and the small pressure variation produced by the wave must drop to zero there. The earlier section of this chapter on the nature of sound established that pressure and displacement are ninety degrees out of phase in a sound wave, which means a pressure node coincides with a displacement antinode. So the open end is a displacement antinode.

The mathematical condition for an antinode of the sine spatial form at $x = L$ is that $\sin(kL)$ equals plus or minus one. This in turn means kL must equal an odd multiple of $\pi/2$:

$$kL = (2n - 1) \frac{\pi}{2}, \quad n = 1, 2, 3, \dots$$

Translating from wave number to wavelength using $k = 2\pi/\lambda$, the allowed wavelengths are:

$$L = (2n - 1) \frac{\lambda_n}{4}, \quad n = 1, 2, 3, \dots$$

and the allowed frequencies, found from $f = v/\lambda$, are:

$$f_n = \frac{(2n - 1) v}{4L}, \quad n = 1, 2, 3, \dots$$

The lowest of these is the fundamental:

$$f_1 = \frac{v}{4L}$$

for which exactly one quarter of a wavelength fits inside the pipe, with a node at the closed end and the nearest antinode at the open end. The next allowed mode has $f_2 = 3f_1$ and is called the **first overtone** or the **third harmonic**, with three-quarters of a wavelength inside the pipe. Above it sit $f_3 = 5f_1$ (second overtone, **fifth harmonic**), $f_4 = 7f_1$ (**third overtone**, **seventh harmonic**), and so on. Only the odd harmonics show up. The even harmonics $2f_1$, $4f_1$, $6f_1$, and so on, are absent from the spectrum of a closed pipe because they cannot satisfy both boundary conditions at the same time. An even harmonic would demand either two nodes at the two ends or two antinodes, but a closed pipe insists on one of each.

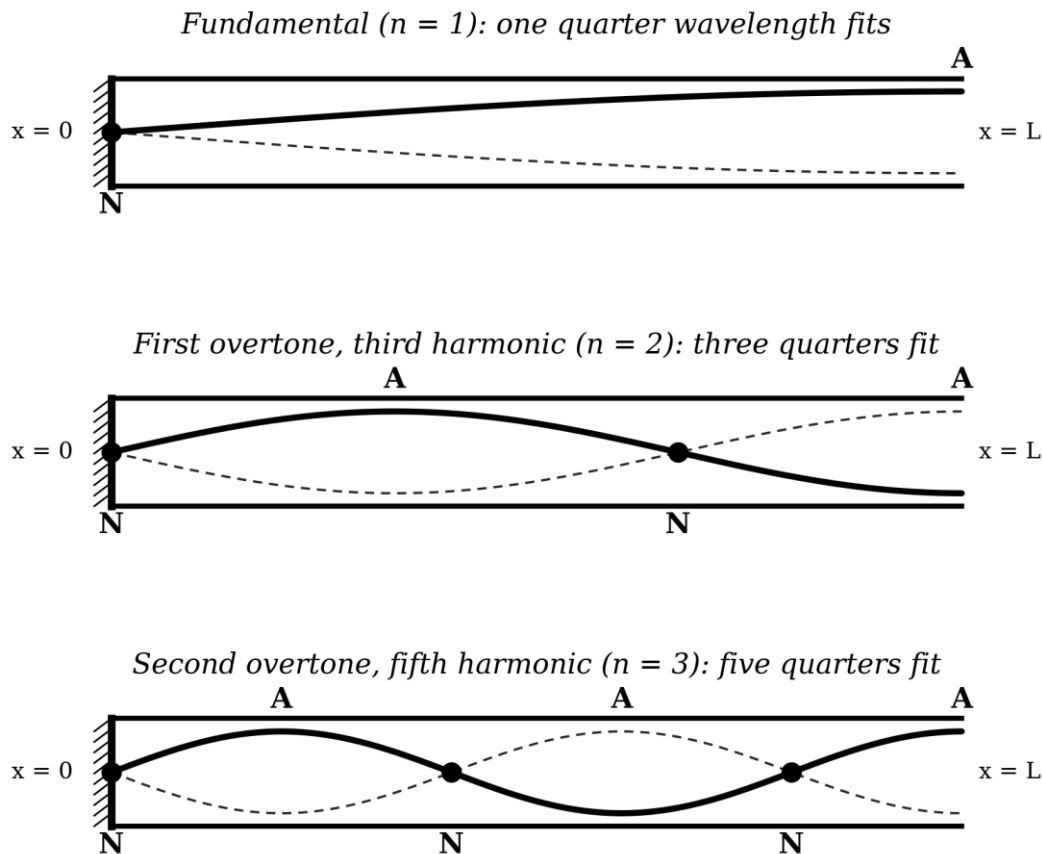


Figure: Standing-wave patterns of the first three modes of a closed pipe. The closed end on the left is always a displacement node (N), and the open end on the right is always a displacement antinode (A).

The open pipe: both ends open

Imagine now a length of brass tubing open at both ends, a piccolo or a flute or the barrel of a horn before anyone has stuck their hand into it. Place the origin $x = 0$ at one open end and let $x = L$ mark the other. Both ends are free; both ends are displacement antinodes; both ends are pressure nodes. The cosine spatial form is the one that puts a displacement antinode at $x = 0$ automatically, so it is the form we adopt:

$$s(x, t) = 2A \cos(kx) \cos(\omega t)$$

The condition for an antinode of the cosine spatial form at $x = L$ is that $\cos(kL)$ equals plus or minus one, which means kL must equal an integer multiple of π :

$$kL = n\pi, \quad n = 1, 2, 3, \dots$$

The allowed wavelengths are:

$$L = \frac{n\lambda_n}{2}, \quad n = 1, 2, 3, \dots$$

and the allowed frequencies are:

$$f_n = \frac{nv}{2L}, \quad n = 1, 2, 3, \dots$$

The fundamental of the open pipe is therefore $f_1 = v/(2L)$, exactly twice the fundamental of a closed pipe of the same length. At the fundamental, half a wavelength fits inside the open pipe, with antinodes at the two ends and a single node halfway between. The next allowed mode is $f_2 = 2f_1$, the first overtone or the second harmonic, with a complete wavelength inside the pipe. After that come $f_3 = 3f_1$, $f_4 = 4f_1$, and so on in unbroken succession. An open pipe is friendlier than a closed pipe, permitting all harmonics, both odd and even. This richer spectrum is the acoustic reason a flute (open at both ends) sounds brighter than a clarinet (closed at the reed end) on the same note: not louder, brighter, because the listener's ear is fed a fuller mix of higher partials, which the auditory cortex processes as timbre.

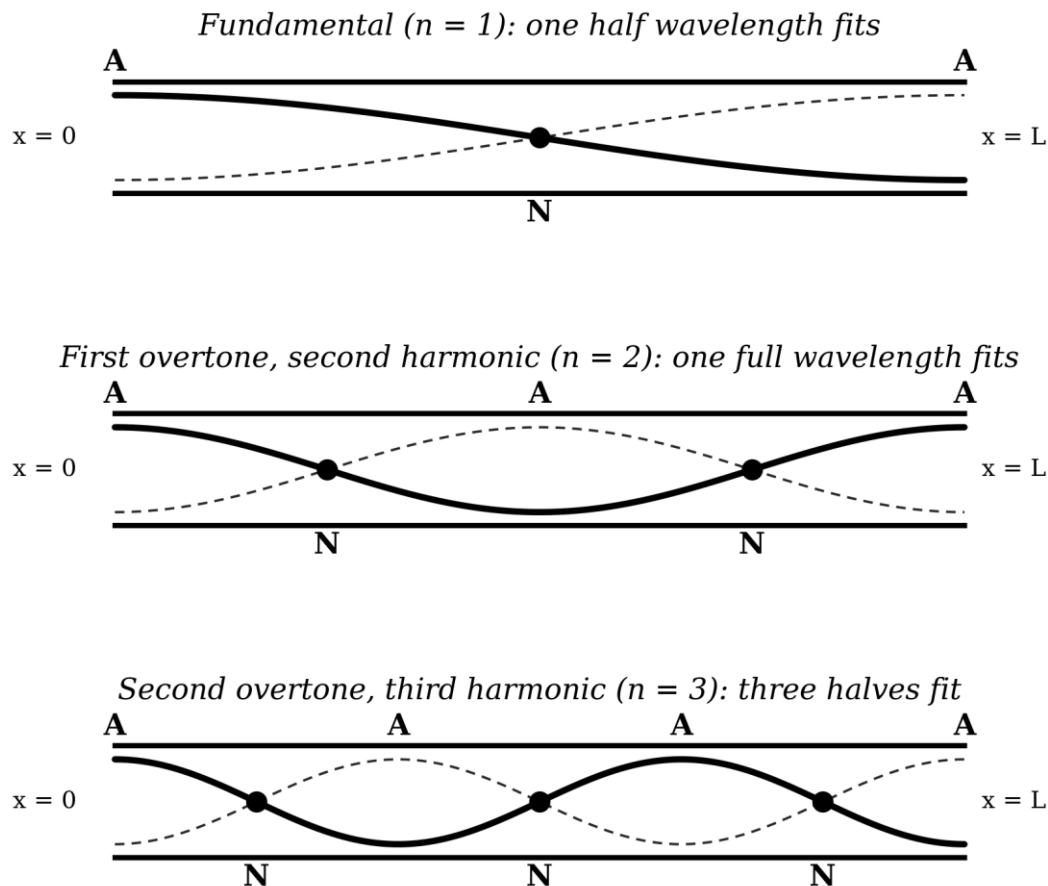


Figure: Standing-wave patterns of the first three modes of an open pipe. Both ends are always displacement antinodes (A), with nodes (N) in between.

Open against closed: the same length, two different worlds

Stand the two formulas side by side and the consequences are immediate. An open pipe and a closed pipe of the same physical length sing two different fundamentals:

$$f_1(\text{open}) = \frac{v}{2L}, \quad f_1(\text{closed}) = \frac{v}{4L}$$

So the open pipe sounds one full octave above the closed pipe of equal length. Turning the comparison around, a closed pipe needs to be only half as long as an open pipe to sound the same fundamental, and that is one practical reason organ builders prefer closed pipes for the very lowest notes of a cathedral organ: a closed 16-foot pipe sounds at the same pitch as an open 32-foot pipe, at half the cost in lead and half the headroom needed in the loft above the choir. The price is the missing even harmonics, which give the closed-pipe stop a slightly hollow timbre next to the open-pipe stop. Different sounds for different musical purposes, but the same physics underwriting both.

The spacing of consecutive harmonics is different in the two cases as well. In an open pipe two consecutive harmonics differ by f_1 , while in a closed pipe two consecutive harmonics differ by $2f_1$. A closed-pipe

spectrum is therefore a sparse comb compared to an open-pipe spectrum, with only every other tooth present. That fact becomes the whole point of the HOT example a few paragraphs from now, in which the second overtone of a closed pipe is matched against the first overtone of an open pipe.

End correction: where the antinode really sits

All of the analysis so far has quietly assumed that the displacement antinode at the open end of a pipe lies exactly at the geometric end of the pipe. In a real pipe it does not. The air at the open end is free to move not only inside the pipe but also into the outside air just beyond the pipe, and so the antinode actually sits a small distance outside the geometric end. The correction is small but not zero, and any laboratory measurement of v using the resonance tube must account for it.

Experiment shows that the antinode of an open end of a pipe of radius r sits, to a good approximation, a distance

$$e \approx 0.6r$$

outside the geometric end. The effective acoustic length of the pipe is therefore longer than its physical length by e at each open end. For a closed pipe, with only one open end, the effective length is:

$$L_{\text{eff}} = L + 0.6r$$

and for an open pipe, with two open ends, the effective length is:

$$L_{\text{eff}} = L + 2(0.6r) = L + 1.2r$$

All the harmonic formulas written down above remain correct, provided L is replaced by the appropriate L_{eff} in each case. A clever way to bypass any uncertainty about e in a laboratory experiment is to determine v from two consecutive resonance lengths rather than one. The fundamental of a closed resonance tube satisfies $L_1 + e = \lambda/4$, and the very next resonance, the first overtone, satisfies $L_2 + e = 3\lambda/4$. Subtracting the first equation from the second makes e vanish entirely:

$$L_2 - L_1 = \frac{\lambda}{2}, \quad \text{so} \quad \lambda = 2(L_2 - L_1), \quad \text{and} \quad v = f\lambda = 2f(L_2 - L_1)$$

Knowing the tuning-fork frequency f and the two resonance lengths L_1 and L_2 , the experimenter computes v without ever having to know e . If the end correction is wanted afterwards, it is recovered by back-substituting into either resonance condition. This trick is what makes the resonance tube the canonical school-laboratory method for measuring the speed of sound in air.

Mistuned pipes: when two notes that should agree disagree

If two pipes that ought to sound the same note are sounded together but their lengths are slightly off, their fundamentals are slightly different and the listener hears beats. The beat frequency is the absolute difference of the two pipe frequencies, $f_b = |f_1 - f_2|$, exactly the formula Chapter 7 derived for two sinusoidal sources of nearly equal frequency. We use it here without re-deriving it and refer the reader who needs a refresher to the earlier chapter. A piano tuner counting beats against a tuning fork at Bagamoyo College of Arts is using the same arithmetic.

We now have five tools at our disposal: the closed-pipe spectrum, the open-pipe spectrum, the octave relationship between same-length pipes, the end-correction shortcut, and the beat formula for mistuned pipes. The worked examples that follow are five carefully chosen pieces of physics theatre designed to put each tool through its paces. Open your notebook; the pipes are about to start singing.

BINDER Example 6

A closed organ pipe is 0.68m long. Taking the speed of sound in air as 340m/s and neglecting end correction, find the wavelengths and frequencies of the three lowest modes of vibration.

Solution

For a closed pipe the spectrum is $f_n = (2n - 1)v/(4L)$, with $n = 1, 2, 3$, and so on. We work out each mode in turn.

Fundamental ($n = 1$):

$$\lambda_1 = 4L = 4 \times 0.68\text{m} = 2.72\text{m}$$

$$f_1 = \frac{v}{4L} = \frac{340\text{m/s}}{4 \times 0.68\text{m}} = 125\text{Hz}$$

First overtone ($n = 2$, third harmonic):

$$\lambda_2 = \frac{4L}{3} = \frac{2.72\text{m}}{3} = 0.907\text{m}$$

$$f_2 = 3f_1 = 3 \times 125\text{Hz} = 375\text{Hz}$$

Second overtone ($n = 3$, fifth harmonic):

$$\lambda_3 = \frac{4L}{5} = \frac{2.72\text{m}}{5} = 0.544\text{m}$$

$$f_3 = 5f_1 = 5 \times 125\text{Hz} = 625\text{Hz}$$

The three lowest modes have wavelengths of 2.72m, 0.907m, and 0.544m, and frequencies of 125Hz, 375Hz, and 625Hz.

Making sense of the answer: *The three frequencies sit in the ratio 1 : 3 : 5, the unmistakable fingerprint of a closed pipe. The pipe itself is only about the height of a school desk, yet it sounds as low as an open pipe more than 1.3m tall would. This is the closed-pipe bonus: by sealing one end, the builder doubles the effective acoustic length and saves half the brass.*

Think like a physicist: *When a closed-pipe problem turns up, write down the ratio 1 : 3 : 5 : 7 : ... before you touch your calculator. Half of every closed-pipe question on an examination paper can be answered from the ratio alone, with no $v/(4L)$ ever computed. The 1 : 3 : 5 ratio is what distinguishes the closed-pipe spectrum from every other harmonic spectrum in physics, and recognising it on sight is worth a free mark in the right exam.*

BINDER Example 7

A closed organ pipe and an open organ pipe are sounded simultaneously, producing five beats per second. The open pipe is 30cm long. Find the change of length the closed pipe must undergo to bring the two into unison. Take the speed of sound in air as 340m/s and neglect end corrections.

Solution

Beats arise because the two fundamentals are close but not equal. The fundamental of the open pipe is fixed at:

$$f_o = \frac{v}{2L_o} = \frac{340\text{m/s}}{2 \times 0.30\text{m}} = 567\text{Hz}$$

The fundamental of the closed pipe must differ from this by 5Hz, so it is either 562Hz or 572Hz. The question does not say which, so we work out both possibilities.

Case 1: the closed pipe sounds 562Hz initially.

$$L_1 = \frac{v}{4f_1} = \frac{340\text{m/s}}{4 \times 562\text{Hz}} = 0.1512\text{m}$$

To bring the closed pipe to 567Hz, its length must be shortened to:

$$L_0 = \frac{v}{4 \times 567\text{Hz}} = \frac{340\text{m/s}}{4 \times 567\text{Hz}} = 0.1499\text{m}$$

The required change of length is:

$$\Delta L = L_1 - L_0 = 0.1512\text{m} - 0.1499\text{m} = 0.0013\text{m} \approx 1.3\text{mm shorter}$$

Case 2: the closed pipe sounds 572Hz initially.

$$L_1 = \frac{v}{4 \times 572\text{Hz}} = \frac{340\text{m/s}}{4 \times 572\text{Hz}} = 0.1486\text{m}$$

To bring the closed pipe to 567Hz, its length must be lengthened to $L_0 = 0.1499\text{m}$:

$$\Delta L = L_0 - L_1 = 0.1499\text{m} - 0.1486\text{m} = 0.0013\text{m} \approx 1.3\text{mm longer}$$

Either way the closed pipe needs adjustment by about 1.3mm.

Hence the change of length is 1.3mm.

Understand: Only the direction (lengthen or shorten) depends on whether the closed pipe was originally too sharp or too flat. To settle which case applies in practice, the player perturbs the closed pipe by a known small amount and listens to the beats. If the beat rate rises, the perturbation is moving away from unison and we are working the wrong way. If the beat rate falls, the perturbation is moving toward unison and we are working the right way.

Alternative solution

The two case-by-case computations both gave $|\Delta L| = 1.3\text{mm}$, hinting that the magnitude of the answer depends only on the magnitude of the beat frequency, not on the direction of the mistuning. This is exactly right and can be obtained in a single step.

For a pipe in any mode the fundamental frequency is inversely proportional to the length, $f = \text{constant} \left(\frac{1}{L}\right)$. Taking the differential of both sides and dividing through, the fractional change in frequency equals the fractional change in length, in magnitude:

$$\frac{|\Delta f|}{f} = \frac{|\Delta L|}{L}$$

The closed pipe, when brought to unison with the open pipe, has length

$$L = \frac{v}{4f} = \frac{340\text{m/s}}{4 \times 567\text{Hz}} = 0.1499\text{m}$$

And the required change of length therefore has magnitude:

$$|\Delta L| = L \times \frac{|\Delta f|}{f} = 0.1499\text{m} \times \frac{5\text{Hz}}{567\text{Hz}} = 1.32 \times 10^{-3}\text{m} \approx 1.3\text{mm}$$

The answer drops out in a single line, with no case branching. The branching is hidden in the absolute-value sign on Δf , which says that the same number of beats per second can arise from a closed pipe that is sharp or from one that is flat; the magnitude of the correction is the same, only its direction differs. To settle the direction in practice, perturb the closed pipe and listen, as in the case-by-case solution above.

Making sense of the answer: Five hertz out of nearly six hundred is less than one per cent of the pitch, and 1.3mm out of nearly 150mm is also just under one per cent of the length. The two percentages agree because for a pipe in any mode the frequency is inversely proportional to the length, so $|\Delta f/f|$ equals $|\Delta L/L|$. The musician adjusting the tuning slide of a trombone is paying his rent on exactly this proportionality.

Think like a physicist: The absolute-value sign in the beat formula $f_b = |f_1 - f_2|$ hides important information: it does not tell the player whether his pipe is sharp or flat. To resolve the sign, perturb one of the two pipes in a known direction and listen to whether the beats speed up or slow down. This trick is the foundation of every piano-tuning method ever invented.

REAL Example 8

A school marimba at Mwakaleli Secondary School in Mbeya is built from a row of closed tubes of varying length. One particular tube is designed to sound the note one octave above middle C, of frequency 523Hz.

- Compute the physical length of the tube required at a Mwakaleli rehearsal-room temperature of 25°C, neglecting end correction. Take the speed of sound at 25°C to be 346m/s.
- The marimba is tuned in Mwakaleli in January at 32°C, where the speed of sound is 350m/s. The same instrument is taken to Iringa in July, where the rehearsal hall sits at 15°C and the speed of sound is 340m/s. By what percentage does the pitch of the tube change, and is the resulting note sharp or flat? A physics teacher warns the class against the seductive but wrong reason.

Solution

(a) For a closed pipe at its fundamental, $L = v/(4f)$. At 25°C:

$$L = \frac{346\text{m/s}}{4 \times 523\text{Hz}} = 0.165\text{m} = 16.5\text{cm}$$

(b) Tuned at 32°C, the same physical length L gives a fundamental $f_m = 350/(4 \times 0.165) = 530\text{Hz}$. Carried to a cold Iringa hall at 15°C, with v dropping to 340m/s, the same length L gives:

$$f_i = \frac{340\text{m/s}}{4 \times 0.165\text{m}} = 515\text{Hz}$$

The pitch falls from 530Hz to 515Hz.

$$\text{Percentage drop} = \frac{f_m - f_i}{f_m} \times 100\% = \frac{530\text{Hz} - 515\text{Hz}}{530\text{Hz}} \times 100\% = 2.83\%$$

The percentage drop is 2.83%.

The note is **flat** in the cold Iringa hall.

Now the physics teacher intervenes.

Student: “Sir, this makes perfect sense. In cold weather the brass tube contracts. A shorter tube gives a higher fundamental from $f = v/(4L)$, so the pitch should rise. But your calculation says the pitch falls. Have we made a sign error somewhere?”

Physics teacher: “Your reasoning is half right. The tube does contract in the cold. But before you cash in on that argument, work out how much. How much does a 16.5cm brass tube actually shorten between 32°C and 15°C?”

Student: “The linear expansivity of brass is about 1.9×10^{-5} per kelvin, from Chapter 2. The temperature drop is 17K. So $\Delta L/L$ is $1.9 \times 10^{-5} \times 17$, which is 3.2×10^{-4} , or about 0.03 per cent of the length.”

Mr. Physics teacher: “Good. And the speed of sound? From the earlier section of this chapter it scales as the square root of absolute temperature. From 305K down to 288K, what does v do?”

Student: “ v drops by the square root of the ratio. $\sqrt{288/305}$ is about 0.9717, so v falls by about 2.83 per cent.”

Physics teacher: “Two effects are pushing on the pitch. v dropping pushes f down directly through $f = v/(4L)$. L contracting pushes f up through the same formula. The first effect weighs in at 2.83 per cent. The second effect weighs in at 0.03 per cent. The first effect is one hundred times bigger than the second. The marimba flattens in the Iringa hall because the speed of sound has dropped, not because the tube has shrunk. Always size up the dominant effect numerically before you let your intuition collect its prize.”

So the marimba arrives flat in Iringa by 2.83 per cent of its design pitch, with v doing almost all of the work. To restore the design pitch, the carpenter would need to trim each tube very slightly shorter to compensate; brass contraction by itself would not save the situation.

Making sense of the answer: The same physics is the reason a wind instrument played in a cold lecture room sounds slightly flat after the air conditioning has been running, and the reason a band tunes up only after the players have warmed their instruments to room temperature. The lazy intuition that says “things shrink in cold weather, so the pitch goes up” is wrong for almost every wind and brass instrument made; the speed-of-sound effect dominates by orders of magnitude every single time.

Think like a physicist: Whenever two effects act on the same quantity, whether they are pulling in the same direction or, as here, in opposite directions, estimate each one numerically before concluding anything from intuition.

REAL Example 9

A Form Five student at Kibaha Secondary School performs the resonance-tube experiment to measure the speed of sound in air. He sounds a tuning fork of frequency 480Hz above an air column whose length is varied by raising and lowering a water reservoir attached by a rubber hose. The first loud resonance is found when the air column is 17.2cm long. He raises the water level no further and instead lowers it to find the next resonance, which occurs at an air-column length of 53.4cm. The internal radius of the glass tube is 1.6cm.

(a) Compute the wavelength of the sound wave in the air column.

(b) Compute the speed of sound in the laboratory air on the day of the experiment.

(c) Compute the end correction, and compare it with the value $0.6r$ predicted from the radius of the tube.

Solution

(a) The first loud resonance of a closed pipe occurs when the air column is one quarter of a wavelength long (ignoring the end correction for a moment), and the next loud resonance occurs when the air column is three quarters of a wavelength long. The two true resonance conditions, including the end correction e at the open top of the tube, are:

$$L_1 + e = \frac{\lambda}{4}, \quad L_2 + e = \frac{3\lambda}{4}$$

Subtracting the first equation from the second wipes the e out cleanly:

$$L_2 - L_1 = \frac{\lambda}{2}$$

So the wavelength is

$$\lambda = 2(L_2 - L_1) = 2 \times (53.4\text{cm} - 17.2\text{cm}) = 72.4\text{cm} = 0.724\text{m}$$

(b) The speed of sound follows from $v = f\lambda$:

$$v = f\lambda = 480\text{Hz} \times 0.724\text{m} = 347.5\text{m/s}$$

(c) Having found λ , the end correction is recovered from either of the two resonance conditions. Using the first:

$$e = \frac{\lambda}{4} - L_1 = \frac{0.724\text{m}}{4} - 0.172\text{m} = 0.009\text{m} = 0.9\text{cm}$$

The $0.6r$ rule predicts:

$$e_{\text{predicted}} = 0.6 \times r = 0.6 \times 1.6\text{cm} = 0.96\text{cm}$$

The experimental and predicted values agree to within about six per cent, well inside the precision of a school laboratory.

Making sense of the answer: *The speed of sound came out at 347.5m/s, comfortably inside the band a Tanzanian classroom should expect, where a temperature around 27°C gives a sound speed close to 347m/s. The measurement would not have been possible at all without the end-correction trick: if the student had used only the first resonance and assumed $L_1 = \lambda/4$, he would have computed $\lambda = 0.688\text{m}$, $v = 330.2\text{m/s}$, and would have been low by about five per cent. Subtracting two resonance lengths to eliminate e buys back exactly that error.*

Think like a physicist: *Whenever an experiment contains a small unknown systematic offset, look for a way to cancel it by taking two measurements and subtracting one from the other. The two-resonance-lengths trick here is a special case of a much more general experimental technique: never measure an absolute quantity if a difference of two absolute quantities will do the same job. A Form Five physics student who internalises this rule once will use it for the rest of his life in every laboratory he ever sets foot in..*

HOT Example 10

A closed organ pipe and an open organ pipe are arranged so that the second overtone of the closed pipe sounds at the same frequency as the first overtone of the open pipe. Find the ratio of the length of the closed pipe to the length of the open pipe. Given that the common frequency is 880Hz and that the speed of sound in air is 340m/s, find each length. Neglect end corrections.

Solution

Write each frequency in turn. The closed-pipe spectrum is $f_n = (2n - 1)v/(4L_c)$, with $n = 3$ for the second overtone, which is the fifth harmonic:

$$f_{\text{2nd overtone, closed}} = \frac{5v}{4L_c}$$

The open-pipe spectrum is $f_n = nv/(2L_o)$, with $n = 2$ for the first overtone, which is the second harmonic:

$$f_{\text{1st overtone, open}} = \frac{2v}{2L_o} = \frac{v}{L_o}$$

Setting the two equal:

$$\frac{5v}{4L_c} = \frac{v}{L_o}$$

$$\frac{L_c}{L_o} = \frac{5}{4}$$

The closed pipe is therefore five-quarters as long as the open pipe. With a common frequency of 880Hz, the open pipe satisfies:

$$L_o = \frac{v}{f} = \frac{340\text{m/s}}{880\text{Hz}} = 0.386\text{m} = 38.6\text{cm}$$

and the closed pipe satisfies:

$$L_c = \frac{5}{4}L_o = 1.25 \times 0.386\text{m} = 0.483\text{m} = 48.3\text{cm}$$

The two pipes have lengths 38.6cm (open) and 48.3cm (closed).

Making sense of the answer: *Half the trick in this question lies in matching the right overtone of each pipe to the right multiple of the fundamental. The second overtone of a closed pipe is the fifth harmonic, not the third, because the closed-pipe spectrum starts at the first harmonic and then skips every even multiple. The first overtone of an open pipe is the second harmonic, because the open-pipe spectrum starts at the first harmonic and then proceeds through every consecutive integer multiple. Mixing up the overtone-versus-harmonic vocabulary is the most common error on this kind of question, and the examiner who set it knew exactly that.*

Think like a physicist: *When a pipe problem mixes a closed pipe and an open pipe, write the two spectra side by side before substituting a single number: closed pipe $f_n = (2n - 1)f_1(\text{closed})$ for $n = 1, 2, 3, \dots$, and open pipe $f_n = n f_1(\text{open})$ for $n = 1, 2, 3, \dots$. The substitutions then become mechanical and almost no algebra is required. The physics is hidden in the boundary conditions the two spectra encode, not in the arithmetic that follows.*

We now have a fairly complete theory of sound. We can compute its speed at any temperature in any gas, predict the harmonic spectrum of any pipe of any length, account for the small end correction at every open end, and read off the beat frequency between any two mistuned pipes that share a stage. But we have not yet asked one question, a question the human ear answers automatically every waking second: *how loud is loud?* The next section introduces the logarithmic scale that engineers, physiologists, and examiners all use for that purpose.

LOUDNESS, INTENSITY, AND THE DECIBEL SCALE

The earlier section of this chapter on the nature of sound listed three things the ear hears: pitch, loudness, and quality. It also paid two of those debts immediately. Pitch was identified with frequency. Quality was identified with the harmonic spectrum. Loudness was given the word “intensity” and then quietly postponed, on the grounds that intensity has its own mathematics and that the mathematics deserves a section of its own. The promise comes due now. This section answers, once and for all, the question the ear answers every waking second without thinking: *how loud is loud?*

The question turns out to be harder than it looks. The human ear can detect a sound as soft as a mosquito flying three metres away in a quiet room and can tolerate, with discomfort but without immediate injury, a sound as loud as a jackhammer one metre away. Measured in watts per square metre, the gap between these two sounds is a factor of a million million. Writing the loud sound as one million million times the soft sound is unwieldy on a page. Asking a Form Five student to feel comfortable with such a number is unreasonable. And the human nervous system itself, conveniently, does not perceive sound on a linear scale; it perceives sound approximately logarithmically, so that successive equal doublings of intensity feel like successive equal steps of loudness.

Engineers, physiologists, and the writers of every reasonable physics syllabus in the world have therefore agreed on a single logarithmic scale for sound: the **decibel**. This section introduces it, derives the two pieces of arithmetic the student must drill until they need no thought, applies it to a point source whose listener walks away from it, and applies it to a roomful of incoherent sources whose intensities add even though their levels do not. Five or six pages from now you will hold the entire decibel toolkit.

Intensity: the physical quantity behind the perception

Chapter 7 introduced the intensity of any wave as the power the wave carries across a unit area perpendicular to its direction of travel, measured in watts per square metre. The same chapter showed two further things that we will use heavily here and will not re-derive. First, the intensity of a sinusoidal wave is proportional to the square of its amplitude, $I \propto A^2$. Second, the intensity of a wave radiating outwards from a small source in all directions falls off as the inverse square of the distance from the source,

$$I = \frac{P}{4\pi r^2}$$

where P is the total power radiated and r is the distance from the source to the listener. Both results are taken as given. The reader who needs a refresher should glance back at Chapter 7 before proceeding; everything in this section sits on top of those two facts.

The softest sound a healthy young human ear can detect at the frequency 1kHz is an intensity of about

$$I_0 = 10^{-12} \text{ W/m}^2$$

This number is called the **threshold of hearing** and is the reference point against which every other sound intensity is to be compared. It is not a fundamental constant of physics; it is a fact about the human ear, measured by carefully controlled experiments on hearing-test volunteers. A different species, a younger ear, a different frequency, or a slightly damaged eardrum would give a slightly different threshold. But the convention of physics is fixed, and the convention is the one above.

The loudest sound the human body tolerates briefly, without immediate injury, is about

$$I_{\text{pain}} \approx 1 \text{ W/m}^2$$

This number is called the **threshold of pain**. Above it, exposure for even a few seconds risks permanent hearing loss. Between the threshold of hearing and the threshold of pain the ear must work over a span of

$$\frac{I_{\text{pain}}}{I_0} = \frac{1}{10^{-12}} = 10^{12},$$

twelve full orders of magnitude. No linear scale can handle that span on a single page without absurdity. The next subsection explains how the logarithmic scale handles it gracefully.

Why a logarithm? The twelve-decade problem

If a single number must describe sounds that vary by a factor of 10^{12} , the number had better be a logarithm. Logarithms convert multiplication into addition and convert wide ratios into narrow numbers. The base-10 logarithm of 10^{12} is exactly 12, a simple value that is easy to handle. The base-10 logarithm of 2 is about 0.301, and the base-10 logarithm of 10 is exactly 1. These three facts, $\log 2 \approx 0.301$, $\log 10 = 1$, and the general identity $\log(ab) = \log a + \log b$, between them, are all the logarithm arithmetic this section will use.

The logarithmic measure of sound intensity goes by the name of Alexander Graham Bell, the inventor of the telephone, who first proposed it for use in long-distance telephone engineering. A factor of ten in intensity is called one bel. The bel itself turns out to be a slightly inconvenient unit, because the practical range of human hearing is only twelve bels wide. A unit one tenth as large gives a more comfortable working range, from 0 to 120, and the unit one tenth as large is called the **decibel**, abbreviated **dB**. The decibel is the working unit of loudness throughout the physics, engineering, and physiology of sound.

The decibel formula

The intensity level L of a sound of intensity I is defined by

$$L = 10 \log_{10} \left(\frac{I}{I_0} \right) \quad \text{decibels}$$

where $I_0 = 10^{-12} \text{ W/m}^2$ is the threshold of hearing introduced above. Several things about this definition deserve immediate comment.

First, L is *dimensionless*. The argument of the logarithm is a ratio of two intensities and so is itself a pure number. The factor of 10 in front is a pure number. The unit “decibel” is an honorary tag, not a physical dimension, and a decibel value cannot be converted to a number of joules or watts the way a kilogram can be converted to a number of newtons. A decibel is a label, not a quantity, in the dimensional sense.

Second, the threshold of hearing sits at $L = 0$ exactly, because $\log(I_0/I_0) = \log 1 = 0$. A sound at the threshold of hearing is, by definition, at 0 decibels. A quieter sound has a negative decibel level, which is a perfectly valid number, though for the human ear at 1kHz it represents a sound that nobody can hear.

Third, the threshold of pain sits near $L = 120$, because $\log(1/10^{-12}) = \log 10^{12} = 12$, and $10 \times 12 = 120$. A sound at the threshold of pain is about a million million (one trillion) times more intense than a sound at the threshold of hearing, and the dB scale lets a Form Five student write this comparison as “120dB” without flinching.

Fourth, the dB scale is not an absolute measure of loudness; it is a measure of intensity relative to the threshold of hearing. A different reference would give different numerical levels for the same physical sounds. In acoustics, in electrical engineering, in optics, and in radio engineering, several other references are in use for different purposes, all of them written with the same formula but with a different I_0 . The reference for sound in air is the one above, and is the only reference this chapter uses.

Two patterns to drill into your fingers

Two arithmetic patterns recur in dB problems often enough that the practised student does them mentally, with no calculator, and the unpractised student loses minutes to them on every examination.

The first pattern is doubling. When the intensity doubles, $I \rightarrow 2I$, the dB level increases by

$$\Delta L = 10 \log \left(\frac{2I}{I} \right) = 10 \log 2 \approx 10 \times 0.301 = 3.01 \text{ dB} \approx 3 \text{ dB}$$

So doubling intensity adds three decibels, and halving intensity subtracts three decibels. This is the most useful single fact in dB arithmetic. A practising audio engineer in a recording studio at Mlimani City, switching from one speaker to two identical speakers, knows in advance that the listening room will get exactly 3dB louder, no measurement required.

The second pattern is ten-fold. When the intensity multiplies by ten, $I \rightarrow 10I$, the dB level increases by

$$\Delta L = 10 \log \left(\frac{10I}{I} \right) = 10 \log 10 = 10 \times 1 = 10 \text{ dB}$$

So a ten-fold intensity adds ten decibels, and a hundred-fold intensity adds twenty decibels, and a thousand-fold intensity adds thirty decibels, and so on, with one extra factor of 10 in intensity worth ten extra dB on the level scale every time.

Combining the two patterns lets the student handle every common intensity ratio without using a calculator. An intensity ratio of 4 is two doublings, worth $2 \times 3 = 6 \text{ dB}$. An intensity ratio of 8 is three doublings, worth $3 \times 3 = 9 \text{ dB}$. An intensity ratio of 20 is a ten-fold (10dB) followed by a doubling (3dB), worth 13dB. An intensity ratio of 100 is two ten-folds, worth 20dB. An intensity ratio of 1000 is three ten-folds, worth 30dB. The student who has these conversions in his fingers will read most dB problems straight off the page.

Drop with distance: the six-decibel rule

A point source radiates power P in all directions equally. From the inverse-square law,

$$I = \frac{P}{4\pi r^2}$$

The intensity at distance r falls off as $1/r^2$. The dB level at distance r relative to the level at some reference distance r_1 is therefore:

$$L_2 - L_1 = 10 \log \left(\frac{I_2}{I_1} \right) = 10 \log \left(\frac{r_1^2}{r_2^2} \right) = 20 \log \left(\frac{r_1}{r_2} \right)$$

If $r_2 > r_1$ then $r_1/r_2 < 1$, the logarithm is negative, and the level drops as expected. If $r_2 < r_1$ the listener has moved closer and the level rises. Either way, the dB change with distance is given by:

$$\Delta L = L_2 - L_1 = 20 \log \left(\frac{r_1}{r_2} \right) = -20 \log \left(\frac{r_2}{r_1} \right),$$

with the sign of the change determined by which way the listener walks.

The single most useful consequence of this formula is the six-decibel rule for point sources. Doubling the distance from a point source, $r_2 = 2r_1$, drops the level by

$$|\Delta L| = 20 \log 2 \approx 20 \times 0.301 = 6.02 \text{ dB} \approx 6 \text{ dB}$$

A pedestrian walking along a quiet street, moving from 10 metres away from a loudspeaker to 20 metres away, will experience a drop in sound level of about six decibels, even though the source of the sound remains unchanged.

The six-decibel rule applies only to idealised point sources in open air with no reflections. A real urban street is full of reflections from buildings on either side, and the rule overestimates how quickly the sound dies with distance. A real concert hall is designed to keep reflections alive so that the audience at the back of the hall hears almost the same loudness as the audience near the front. Reverberation and reflection are a topic for an applied acoustics chapter that this book will not write; for our purposes, the six-decibel rule is a good first approximation and an excellent examination tool.

Combining incoherent sources: intensities add, levels do not

If two sound sources are sounded together and are incoherent, meaning that their phases bear no fixed relationship to each other, their intensities add. This is one of the basic results of Chapter 7 on interference and superposition. We use it here without re-deriving it.

For two incoherent sources of intensities I_1 and I_2 at the listener's ear, the total intensity is:

$$I_{\text{total}} = I_1 + I_2$$

The total level L_{total} is then computed from this total intensity through the usual formula. The levels themselves do not add. A common student error is to say, of two simultaneous 50dB sources, that the combined level is 100dB. The combined level is in fact 53dB. Here is why: two equal-intensity sources have a combined intensity of $2I$, which is a doubling, which adds 3dB to the level. Two equal sources at 50dB combine to a single level of 53dB.

More generally, n equal-intensity sources combine to a total intensity of nI and a total level higher by:

$$\Delta L = 10 \log n$$

decibels. So three identical sources add $10 \log 3 \approx 4.8$ dB. Ten identical sources add exactly 10dB. A hundred identical sources add exactly 20dB. The rule scales smoothly with n and applies in every situation where the sources are incoherent, which in practice is every situation outside of carefully arranged laboratory interference experiments.

Equal-loudness contours and the dBA tag

A complication the dB scale does not capture is that the ear is not equally sensitive to all frequencies. At the same physical intensity, a 1kHz tone sounds substantially louder than a 30Hz tone, because the ear is most sensitive in the speech-frequency range from about 1 to 4kHz and rolls off in sensitivity at both lower and higher frequencies. The standard plot of equal perceived loudness versus frequency and intensity, called the **equal-loudness contour**, looks like a stack of U-shaped curves, with each curve carrying its own loudness label. The Form Five student is not required to draw or work with these curves but should know they exist.

To handle this frequency dependence approximately in noise-control engineering, a frequency-weighted version of the decibel scale is used. It is called the A-weighted decibel and is denoted dBA. The dBA value of a sound is its physical dB value, adjusted by the standardised A-weighting filter, so that a 60dBA reading corresponds to a sound that a typical listener would perceive as if it were 60dB at 1kHz. Industrial hearing-protection rules, city noise ordinances, and most legal noise limits are written in dBA, not in dB. A typical Tanzanian municipal council writes residential noise limits as “60dBA during the day, 45dBA at night.” For the purposes of this book, the dBA is mentioned only so that students are not surprised by it when they encounter it.

A reference table for the loud and the soft

The table below anchors the dB scale to sounds the student will recognise. Memorising the table is not necessary; recognising at a glance whether a quoted level is realistic is.

Sound	Approximate intensity (W/m^2)	Level (dB)
Threshold of hearing (1kHz, healthy young ear)	10^{-12}	0
Rustle of leaves at 5m	10^{-10}	20
Quiet conversation at 1m	10^{-7}	50
Daladala at full speed, passenger view	10^{-4}	80
Construction site at 5m	10^{-2}	100
Threshold of pain	1	120
Saba-Saba exhibition loudspeaker stack at 1m	10^2	140 (instant injury)

A passenger on a daladala in afternoon traffic on Morogoro Road is sitting in a sound field of about 80dB. A conversation across the aisle takes place against a background that is one hundred million times more intense than what the same passenger would call quiet, yet by the logarithmic standard of the dB scale it is “only” 80dB. The scale’s compression is what makes it usable at all.

We now have five pieces of equipment in the dB toolkit. The worked examples that follow are three carefully chosen pieces of physics theatre designed to exercise each of those pieces in turn, with the third example landing on a surprise that catches more students by misjudgement than any of the previous five worked examples in this chapter combined.

BINDER Example 11

The intensity of a sound at a particular point in a Mwenge auditorium is $5.0 \times 10^{-5} \text{ W/m}^2$. Find the intensity level in decibels. Take the threshold of hearing as $I_0 = 10^{-12} \text{ W/m}^2$.

Solution

Apply the decibel formula directly:

$$L = 10 \log \left(\frac{I}{I_0} \right) = 10 \log \left(\frac{5.0 \times 10^{-5} \text{ W/m}^2}{10^{-12} \text{ W/m}^2} \right) = 76.99 \text{ dB} \approx 77 \text{ dB}$$

The intensity level is therefore about 77dB, in the same neighbourhood as a daladala or a noisy classroom.

Making sense of the answer: The student who has internalised the two arithmetic patterns can verify this answer without ever using a calculator. The intensity ratio 5.0×10^7 is half of 10^8 , which would be 80dB exactly. Half of an intensity is a loss of 3dB. So 5.0×10^7 is $80 - 3 = 77 \text{ dB}$. Two patterns and one subtraction, no logarithm key pressed.

Think like a physicist: Always estimate before computing. A 77dB answer is reasonable for the inside of a busy room; a 7dB or a 770dB answer would be obviously wrong. The simple check “where does this answer lie between 0 and 120?” catches almost every such error before any marks are lost.

REAL Example 12

A road construction site near Magomeni produces a sound level of 95dB at a distance of 10m from its centre. Tanzanian occupational safety guidelines require workers to wear hearing protection at any location where the sound level reaches or exceeds 85dB.

- Find the minimum distance from the centre of the site at which a worker without hearing protection just complies with the rule, assuming the site behaves as a point source in open air.
- Show that doubling the distance from a point source drops the level by approximately 6dB.

Solution

(a) The dB level at distance r_2 relative to the level at a reference distance r_1 follows the inverse-square law applied to dB. The change in level is

$$\Delta L = L_2 - L_1 = -20 \log \left(\frac{r_2}{r_1} \right)$$

Substitute $L_1 = 95$ dB at $r_1 = 10$ m, and require $L_2 = 85$ dB at the unknown r_2 :

$$\Delta L = 85 \text{ dB} - 95 \text{ dB} = -10 \text{ dB}$$

Hence:

$$\begin{aligned} -10 \text{ dB} &= -20 \log \left(\frac{r_2}{10 \text{ m}} \right) \\ r_2 &= 131.62 \text{ m} \end{aligned}$$

A worker without hearing protection must stand at least 31.6m from the centre of the site to comply with the rule.

(b) Doubling the distance means $r_2 = 2r_1$. Substitute into the formula:

$$\Delta L = -20 \log(2) = -20 \times 0.301 = -6.02 \text{ dB} \approx -6 \text{ dB}$$

Hence doubling the distance drops the level by approximately 6 decibels. (This is the six-decibel rule for point sources.)

Making sense of the answer: Bringing the level down by 10dB requires the ratio of intensities to drop by a factor of 10, which by the inverse-square law requires the distance to grow by a factor of $\sqrt{10} \approx 3.16$. The factor of $\sqrt{10}$ arises because intensity goes as $1/r^2$, so a ten-fold reduction in intensity is a $\sqrt{10}$ -fold growth in distance. The numerical answer of 31.6m is reasonable for a road construction site; it agrees with the rule of thumb that safety engineers in noisy industries use mentally without recourse to a calculator.

Think like a physicist: The six-decibel rule is not a coincidence of arithmetic; it is the inverse-square law dressed up in logarithmic clothing. Whenever a question about a point source mentions doubling, halving, or a factor of any power of two in either distance or intensity, the answer will involve an integer multiple of 3dB (for intensities) or an integer multiple of 6dB (for distances). Reaching for these mental shortcuts before reaching for the calculator is the difference between a student who finishes the dB question with time to spare and a student who runs out of paper at the end.

HOT Example 13

A primary-school classroom at Mawenzi Primary School in Moshi has an average background sound level of 35dB when the children are quiet. With 36 children present and chattering, the average level rises to 65dB. Treating each child as an independent, incoherent source of equal average intensity, and ignoring the small contribution of the background sound to the chattering total,

- estimate the average sound intensity contributed by one child, in W/m^2 , averaged over the room;
- predict the sound level the teacher would measure if the class size doubled to 72 children, with each child contributing the same intensity as before.

Solution

(a) The two given dB levels must be converted back to intensities. The decibel formula $L = 10 \log(I/I_0)$ can be solved for I by taking the inverse logarithm of both sides.

$$\begin{aligned} L &= 10 \log \left(\frac{I}{I_0} \right) \\ I &= I_0 \log^{-1} \left(\frac{L}{10} \right) \end{aligned}$$

This is the inverse-log form of the decibel formula and is the form a student must reach for whenever a dB level is given and an intensity is wanted. Apply it to the background sound at $L = 35$ dB:

$$I_{\text{background}} = 10^{-12} \text{ W/m}^2 \log^{-1} \left(\frac{35}{10} \right) = 3.16 \times 10^{-9} \text{ W/m}^2$$

Apply it again to the total sound at $L = 65 \text{ dB}$:

$$I_{\text{total}} = 10^{-12} \text{ W/m}^2 \log^{-1} \left(\frac{65}{10} \right) = 3.16 \times 10^{-6} \text{ W/m}^2$$

The intensity coming from the children alone is the total intensity minus the background intensity:

$$I_{\text{children}} = I_{\text{total}} - I_{\text{background}} = 3.16 \times 10^{-6} \text{ W/m}^2 - 3.16 \times 10^{-9} \text{ W/m}^2$$

The background contributes one part in a thousand to the total and can safely be neglected:

$$I_{\text{children}} \approx 3.16 \times 10^{-6} \text{ W/m}^2$$

Divide by the number of children:

$$I_{\text{per child}} = \frac{I_{\text{children}}}{36} = \frac{3.16 \times 10^{-6} \text{ W/m}^2}{36} = 8.78 \times 10^{-8} \text{ W/m}^2$$

Each child contributes, on average, about $8.78 \times 10^{-8} \text{ W/m}^2$ of intensity to the room.

(b) With 72 (2×36) children, each contributing the same average intensity as before, the combined intensity from the children alone doubles:

$$I_{\text{children, new}} = 2 \times I_{\text{children}} = 2 \times 3.16 \times 10^{-6} \text{ W/m}^2 = 6.32 \times 10^{-6} \text{ W/m}^2$$

Now go back through the forward decibel formula to find the corresponding level. Ignoring the negligible background,

$$L_{\text{total, new}} = 10 \log \left(\frac{I_{\text{children, new}}}{I_0} \right) = 10 \log \left(\frac{6.32 \times 10^{-6} \text{ W/m}^2}{10^{-12} \text{ W/m}^2} \right) = 68.01 \text{ dB} \approx 68 \text{ dB}$$

Doubling the class size raises the level by only 3dB, from 65dB to 68dB.

Mr. Akilikubwa walks in at this moment and asks the class, with a face he uses when he expects somebody to be surprised.

Mr. Akilikubwa: “Kipanga, are you content with this answer?”

Kipanga: “Sir, I am not. We have doubled the children. Why has the level risen by only 3dB? Surely twice as many children should be twice as loud.”

Mr. Akilikubwa: “And what does ‘twice as loud’ mean to the ear, Kipanga? On the dB scale, twice the intensity adds 3dB. We doubled the intensity, so we added 3dB. The ear, on the other hand, does not perceive 3dB as a doubling of loudness; it perceives a doubling of loudness as roughly 10dB, which requires not a doubling of children but a tenfold multiplication of children. To make the classroom sound twice as loud to the teacher, you would need not 72 children but about 360 of them.”

Kipute: “So the head teacher who complains that the school is twice as noisy as last year is probably wrong by a factor of about five.”

Mr. Akilikubwa: “She is probably wrong by a factor of about five, yes, but only because she is using her ear, which is logarithmic, while her impression speaks in linear language. The dB scale is closer to what the ear actually feels; the linear count of children is closer to what the eye actually counts. The two scales disagree by exactly the amount the logarithm compresses.”

Making sense of the answer: The answer follows the doubling-adds-3-dB rule exactly, applied not to a single source but to a collection of equal incoherent sources. The general result is that n equal sources combine to a level $10 \log n \text{ dB}$ above the level of one such source, and so 72 children sit $10 \log(72/36) = 10 \log 2 = 3.01 \text{ dB}$ above 36 children, regardless of what the individual per-child intensity is. The answer would have been the same if each child had been twice as loud or half as loud, because the doubling-rule depends only on the ratio of the numbers of sources, not on the absolute intensities. This is one of the cleanest demonstrations in the chapter of why the dB scale is the right scale: it makes the physically relevant ratios visible without forcing the student to track unwieldy absolute numbers.

Think like a physicist: The classroom example is the simplest case of a result that scales to large auditoria, traffic streams, factories, and music festivals. Whenever the engineer wants to halve the perceived noise of a workplace, halving the number of noise sources will not do; the engineer must reduce the number of

sources by a factor of about ten, or alternatively reduce the intensity of each source by the same factor. This is why noise-control engineering in industry focuses overwhelmingly on quieter individual machines rather than on fewer machines: linear cuts in count do not pay for themselves in dB savings, but logarithmic cuts in intensity per machine do. Once you see the world in dB you cannot un-see it, and the world is full of acoustic surprises waiting for somebody to notice them.

We have now answered, in physical terms, the question the ear answers every waking second without thinking. We can compute the intensity level of any sound from its intensity, predict how its level falls off as the listener walks away from the source, and combine the levels of many incoherent sources without confusing dB with intensity. The next section turns away from sound altogether and goes to the other great family of waves the chapter must cover: the electromagnetic family, in which the speed is not 340m/s but 3×10^8 m/s, the medium is not air but the vacuum of empty space, and the wave variable is not pressure but electric and magnetic field. Sound has been our friend for four sections; it is time to meet its faster, stranger cousin.

ELECTROMAGNETIC WAVES (EM) AND THE SPEED OF LIGHT (c)

The first four sections of this chapter belonged to sound. We described its nature, found its speed in air, identified the resonant frequencies of any pipe of any length, and converted intensity into the logarithmic units the human ear actually uses. Sound, however, is not the only wave the syllabus knows. The wave that lets the reader see the page in front of him, the wave that carries the voice of a TBC Taifa newsreader from Dar es Salaam to a receiver in Musoma, the wave that lets a doctor at Muhimbili photograph a patient's broken bone, and the wave that defines the upper speed limit of the universe, all belong to a single family. The family is called the **electromagnetic** family. Unlike sound, its members do not need a medium to propagate; they travel through the vacuum of empty space at a fixed speed, the same speed in every direction, the same speed for every observer, and the same speed for every member of the family from radio waves to gamma rays. The chapter that gave four sections to sound must now give equal billing to its faster, stranger cousin.

In this section we do six things. We state Maxwell's great prediction of 1865, namely that electric and magnetic fields can propagate through empty space as a coupled wave at a speed determined entirely by the two constants μ_0 and ϵ_0 . We compute that speed. We list the properties of the resulting wave, which the student will find both familiar (the wave equation is the same form Chapter 7 derived for strings) and strange (no medium is required, and the wave behaves identically in glass and in water with only the constants adjusted). We write down the sine forms of E and B that show how the wave actually propagates through space and time. We note how the speed adjusts when the wave passes from vacuum into a material. We work through the experimental measurement of c performed by Léon Foucault in 1862, the cleanest school-laboratory method ever devised. By the end of the section the student will be able to compute c from μ_0 and ϵ_0 , identify the structure of a plane electromagnetic wave, write its sine-form expressions in time and space, compute its intensity from the amplitude of its electric field, find the speed of light in a transparent material from its refractive index, and read a Foucault apparatus question off the page.

Maxwell's prediction: light is an electromagnetic wave

In 1865 the Scottish physicist James Clerk Maxwell published a paper called *A Dynamical Theory of the Electromagnetic Field*. In it he showed that the four equations governing electric and magnetic fields, today known as Maxwell's equations (Gauss's law for electricity, Gauss's law for magnetism, Faraday's law of induction, and Ampère's law with Maxwell's own correction term), admit wave solutions in vacuum. The full derivation of these wave solutions requires Maxwell's equations in differential form, which lies beyond the scope of this book. Instead we state the result and identify, by direct comparison with a wave equation the student already knows, the speed of the resulting wave.

We begin with the wave equation the student already knows. Chapter 7 showed that a small transverse displacement $y(x, t)$ on a stretched string obeys:

$$\frac{\partial^2 y}{\partial t^2} = v^2 \frac{\partial^2 y}{\partial x^2}$$

The coefficient sitting in front of the spatial second derivative on the right-hand side is the squared wave speed.

Maxwell's equations, when manipulated in a region of empty space where no charges and no currents are present, produce a wave equation of exactly this same form for the electric field. The result is:

$$\frac{\partial^2 E}{\partial t^2} = \frac{1}{\mu_0 \epsilon_0} \frac{\partial^2 E}{\partial x^2},$$

with an identical equation holding for the magnetic field. Comparing this electromagnetic wave equation with the string wave equation term by term, the coefficient sitting in front of the spatial second derivative on the right-hand side is now $1/(\mu_0 \epsilon_0)$ in place of v^2 . The two equations have the same form, so the role of the coefficient is the same in both: the coefficient is the squared wave speed of whatever wave the equation describes. We therefore read off:

$$v^2 = \frac{1}{\mu_0 \epsilon_0}$$

Taking the positive square root gives the speed of an electromagnetic wave in vacuum:

$$c = \frac{1}{\sqrt{\mu_0 \epsilon_0}}$$

In the unified-wave-speed picture of Chapter 7, every wave speed comes from the square root of an elastic property divided by an inertial property. The same structure repeats here, with the **permeability** μ_0 playing the role of an *inertial property* of the vacuum and the **reciprocal of the permittivity** $1/\epsilon_0$ playing the role of an *elastic property*. Both are properties of the vacuum itself rather than of a string or a gas. The vacuum is the medium that supports the wave, though the word “medium” must be taken loosely: there is nothing in the vacuum to shake, only fields that propagate through it.

The numerical value of c follows by substituting the modern measured values $\mu_0 = 4\pi \times 10^{-7} \text{ T} \cdot \text{m/A}$ and $\epsilon_0 = 8.854 \times 10^{-12} \text{ F/m}$.

$$c = \frac{1}{\sqrt{(4\pi \times 10^{-7})(8.854 \times 10^{-12})}} = 2.998 \times 10^8 \text{ m/s} \approx 3 \times 10^8 \text{ m/s}$$

When Maxwell arrived at this number in 1865 he stared at it. The value he had obtained from purely electromagnetic constants was the same, to within the precision of the measurements then available, as the experimentally known speed of light in vacuum, measured by Hippolyte Fizeau in 1849 and by Léon Foucault in 1862.

The conclusion was unavoidable. *Light is an **electromagnetic wave***. The visible light that warms a face in the sun, the green of a banana leaf, the blue of the Indian Ocean off Bagamoyo, are oscillating electric and magnetic fields propagating through space at the speed Maxwell’s equations predicted. Forty years later Einstein would build his special relativity on the foundation that this speed is the same for every observer; ninety years later quantum electrodynamics would describe the same fields as photons. But Maxwell saw the wave nature of light first, and saw it with nothing more than a pencil, a sheet of paper, and the equations of electromagnetism.

Properties of an electromagnetic wave

A plane electromagnetic wave propagating in vacuum has four essential properties.

1) The wave is transverse

Both the **electric field** $\mathbf{E}$ and the **magnetic field** $\mathbf{B}$ oscillate in directions perpendicular to the direction of propagation. There is no analogue here of the longitudinal compression that defined sound earlier in this chapter, because there is no medium to compress. The transverseness is intrinsic to the electromagnetic field itself.

2) The two fields are perpendicular to each other

At any point in space and at any instant, $\mathbf{E}$ and $\mathbf{B}$ are perpendicular both to the direction of propagation and to each other. The three directions, that of $\mathbf{E}$, that of $\mathbf{B}$, and that of propagation, form three mutually perpendicular axes, with the direction of propagation given by the right-hand rule applied from $\mathbf{E}$ to $\mathbf{B}$.

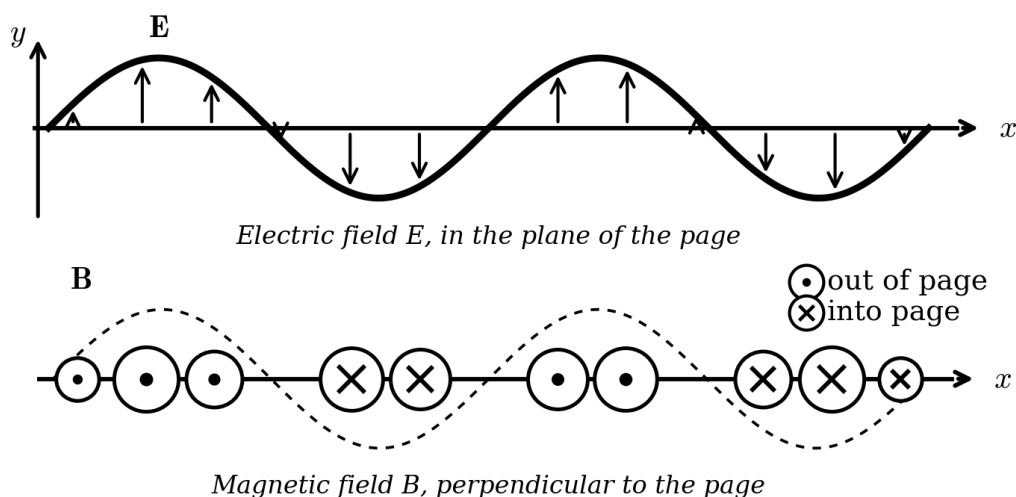


Figure: Structure of a plane electromagnetic wave. The electric field E oscillates in the y -direction (in the plane of the page); the magnetic field B oscillates in the z -direction (perpendicular to the page), shown using the standard dot-and-cross symbols. Both fields reach their maxima at the same x and are in phase with each other.

3) The two fields are in phase

They reach their maxima at the same place and at the same time and pass through zero together, in lock step. This is a consequence of the coupled structure of Maxwell's equations: each field generates a changing component of the other as the wave propagates, and the only consistent solution has the two fields oscillating together.

4) The magnitudes are linked by the wave speed

At every point in space and every instant, the amplitudes of the two fields satisfy

$$E = cB$$

so that a wave with electric-field amplitude $E_0 = 1 \text{ V/m}$ has magnetic-field amplitude $B_0 = \frac{E_0}{c} = 3.33 \times 10^{-9} \text{ T}$, a very small number. EM waves in practice are usually described by their electric-field amplitude, because the corresponding magnetic-field amplitude is so small numerically as to be inconvenient to write down.

One immediate consequence of the transverseness of an electromagnetic wave is that the wave can be polarised. The direction in which the electric field oscillates from cycle to cycle is called the polarisation direction of the wave, and a wave whose electric field stays in one plane is said to be plane-polarised. A longitudinal wave, like sound, has no polarisation direction to speak of, because its displacement is along the direction of travel and no transverse choice is to be made. Polarisation is therefore a property unique to transverse waves, and the full treatment of how it is produced, manipulated, and detected was given in Chapter 8 on physical optics.

Propagation: sine-form description of E and B

The four geometric properties just listed, describe the wave at a single instant of time. They do not yet describe how the wave actually propagates: how the pattern of E and B shifts down the x -axis as time passes, and how a single fixed point in space sees its E and B oscillate up and down with time. The propagation picture is supplied by the sine forms of E and B , which are the explicit time-and-space solutions of the wave equations Maxwell's prediction gave us.

Consider a plane electromagnetic wave travelling in the positive x -direction with angular frequency ω and wavenumber k . The electric field, polarised in the y -direction, is

$$\mathbf{E}(\mathbf{x}, t) = E_0 \sin(kx - \omega t)$$

and the magnetic field, polarised in the z -direction, is

$$\mathbf{B}(\mathbf{x}, t) = B_0 \sin(kx - \omega t)$$

where E_0 and B_0 are the amplitudes of the two fields. Four features of these sine forms deserve attention.

- **Same argument $kx - \omega t$ for both fields.** Both E and B depend on the same combination $kx - \omega t$ of position and time. This is exactly the in-phase property listed earlier: at every x and every t , the argument of the sine is the same for the two fields, so they peak together, pass through zero together, and reach their troughs together. The two oscillations are locked in step, not by accident but by the structure of Maxwell's equations.
- **Two amplitudes, linked by $E_0 = cB_0$.** The amplitudes E_0 and B_0 are not independent; they are tied by the magnitude relation $E = cB$ from the properties list, applied to the peak values. The wave's strength is fully specified by either of E_0 or B_0 alone, and standard practice is to specify E_0 because the corresponding B_0 is so small numerically.
- **The wavenumber k and angular frequency ω .** The wavenumber k , measured in radians per metre, governs how rapidly the field changes with position at a fixed time: a complete cycle takes a distance $\lambda = 2\pi/k$, the wavelength. The angular frequency ω , measured in radians per second, governs how rapidly the field changes with time at a fixed position: a complete cycle takes a time $T = 2\pi/\omega$, the period. The ordinary frequency f and the ordinary wavelength λ are related to ω and k by:

$$\omega = 2\pi f \quad \text{and} \quad k = \frac{2\pi}{\lambda}$$

- **The wave speed is ω/k , and it equals c .** Substituting the sine forms into the wave equation derived earlier in this section, $\partial^2 E / \partial t^2 = (1/\mu_0 \epsilon_0) \partial^2 E / \partial x^2$, the time derivative produces a factor of ω^2 and the spatial derivative a factor of k^2 . The wave equation is satisfied provided

$$\omega^2 = \frac{k^2}{\mu_0 \epsilon_0}$$

Taking the positive square root and rearranging,

$$\frac{\omega}{k} = \frac{1}{\sqrt{\mu_0 \epsilon_0}} = c$$

The ratio ω/k is the speed at which the pattern of the wave moves along the x -axis, and that speed comes out to be exactly the c derived earlier from the coefficient of the wave equation. The same value of c that the wave equation predicted in the abstract is now confirmed concretely by the propagation of the sine pattern itself. The same conclusion is reached from the more familiar pair of relations $c = f\lambda$ and $c = \omega/k$, which the student has met in Chapter 7 for mechanical waves; the new content here is that the value of c comes out to be $1/\sqrt{\mu_0 \epsilon_0}$ for any electromagnetic wave, irrespective of its frequency or wavelength.

The student should imagine the sine forms vividly. At a fixed moment in time, E traces out a sinusoid in space along the x -axis (the snapshot the figure shows); as time advances by one period T , the entire sinusoid slides forward by exactly one wavelength λ at speed $c = \lambda/T = \omega/k$. A fixed point in space, like a TV antenna or a retinal cell in a human eye, watches the field at that location oscillate sinusoidally in time with period T ; this is the time-domain picture. The two pictures, the spatial snapshot and the time-domain oscillation at one point, are the same wave seen from two different angles.

Speed of light in a material medium

In a transparent material such as glass or water, the speed of an electromagnetic wave is not c . The material's atoms respond to the passing wave by polarising and re-radiating in step with it, and the net effect is to slow the wave down compared with its vacuum speed. The slowdown is captured by a single number, the **refractive index n** of the material, defined by:

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

The same number n is the one introduced in Chapter 8 in connection with Snell's law of refraction at a boundary between two media. The two definitions are consistent: a wave that slows down on entering a denser material also bends towards the normal at the boundary, and the same ratio n governs both effects.

A few representative values, which a Tanzanian A-level student should recognise on sight, are $n = 1.00$ for vacuum (exactly by definition), $n = 1.00029$ for dry air at sea-level pressure (almost equal to vacuum), $n = 1.33$ for fresh water, $n = 1.50$ for ordinary crown glass, $n = 1.52$ for the dense flint glass used in optical instruments, and $n = 2.42$ for diamond. **Diamond's high refractive index** is what gives the polished stone

its characteristic sparkle, because light entering it slows down dramatically and bounces back and forth inside the stone many times before escaping.

The refractive index of a material in turn comes from its electromagnetic properties, in exact analogy with $c = 1/\sqrt{\mu_0 \epsilon_0}$ in vacuum. If the material has absolute permeability μ and absolute permittivity ϵ , then

$$v = \frac{1}{\sqrt{\mu \epsilon}}$$

and the refractive index is:

$$n = \frac{c}{v} = \sqrt{\frac{\mu \epsilon}{\mu_0 \epsilon_0}}$$

For most non-magnetic materials μ is very close to μ_0 , so the formula simplifies to:

$$n \approx \sqrt{\frac{\epsilon}{\epsilon_0}},$$

and the optical behaviour of the material is determined by its dielectric properties alone.

Energy and intensity of an electromagnetic wave

Just like the mechanical waves of Chapter 7, an EM wave carries energy. The energy is stored partly in the electric field and partly in the magnetic field, both expressed per unit volume. The energy density in the electric field is

$$u_E = \frac{1}{2} \epsilon_0 E^2$$

and the energy density in the magnetic field is

$$u_B = \frac{1}{2\mu_0} B^2$$

In an EM wave the two energy densities are equal at every instant, because $E = cB$ gives $B^2 = E^2/c^2 = \mu_0 \epsilon_0 E^2$, so that $u_B = (1/(2\mu_0)) \times \mu_0 \epsilon_0 E^2 = 1/2 \epsilon_0 E^2 = u_E$. The energy is split exactly half-and-half between the electric and magnetic fields. The total energy density is:

$$u = u_E + u_B = \frac{1}{2} \epsilon_0 E^2 + \frac{1}{2} \epsilon_0 E^2 = \epsilon_0 E^2$$

Averaged over one full cycle of oscillation, the squared field averages to half its peak value, so the time-averaged total energy density is:

$$\bar{u} = \frac{1}{2} \epsilon_0 E_0^2$$

The intensity of the wave is the rate at which energy crosses a unit area perpendicular to the propagation direction. Energy is carried at the wave speed c , so the intensity is

$$I = \bar{u} \cdot c = \frac{1}{2} \epsilon_0 c E_0^2$$

This is the formula a student needs whenever an EM wave's intensity must be related to the amplitude of its electric field.

Foucault's rotating mirror: measuring c on a laboratory bench

Maxwell predicted c from theory; Léon Foucault, in 1862, measured it on a laboratory bench in Paris. His method was clever enough to deserve a place in every undergraduate physics curriculum since, and clean enough that a Form Five student with access to a fast-spinning mirror, a long room, and a graduated screen could replicate it in principle.

The apparatus, illustrated in the following figure, consists of four optical elements. A small light source S emits a thin pencil of light. The light reflects off a small rotating mirror M_R that spins about a vertical axis

at a known angular speed ω . The reflected beam travels a long horizontal distance D and reflects off a fixed concave mirror M_F mounted at the far end of the apparatus. The returning beam strikes the rotating mirror a second time. Because the rotating mirror has rotated through a small angle in the time the light took to make its round trip, the second reflection does not retrace the outgoing path. Instead, the returning beam is deflected by a measurable amount onto a screen placed at a known distance r from M_R , and the deflection is read off the screen.

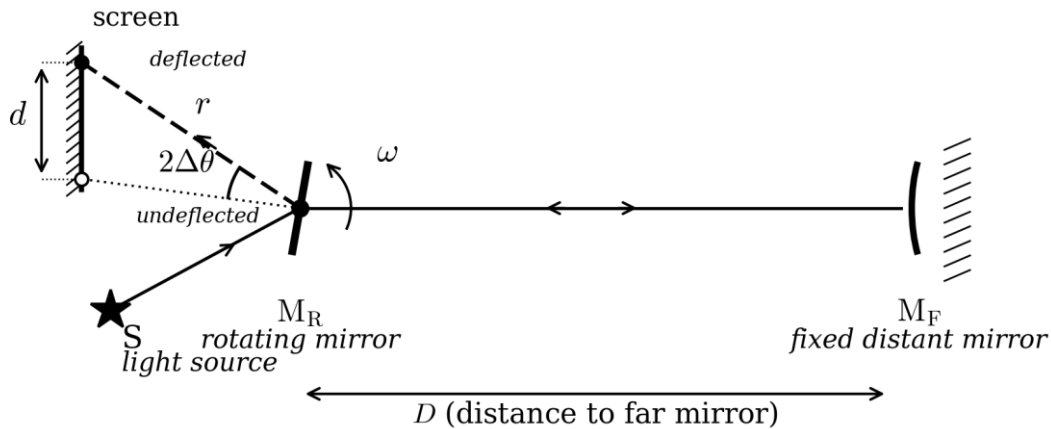


Figure: Schematic of Foucault's rotating-mirror apparatus for measuring c . Light from source S reflects off the rotating mirror M_R , travels distance D to the fixed mirror M_F , and returns to M_R . The dotted line shows the path the returning beam would take if M_R had not rotated during the round trip; the dashed line shows the actual path, deflected by an angle $2\Delta\theta$ because M_R has rotated by $\Delta\theta$ during the time the light was in flight. The deflected beam strikes the screen at distance r from M_R , displaced by d from the spot at which the undeflected beam would have arrived.

The derivation is short. Let the round-trip time for the light be t . Then:

$$t = \frac{2D}{c}$$

During this time the rotating mirror sweeps through an angle,

$$\Delta\theta = \omega t = \frac{2D\omega}{c}$$

By the law of reflection, a rotation of the mirror by $\Delta\theta$ rotates the reflected beam by $2\Delta\theta$. The returning beam is therefore deflected relative to the original outgoing direction by:

$$2\Delta\theta = \frac{4D\omega}{c}$$

The displacement d of the spot on the screen at distance r from M_R , for small angles, is the deflection angle times the screen distance:

$$d = r \times 2\Delta\theta = \frac{4Dr\omega}{c}$$

Solving for c :

$$c = \frac{4Dr\omega}{d}$$

This is Foucault's formula. The right-hand side contains only quantities the experimenter measures with care on a laboratory bench: the two distances D and r , the angular speed ω of the rotating mirror, and the displacement d of the returning spot on the screen. Foucault's 1862 measurement obtained $c = 2.98 \times 10^8$ m/s, within 0.7 per cent of the modern accepted value, using a rotating mirror driven by a falling weight and a distant fixed mirror 20 metres down the corridor of an observatory.

Foucault's method is one of two classic nineteenth-century laboratory measurements of c . The other is **Fizeau's toothed-wheel method** of 1849, in which a beam of light is chopped into pulses by a rotating

cogwheel, sent to a distant mirror, and chopped a second time on its return; the speed of light is recovered from the rotational speed at which the second chop just lets the returning pulse through, given the gap between adjacent teeth and the distance to the distant mirror. We restrict ourselves to Foucault's apparatus here because it is the simpler of the two to derive from first principles, and it is more common.

We now have the six pillars of the electromagnetic chapter. The worked examples that follow exercise four of these pillars in turn, beginning with a direct numerical application of Maxwell's formula and ending with a question that asks the student to compute the electric and magnetic fields of sunlight at the Earth's surface in Tanzania on a clear noon.

BINDER Example 14

- (a) Taking the permeability of free space as $\mu_0 = 4\pi \times 10^{-7} \text{ T} \cdot \text{m/A}$ and the permittivity of free space as $\epsilon_0 = 8.854 \times 10^{-12} \text{ F/m}$, compute the speed of an electromagnetic wave in vacuum to four significant figures.
- (b) The refractive index of ordinary crown glass is $n = 1.50$. Compute the speed of light in this glass.

Solution

- (a) Apply Maxwell's prediction directly.

$$c = \frac{1}{\sqrt{\mu_0 \epsilon_0}} = \frac{1}{\sqrt{(4\pi \times 10^{-7} \text{ T} \cdot \text{m/A}) \times (8.854 \times 10^{-12} \text{ F/m})}} = 2.998 \times 10^8 \text{ m/s}$$

To four significant figures the speed of light in vacuum is $c = 2.998 \times 10^8 \text{ m/s}$.

- (b) For glass, $v = c/n$:

$$v = \frac{c}{n} = \frac{2.998 \times 10^8 \text{ m/s}}{1.50} = 1.999 \times 10^8 \text{ m/s}$$

The speed of light in the glass is $1.999 \times 10^8 \text{ m/s}$.

Making sense of the answer: Two universal constants of electromagnetism, neither of which has anything obvious to do with light, combine in part (a) to give the speed at which light travels. The numerical coincidence is so striking that Maxwell took it as proof that light is itself an electromagnetic phenomenon, twenty-six years before Heinrich Hertz produced electromagnetic waves at radio frequencies and verified that they had every property Maxwell had predicted. The slowdown in glass in part (b) is large enough to matter for any optical instrument: it is the reason a glass prism bends light into a spectrum, the reason a swimming pool looks shallower than it is, and the reason a lens focuses parallel rays to a point.

Think like a physicist: Whenever a formula in physics combines two constants from different branches of the subject and reproduces a known result from a third branch, sit up and pay attention. The unification of optics and electromagnetism by Maxwell, of mechanics and gravitation by Newton, of electricity and magnetism by Faraday and Ampère, and of matter and energy by Einstein, are all of this form. Most progress in physics is the unification of two things that were previously thought to be three.

REAL Example 15

TANESCO uses a long-distance microwave radio link between two relay towers to coordinate power-grid operations. The link operates at a frequency of 6.0GHz.

- (a) Compute the wavelength of the microwave signal in air.
- (b) Why is microwave preferred over visible light for line-of-sight communication between two relay towers?
- (c) Why is microwave not used to illuminate a stage at a concert, even though it would carry the same energy a stage spotlight does?

Solution

- (a) Treating the air as effectively a vacuum ($n_{\text{air}} \approx 1.000$), the speed of the microwave is essentially $c = 3.0 \times 10^8 \text{ m/s}$. From the basic wave relation $c = f\lambda$,

$$\lambda = \frac{c}{f} = \frac{3.0 \times 10^8 \text{ m/s}}{6.0 \times 10^9 \text{ Hz}} = 0.050 \text{ m} = 5.0 \text{ cm}$$

The signal has a wavelength of 5 centimetres in air, about the width of a smartphone.

(b) Microwave is preferred over visible light for line-of-sight tower-to-tower communication for three connected reasons.

First, the 5cm microwave wavelength is short enough that a dish antenna of practical size (a metre or so across) can focus the signal into a tight, narrow beam aimed precisely at the receiving tower, which keeps the transmitted power on target and reduces interference with neighbouring links. A visible-light beam over the same range would also be directional, but to focus visible light into a narrow beam over kilometres of atmosphere requires a laser, not a passive dish.

Second, microwave at 6GHz penetrates rain, fog, mist, and dust far better than visible light does. The rainy season in Mwanza or Mbeya routinely blocks visible light over a few hundred metres, but a microwave link continues to operate.

Third, the manufacture of high-power microwave transmitters (magnetrons, klystrons, and modern solid-state amplifiers) is mature and economical, whereas high-power visible-light transmitters that could carry equivalent data rates over equivalent distances would be substantially more expensive and less reliable.

(c) Microwave is not used to illuminate a stage because the human eye **does not detect** microwave at all. The retina is sensitive only to electromagnetic waves with wavelengths in the band from about 400nm to 700nm, and 5cm microwave is some hundred thousand times longer than the longest wavelength the eye can register. Replacing a stage spotlight with a microwave transmitter of the same power would leave the stage in pitch darkness, while **heating** the audience uncomfortably. The same energy can be delivered by either wave, but the perception of the energy depends on its wavelength. A successful stage spotlight must lie inside the band the human eye can detect, and that band is the band we call visible light.

Making sense of the answer: *The whole microwave band (1mm to 1m in wavelength) sits between the radio band and the infrared band on the electromagnetic spectrum. At 5cm it sits well inside the part of the band used for telecommunications, broadcast satellite, and weather radar. The Vodacom mobile-phone band at 900MHz uses 33cm wavelengths, longer than the TANESCO link in this question; the Wi-Fi router in a Dar es Salaam student's home at 2.4GHz uses 12.5cm; and a kitchen microwave oven at 2.45GHz also uses 12.5cm. The choice of a particular frequency for a particular application is always a compromise between antenna size, atmospheric transparency, and bandwidth, and the engineer's job is to find the band that wins on all three.*

Think like a physicist: *Every part of the electromagnetic spectrum can carry energy, can be modulated to carry information, and can be focused by a properly designed antenna. What changes from band to band is which technological problems become easy and which become hard. A radio engineer designing a link of any kind asks four questions before choosing a frequency: how directional must the beam be (which sets the smallest workable antenna at any given size), how much rain and fog must it punch through (which sets the highest workable frequency), what data rate is required (which sets the lowest workable frequency through the bandwidth-per-cycle argument), and how much licensing the regulator will charge. The Tanzania Communications Regulatory Authority (TCRA) controls the answer to the fourth.*

REAL Example 16

A physics teacher at Mzumbe Secondary School sets up a Foucault apparatus on the long corridor outside the laboratory. A small rotating mirror, spinning at 400 revolutions per second about a vertical axis, sends a thin beam of light to a fixed concave mirror at a distance of 35m down the corridor. The light returns to the rotating mirror and is reflected onto a screen placed at a perpendicular distance of 5.0m from the rotating mirror. The displacement of the returning spot on the screen, relative to its position when the rotating mirror is stationary, is measured to be 5.87mm.

- Use this data to compute the speed of light in air, and compare your result with the modern accepted value.
- Without redoing the experiment, predict the displacement d that would be observed if the rotation rate were doubled to 800 revolutions per second, with everything else unchanged. Justify your prediction with reference to the formula derived in this section.

Solution

(a) Foucault's formula is $c = 4D\omega/d$, where ω is the angular speed of the rotating mirror in radians per second. The rotation rate is given as 400 revolutions per second, and each revolution sweeps an angle of 2π radians, so:

$$\omega = 2\pi f = 2\pi \times 400 \text{ Hz} = 2513 \text{ rad/s}$$

Substitute the four measured quantities into Foucault's formula:

$$c = \frac{4D\omega}{d} = \frac{4 \times 35 \text{ m} \times 5.0 \text{ m} \times 2513 \text{ rad/s}}{5.87 \times 10^{-3} \text{ m}} = 2.997 \times 10^8 \text{ m/s}$$

The measured speed of light is $c = 2.997 \times 10^8 \text{ m/s}$, which agrees with the modern accepted value of $2.998 \times 10^8 \text{ m/s}$ to four significant figures. The small discrepancy of about 0.03 per cent could come from rounding the rotation rate to 400Hz exactly, from a small misalignment of the screen, or from the slightly slower speed of light in air than in vacuum (in air, c is about 0.03 per cent lower than in vacuum, which would account for the entire discrepancy on its own).

(b) The Foucault formula in its forward form is $d = 4D\omega/c$. Holding D , r , and c fixed, the displacement d is directly proportional to ω :

$$d \propto \omega$$

If ω doubles, d doubles. The predicted displacement at 800 revolutions per second is therefore

$$d_{\text{new}} = 2 \times 5.87 \text{ mm} = 11.7 \text{ mm}$$

The experimenter would observe the spot move twice as far on the screen as in part (a).

Making sense of the answer: *That a corridor of 35m, a mirror spinning at 400 revolutions per second, and a deflection measured in millimetres, can deliver the speed of light to four significant figures, is one of the great miracles of nineteenth-century experimental physics. The geometric amplification is what makes it work: a tiny rotation angle of the mirror ($2\Delta\theta$ in this experiment is about 1.2×10^{-3} radians, or about a fifteenth of a degree) becomes a millimetres-large displacement on the screen because the screen is several metres away from the mirror. Tiny angles become measurable distances when the lever arm is long enough. Part (b) of the question turns the formula around and uses it predictively: once the relationship between input and output is established, the experimenter can predict outputs for new inputs without measuring afresh.*

Think like a physicist: *Foucault's apparatus is an early example of what experimental physicists call a leverage arm: an arrangement in which a small quantity of interest is amplified into a much larger quantity that can be measured directly. The same principle appears again and again throughout experimental physics: the challenge is always to arrange the system so that an extremely small effect becomes large enough to detect before it disappears into background noise.*

HOT Example 17

The intensity of sunlight at the Earth's surface in Tanzania on a clear day at noon is approximately 1000 W/m^2 . Treating the sunlight as a plane electromagnetic wave with a single dominant frequency, find:

- the amplitude E_0 of the electric field at the surface;
- the amplitude B_0 of the corresponding magnetic field;
- the total electromagnetic energy striking a one-square-metre solar panel mounted perpendicular to the sun's rays in one hour of clear noon.

Take $\epsilon_0 = 8.854 \times 10^{-12} \text{ F/m}$ and $c = 3.00 \times 10^8 \text{ m/s}$.

Solution

(a) Apply the intensity formula $I = 1/2 \epsilon_0 c E_0^2$. Solving for E_0 ,

$$E_0^2 = \frac{2I}{\epsilon_0 c}$$

$$E_0 = \sqrt{\frac{2I}{\epsilon_0 c}} = \sqrt{\frac{2 \times 1000 \text{ W/m}^2}{(8.854 \times 10^{-12} \text{ F/m}) \times (3.00 \times 10^8 \text{ m/s})}} = 868 \text{ V/m}$$

The electric-field amplitude of sunlight at the surface is 868V/m.

(b) The corresponding magnetic-field amplitude follows from $B_0 = E_0/c$:

$$B_0 = \frac{E_0}{c} = \frac{868 \text{ V/m}}{3.00 \times 10^8 \text{ m/s}} = 2.89 \times 10^{-6} \text{ T} \approx 2.9 \mu\text{T}$$

The magnetic-field amplitude is therefore about 2.9 microtesla, comparable to the Earth's magnetic field but oscillating at the frequency of visible light.

(c) The total energy striking a one-square-metre solar panel in one hour is the intensity times the area times the time:

$$\text{Energy} = I \times A \times t = (1000 \text{ W/m}^2) \times (1 \text{ m}^2) \times (3600 \text{ s}) = 3.60 \times 10^6 \text{ J} = 3.6 \text{ MJ}$$

A one-square-metre solar panel exposed to clear noon sunlight in Dar es Salaam receives 3.6 megajoules of energy in one hour. At a typical solar-panel conversion efficiency of 20 per cent, this corresponds to about 720 kilojoules of useful electrical energy, or roughly 0.2 kilowatt-hours per square metre per hour of clear noon. A 5-square-metre roof installation, exposed to the equivalent of 5 hours of clear noon per day, would generate about 5 kilowatt-hours per day, which is roughly the daily electricity needs of a small Tanzanian household.

Making sense of the answer: *The electric-field amplitude of sunlight is about 870 V/m, which sounds large until one remembers that the dielectric strength of dry air is about 3 million V/m, so sunlight is operating at less than one part in 3000 of the field strength needed to ionise air. The magnetic-field amplitude of 2.9 microtesla is comparable to Earth's magnetic field, but the sunlight oscillates at frequencies near 5×10^{14} Hz, while Earth's field is static. The 3.6 megajoules of energy in one hour is enough to heat about 10 kilograms of water by 86 degrees, which is why a black-painted water tank on a tin roof in Dodoma produces hot water all day long without any external power.*

Think like a physicist: *Every wave that carries energy can be characterised by the amplitude of the quantity that oscillates: pressure for sound, displacement for a wave on a string, electric field for light. The intensity formula always has the same structure: half times something times something times amplitude squared, where the two "somethings" are the medium's response constants. For light it is $1/2 \epsilon_0 c E_0^2$. For sound in a fluid it is $1/2 \rho v \omega^2 s_0^2$ where s_0 is the displacement amplitude. The pattern repeats across the entire wave physics of the universe.*

We have now answered, in physical terms, every question the chapter set for itself about electromagnetic waves. But the family of electromagnetic waves is broad, and the visible light that has occupied our discussion so far is only one narrow band in a spectrum that extends from radio waves at one extreme to gamma rays at the other. The next section catalogues the electromagnetic family by frequency, describes how each band is produced and detected, and lists the applications by which each band touches the daily life of an A-level student, from the morning radio bulletin to the evening television and from the X-ray of a broken arm to the microwave that warms last night's leftovers.

THE ELECTROMAGNETIC SPECTRUM

The previous section closed with the observation that visible light is a single narrow band inside a much broader electromagnetic family. The family runs from radio waves of kilometre wavelength at one end to gamma rays of sub-picometre wavelength at the other end, and every member shares the same fundamental properties: a transverse oscillation of electric and magnetic fields, propagation through vacuum at speed c , the in-phase relationship $E = cB$ between amplitudes, and the intensity formula $I = 1/2 \epsilon_0 c E_0^2$. What changes from one member to the next is wavelength, and with wavelength every practical consequence: how the wave is produced, how it is detected, what it does on contact with matter, and what technology, medicine, and engineering have built around it. This section catalogues the seven bands the syllabus recognises, gives a Tanzanian instance for each, and ends with one worked example on the most familiar microwave instrument the student has met, the kitchen microwave oven at 2.45 GHz.

The spectrum at a glance

The figure below shows the entire electromagnetic spectrum on a single logarithmic frequency axis. Frequency increases from left to right, by two powers of ten per tick, from 10^4 Hz on the far left (very-low-frequency radio) to 10^{22} Hz on the far right (high-energy gamma rays from nuclear processes). Wavelength, plotted along the top of the same axis, decreases correspondingly from 30 km on the far left to 30 fm (femtometres) on the far right, an overall span of eighteen orders of magnitude. Three Tanzanian-context frequencies are marked above the bar: the TBC FM band at 100 MHz, the kitchen microwave oven at

2.45GHz, and a medical X-ray at the Muhimbili National Hospital at about 0.1nm. The visible band is so narrow on this scale that it is shown by hatching and a leader line, with a magnified inset below the main bar showing the seven colours of the visible rainbow.

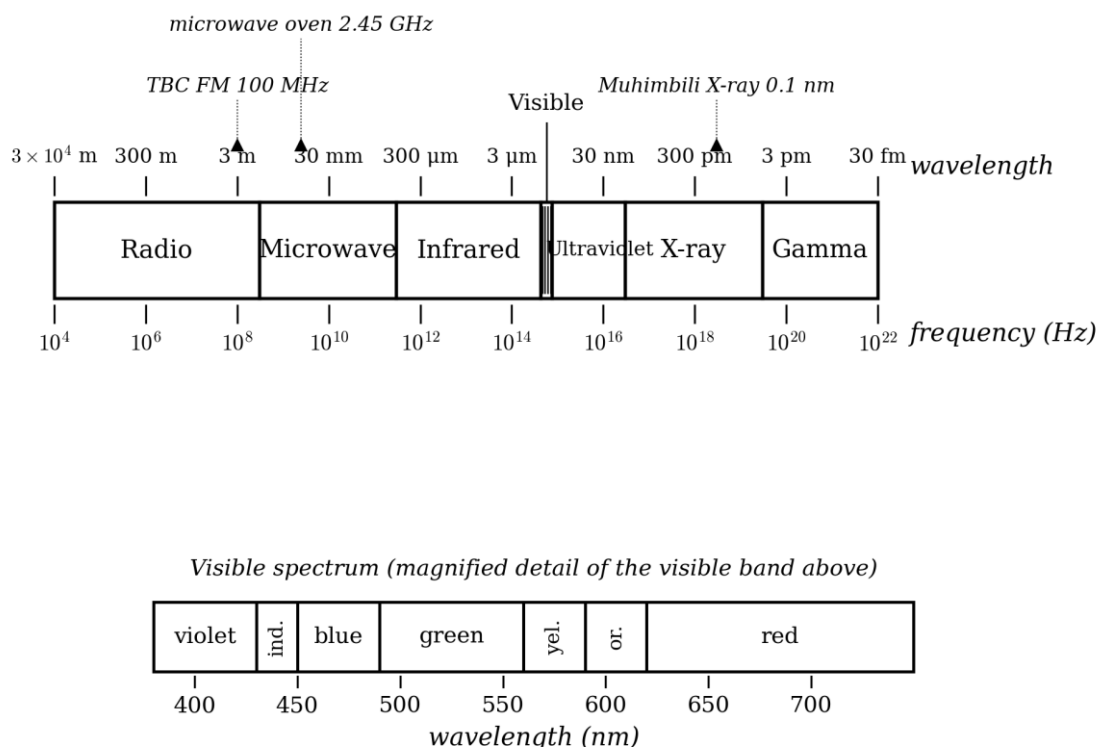


Figure: The electromagnetic spectrum. Top: the full spectrum from radio to gamma, on a logarithmic axis spanning eighteen orders of magnitude in frequency. Three Tanzanian-context frequencies are marked above the bar. Bottom: a magnified inset of the visible band (380–750nm), showing the seven colours violet, indigo, blue, green, yellow, orange, red.

The student should pause and stare at this figure for a moment. The visible light that drives every act of human sight, from reading this page to watching a sunset over the Indian Ocean, occupies less than one twentieth of one decade on a logarithmic scale eighteen decades wide. Everything else, from the FM radio in a car's dashboard to the cobalt-60 source at the Ocean Road Cancer Institute, is invisible to the eye yet identical in nature to the light the eye does see. Maxwell's prediction unified them all, and the seven bands below are a catalogue of the same wave, partitioned only by what humans have learnt to do with it at each scale.

Radio waves occupy the longest wavelengths, from kilometres at the very-low-frequency end down to about one metre at the boundary with microwave. They are produced by accelerating electrons along the length of a metal antenna, driven by an LC oscillator circuit at the desired frequency. They are detected by a receiving antenna, tuned by a resonant LC circuit at the listener's end, and demodulated by a detector. In Tanzania, AM radio operates in the kilohertz band, FM radio (including the Tanzania Broadcasting Corporation, TBC, and Clouds FM) operates around 100MHz, and short-wave radio reaches the rest of East Africa from transmitters in Dar es Salaam. The same band carries television broadcast, mobile-phone signals at the low end of the microwave boundary, and amateur radio.

Microwaves occupy wavelengths from 1m down to 1mm, equivalently frequencies from 300MHz to 300GHz. They are produced by specialised oscillators such as klystrons and magnetrons (the latter is the source inside every kitchen microwave oven) or by modern solid-state circuits. They are detected by crystal diodes paired with horn antennas or parabolic dishes. Tanzanian instances are abundant: Vodacom and Airtel mobile-phone networks operate at 900, 1800, 2100, and 2600MHz; the SEACOM undersea optical cable lands microwave-relayed traffic in Dar es Salaam; the TANESCO long-distance microwave radio link of

REAL Example 15 in the previous section operates at 6GHz; satellite television uses the 4GHz C-band; and the kitchen microwave oven, the subject of the worked example at the end of this section, operates at 2.45GHz.

Infrared occupies wavelengths from 1mm down to about 700nm at the boundary with visible red, equivalently frequencies from 300GHz up to about 430THz. Infrared is produced by every warm object in the universe simply by virtue of being warm; the spectrum a body emits is determined by its temperature, and the warmer the body the shorter the peak wavelength. A human body at 37°C radiates predominantly in the infrared band centred around 10μm. Infrared is detected by bolometers (which measure the heat the radiation deposits), thermopiles, and the silicon-array infrared cameras now built into every smartphone selfie camera in altered form. Tanzanian instances: every television remote control uses an 850nm infrared LED to send a command to the television; airport thermal scanners use infrared cameras to screen arriving passengers for fever; night-vision goggles, used by anti-poaching units patrolling the Selous Game Reserve, see the warm bodies of poachers and wildlife against the cool background of bush at night.

Visible light occupies the narrow band from 700nm (red) down to 400nm (violet), equivalently a frequency window from 430THz to 750THz. The seven colours of the rainbow violet, indigo, blue, green, yellow, orange, red sit inside this window, each occupying a slice perhaps 30 to 50nm wide. Visible light is produced by electronic transitions in atoms, including the heated filament of an incandescent bulb (giving a continuous spectrum of blackbody emission), the discrete spectral lines of a sodium-vapour streetlamp, the white light of a light-emitting diode, and the broadband emission of the Sun. The eye is the most sensitive detector of visible light known, with the rod cells of the dark-adapted retina capable of registering individual photons; photographic film, semiconductor photodiodes, and CCD arrays in digital cameras are the engineered equivalents. Tanzanian instances are everywhere: sight itself, photography, fibre-optic telecommunications at the special wavelengths 1310nm and 1550nm (technically in the near infrared, but the same physics: the SEACOM cable that lands at Dar es Salaam carries internet traffic at these wavelengths). The seven colours of the rainbow that follow a Dar es Salaam thunderstorm are this same visible band, dispersed by the prism action of the raindrops.

Ultraviolet occupies the band from 400nm down to about 10nm, equivalently 750THz up to 3×10^{16} Hz. Ultraviolet is produced by hotter sources than visible light: the mercury arc lamp, the very hot surface of the Sun, and electronic transitions involving larger energy gaps. It is detected by fluorescent screens (which absorb ultraviolet and re-emit it as visible light, the same physics that makes a fluorescent banknote feature glow under a security checker), and by specialised photographic film. The Sun radiates substantial ultraviolet, most of which is absorbed by the ozone layer of the upper atmosphere; the small fraction that reaches the surface of Tanzania is responsible both for the tan a holidaymaker acquires on the beaches of Zanzibar and for the vitamin D the same holidaymaker's skin synthesises in response. The harder ultraviolet ionises atoms and can damage DNA, which is why excessive sun exposure carries a cancer risk and why ultraviolet lamps are used to sterilise surgical instruments and drinking water.

X-rays occupy the band from 10nm down to 0.01nm, equivalently 3×10^{16} Hz up to 3×10^{19} Hz. X-rays are produced by inner-shell electron transitions in heavy atoms and by bremsstrahlung (literally “braking radiation”) emitted when a fast electron is suddenly decelerated by hitting a metal target; this is the mechanism inside every medical X-ray tube. They are detected by photographic film, by image intensifiers in modern hospital systems, and by ionisation chambers. The wavelength of an X-ray is comparable to the spacing of atoms in a crystal, which is why X-ray diffraction is the standard tool for finding atomic positions in solids. Tanzanian instances: every chest X-ray taken at Muhimbili National Hospital, every dental X-ray at a clinic in Mwanza, every airport luggage scanner at Julius Nyerere International Airport that checks for contraband under bags.

Gamma rays occupy the shortest wavelengths, below 0.01nm and equivalently above 3×10^{19} Hz. They are produced by nuclear transitions, by the annihilation of an electron with a positron, and by cosmic-ray showers in the upper atmosphere. They are detected by Geiger counters, scintillation detectors, and modern semiconductor gamma cameras. Gamma rays carry enough energy per photon to penetrate substantial thicknesses of lead, which is why gamma sources are stored in heavy lead-lined containers and why gamma therapy is among the most carefully controlled procedures in medicine. The Ocean Road Cancer Institute in Dar es Salaam uses a cobalt-60 gamma source to deliver radiation therapy to cancer patients; the same isotope is used worldwide to sterilise surgical equipment and food. The natural background of gamma rays in Tanzania, from radon in soil and cosmic rays from space, is small but never zero, and adds a few millisieverts per year to every Tanzanian's lifetime radiation exposure.

Each band above is the same wave, propagating at the same speed c , differing only in wavelength. The choice of which technological apparatus to use, which biological effect to expect, and which name to assign, depends on the wavelength alone. Once the wavelength is given, the band is fixed, the production and detection methods follow, and the applications fall into place. The worked example below applies this picture to the most familiar microwave appliance in any Tanzanian kitchen, and asks three questions every student should be able to answer.

REAL Example 18

A kitchen microwave oven operates at a frequency of 2.45GHz, the standard frequency for domestic microwave ovens worldwide.

- Compute the wavelength of the microwave radiation inside the oven cavity, assuming the cavity is essentially empty (filled with air, with the small water-containing meal occupying a small fraction of the cavity volume).
- Explain why 2.45GHz, rather than some other microwave frequency, is used in domestic microwave ovens.
- Inside the oven cavity, standing electromagnetic waves form between the conducting metal walls. Estimate the spacing between adjacent nodes of the standing wave, and explain why the oven includes a rotating turntable.

Solution

- (a) Using $c = f\lambda$ with $c = 3.0 \times 10^8$ m/s and $f = 2.45 \times 10^9$ Hz,

$$\lambda = \frac{c}{f} = \frac{3.0 \times 10^8 \text{ m/s}}{2.45 \times 10^9 \text{ Hz}} = 0.122 \text{ m} = 12.2 \text{ cm}$$

The microwave inside the oven has a wavelength of about 12cm, comparable to the width of a slice of bread.

- (b) Two reasons drive the choice of 2.45GHz, neither involving the resonant frequency of any water molecule.

First, 2.45GHz lies inside an internationally protected slice of the radio spectrum called the **ISM band** (Industrial, Scientific, Medical), which is reserved by international agreement for free, unlicensed use by domestic and industrial appliances that radiate electromagnetic energy. Designing microwave ovens for the ISM band means manufacturers can sell ovens worldwide without negotiating a separate broadcast licence in every country.

Second, water absorbs microwave radiation over a broad range of frequencies, with absorption increasing as the frequency rises. At very high microwave frequencies, near 20GHz, water absorbs so strongly that the radiation penetrates only the outer few millimetres of the food and the inside of the food stays cold; at very low microwave frequencies, near 100MHz, water absorbs so weakly that the radiation passes through the food without depositing significant energy. The frequency 2.45GHz is a compromise: water absorbs enough that a substantial fraction of the radiated power is deposited inside the food, but not so much that the deposition is confined to a thin surface layer. The penetration depth at 2.45GHz is several centimetres, well matched to the typical size of a meal.

It is worth correcting a textbook myth at this point. Many physics books claim that 2.45GHz is the “resonant frequency” of the water molecule, suggesting some sharp peak in absorption at exactly this frequency. This is **not** true. Liquid water has no sharp resonance in the microwave band; instead, it has a broad rotational and dipolar relaxation response that peaks near 20GHz and falls smoothly to zero at much lower frequencies. 2.45GHz sits well below the peak, in the low-frequency tail of that broad band. The choice of 2.45GHz is an engineering compromise between ISM-band availability and useful penetration depth, not a tuning to any resonance.

- (c) Inside the oven cavity, the microwave reflects off the metal walls, and the incident and reflected waves combine to form a standing wave with nodes (zero electric field) at fixed positions and antinodes (maximum electric field) between them. As established in the section on resonance in pipes, the spacing between adjacent nodes of a standing wave is half the wavelength of the travelling wave:

$$\text{node spacing} = \frac{\lambda}{2} = \frac{0.122 \text{ m}}{2} = 0.061 \text{ m} \approx 6 \text{ cm}$$

The cavity therefore has hot spots (at the antinodes, where the electric field is large and the heating intense) separated by cold spots (at the nodes, where the field is zero and no heating occurs) approximately 6cm apart. A piece of food sitting in a single position in the oven would have parts inside hot spots and parts inside cold spots, and would heat unevenly. The standard engineering fix is to install a rotating turntable, which spins the food slowly during cooking so that every part of the food spends roughly equal time in hot and cold spots, and the heating averages out across the meal. A microwave oven without a turntable would cook food in a polka-dot pattern, a fact that can be demonstrated with chocolate or marshmallows in any oven by switching the turntable off and observing where the melt-spots appear.

Making sense of the answer: Every part of this question is grounded in physics already established in this chapter. Part (a) is the basic wave relation $c = f\lambda$, the same relation that gave the wavelength of the TANESCO 6GHz microwave in REAL Example 15. Part (b) corrects a common physics misconception by appealing to penetration depth. Part (c) reuses the standing-wave node spacing rule $\lambda/2$ from earlier in this chapter on resonance in pipes, applied to a closed metal cavity instead of an open pipe. A microwave oven turns out to be one of the most thoroughly Form-Five-physics-explainable appliances in a Tanzanian kitchen.

Think like a physicist: When a physics question describes a real device and asks why it is built the way it is, the answer is almost never “because that is the resonant frequency.” Engineers think in terms of compromises across several constraints: cost, availability of components, regulatory permission, performance, safety, and ease of manufacture. The 2.45GHz choice for microwave ovens is a textbook example of an engineering compromise being mis-told as a physics resonance by simplification. Whenever you read a one-line explanation of a real engineered device, ask what compromise it actually represents. The honest physics answer is usually more interesting than the textbook simplification.

This section has catalogued the entire electromagnetic family by frequency, named each band, identified the typical production and detection methods, and grounded each band in at least one Tanzanian context. The next section turns from the abstract spectrum back to a concrete acoustic question that Chapter 9 has been carrying since its opening dialogue: *when an ambulance passes the listener at speed, sounding its siren, what exactly does the siren's frequency do, and why?* The answer is the Doppler effect, and the next section will derive it from first principles for sound, before applying the same principle to electromagnetic waves and to the redshift of distant galaxies in the section after that.

THE DOPPLER EFFECT FOR SOUND

The opening dialogue of this chapter ended with Mr. Akilikubwa writing a single equation on the blackboard at Jangwani Secondary School,

$$f' = f \times \frac{v \pm u_o}{v \mp u_s}$$

and promising that the next seventy pages would unpack it. The previous four sections have built the toolkit. Sound is a longitudinal mechanical wave in air at a speed of about 340m/s; the speed is set by the medium alone, not by the source; air columns ring at their natural frequencies; the ear hears intensity logarithmically and the eye sees a band of electromagnetic radiation that obeys, in vacuum, exactly the same wave equation. With the toolkit in hand, the chapter must now keep its promise and explain Mr. Akilikubwa's equation, case by case, from the principle that decides it. It must also answer the question Kipute left hanging on the table at the end of the introduction: *when the ambulance siren produces the same number of beeps per second whether the vehicle is parked outside the school gate or racing to Muhimbili, what exactly is the universe doing to make the note rise as it comes and drop as it goes?*

The answer is the **Doppler effect**, named after the Austrian physicist Christian Doppler who described it for sound and for starlight in a paper of 1842. The principle behind every form of the effect, from the ambulance siren on United Nations Road to the redshift of a galaxy in the constellation of Centaurus, can be stated in one statement. Read it once, slowly, before going on.

*In the rest frame of the medium, the wave speed is fixed by the medium alone. Source motion changes the **wavelength** the source lays down behind it. Observer motion changes the **encounter rate** at which the observer meets the wavefronts. The apparent frequency is the wave speed seen by the observer divided by the wavelength as set by the source.*

That statement is the entire physics of the Doppler effect. Every formula in this section follows from it by careful bookkeeping. We will derive the five primary cases, list the three no-Doppler cases (where source

motion and observer motion conspire to leave the apparent frequency unchanged), then return to Mr. Akilikubwa's equation and show how the same single formula contains all of them.

The source moves, the observer is at rest

Imagine a stationary source S of sound in still air, emitting a steady tone of frequency f . The source plants a new wavefront into the air every $T = 1/f$ seconds. Each wavefront then propagates outwards through the air at speed v , expanding as a sphere centred on the point where the source emitted it. Adjacent wavefronts are separated by one wavelength,

$$\lambda = vT = \frac{v}{f}$$

so that exactly f wavefronts pass any stationary point in the air every second. The picture is symmetric: the wavefronts form a set of concentric circles around the source, in any plane through the source.

Now let the source move. Let the source travel in a straight line through the air at constant speed u_s , while still emitting wavefronts at the same rate f per second. The source is no longer at the same place when it emits the second wavefront as it was when it emitted the first. In the time interval T between two successive emissions, the source has moved a distance $u_s T$ in the direction of its motion. The second wavefront is therefore emitted from a point that is closer to any observer directly in front of the source by exactly that distance.

The geometric consequence is immediate. In front of the source, the wavefronts are closer together than they would have been from a stationary source. Behind the source, they are farther apart. The wave speed in the air is unchanged at v (the medium decides the speed; the source's motion does not), but the wavelength in front is

$$\lambda' = \lambda - u_s T = \frac{v}{f} - \frac{u_s}{f} = \frac{v - u_s}{f}$$

and the wavelength behind is

$$\lambda' = \lambda + u_s T = \frac{v + u_s}{f}$$

A stationary observer directly ahead of the approaching source therefore meets wavefronts of wavelength $(v - u_s)/f$, propagating at the air speed v , and so hears the apparent frequency:

$$f' = \frac{v}{\lambda'} = \frac{v}{(v - u_s)/f} = f \times \frac{v}{v - u_s}$$

A stationary observer directly behind the receding source meets wavefronts of wavelength $(v + u_s)/f$, propagating at the same air speed, and hears:

$$f' = f \times \frac{v}{v + u_s}$$

In words: a source moving towards an observer raises the apparent frequency (the denominator shrinks), and a source moving away from an observer lowers it (the denominator grows). The siren of the ambulance racing from Kariakoo to Muhimbili rises in pitch as it approaches the Jangwani classroom and drops as it recedes towards the hospital, by exactly this mechanism. The “bunching of wavefronts in front” that Khadija reached for in the chapter's opening dialogue was the correct intuition. What Kipute pressed her for, and what Khadija could not yet quite put into words, was the bookkeeping: *the source moves between two successive emissions, and the wavefront the source has just emitted is therefore born from a point closer to the observer than the previous one was.*

The observer moves, the source is at rest

Now reverse the picture. Let the source be at rest in still air, emitting wavefronts at rate f per second; the wavefronts have wavelength $\lambda = v/f$ and propagate outwards at the air speed v in all directions. The observer, no longer stationary, walks towards the source at speed u_o .

The wavelength of the wave in the air is what the source laid down, $\lambda = v/f$, and the source's stationary condition means that wavelength is the same in every direction. What changes is the speed at which the observer meets successive wavefronts. The observer moving towards the source at u_o sees each wavefront approaching her at $v + u_o$ (the wavefront's speed in the air plus the observer's speed through the air, both

measured in the same direction). The rate at which she crosses wavefronts is the encounter speed divided by the spacing,

$$f' = \frac{v + u_o}{\lambda} = \frac{v + u_o}{v/f} = f \times \frac{v + u_o}{v}$$

If the observer instead walks away from the source at speed u_o , each wavefront catches her at the slower speed $v - u_o$, and she meets them at the lower rate:

$$f' = f \times \frac{v - u_o}{v}$$

An observer moving towards a stationary source therefore hears a raised frequency, and an observer moving away from a stationary source hears a lowered frequency, by the same physical sign as for a moving source. But the algebra is now different: the moving observer changes the **encounter rate** rather than the wavelength.

The general formula: both source and observer in motion

When both source and observer move through the air, both effects operate simultaneously. The source, moving at u_s relative to the air, alters the wavelength it lays down; the observer, moving at u_o relative to the air, alters the encounter speed at which she meets those altered wavefronts. To arrive at the general formula, we first re-state each of the two component effects on its own, and then combine them step by step.

Source motion alters the wavelength. From the source-moving subsection, the wavelength the source plants in the air, measured in the rest frame of the air, is:

$$\lambda' = \frac{v - u_s}{f} \quad \text{when the source moves towards the observer}$$

$$\lambda' = \frac{v + u_s}{f} \quad \text{when the source moves away from the observer}$$

These two cases can be written as a single expression using a paired-sign symbol,

$$\lambda' = \frac{v \mp u_s}{f}$$

in which the **upper sign (the minus)** applies when the source moves *towards* the observer, shortening the wavelength, and the **lower sign (the plus)** applies when the source moves *away* from the observer, lengthening it.

Observer motion alters the encounter speed. From the observer-moving subsection, the speed at which the observer meets successive wavefronts in the air is:

$$v' = v + u_o \quad \text{when the observer moves towards the source}$$

$$v' = v - u_o \quad \text{when the observer moves away from the source}$$

Combining the two cases into a single expression,

$$v' = v \pm u_o$$

in which the **upper sign (the plus)** applies when the observer moves *towards* the source, raising the encounter speed, and the **lower sign (the minus)** applies when the observer moves *away* from the source, lowering it.

Combining the two effects: The apparent frequency the observer hears is the encounter speed divided by the wavelength, both measured in the rest frame of the air:

$$f' = \frac{v'}{\lambda'} = \frac{v \pm u_o}{(v \mp u_s)/f} = f \times \frac{v \pm u_o}{v \mp u_s}$$

This is the equation Mr. Akilikubwa wrote on the Jangwani blackboard, derived now from the principle that opened this section.

How to apply the sign convention: The notation $\pm$ and $\mp$ encodes two independent sign choices, one for the observer in the numerator and one for the source in the denominator. The two choices are read separately, from the motion of each.

In the **numerator**, take the **plus** sign if the observer moves *towards* the source, and the **minus** sign if the observer moves *away* from the source. In the **denominator**, take the **minus** sign if the source moves *towards* the observer, and the **plus** sign if the source moves *away* from the observer. The single mnemonic that captures both rules is: *towards* always pushes the apparent frequency upwards (whether it is the source or the observer doing the moving), and *away* always pushes it down.

The choice of $\pm$ in the numerator and $\mp$ in the denominator is the formula's quiet way of reminding the student that the two halves of the bar are inverted with respect to each other. When the numerator is positive ($v + u_o$, observer approaching), the denominator is negative ($v - u_s$, source approaching); the two choices then push the apparent frequency upwards together. When the numerator is negative ($v - u_o$, observer receding), the denominator is positive ($v + u_s$, source receding); the two choices then push the frequency downwards together. The four physical scenarios are shown in the table below.

Observer motion	Source motion	Numerator	Denominator	Effect on f'
towards	towards	$v + u_o$	$v - u_s$	highest possible
towards	away	$v + u_o$	$v + u_s$	depends on speeds
away	towards	$v - u_o$	$v - u_s$	depends on speeds
away	away	$v - u_o$	$v + u_s$	lowest possible

For the two “same direction” cases (both towards each other, or both away from each other) the numerator and denominator pull the apparent frequency in the same direction and the answer is decided in sign. For the two “mixed” cases (one towards, one away), the two effects compete and the answer depends on the relative size of u_s and u_o .

Five further features of the general formula deserve emphasis.

1) The source and observer speeds are measured against the medium, not against each other

The Doppler effect for sound is not a function of relative velocity alone; it is a function of source speed through the air and observer speed through the air, separately. Two cases that look the same from a “relative motion” perspective give different apparent frequencies. (This will become important in REAL Example 21, part (c).) The medium decides everything in the rest frame of the medium.

2) The wave speed v that appears in the formula is the wave speed in the rest frame of the medium

This is the same speed that featured in $c = f\lambda$ throughout Chapter 7. In still air at 20°C that speed is approximately 340m/s. If the air itself moves (a wind blows), the wave speed in the ground frame changes, and a correction is needed; we will examine the wind correction later.

3) The formula applies along the source-to-observer line

If the source moves at an angle to that line, the relevant velocity is the **component along the line**, $u_s \cos\theta$. The same applies to the observer. We treat this oblique case in worked Example 22.

4) The formula is exact only for source and observer speeds well below the wave speed, $u_s, u_o \ll v$

At source speeds approaching v (the speed of sound), the wavefronts pile up into a shock wave and the formula breaks down; at $u_s = v$ the formula's denominator vanishes and the apparent frequency would diverge. This regime, the “sonic boom” of a supersonic aircraft, is qualitatively different physics and is not part of this section.

5) The formula reduces to its four simpler limiting cases by setting one of u_s, u_o to zero, and recovers $f' = f$ in the trivial case $u_s = u_o = 0$

A useful sanity check on any calculation is to set the wrong speed to zero and confirm that the remaining formula matches the four sub-cases of the previous two subsections.

The three no-Doppler cases

You may sometimes encounter a situation involving motion and wonder whether a Doppler shift will be heard. The three principal scenarios in which there is **no** Doppler shift, in spite of motion being present somewhere, are these.

1) Both source and observer at rest in the medium

The trivial case. Wavefronts are emitted at the source's rate, propagate at the medium's speed, and meet the observer at exactly the same rate. $u_s = u_o = 0$ and the formula gives $f' = f$.

2) Source and observer moving at the same velocity through the medium

The geometric consequence is that in the rest frame of either of them, both are stationary; the medium is what moves, and the formulas of the previous subsections, derived in the medium's rest frame, give a non-trivial apparent frequency only when source and observer move differently through the medium. The cleanest example is a passenger sitting inside a moving daladala that is sounding its horn: the source (the horn) and the observer (the passenger) both move with the daladala at the same speed through the air, so neither experiences a Doppler shift relative to the other. The passenger inside the daladala hears the horn at exactly its source frequency, regardless of how fast the daladala travels. We exercise this insight in REAL Example 20, part (c).

3) Source moving along a circle of which the observer is the centre

The source's distance from the observer is constant; the wavefronts the source emits travel outwards along radii of equal length and arrive at the observer at the same rate they are emitted. There is no Doppler shift. (Equivalently, the source's velocity at every instant is perpendicular to the source-to-observer line, so the component along that line is zero, and the oblique formula gives $f' = f$.)

A common student trap is the case of an observer running across a railway platform parallel to a train that is whistling and standing still. The observer's velocity has a component along the source-to-observer line (zero if the observer is exactly abreast of the train, but nonzero when the observer is upstream or downstream). The shift is therefore nonzero at most points along the trajectory and only vanishes at the single instant when the observer is exactly perpendicular to the source. The "perpendicular at closest approach" version of no-Doppler is an instantaneous condition, not a sustained one.

We have now derived every formula needed for the first three worked examples. The fourth covers oblique Doppler, which uses the same source-moving formula with the source velocity component replaced by $u_s \cos \theta$.

BINDER Example 19

A train sounds a whistle of frequency 400Hz, approaching a railway platform at 10m/s. Take the speed of sound in still air as 340m/s. Find the frequency a stationary platform observer hears when the train is:

- (a) approaching,
- (b) receding.

Solution

(a) Apply the source-approaching formula with $u_s = 10\text{m/s}$, $f = 400\text{Hz}$, and $v = 340\text{m/s}$:

$$f' = f \times \frac{v}{v - u_s} = 400\text{Hz} \times \frac{340\text{m/s}}{340\text{m/s} - 10\text{m/s}} = 412.1\text{Hz}$$

(b) Apply the source-receding formula with the same numbers, switching the sign in the denominator:

$$f' = f \times \frac{v}{v + u_s} = 400\text{Hz} \times \frac{340\text{m/s}}{340\text{m/s} + 10\text{m/s}} = 388.6\text{Hz}$$

Dig deeper: The total swing in pitch as the train passes, from the approaching note to the receding note, is

$$412.1\text{Hz} - 388.6\text{Hz} = 23.5\text{Hz}$$

a clearly audible step that the platform observer would describe as the whistle "dropping its tone" the instant the train passes.

Making sense of the answer: The shifts are roughly ± 3 per cent of the source frequency, the size of a small musical interval. A 23.5Hz swing on a 400Hz tone is about three semitones, just under a quarter-tone, well inside the resolution of any human ear. The asymmetry between the approaching and receding values is a hint that source motion is not purely a relative-velocity effect: $412.1\text{Hz} - 400\text{Hz} = 12.1\text{Hz}$ upwards on approach is not quite the same as $400\text{Hz} - 388.6\text{Hz} = 11.4\text{Hz}$ downwards on recede. The two differ at second order in u_s/v , an asymmetry we return to in REAL Example 21.

Think like a physicist: Whenever the source speed u_s is small compared with the wave speed v , both formulas reduce to the same first-order approximation $f' \approx f(1 \pm u_s/v)$. The student should learn to

recognise this regime: at speeds well below the speed of sound (which is to say almost every speed encountered in everyday life), Doppler shifts are linear in source speed and approximately symmetric in sign. The non-linear regime, where the approach and recede shifts differ noticeably, sets in only when source speed becomes a substantial fraction of the wave speed, and is the regime where shock waves and sonic booms eventually live.

REAL Example 20

A daladala on the Nyerere road sounds its horn, of true frequency 600Hz, while travelling at 25m/s along a straight section of road. A pedestrian standing on the grass verge hears the horn as the daladala approaches her and again as it recedes. Take the speed of sound in still air as 340m/s.

- Compute the apparent frequency the pedestrian hears before the daladala passes her, and the apparent frequency she hears after it has passed.
- Find the swing in pitch, and express it as a percentage of the true frequency of the horn.
- A passenger sitting inside the daladala hears the horn the whole time. Explain to the passenger, using the principle that governs Doppler effect, why she hears the horn at its true frequency of 600Hz at every moment, regardless of how fast the daladala is travelling.

Solution

- Apply the source-moving formulas to a stationary observer.

Before the daladala passes the pedestrian, the source approaches her at $u_s = 25\text{m/s}$:

$$f' = f \times \frac{v}{v - u_s} = 600\text{Hz} \times \frac{340\text{m/s}}{340\text{m/s} - 25\text{m/s}} = 647.6\text{Hz}$$

After the daladala passes, the source recedes from her at the same $u_s = 25\text{m/s}$:

$$f' = f \times \frac{v}{v + u_s} = 600\text{Hz} \times \frac{340\text{m/s}}{340\text{m/s} + 25\text{m/s}} = 558.9\text{Hz}$$

- The swing is the difference between the two apparent frequencies:

$$\Delta f = 648\text{Hz} - 559\text{Hz} = 89\text{Hz}$$

As a percentage of the true source frequency,

$$\frac{\Delta f}{f} = \frac{89\text{Hz}}{600\text{Hz}} = 0.148 = 14.8\%$$

The pedestrian hears the horn drop by almost fifteen per cent of its source frequency the instant the daladala passes her, the largest single-event pitch change a pedestrian standing on a Tanzanian road verge ever hears.

- The passenger inside the daladala shares the daladala's motion through the air. The horn, mounted on the daladala, also shares that motion. The horn and the passenger are therefore in the same rest frame: in their shared frame, neither is moving. The principle states that the apparent frequency is the wave speed seen by the observer divided by the wavelength as set by the source, both measured in the rest frame of the medium. In the passenger's frame, source and observer are at rest relative to each other, the air is what moves (backwards at 25m/s relative to the daladala), and the formal sub-case, "source and observer at the same velocity through the medium" forms no-Doppler effect scenario. The result is:

$$f' = f \times \frac{v + u_o}{v + u_s} = f \times \frac{v - u_s}{v - u_s} = f$$

(where the same speed has been substituted for u_s and u_o , with the same sign, because source and observer move identically through the air). The passenger hears 600Hz, not 648Hz or 559Hz, regardless of how fast the daladala travels. The Doppler effect requires that source and observer move **differently** through the medium; identical motion through the medium produces no shift.

Making sense of the answer: Parts (a) and (b) give the largest pitch change in any of this section's worked examples, because the source speed of 25m/s is a sizeable seven per cent of the speed of sound. The asymmetry between the approach and recede shifts is now visible: $648\text{Hz} - 600\text{Hz} = 48\text{Hz}$ upward versus $600\text{Hz} - 559\text{Hz} = 41\text{Hz}$ downward. Part (c) is the conceptual punchline. A common student instinct is to think the passenger should hear a Doppler shift because the daladala is moving, but the moving daladala

only matters relative to a stationary observer in the air outside; relative to the passenger, the horn is stationary. The Doppler effect is about motion through the medium, not about motion itself.

Think like a physicist: Two motions that produce the same relative velocity between source and observer can produce different Doppler shifts. The Doppler effect is asymmetric in source motion versus observer motion for a wave that travels in a medium; symmetry is only restored for a wave that travels in vacuum, namely electromagnetic radiation, which we treat in the next section. The asymmetry is a fingerprint of the medium: any wave that propagates in a medium carries this fingerprint, and any wave that propagates in vacuum does not.

REAL Example 21

A Form Five student is cycling along the bus turnaround at Jangwani Secondary School at a steady 5m/s. Parked at the far end of the turnaround is a stationary Land Cruiser ambulance whose siren is set to a steady test tone of 800Hz for an electrical check. The student cycles directly towards the parked ambulance, passes it without slowing, and continues directly away from it on the other side. Take the speed of sound in still air as 340m/s.

- Compute the apparent frequency the cyclist hears while approaching the ambulance.
- Compute the apparent frequency she hears while receding from the ambulance.
- Compute, in two ways, the apparent frequency a stationary observer on the school field would hear if instead the ambulance were **moving** towards the field at 5m/s while the observer stood still. Compare the two predictions and comment on whether they agree.

Solution

(a) The source (the parked ambulance) is at rest in the air. The observer (the cyclist) moves towards the source at $u_o = 5\text{m/s}$. Apply the observer-approaching formula:

$$f' = f \times \frac{v + u_o}{v} = 800\text{Hz} \times \frac{340\text{m/s} + 5\text{m/s}}{340\text{m/s}} = 811.76\text{Hz}$$

(b) The cyclist continues past the ambulance and recedes from it at the same speed. Apply the observer-receding formula:

$$f' = f \times \frac{v - u_o}{v} = 800\text{Hz} \times \frac{340\text{m/s} - 5\text{m/s}}{340\text{m/s}} = 788.24\text{Hz}$$

(c) Now reverse the geometry. The ambulance moves towards the observer on the field at $u_s = 5\text{m/s}$, while the observer stands still. Apply the source-approaching formula:

$$f'_{\text{source-moving}} = f \times \frac{v}{v - u_s} = 800\text{Hz} \times \frac{340\text{m/s}}{340\text{m/s} - 5\text{m/s}} = 811.94\text{Hz}$$

Compare with part (a):

$$f'_{\text{observer-moving}} = 811.76\text{Hz}$$

$$f'_{\text{source-moving}} = 811.94\text{Hz}$$

The two answers are in close agreement, **differing by about 0.2Hz**. The source-moving case gives a slightly larger upward shift than the observer-moving case, for the same relative speed of 5m/s. The percentage difference is:

$$\frac{0.2}{812} = 0.00022 = 0.022\%$$

So the two scenarios are **not** exactly equivalent.

Making sense of the answer: The 0.2Hz discrepancy is a real physical fact, not a rounding error. The source-moving formula has v in the numerator and $v - u_s$ in the denominator; the observer-moving formula has $v + u_o$ in the numerator and v in the denominator. The ratio between them, holding $u_s = u_o = u$, is

$$\frac{f'_{\text{source-moving}}}{f'_{\text{observer-moving}}} = \frac{v/(v - u)}{(v + u)/v} = \frac{v^2}{(v - u)(v + u)} = \frac{v^2}{v^2 - u^2}$$

which exceeds 1 by approximately u^2/v^2 for $u \ll v$. With $u = 5\text{m/s}$ and $v = 340\text{m/s}$, the discrepancy is $(5/340)^2 \approx 2.2 \times 10^{-4}$, exactly the 0.022 per cent we found. The Doppler effect for sound is genuinely asymmetric in source motion versus observer motion, even at the same relative speed.

Think like a physicist: When two formulas almost agree but differ at second order, the disagreement carries information. The information here is that sound has a preferred frame: the rest frame of the air. Source motion and observer motion are physically distinguishable, and a careful measurement of f' at known u can in principle tell the experimenter whether it is the source or the observer that is moving. For light propagating in vacuum, by contrast, no such preferred frame exists; the Doppler formula for light is symmetric in source-versus-observer.

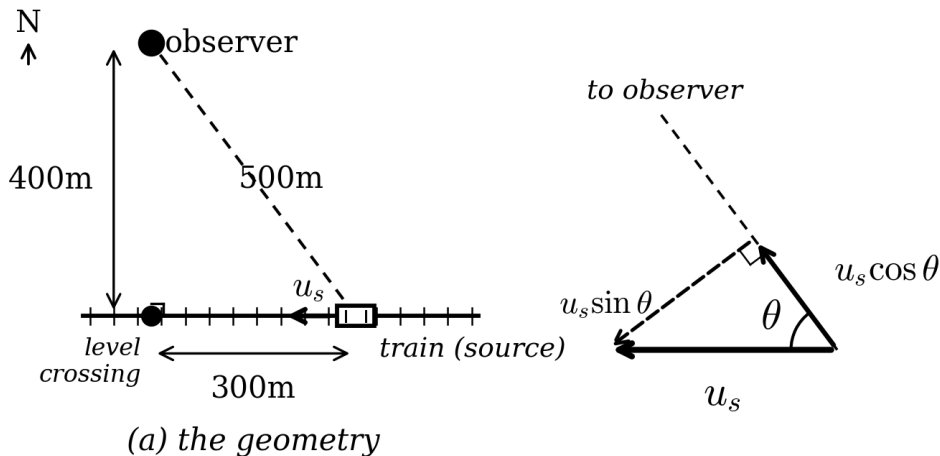
HOT Example 22

A train approaches a level crossing at 120km/h along an east-west track. At the instant of interest, the train is exactly 300m east of the level crossing, with its whistle blowing a steady 640Hz tone. An observer is standing 400m north of the level crossing on the access road. Take the speed of sound in still air as 340m/s .

Find the apparent frequency the observer hears at this instant.

Solution

The figure below shows the setup.



(b) the velocity decomposition

Figure: Geometry for HOT Example 22. Panel (a): the train (source) is at 300m east of the level crossing; the observer is at 400m north of the level crossing; the source-to-observer line is the hypotenuse of a 3-4-5 right triangle of length 500m . Panel (b): the train's velocity u_s decomposes into a component $u_s \cos \theta$ along the source-to-observer line (the only component that affects the Doppler shift) and a component $u_s \sin \theta$ perpendicular to it (which does not).

Convert the train's speed to SI units:

$$u_s = 120\text{km/h} = \frac{120000\text{m}}{3600\text{s}} = 33.33\text{m/s}$$

Find the source-to-observer distance:

From panel (a), the train, the level crossing, and the observer form a right triangle with legs 300m (east-west) and 400m (north-south). By Pythagoras, the hypotenuse is:

$$r = \sqrt{(300\text{m})^2 + (400\text{m})^2} = 500\text{m}$$

This is a 3-4-5 right triangle; the source-to-observer distance is 500m .

Find $\cos\theta$:

The angle θ in panel (b) is the angle between the train's velocity (which points along the rails, westwards) and the source-to-observer line (which points from the train towards the observer). By inspecting the triangle in panel (a), θ is also the angle at the train inside the right triangle, between the east-west rail line and the hypotenuse. The cosine of that angle is the ratio of the adjacent leg (300m, along the rails) to the hypotenuse (500m, to the observer):

$$\cos\theta = \frac{\text{adjacent}}{\text{hypotenuse}} = \frac{300\text{m}}{500\text{m}} = 0.6$$

Find the component of velocity along the source-to-observer line:

Multiply the train's full speed by $\cos\theta$:

$$u_s \cos\theta = 33.33\text{m/s} \times 0.6 = 20.00\text{m/s}$$

This is the rate at which the train is closing the distance to the observer at this instant. The other component of the train's velocity, namely $u_s \sin\theta = 33.33\text{m/s} \times 0.8 = 26.66\text{m/s}$ (shown dashed in panel b), points sideways relative to the source-to-observer line. This perpendicular component carries the train across the line of sight rather than along it, so it does **not** contribute to the Doppler shift.

Apply the source-approaching Doppler formula:

Use $u_s \cos\theta$ in place of u_s :

$$f' = f \times \frac{v}{v - u_s \cos\theta} = 640\text{Hz} \times \frac{340\text{m/s}}{340\text{m/s} - 20.00\text{m/s}} = 680\text{Hz}$$

The observer hears the train's whistle at exactly 680Hz at this instant.

Making sense of the answer: *The component of the train's velocity along the source-to-observer line at this instant is 20m/s. That component, not the full 33.33m/s, is what enters the Doppler formula. The train is at 120km/h on the track, but only 72km/h "towards the observer" at this moment, because of the angle. The Doppler shift is therefore smaller than it would be if the observer stood directly on the track.*

Think like a physicist: *Doppler shift for an oblique source is governed by the velocity component along the source-to-observer line, not by the total speed. This generalises to a useful one-liner: the part of any motion that does not bring source and observer closer or further apart does not contribute to the Doppler shift. At the single instant when source and observer move exactly perpendicular to the line between them, the component along the line is zero and the apparent frequency equals the source frequency. This is the third of the three no-Doppler scenarios listed above, expressed in geometric language.*

This concludes the primary cases and their worked examples. The second half of this section turns to the two remaining complications, namely a wind blowing through the medium and a wave reflecting off a moving target.

Wind: the medium itself in motion

So far we have assumed the air stands still. Air rarely does. A sea breeze pushes inland from Bagamoyo every afternoon; the Maasai Mara highlands feed cool air down the Rift Valley by evening; the long-range monsoon flows north over Tanzania in summer and south in winter. Whenever the medium itself moves, the Doppler formula must be corrected, but the correction is small, intuitive, and easy to apply once stated properly.

The principle behind the correction is this. The wave speed v that appears throughout the still-air formula is the wave speed in the **rest frame of the medium**. When wind blows, the air is no longer at rest with respect to the ground, and the wave speed in the **rest frame of the ground**, where source and observer actually live and where their speeds are measured, is no longer v . The wind sweeps the air along with itself, carrying the wave with it, and the wave's ground speed in the source-to-observer direction picks up an extra contribution from the wind's velocity along that same direction.

Let $\mathbf{w}$ denote the component of the wind's velocity along the line from source to observer, signed so that $\mathbf{w} > 0$ means the wind blows *from the source toward the observer* and $\mathbf{w} < 0$ means the wind blows from the observer toward the source. The wave's ground speed in the source-to-observer direction is then:

$$\mathbf{V} = \mathbf{v} + \mathbf{w}$$

This V , not the still-air v , is what should appear in the Doppler formula when the experimenter chooses to measure the source and observer speeds in the ground frame (which is the natural choice). The source and observer velocities u_s and u_o continue to be measured as ground speeds, with the same sign convention as before. The Doppler formula with wind correction becomes:

$$f' = f \times \frac{V \pm u_o}{V \mp u_s}, \quad \text{where: } V = v + w$$

A wind of 30m/s in the source-to-observer direction lifts the effective wave speed from 330m/s to 360m/s. A wind of 30m/s from the observer toward the source drops the effective wave speed to 300m/s. A wind blowing perpendicular to the source-to-observer line has zero component along that line, gives $w = 0$, and leaves the formula unchanged. Only the parallel component matters.

A subtle and counter-intuitive consequence of the wind correction is that a wind blowing in the same direction as the source's motion (a tailwind for the source) actually **reduces** the size of the Doppler shift. The reason is that the *wind makes wavefronts travel faster in the ground frame, so the source has less time to "catch up" with the wavefront it has just emitted, and the wavelength compression is therefore less severe.* Conversely, a headwind blowing back against the source's motion **increases** the Doppler shift, because the *wave moves more slowly in the ground frame and the source catches up with its own wavefronts more aggressively.* HOT Example 23 below exercises this subtlety with a worked numerical example at the Tabora railway station.

Round-trip Doppler: when the wave comes back

A different complication arises when the wave does not simply travel from source to observer but bounces off a reflector and returns. The reflector might be a cliff face (for the wave from a moving siren), a moving moth (for the chirp from a bat), a speeding car (for the radio pulse from a police speed gun), a moving submarine (for the pulse from a ship's sonar), or a moving red blood cell (for the ultrasound pulse from a medical Doppler probe). In each of these scenarios the original emitter eventually hears its own wave again, modified twice over.

The procedure is to apply the Doppler formula **twice in succession**, treating the reflector as a relay.

Step 1: *The reflector is the receiver of the wave coming from the original source.*

Apply the Doppler formula with the original source as the source and the reflector as the observer. Call the apparent frequency the reflector "hears" f_1 .

Step 2. *The reflector immediately re-emits whatever wave it has received, at the same frequency f_1 , in the reverse direction.*

The reflector is now the **source**, emitting f_1 , and the original emitter (now playing the role of receiver) is the **observer**. Apply the Doppler formula again with the reflector as the source emitting f_1 and the original emitter as the observer. The frequency the original emitter ultimately hears is the round-trip Doppler-shifted frequency.

The reflector itself need not be stationary. If it is moving, then both Step 1 and Step 2 involve a moving observer (Step 1) or moving source (Step 2), and the formula must be applied with the reflector's velocity in mind for each step. The chain double-Doppler shift is approximately twice as large as a single-source Doppler shift for the same target velocity, a fact which the bat in HOT Example 24 will exploit to identify the velocity of its dinner with surprising precision, and which the police-speed-gun operator on the Morogoro Road in the applications section coming after the next one will exploit to issue the speeding ticket the chapter introduction warned the student about.

We are now equipped with all the tools required for the worked examples that follow.

HOT Example 23

A locomotive approaching the Tabora railway station sounds a horn of true frequency 660Hz, travelling at a ground speed of 30m/s towards a stationary passenger waiting on the platform. Take the speed of sound in still air as 330m/s.

(a) Compute the apparent frequency the passenger hears in still air.

- (b) On a different day, a steady tailwind blows at 30m/s in the direction from the locomotive towards the platform, while the rest of the situation is unchanged. Compute the apparent frequency the passenger hears.
- (c) On a third day, a steady crosswind of 30m/s blows perpendicular to the rail line, while the rest of the situation is again unchanged. Compute the apparent frequency the passenger hears, and explain in one sentence why.

Solution

(a) Still air, source approaching observer:

$$f' = f \times \frac{v}{v - u_s} = 660\text{Hz} \times \frac{330\text{m/s}}{330\text{m/s} - 30\text{m/s}} = 726\text{Hz}$$

The passenger hears the horn at 726Hz.

(b) The wind blows from the locomotive (the source) towards the platform (the observer), so the wind's component along the source-to-observer direction is $w = +30\text{m/s}$. The effective wave speed in the ground frame is:

$$V = v + w = 330\text{m/s} + 30\text{m/s} = 360\text{m/s}$$

Apply the wind-corrected Doppler formula, with the source speed and observer speed unchanged from part (a):

$$f' = f \times \frac{V}{V - u_s} = 660\text{Hz} \times \frac{360\text{m/s}}{360\text{m/s} - 30\text{m/s}} = 720\text{Hz}$$

The apparent frequency in the tailwind is 720Hz, only 6Hz **lower** than the still-air answer. The tailwind has **reduced** the Doppler shift from 66Hz above the source to 60Hz above.

(c) A crosswind blows perpendicular to the rail line, that is to say perpendicular to the source-to-observer direction. Its component along the source-to-observer direction is zero:

$$w = 0, \quad V = v + 0 = 330\text{m/s}$$

The wind-corrected formula therefore reduces to the still-air formula, and the apparent frequency the passenger hears is:

$$f' = 726\text{Hz}$$

exactly the same as in part (a).

Only the wind's parallel component contributes to the Doppler correction; the perpendicular component has no effect.

Making sense of the answer: Part (a) gives the baseline still-air Doppler shift, +66Hz upward. Part (b) gives the surprising result that a tailwind in the direction the source is moving reduces the shift, to +60Hz. The intuition is that the tailwind speeds up the wavefronts in the ground frame (they now propagate at 360m/s instead of 330m/s), so the source has less time to "catch up with" the wavefront it has just emitted, and the resulting wavelength compression is less severe. A headwind would do the opposite: the wave slows down in the ground frame, the source catches up more aggressively, and the Doppler shift grows. Part (c) confirms that only the parallel component of the wind matters; a perpendicular wind, however strong, has no effect on the apparent frequency.

Think like a physicist: The wind correction in this section is the simplest possible: the wind blows along or across the source-to-observer line, and the velocity decomposes cleanly into one parallel component (which contributes) and one perpendicular component (which does not). **For winds at intermediate angles, the same principle applies with $\cos \theta$: the contributing component of the wind is $w \cos \theta$, where θ is the angle between the wind direction and the source-to-observer line.** A wind blowing at 45 degrees to the rail line, at the same 30m/s magnitude, would contribute about $30 \cos 45^\circ = 21\text{m/s}$ of effective ground-speed adjustment, giving an apparent frequency somewhere between the parallel-wind and zero-wind results.

HOT Example 24

A bat hunting inside one of the limestone caves at Kondoa emits an ultrasonic chirp at a frequency of 40kHz. A moth, the bat's intended dinner, is flying through the cave at 5m/s. The cave air is still and at body temperature, with the speed of sound 340m/s.

- The moth is flying directly towards the bat at 5m/s. Find the frequency at which the bat receives its own chirp back after the chirp has reflected from the moth. State the chain of Doppler shifts that produces the answer.
- The bat now hovers in place and watches a second moth fly directly away from it at 5m/s. Find the frequency at which the bat receives the reflected chirp from this second moth. Compare with part (a) and explain how the bat tells the two moths apart.

Solution

(a) Apply the Doppler formula twice in succession, with the moth acting as a relay.

Step 1: The bat is the source (stationary in the cave air), the moth is the observer (moving towards the bat at $u_m = 5\text{m/s}$).

The moth's velocity is along the source-to-observer line, so the simple observer-moving formula applies:

$$f_1 = f_0 \times \frac{v + u_m}{v} = 40000\text{Hz} \times \frac{340\text{m/s} + 5\text{m/s}}{340\text{m/s}} = 40588\text{Hz}$$

This is the frequency the moth, were it equipped with a bat-ear, would "hear" the chirp at.

Step 2: The moth re-emits the chirp at f_1 , this time playing the role of source moving towards the bat at the same $u_m = 5\text{m/s}$.

The bat is now the stationary observer. Apply the source-moving formula:

$$f_2 = f_1 \times \frac{v}{v - u_m} = 40588\text{Hz} \times \frac{340\text{m/s}}{340\text{m/s} - 5\text{m/s}} = 41194\text{Hz}$$

The bat receives the reflected chirp at 41194Hz, which is 1194Hz above the chirp it originally emitted.

Alternative solution

The two factors combine into a single expression for the *round-trip Doppler shift*,

$$f_2 = f_1 \times \frac{v}{v - u_m} = f_0 \times \frac{v + u_m}{v} \times \frac{v}{v - u_m} = f_0 \times \frac{v + u_m}{v - u_m} = 40000\text{Hz} \times \frac{345\text{m/s}}{335\text{m/s}} = 41194\text{Hz}$$

a one-line formula whose numerator and denominator both push the apparent frequency upwards (numerator by the moth-as-observer effect, denominator by the moth-as-source effect).

(b) The moth is now receding from the bat at $u_m = 5\text{m/s}$, the same speed in the opposite direction. Both halves of the round-trip formula flip sign:

$$f_2 = f_0 \times \frac{v - u_m}{v + u_m} = 40000\text{Hz} \times \frac{340\text{m/s} - 5\text{m/s}}{340\text{m/s} + 5\text{m/s}} = 38841\text{Hz}$$

The bat receives the reflected chirp at 38841Hz, which is 1159Hz **below** the chirp it originally emitted.

Comparing the two moths: the approaching moth shifts the returned chirp **up** by about 1.2kHz, and the receding moth shifts the returned chirp **down** by about 1.2kHz. The bat's auditory cortex, which is exquisitely sensitive to small frequency differences in the ultrasonic band, hears the two moths as two distinguishable tones: one above the 40kHz chirp it emitted, one below. The bat reads the **sign** of the shift to know which moth is coming towards it, and reads the **size** of the shift to know how fast.

Making sense of the answer: The round-trip Doppler shift is approximately twice the single-Doppler shift for the same target speed, because the two factors in $(v + u_m)/(v - u_m)$ multiply together. For a moth at 5m/s, the round-trip shift is about 3 per cent of the carrier frequency; for a moth at 1m/s it is about 0.6 per cent. The bat's auditory system can resolve differences in pitch much finer than 0.1 per cent at ultrasonic frequencies, which is why bats can identify a hovering moth from a fast-darting moth in pitch darkness.

Think like a physicist: Every round-trip Doppler problem reduces to "apply the formula twice, with the reflector as a relay." The pattern repeats from acoustics (the bat in this example) to radar (the police speed gun in the applications section to come) to medical ultrasound (the blood-flow probe at Muhimbili). The

mathematics is identical in each case; the wave is different. This is one of the best illustrations in classical physics of why the abstract formalism is worth learning: the formula derived for one situation transfers, intact and unmodified, to a dozen seemingly unrelated technologies.

HOT Example 25

A police car chases a speeder along a straight section of the Morogoro Road towards a tall cliff face that rises perpendicular to the road. Both vehicles travel at 160km/h towards the cliff. The police-car siren emits a steady tone at 100Hz. A second car, driven by an off-duty inspector, follows the police car along the same straight section at 120km/h. Take the speed of sound in air as 330m/s.

Find:

- the frequency at which the inspector hears the police-car siren **directly**, by sound travelling from siren to inspector through the air,
- the frequency at which the inspector hears the **cliff-echo** of the same siren, by sound reflecting off the cliff face and travelling back to the inspector,
- the beat frequency between the two tones the inspector simultaneously hears.

Solution

Convert all speeds to SI units. The police car and the speeder both travel at:

$$u_p = 160\text{km/h} = \frac{160000\text{m}}{3600\text{s}} = 44.44\text{m/s}$$

The inspector's car travels at:

$$u_i = 120\text{km/h} = \frac{120000\text{m}}{3600\text{s}} = 33.33\text{m/s}$$

The geometry: all three vehicles move in the same direction along the road towards the cliff. The inspector's car is behind the police car (because the police car is the chaser, the speeder is ahead of the police, the inspector follows). Sound from the police-car siren travels backwards through the air to reach the inspector, and forwards through the air to reach the cliff.

(a) Direct sound from siren to inspector. The siren (source) is in front of the inspector (observer), and both move in the same direction. The police car, being faster, recedes from the inspector at the ground speed u_p . The inspector, moving in the same direction as the wave's source, approaches the source at the ground speed u_i .

Apply the general Doppler formula. Observer approaches the source: numerator carries a plus sign. Source recedes from the observer: denominator carries a plus sign.

$$f_{\text{direct}} = f \times \frac{v + u_i}{v + u_p} = 100\text{Hz} \times \frac{330\text{m/s} + 33.33\text{m/s}}{330\text{m/s} + 44.44\text{m/s}} = 97.03\text{Hz}$$

The inspector hears the direct siren at about 97Hz, that is to say slightly **below** the source frequency of 100Hz, because the source recedes faster than the observer pursues.

(b) The cliff echo arrives by reflection. Apply Doppler twice with the cliff as a stationary relay.

Step 1: The police car emits the siren and the cliff face is the stationary observer in the air.

Apply the source-approaching formula:

$$f_1 = f \times \frac{v}{v - u_p}$$

Step 2: The cliff re-emits the wave at f_1 , acting as a stationary source. The inspector, now playing the role of observer. Apply the observer-approaching formula:

$$f_{\text{echo}} = f_1 \times \frac{v + u_i}{v} = f \times \frac{v}{v - u_p} \times \frac{v + u_i}{v} = f \times \frac{v + u_i}{v - u_p} = 100\text{Hz} \times \frac{330\text{m/s} + 33.33\text{m/s}}{330\text{m/s} - 44.44\text{m/s}} = 127.21\text{Hz}$$

The cliff-echo arrives at the inspector at 127.21Hz, well above the source frequency of 100Hz. The two factors (police approaching cliff, inspector approaching cliff) both push the echo upwards.

(c) The inspector simultaneously hears two pure tones: the direct siren at 97.0Hz and the cliff-echo at 127.2Hz. Two pure tones close in frequency superpose to produce beats. The beat frequency is the difference between the two tones:

$$f_{\text{beat}} = |f_{\text{echo}} - f_{\text{direct}}| = |127.21\text{Hz} - 97.03\text{Hz}| = 30.18\text{Hz}$$

The inspector hears, riding atop the average tone of $(97 + 127)/2 \approx 112\text{Hz}$, a slow throbbing at 30Hz, that is to say 30 loud-quiet cycles per second, exactly as Chapter 7 derived for the superposition of two close-frequency waves.

Making sense of the answer: *Three physical effects interact in this single problem. First, the direct sound is shifted downwards from 100Hz to 97Hz, because the source recedes from the observer faster than the observer chases the source. Second, the cliff-echo is shifted upwards from 100Hz to 127Hz, because both the original emission and the reflected arrival involve approach motions. Third, the two tones simultaneously incident on the inspector's ear produce a beat at the difference frequency of 30Hz. The Doppler effect, source reflection, and the principle of superposition (the last from Chapter 7) all play together in a single auditory experience: a continuous deep buzz at about 112Hz, throbbing thirty times a second. The inspector, if she is paying attention to the physics, can read the beat rate alone to deduce the closing speed of the police car towards the cliff.*

Think like a physicist: *The richest physical questions are the ones in which three or four pieces of physics that the student has learned separately suddenly turn out to be three or four pieces of the same phenomenon. This is one such question. The Doppler effect of this chapter, the round-trip relay of this section, the beat phenomenon from Chapter 7, all converge in the inspector's car on the Morogoro Road. Real-world acoustic environments are almost always multi-physics: the listener at the side of a busy highway hears direct engines, reflected engines, Doppler-shifted horns, beats between the Doppler-shifted tones, and ambient noise from many other sources, all simultaneously. Disentangling them mathematically is the whole point of having learned the physics. The exam question is a model of the world; the world is a richer version of the exam question.*

This section has done five things. It has derived Mr. Akilikubwa's blackboard equation from one principle, namely that the medium fixes the wave speed in its own rest frame and that source motion compresses wavelength while observer motion alters encounter rate. It has stated the three no-Doppler cases and shown that they all reduce to the same condition that source and observer share their velocity through the medium. It has corrected the formula for wind, that is to say for a medium that itself moves with respect to the ground. It has handled round-trip Doppler, where the wave bounces off a reflector and returns to be detected by the original emitter. And it has paid one of the four promises Mr. Akilikubwa made in the opening dialogue, by explaining how a bat in a Kondo cave finds a moth in pitch darkness.

Three promises remain. Two of them, the colour of a distant galaxy and the speed of a car caught by a speed camera on the Morogoro Road, depend on Doppler shifts for **electromagnetic** waves rather than for sound. Electromagnetic waves, as the previous two sections established, propagate through vacuum at the fixed speed c and require no medium at all to do it. Without a medium there is no rest frame for the medium, and the asymmetry between source motion and observer motion that this section has just exploited disappears. The Doppler formula for light is symmetric in source and observer velocity, depending only on their relative velocity. The next section derives this symmetric formula, applies it to the redshift of distant galaxies (the first promise paid), introduces Hubble's law connecting redshift to distance, and treats the Doppler broadening of spectral lines. The final promise, about TBC radio at night, depends not on Doppler but on atmospheric propagation, and is treated in the section after that.

THE DOPPLER EFFECT FOR LIGHT, REDSHIFT, AND HUBBLE'S LAW

The previous section closed with the observation that the Doppler effect for light is fundamentally different from the Doppler effect for sound, because light does not require a medium in which to travel. The wave speed c is the same in every inertial frame; no preferred frame for the electromagnetic field has ever been found; and the asymmetry between source motion and observer motion, which we exploited in REAL Example 21 to expose the slight non-equivalence of approaching-source and approaching-observer scenarios at second order in u/v , disappears entirely for light. What is left is a Doppler formula that depends only on the **relative** velocity of source and observer, and that takes a particularly clean form for source-observer relative speeds much smaller than c .

This section does four things. It states the first-order Doppler formula for light, $\Delta\lambda/\lambda = v/c$, and reassures the student honestly that the rigorous relativistic formula reduces to this approximation for $v \ll c$ but requires

special relativity to derive. It applies the formula to the redshift of a star, paying one of the four promises Mr. Akilikubwa made in the chapter's opening dialogue about the colour of a distant galaxy. It introduces Hubble's law, which links the recession velocity of a galaxy to its distance from us and gives the order-of-magnitude age of the universe almost as a free bonus. And it discusses the Doppler broadening of spectral lines, the phenomenon by which a hot gas of atoms in random thermal motion produces a smeared rather than a sharp emission line, and explains why the laboratory spectroscope at Mzumbe Secondary School cannot resolve atomic lines below a certain limit set entirely by temperature and atomic mass.

The first-order Doppler formula for light

When a source of light and an observer move with relative speed v along the line between them, with v measured as **positive for recession** and **negative for approach**, the wavelength of the light the observer measures is shifted from the wavelength the source emits according to:

$$\frac{\Delta\lambda}{\lambda} = \frac{v}{c},$$

where $\Delta\lambda = \lambda_{\text{observed}} - \lambda_{\text{source}}$ is the shift in wavelength and λ is the source wavelength. A **positive $\Delta\lambda$** means the *wavelength has stretched (the source recedes)*, and a **negative $\Delta\lambda$** means the *wavelength has compressed (the source approaches)*. Equivalently, the corresponding shift in frequency is

$$\frac{\Delta f}{f} = -\frac{v}{c},$$

with the opposite sign because frequency and wavelength are inversely related through $c = f\lambda$.

The formula is an approximation. The exact, rigorous formula that special relativity supplies is:

$$\frac{\lambda_{\text{observed}}}{\lambda_{\text{source}}} = \sqrt{\frac{1 + \frac{v}{c}}{1 - \frac{v}{c}}}$$

with v positive for recession. For $v \ll c$ this exact formula reduces to $1 + v/c$, which gives $\Delta\lambda/\lambda = v/c$, the approximation above. The error in the approximation is of order $(v/c)^2$, which for stars in our galaxy (typical speeds 100 to 1000 km/s, so v/c of order 10^{-3}) is one part in a million, beyond the resolution of any school spectroscope. The approximation breaks down only when v becomes a substantial fraction of c , which happens for distant quasars and for cosmological redshifts above about 0.1. For everything in this section we use the first-order formula and flag, in the closing paragraph, where the approximation begins to bite.

Three differences between the light formula and the sound formula deserve emphasis.

1) Only relative velocity matters for light

There is no rest frame of the medium for an electromagnetic wave, because there is no medium. The Doppler shift for light depends only on the **relative** velocity of source and observer, not on either velocity separately. A source moving at speed v towards a stationary observer and a stationary source watched by an observer moving at speed v towards it produce **exactly** the same Doppler shift, to all orders in v/c . This is unlike the case of sound, where the air sets a preferred rest frame and source-vs-observer asymmetries appear at order $(v/c)^2$.

2) The formula is symmetric in v

Swap the roles of source and observer and the sign of v flips, but the formula keeps its form. This is a direct consequence of Einstein's principle that the laws of physics look the same in every inertial frame.

3) The Doppler shift for light has a real "limiting case" at $v = c$

At this limiting case, the exact formula gives $\Delta\lambda/\lambda \rightarrow \infty$ for recession and $\Delta\lambda/\lambda \rightarrow -1$ for approach. The first signals that a source receding at the speed of light would have its wavelength stretched without limit; the second, that an approaching source would have its wavelength contracted to zero. Neither case occurs in nature (no massive object can travel at c), but the closer one comes to c , the more dramatic the shifts become.

Redshift and blueshift: astronomy meets Doppler

The diagram below shows the same atomic spectral lines as they would appear from three different sources: a laboratory reference lamp at rest in the spectroscope, a star approaching the Earth at modest speed (the

lines are shifted towards shorter wavelengths, hence the term **blueshift**), and a star receding from the Earth (the lines are shifted towards longer wavelengths, **redshift**). The dashed line traces the sodium D line at $5896\text{\AA}$ in the laboratory reference and shows where the same physical line falls in each of the moving-source spectra.

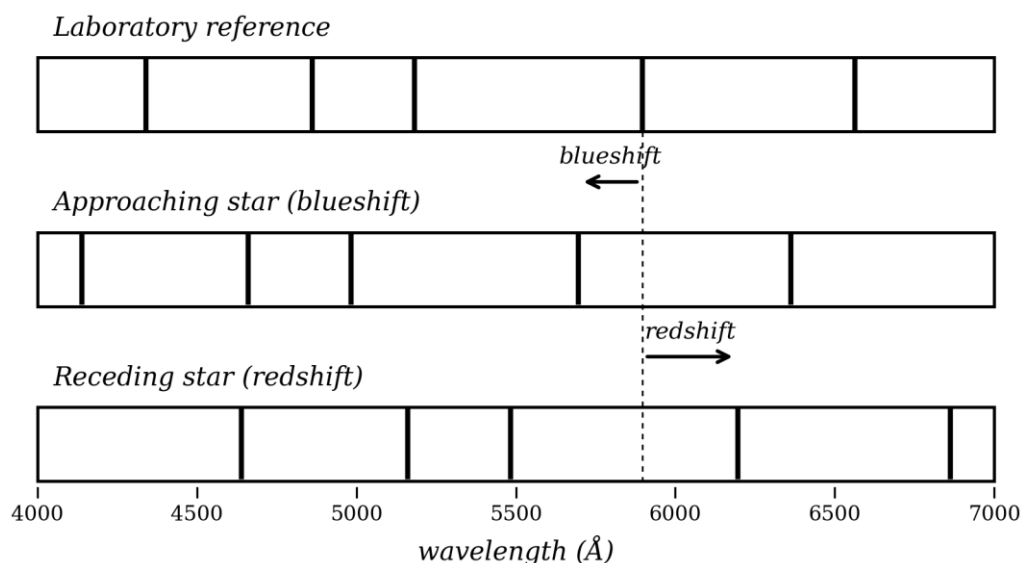


Figure: Spectral lines from three sources. Top: laboratory reference lamp at rest. Middle: an approaching star, whose lines are blueshifted to shorter wavelengths. Bottom: a receding star, whose lines are redshifted to longer wavelengths. The dashed vertical line at $5896\text{\AA}$ (the sodium D line in the laboratory) shows where the same line falls in each spectrum.

The interpretation is straightforward. A star whose spectrum is blueshifted is approaching the Earth; a star whose spectrum is redshifted is receding. The size of the shift fixes the radial velocity through $v = c \Delta\lambda/\lambda$. The direction is read off the sign of $\Delta\lambda$.

There is one large fact that runs through every modern observation of the cosmos. *Almost every galaxy beyond our local group is redshifted, not blueshifted.* This means almost every galaxy is moving away from us. The few blueshifts that astronomers do measure correspond to nearby galaxies in our own gravitational neighbourhood (most famously the Andromeda galaxy, which is approaching us and will collide with the Milky Way in about four billion years). Beyond that nearby cluster, redshift dominates because the universe itself is expanding, a fact discovered by Edwin Hubble in 1929 and explained by the next subsection of this section.

The numerical worked example that follows is the textbook calculation of a stellar redshift.

BINDER Example 26

A spectral line from a distant star is identified as a sodium line. The same line, measured in a laboratory spectroscope at rest, has wavelength $\lambda_{\text{lab}} = 5896\text{\AA}$. When the spectroscope is pointed at the star, the same line is observed at $\lambda_{\text{obs}} = 6010\text{\AA}$. Take the speed of light to be $c = 3.0 \times 10^8\text{m/s}$.

- Is the star approaching the Earth or receding from it?
- Compute the radial velocity of the star, that is to say the component of its velocity along the line of sight from the star to the Earth.

Solution

- Compute the shift:

$$\Delta\lambda = \lambda_{\text{obs}} - \lambda_{\text{lab}} = 6010\text{\AA} - 5896\text{\AA} = +114\text{\AA}$$

The shift is positive: the observed wavelength is **longer** than the laboratory wavelength. This is a redshift, and the star is therefore **receding** from the Earth.

- Apply the first-order Doppler formula for light, $\Delta\lambda/\lambda = v/c$, solved for v :

$$v = c \times \frac{\Delta\lambda}{\lambda_{\text{lab}}} = 3.0 \times 10^8 \text{ m/s} \times \frac{114 \text{ \AA}}{5896 \text{ \AA}}$$

$$v = 3.0 \times 10^8 \text{ m/s} \times 0.0193 = 5.80 \times 10^6 \text{ m/s}$$

Converting to the unit astronomers prefer,

$$v = 5800 \text{ km/s}$$

The star is receding from the Earth at about 5800 km/s (roughly 1.9 per cent of the speed of light).

This is fast in everyday terms (twenty thousand times the speed of a passenger jet) but still well below c , so the first-order formula is reliable here.

Making sense of the answer: The sign of $\Delta\lambda$ tells the direction of motion: positive means recession, negative means approach. The size of $\Delta\lambda$ as a fraction of λ gives the speed as a fraction of c . The student should learn to read a spectroscope output in exactly this two-step way. The 1.9 per cent of c is a typical figure for a galaxy in the nearer half of the observable universe; nearby stars in our own galaxy are usually shifted by far less than this, by a fraction v/c of order 10^{-4} , with corresponding shifts of fractions of an Ångström.

Think like a physicist: Notice that the radial velocity is what the Doppler shift gives, not the full velocity. A star could be moving transversely across the sky at any speed without any Doppler signature at all, just as the train in HOT Example 22 produced no Doppler shift from the perpendicular component of its velocity. To get the full velocity, an astronomer combines the Doppler-derived radial velocity with the transverse velocity (measured by tracking the star's apparent motion across the sky over many years) into a single three-dimensional velocity vector. This is the same vector-decomposition principle as the train at the level crossing, scaled up by a factor of ten quintillion.

The first-order formula $\Delta\lambda/\lambda = v/c$ has served BINDER Example 26 well, because the star was receding at about 2 per cent of the speed of light, well inside the safe regime of the approximation. But astronomers regularly study sources whose recession velocity is a substantial fraction of c , and for such sources the approximation begins to break down noticeably. The worked example that follows takes a high-redshift quasar (a very luminous active galactic nucleus, named “quasi-stellar object” historically because the earliest examples looked like point-like stars in their photographs) and compares the first-order prediction against the exact relativistic Doppler formula. The discrepancy is large enough to matter for any quantitative work at that redshift.

HOT Example 27

A distant quasar in the constellation of Centaurus is observed with a sodium-D spectral line that is redshifted by the fraction $z = \Delta\lambda/\lambda_{\text{lab}} = 0.20$ from its laboratory wavelength.

- Use the first-order Doppler formula to estimate the recession velocity of the quasar.
- Use the exact relativistic Doppler formula to compute the exact recession velocity.
- Compare the two answers. By what percentage does the first-order formula overestimate the recession velocity?

Solution

- The first-order formula gives directly,

$$\frac{v}{c} \approx \frac{\Delta\lambda}{\lambda_{\text{lab}}} = z = 0.20$$

So:

$$v_{\text{first-order}} = 0.20 \times c = 0.20 \times 3.0 \times 10^8 \text{ m/s} = 6.0 \times 10^7 \text{ m/s} = 60000 \text{ km/s}$$

The first-order formula predicts that the quasar is receding at 20 per cent of the speed of light.

- The exact relativistic Doppler formula:

$$\frac{\lambda_{\text{obs}}}{\lambda_{\text{lab}}} = \sqrt{\frac{1 + \frac{v}{c}}{1 - \frac{v}{c}}}$$

From the relativistic formula and the definition $z = \Delta\lambda/\lambda_{\text{lab}}$:

$$z = \frac{\lambda_{\text{obs}} - \lambda_{\text{lab}}}{\lambda_{\text{lab}}} = \frac{\lambda_{\text{obs}}}{\lambda_{\text{lab}}} - 1$$

$$\frac{\lambda_{\text{obs}}}{\lambda_{\text{lab}}} = z + 1 = 0.20 + 1 = 1.20$$

Let $\beta = v/c$.

Substitute into the exact formula and square both sides:

$$(1.20)^2 = \frac{1 + \beta}{1 - \beta}$$

$$\beta = 0.1803$$

The exact recession velocity is therefore:

$$v_{\text{exact}} = \beta \times c = 0.1803 \times 3.0 \times 10^8 \text{ m/s} = 5.41 \times 10^7 \text{ m/s} = 54100 \text{ km/s}$$

The quasar is receding at 18.03 per cent of the speed of light, not the 20 per cent the first-order formula gave.

(c) The first-order formula gave 60000 km/s; the exact formula gives 54100 km/s. The first-order formula overestimates the true recession velocity by:

$$\Delta v = 60000 \text{ km/s} - 54100 \text{ km/s} = 5900 \text{ km/s}$$

As a percentage of the exact value,

$$\frac{\Delta v}{v_{\text{exact}}} = \frac{5900}{54100} = 0.109 = 10.9\%$$

The first-order formula overestimates the recession velocity by about 11 per cent at this redshift. For accurate quasar kinematics, the exact relativistic formula must be used.

Making sense of the answer: *The size of the discrepancy is the practical statement of “where the approximation breaks down.” A 10 per cent error is too large to ignore in any quantitative astronomical work. The error in the first-order formula scales roughly as $z^2/2$: at $z = 0.01$ the error is 5×10^{-5} (negligible), at $z = 0.10$ it is 5×10^{-3} (still tolerable for most work), at $z = 0.20$ it is 2×10^{-2} (the regime of this example, becoming uncomfortable), and at $z = 1$ the first-order formula fails completely. Modern cosmological surveys routinely deal with redshifts up to $z \approx 7$ for the most distant galaxies known, where the relativistic correction is dominant.*

Think like a physicist: *The Doppler formula for light has the same general status as Newtonian mechanics: it is a low-velocity approximation to a more general theory that becomes essential at high velocity. The student should not despair that physical formulas have limited regimes of validity; this is the rule, not the exception. The skill the student must cultivate is recognising the regime of validity for each formula and switching to the more general formula when the regime breaks down. Newtonian mechanics is exact below 1 per cent of c ; the first-order Doppler formula is good to 1 per cent below $z = 0.05$, to 10 per cent below $z = 0.20$, and useless above $z = 1$. Each formula carries its own boundary, and physics is the art of staying on the right side of the boundary or correctly stepping past it.*

Hubble's law

In 1929 the American astronomer Edwin Hubble compiled the recession velocities (measured by Doppler shift) and distances (estimated by various means, including the brightness of standard-candle stars called Cepheid variables) of many galaxies and noticed a striking pattern: *the recession velocity of a galaxy is approximately proportional to its distance from Earth*. Modern data, refined over nearly a century since Hubble's original work, confirm this relationship across hundreds of millions of parsecs and assemble it into what is now universally called **Hubble's law**:

$$v = H_0 d$$

where v is the recession velocity of a galaxy, d is its distance from Earth, and H_0 is **Hubble's constant**, a fundamental cosmological parameter with current best-measured value approximately

$$H_0 \approx 70 \text{ km/s per Mpc}$$

The unit “km/s per Mpc” reads as “every additional megaparsec (one Mpc = 3.086×10^{22} metres, or about 3.26 million light-years) of distance from us adds about 70km/s of recession velocity.” In more student-friendly units, H_0 corresponds to about 22km/s per million light-years.

The figure below plots recession velocity against distance for a representative selection of galaxies. The straight line through the origin is Hubble’s law. The data points scatter modestly around the line, in part because of measurement uncertainty and in part because of local gravitational motions superimposed on the cosmic expansion. The strong linear relationship dominates the scatter, especially at larger distances where the cosmic expansion overwhelms the local motions.

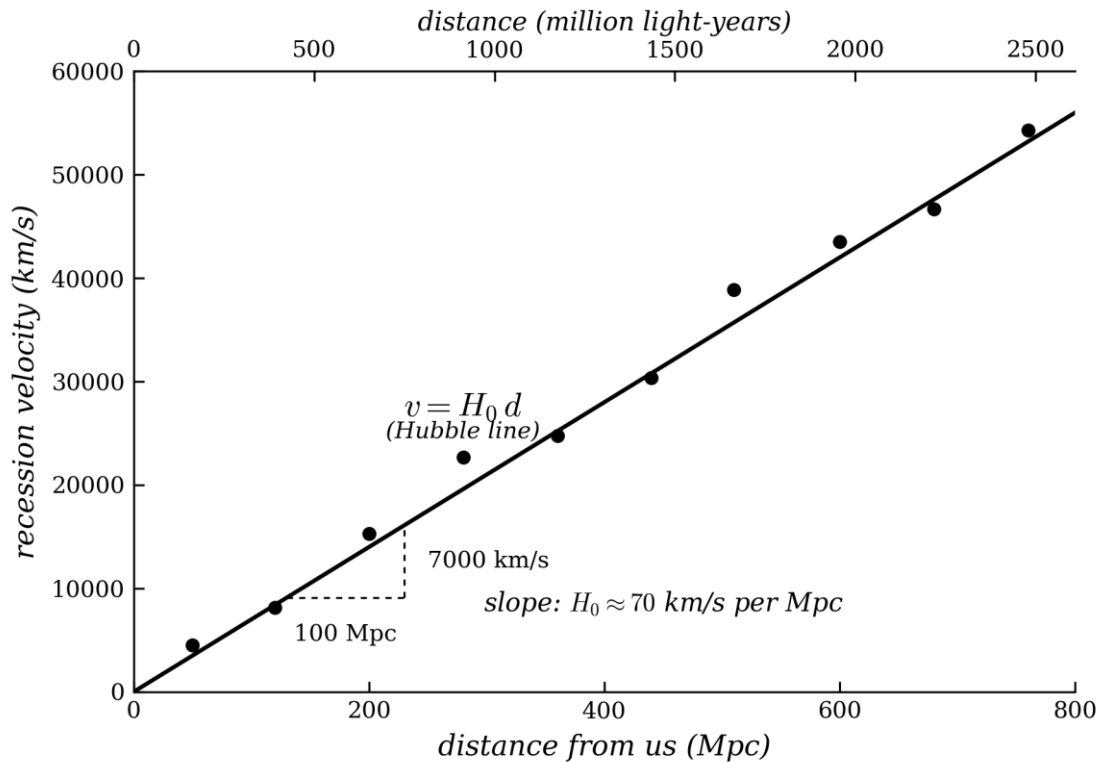


Figure: Recession velocity versus distance for a sample of galaxies. The straight line is Hubble’s law $v = H_0 d$ with $H_0 \approx 70$ km/s per Mpc. The slope-triangle shows the meaning of the constant: every 100Mpc of additional distance corresponds to an additional 7000km/s of recession velocity. The upper axis converts distance from Mpc to millions of light-years using the conversion $1 \text{ Mpc} \approx 3.262 \text{ million light-years}$.

The most striking property of Hubble’s law, and the one that took half a century to understand, is what it says about the universe itself. If every galaxy is receding from us at a rate proportional to its distance, the natural conclusion is that the universe is **expanding uniformly**. This is not, as the student might first think, the consequence of us sitting at the centre of an explosion. Instead, it is the geometric consequence of space itself stretching out: every point in the universe sees every other point receding from it, with a recession speed proportional to the distance between them, exactly as Hubble’s law describes. The student should picture not a bomb exploding outwards from a single centre, but a loaf of raisin bread rising in an oven: every raisin sees every other raisin receding, and the further away the receding raisin, the faster it recedes.

A useful by-product of Hubble’s law is an estimate of the age of the universe. The quantity $1/H_0$, called the **Hubble time**, has units of time. Plugging in $H_0 \approx 70$ km/s per Mpc:

$$\frac{1}{H_0} \approx \frac{1 \text{ Mpc}}{70 \text{ km/s}} = \frac{3.086 \times 10^{19} \text{ km}}{70 \text{ km/s}} \approx 4.4 \times 10^{17} \text{ s} \approx 14 \text{ billion years}$$

The Hubble time is the time it would take for the present rate of expansion to have produced the universe’s present size, if the expansion has always proceeded at the same rate. The actual age of the universe, determined by more careful analysis of how the expansion rate has changed over cosmic history, comes out to be approximately 13.8 billion years, remarkably close to the Hubble-time estimate. The order of magnitude is set, and the precise correction requires detailed cosmological modelling that is well beyond the scope of this book.

The next worked example exercises Hubble's law on a concrete galaxy, the Virgo cluster of galaxies in the constellation of Virgo, the nearest large galaxy cluster to our own and a standard textbook target.

REAL Example 28

The Virgo cluster of galaxies sits at a measured distance of $d = 16.5 \text{ Mpc}$ from the Milky Way. A spectral line of hydrogen ($\text{H}\alpha$, $\lambda_{\text{lab}} = 6563 \text{ \AA}$) emitted by a galaxy in the Virgo cluster is observed on Earth at $\lambda_{\text{obs}} = 6587 \text{ \AA}$. Take $H_0 = 70 \text{ km/s per Mpc}$ and $c = 3.0 \times 10^8 \text{ m/s}$.

- Use Hubble's law to predict the recession velocity of the Virgo cluster.
- Compute the observed recession velocity directly from the Doppler shift of the $\text{H}\alpha$ line.
- Compare the two answers. Comment on the size of the discrepancy.

Solution

- (a) Hubble's law predicts $v = H_0 d$. Substituting:

$$v_{\text{predicted}} = 70 \text{ km/s per Mpc} \times 16.5 \text{ Mpc} = 1155 \text{ km/s}$$

If the universe expanded uniformly with no other effects superimposed, the Virgo cluster would be receding from us at 1155 km/s .

- (b) Compute the wavelength shift:

$$\Delta\lambda = \lambda_{\text{obs}} - \lambda_{\text{lab}} = 6587 \text{ \AA} - 6563 \text{ \AA} = +24 \text{ \AA}$$

Apply the first-order Doppler formula:

$$v_{\text{observed}} = c \times \frac{\Delta\lambda}{\lambda_{\text{lab}}} = 3.0 \times 10^8 \text{ m/s} \times \frac{24 \text{ \AA}}{6563 \text{ \AA}} = 1.097 \times 10^6 \text{ m/s} = 1097 \text{ km/s}$$

- (c) The two values are:

$$v_{\text{predicted}} = 1155 \text{ km/s}, \quad v_{\text{observed}} = 1097 \text{ km/s}$$

They differ by $\Delta v = 1155 \text{ km/s} - 1097 \text{ km/s} = 58 \text{ km/s}$, which is about 5 per cent of either value. The Hubble-law prediction is slightly higher than the actual observed recession.

Making sense of the answer: *The 5 per cent discrepancy is real, and it is the signature of what cosmologists call a **peculiar velocity**. The Virgo cluster does not move only with the smooth cosmic expansion; it also has a small additional motion through space, driven by the local gravitational pull of nearby mass concentrations. For Virgo specifically, this peculiar velocity points partly back towards us, slowing the cluster's net recession below the pure-Hubble value. Hubble's law is therefore strictly accurate only on average over many galaxies and over distances large enough that local gravitational effects average out. A single galaxy's redshift can deviate from the Hubble prediction by hundreds of km/s in either direction, but the trend across hundreds of galaxies is the clean linear law shown in the figure above.*

Think like a physicist: *The clean linear law sits underneath a noisy reality. Every individual galaxy carries a peculiar velocity on top of its cosmic-expansion velocity, and the peculiar velocity can point in any direction. Astronomers extract H_0 from large samples of galaxies, fitting a straight line through the cloud of data points, not from any single galaxy in isolation. This is the same principle that drives every physical-law measurement: the law is an idealisation, the data are noisy, and the constant that comes out of the linear regression. The student should always ask, on encountering a physical law, "what is the noise that this law hides on top of?", and seek out at least one source of that noise. For Hubble's law, the noise is peculiar velocities. For Ohm's law, it is the temperature dependence of resistance. For the ideal-gas law, it is intermolecular forces. Every clean physical law has its noise; the law is the average, not the individual case.*

Doppler broadening of spectral lines

The third application of the light Doppler formula is to a single body of gas, hot enough to emit a spectral line, but with the atoms inside the gas moving with thermal velocities in random directions. Some atoms move towards the spectroscope and emit Doppler-blueshifted light; some move away and emit Doppler-

redshifted light; most move at intermediate angles and produce intermediate shifts. The observed line is therefore not a single sharp wavelength but a **smeared** distribution centred on the laboratory wavelength, with a width controlled by the typical atomic speed.

The kinetic theory of gases (Chapter 4) gives the root-mean-square speed of atoms in an ideal gas at absolute temperature T as:

$$v_{\text{rms}} = \sqrt{\frac{3RT}{M}},$$

where $R = 8.314\text{J}/(\text{mol}\cdot\text{K})$ is the gas constant and M is the molar mass of the gas in kg/mol . Atoms move at this typical speed in random directions: about half of them have a component of velocity towards the spectroscope of order v_{rms} , and about half a component away of order v_{rms} . The observed spectral line is broadened by the combined Doppler shift, of magnitude approximately:

$$\Delta\lambda \approx \frac{2\lambda v_{\text{rms}}}{c}$$

This is the **Doppler width** of the spectral line. It depends only on three things: the wavelength of the line itself (λ), the absolute temperature of the gas (T), and the molar mass of the emitting atom (M). For a given line, lighter atoms move faster at any given temperature, and hotter gases move faster at any given mass; both effects broaden the line.

The next worked example exercises this formula. The example illustrates an important and counter-intuitive point: the **colder** gas can produce **broad**er spectral lines, provided the atoms inside it are light enough.

HOT Example 29

Two laboratory discharge lamps emit the same green spectral line at $\lambda = 5461\text{\AA}$. The first lamp contains mercury vapour at 200°C , the second krypton gas at 0°C . The molar masses are $M_{\text{Hg}} = 200\text{g}/\text{mol}$ for mercury and $M_{\text{Kr}} = 84\text{g}/\text{mol}$ for krypton. The gas constant is $R = 8.314\text{J}/(\text{mol}\cdot\text{K})$ and the speed of light is $c = 3.0 \times 10^8\text{m/s}$.

- Compute the root-mean-square atomic speed in each lamp.
- Compute the Doppler width of the spectral line in each lamp.
- Which lamp produces the sharper spectral line, and by what factor?

Solution

- Convert temperatures to absolute scale:

$$T_{\text{Hg}} = 200^\circ\text{C} + 273 = 473\text{K}$$

$$T_{\text{Kr}} = 0^\circ\text{C} + 273 = 273\text{K}$$

Convert molar masses to kg/mol :

$$M_{\text{Hg}} = 0.200\text{kg}/\text{mol}, \quad M_{\text{Kr}} = 0.084\text{kg}/\text{mol}$$

Apply $v_{\text{rms}} = \sqrt{\frac{3RT}{M}}$ for each:

Mercury:

$$v_{\text{rms,Hg}} = \sqrt{\frac{3 \times 8.314\text{J}/(\text{mol}\cdot\text{K}) \times 473\text{K}}{0.200\text{kg}/\text{mol}}} = 243\text{m/s}$$

Krypton:

$$v_{\text{rms,Kr}} = \sqrt{\frac{3 \times 8.314\text{J}/(\text{mol}\cdot\text{K}) \times 273\text{K}}{0.084\text{kg}/\text{mol}}} = 285\text{m/s}$$

The krypton atoms move faster than the mercury atoms even though the mercury lamp is far hotter, because the krypton atoms are so much lighter.

(b) Apply the Doppler-broadening formula $\Delta\lambda = \frac{2\lambda v_{rms}}{c}$ for each. Convert $\lambda = 5461 \text{ \AA}$ to metres: $\lambda = 5461 \times 10^{-10} \text{ m} = 5.461 \times 10^{-7} \text{ m}$.

Mercury:

$$\Delta\lambda_{\text{Hg}} = \frac{2 \times 5.461 \times 10^{-7} \text{ m} \times 243 \text{ m/s}}{3.0 \times 10^8 \text{ m/s}} = 8.85 \times 10^{-13} \text{ m} = 0.00885 \text{ \AA}$$

Krypton:

$$\Delta\lambda_{\text{Kr}} = \frac{2 \times 5.461 \times 10^{-7} \text{ m} \times 285 \text{ m/s}}{3.0 \times 10^8 \text{ m/s}} = 1.038 \times 10^{-12} \text{ m} = 0.01038 \text{ \AA}$$

(c) The sharper line has the smaller Doppler width. Comparing:

$$\frac{\Delta\lambda_{\text{Hg}}}{\Delta\lambda_{\text{Kr}}} = \frac{0.00885 \text{ \AA}}{0.01038 \text{ \AA}} = 0.853$$

The mercury lamp at 200°C produces the **sharper** line, narrower than the krypton lamp at 0°C by a factor of 0.853, equivalently the krypton line is broader by a factor of $1/0.853 = 1.17$.

Making sense of the answer: The counter-intuitive result is that the cooler lamp (krypton at 0°C) produces broader spectral lines than the hotter lamp (mercury at 200°C), simply because krypton atoms are roughly 2.4 times lighter than mercury atoms. The line width scales as $\sqrt{T/M}$, and the krypton lamp wins on $1/M$ by a larger factor than it loses on T :

$$\frac{\Delta\lambda_{\text{Hg}}}{\Delta\lambda_{\text{Kr}}} = \sqrt{\frac{T_{\text{Hg}}/M_{\text{Hg}}}{T_{\text{Kr}}/M_{\text{Kr}}}} = \sqrt{\frac{473/200}{273/84}} = \sqrt{\frac{2.365}{3.250}} = 0.853$$

confirming the result. The student should learn to recognise this trade-off: temperature and molar mass enter the line width through the single combination T/M , and a hot lamp made of heavy atoms can be just as quiet, spectroscopically speaking, as a cold lamp made of light atoms.

Think like a physicist: Doppler broadening sets a fundamental limit on the resolution of any spectroscopic measurement performed on a hot gas. Higher-resolution spectroscopy at fixed wavelength requires cooling the gas (to slow the atoms) or selecting heavier atoms (so the random thermal motion is slower at any given temperature). The same principle drives modern atomic clocks: caesium atomic-clock laboratories cool their atoms to within a millionth of a kelvin of absolute zero so that Doppler broadening contributes negligibly to the precision of the frequency reference. Without cooling, the world's atomic clocks would lose seconds per day instead of seconds per million years.

This section has paid the second of Mr. Akilikubwa's four chapter-opening promises by showing how the colour of a distant galaxy reveals its recession velocity through the redshift of its spectral lines, and how Hubble's law turns those redshifts into a panoramic picture of cosmic expansion. The third promise, the speed of a car caught by a speed camera on the Morogoro Road, depends on round-trip Doppler shifting of a radar pulse, the same physics the bat in Kondoa cave used to find its moth, but applied this time to electromagnetic waves and exploited by traffic police to issue speeding tickets. The fourth, the TBC radio at night, depends on atmospheric propagation, the way long-wavelength radio waves bend around the curvature of the Earth and bounce off the ionosphere, a quite different topic from Doppler that nevertheless rounds out the chapter. Both are brought together in the applications section that follows.

APPLICATIONS OF SOUND, EM WAVES, AND THE DOPPLER EFFECT

This final section of Chapter 9 introduces no new physics. Its purpose is to show the chapter's machinery, all of which the student has now derived and exercised, doing useful work in the real world. The seven applications below are arranged in the order the physics behind them appeared in this chapter, so a reader can quickly trace each one back to its theoretical home.

1. Ultrasound imaging in obstetrics

A pregnant woman attending her antenatal check at KCMC Hospital sees the inside of her own womb on a screen: the foetus, its heart beating, its head and limbs measurable to the millimetre. The image is constructed entirely from acoustic echoes. A piezoelectric transducer, held against the mother's abdomen, emits short pulses of ultrasound at a frequency between 2MHz and 15MHz, well above the audible range. The pulses travel into the body at the speed of sound in soft tissue, approximately 1540m/s, and partially reflect at every internal boundary where the acoustic impedance changes: skin-fat, fat-muscle, muscle-fluid, fluid-bone. Each reflected pulse returns to the transducer after a time proportional to twice the distance to the boundary that produced it. The machine measures this time, multiplies by half the wave speed, and assembles a two-dimensional image of the internal structures.

The choice of frequency in the megahertz range is a deliberate compromise. Higher frequencies give shorter wavelengths (at 15MHz the wavelength in tissue is about 0.1mm, smaller than the thickness of a human hair) and therefore sharper resolution, but they also attenuate more rapidly in tissue and so probe to shallower depths. Lower frequencies penetrate deeper but resolve less. Obstetric scans of a foetus deep inside the uterus typically use 3–5MHz; superficial scans of the thyroid or breast use 7–15MHz. The pedagogy connects to acoustic impedance from Chapter 9's opening section: only when impedance changes at a boundary does a reflection occur, and the brightness of the boundary on the screen is proportional to the impedance mismatch.

2. Sonar

The Tanzania People's Defence Force naval division operates sonar systems on its patrol boats off the coast at Dar es Salaam, mapping the seabed for navigation and detecting submerged objects. Civilian sonar serves the same physics in fishing trawlers off the Mafia archipelago, oceanographic survey vessels at the Pemba channel, and the harbour authorities at the Dar es Salaam port. A sonar transducer mounted on the underside of the vessel emits pulses of sound at frequencies between 20kHz and 100kHz, depending on the application, into the water. The pulses travel through seawater at approximately 1500m/s, reflect off the seabed, fish schools, or any other object with a different acoustic impedance from the surrounding water, and return to the transducer.

The depth of a target is read off the round-trip time exactly as in obstetric ultrasound: a depth of 1km corresponds to a round-trip time of $2 \times 1000\text{m} / 1500\text{m/s} \approx 1.33\text{s}$. Doppler sonar, an upgrade of the basic system, also measures the relative speed of any moving target the pulse bounces off, by reading the Doppler shift between the emitted and reflected frequencies. The chapter's round-trip Doppler formula, derived for the bat in the Kondo cave, applies to the sonar without modification: only the medium and the wave are different.

3. Police radar speed gun

The road traffic police on the Morogoro Road operate radar speed guns, often hand-held devices held to the eye like a small camera. The device emits a continuous beam of microwave radiation at a fixed frequency (usually around 10GHz or 24GHz, in the so-called X-band or K-band of the radio spectrum) directly down the road, picks up the reflection from any oncoming or receding vehicle, and computes the vehicle's speed from the Doppler shift between the emitted and reflected frequencies. The round-trip relay used by the bat in HOT Example 24 of this chapter is exactly the principle in play.

The numerical scale is worth noting. At 24GHz, a car approaching at 100km/h produces a Doppler shift of,

$$\Delta f = \frac{2fv}{c} = \frac{2 \times 24 \times 10^9 \text{Hz} \times 27.8 \text{m/s}}{3 \times 10^8 \text{m/s}} \approx 4400 \text{Hz}$$

a frequency difference well within the resolving power of any modern radio receiver. The speed gun reads this 4400Hz shift, divides by twice the carrier frequency, multiplies by the speed of light, and displays the answer in km/h to the operator in roughly a hundredth of a second. This is the third of Mr. Akilikubwa's chapter-opening promises paid: the speed of a car caught by a speed camera on the Morogoro Road is read directly from the same round-trip Doppler formula the bat used to find its moth.

4. Doppler weather radar

The Tanzania Meteorological Authority (TMA) operates weather radar at sites across the country, and the colour-coded weather maps shown on ITV nightly news bulletins are produced by these radars. The principle is the same as the police radar gun, but the targets are raindrops and ice particles inside clouds rather than vehicles on a road. A microwave pulse at around 5GHz (the C-band) is emitted from a ground-based dish, reflects off whatever raindrops or ice particles it meets, and returns to the receiver. The total power of the

returned signal indicates how much water the radar beam encountered, which translates directly into rainfall intensity. The Doppler shift of the returned signal indicates the radial velocity of the rain or ice particles: positive shift means raindrops moving towards the radar, negative shift means moving away.

The colour-coding on the ITV weather forecast follows the velocity field: red colours indicate fast-approaching wind systems (rain bands closing in on the viewer), blue colours indicate fast-receding systems (rain moving away), and a rotational pattern of red against blue across a small region indicates a cyclonic vortex. Tropical cyclone tracking in the Indian Ocean basin, which affects the Tanzanian coast roughly twice per decade, relies critically on this Doppler velocity signature, since the rotational pattern reveals the cyclone's centre and the wind speeds at its eyewall.

5. Blood-flow ultrasound

A medical specialist at the Aga Khan Hospital who wishes to measure the blood flow through a patient's carotid artery applies an ultrasound transducer to the patient's neck, but adds a Doppler-shift readout to the basic imaging procedure described in the obstetric subsection above. The transducer emits a continuous pulse at typically 5MHz, and the red blood cells moving through the artery reflect a Doppler-shifted version of the pulse back to the transducer.

For a typical arterial blood flow of 1m/s and a 5MHz carrier, the round-trip Doppler shift is

$$\Delta f = \frac{2fv}{v_{\text{tissue}}} = \frac{2 \times 5 \times 10^6 \text{Hz} \times 1 \text{m/s}}{1540 \text{m/s}} \approx 6500 \text{Hz},$$

a shift easily resolved by the receiver. The doctor reads off blood flow velocity in real time, can spot a partial blockage (the blood downstream of an obstruction speeds up to maintain volume flow, by the continuity principle from Chapter 10 of *Foundations of Physics*), and can confirm the direction of flow (arterial flow into the head should always be in the same direction; a Doppler shift of the wrong sign is an immediate warning sign of vascular disease). The same physics, applied to the umbilical cord during pregnancy, monitors the blood supply between mother and foetus throughout pregnancy.

6. GPS as a Doppler-corrected positioning system

The Global Positioning System (GPS) used by every smartphone in Tanzania to find its location on a map depends critically on signals broadcast from a constellation of about 30 satellites orbiting the Earth at approximately 20200km altitude. Each satellite continuously broadcasts a microwave signal at 1575MHz (the so-called L1 frequency) carrying a precisely timed message identifying the satellite and the moment of transmission. A GPS receiver on the ground listens for these signals from at least four satellites simultaneously, computes the time delay between transmission and reception for each, and from those four delays solves for its own position in three dimensions plus the receiver clock offset.

Two corrections are essential, and without either the system would be useless. The first is the Doppler correction. Each satellite moves at about 3.9km/s in its orbit, with a radial component (the component of velocity along the line to the receiver) that can reach several hundred metres per second as the satellite rises or sets across the receiver's sky. At the 1575MHz L1 frequency, this radial motion produces a Doppler shift of a few kilohertz, well within the receiver's capability to measure. Without correcting for this shift, the receiver would misread the timing of the signal by an amount that translates, after the time-of-flight calculation, into a position error of several hundred metres at every measurement. Over the course of a day, uncorrected positions would drift by kilometres. The second correction is from the general theory of relativity whose discussion is beyond the scope of this book.

7. TBC radio at night: ionospheric skywave propagation

The fourth of Mr. Akilikubwa's chapter-opening promises was about TBC radio. A listener tuning a portable AM radio in Dar es Salaam at noon hears only the local Tanzanian stations broadcasting from transmitters within a few hundred kilometres. The same listener tuning the same radio at midnight often picks up stations broadcasting from Maputo, from Nairobi, from as far as Cape Town and Cairo. The same physical receiver, the same station antennas, but a very different reception. The reason has nothing to do with Doppler. It has to do with the ionosphere.

The Earth's upper atmosphere, between roughly 60km and 1000km altitude, is partially ionised by the Sun's ultraviolet and X-ray radiation. Free electrons in this ionised region (the **ionosphere**) act as a partial reflector for radio waves of the right frequency. The reflection is most efficient for AM-band frequencies (530–1700kHz, wavelengths of order a few hundred metres) bouncing off the so-called F-layer, the highest part

of the ionosphere at about 250–400km altitude. During the day, the lower D-layer (around 60–90km altitude) absorbs these AM waves before they can reach the F-layer, and only line-of-sight propagation is possible: the listener hears only nearby transmitters. At night, the D-layer disappears (the Sun no longer ionises it, and the electrons recombine), the AM waves climb unimpeded to the F-layer, reflect, and return to the ground hundreds or thousands of kilometres from where they were transmitted. The listener in Dar es Salaam at midnight is receiving signals that have travelled by a single skip off the F-layer, sometimes by two skips with an intermediate ground reflection, from transmitters in southern Africa and beyond.

The physics is the same wave-propagation physics this chapter has used throughout: refraction (the ionosphere has a frequency-dependent refractive index that effectively bends the wave back down at appropriate angles), transmission (waves above a critical frequency penetrate the layer rather than reflecting), and reflection (waves below the critical frequency bounce back). Higher frequencies in the shortwave band (3–30MHz) reflect off lower layers and over shorter distances; lower frequencies penetrate higher. The student should not be surprised that this final phenomenon, paying the last promise of the opening dialogue, reuses only the wave physics already established earlier in the chapter; the chapter began with the observation that all waves obey one set of rules, and the chapter ends with a list of seven distinct technologies, each one a different combination of those same rules.

Mr. Akilikubwa's blackboard equation, the four promises he made in the chapter's opening dialogue, the eight sections that derived the physics, and the seven applications that completed it, now together form a single integrated treatment of sound waves, electromagnetic waves, and the Doppler effect for both. What remains is to see the machinery operating in full: mixed situations, hidden traps, numerical subtleties, sign-convention decisions, and real physical interpretations brought together in a final collection of miscellaneous worked examples.

MISCELLANEOUS WORKED EXAMPLES ON SOUND, ELECTROMAGNETIC WAVES, AND THE DOPPLER EFFECT

Example 30

- Explain why a person swimming below the surface of the sea at Coco Beach cannot hear a friend who is shouting from a boat in the air above, even though the swimmer's ears are working perfectly.
- An organ pipe has two consecutive resonant frequencies of 1372Hz and 1764Hz. Taking the speed of sound in air as 340m/s and neglecting end correction, determine whether the pipe is open or closed, calculate its fundamental frequency, and find its length.

Solution

(a) When a sound wave meets the boundary between two media, the fraction of its power that crosses depends on how closely the acoustic impedances $Z = \rho v$ of the two media match. Air has a very low acoustic impedance and water a very high one, the two differing by a factor of several thousand. Because the mismatch is so large, almost all the sound power arriving at the water surface from the air is reflected back into the air, and only about one part in a thousand is transmitted into the water. The small fraction that does enter the water spreads and weakens further, so by the time it reaches the submerged swimmer its intensity is far below the threshold of hearing.

(b) The difference of the two frequencies is:

$$\Delta f = 1764\text{Hz} - 1372\text{Hz} = 392\text{Hz}$$

Test for an open pipe:

An open pipe has consecutive resonances separated by the fundamental, so the fundamental frequency would be 392Hz. But $\frac{1372}{392} = 3.5$, which is not a whole number, so the pipe is not open.

Test for a closed pipe:

A closed pipe has consecutive resonances separated by $2f_1$, so $f_1 = \frac{392\text{Hz}}{2} = 196\text{Hz}$.

Then $\frac{1372}{196} = 7$ and $\frac{1764}{196} = 9$, both odd whole numbers. Hence the pipe is closed, sounding its seventh and ninth harmonics.

The fundamental frequency is 196Hz.

For a closed pipe the fundamental is $f_1 = \frac{v}{4L}$, so the length is:

$$L = \frac{v}{4f_1} = \frac{340\text{m/s}}{4 \times 196\text{Hz}} = 0.434\text{m} = 43.4\text{cm}$$

Example 31

- (a) Sir Isaac Newton's calculation of the speed of sound in air gave a value about sixteen per cent below the measured value. State the physical assumption Newton made that was wrong, identify the correction supplied by Laplace, and explain why the correction was necessary.
- (b) An observer walks along the straight line joining two identical stationary sources, each emitting a steady note of frequency 500Hz. She hears 4.0 beats per second. Taking the speed of sound in still air as 340m/s, calculate her walking speed.

Solution

(a) Newton assumed that the compressions and rarefactions of a sound wave take place isothermally, that is, slowly enough for heat to flow and keep the gas at the temperature of its surroundings. This made the elastic property of the gas equal to its pressure P . Laplace recognised that a sound wave oscillates far too rapidly for any appreciable heat to flow during a single cycle, so each compression is in fact adiabatic, and the correct elastic property is the adiabatic bulk modulus γP rather than P . Replacing P by γP inside the square root multiplies the predicted speed by $\sqrt{\gamma}$, and for air $\sqrt{1.40} \approx 1.18$, which closes the sixteen-per-cent gap and brings the formula into agreement with experiment.

(b) As the observer walks along the line, she approaches one source and recedes from the other at the same speed u . For the source she approaches, the apparent frequency is raised:

$$f_1 = \frac{f(v + u)}{v}$$

For the source she recedes from, the apparent frequency is lowered:

$$f_2 = \frac{f(v - u)}{v}$$

The two notes are close in frequency and superpose to give beats at the difference frequency:

$$f_{\text{beat}} = f_1 - f_2 = \frac{f(v + u)}{v} - \frac{f(v - u)}{v} = \frac{2fu}{v}$$

Solve for u :

$$u = \frac{f_{\text{beat}} \times v}{2f} = \frac{4.0\text{Hz} \times 340\text{m/s}}{2 \times 500\text{Hz}} = 1.36\text{m/s}$$

The observer is walking at 1.36m/s.

Example 32

- (a) Explain why a closed organ pipe sounds a note one octave lower than an open organ pipe of the same length, and explain why a closed pipe produces only the odd harmonics while an open pipe produces all of them.
- (b) The speed of sound in a certain gas is measured as 353m/s at a temperature of 27°C. The molar mass of the gas is 28.0g/mol and the molar gas constant is $R = 8.314\text{J}/(\text{mol} \cdot \text{K})$. Calculate the ratio of the principal specific heat capacities of the gas, and state, with a reason, whether the gas is monatomic or diatomic.

Solution

(a) A closed pipe has a displacement node at its closed end and a displacement antinode at its open end, so the shortest standing wave that fits has just one quarter of a wavelength inside the pipe, giving a fundamental frequency $f = \frac{v}{4L}$. An open pipe has a displacement antinode at each end, so its shortest standing wave has half a wavelength inside, giving a fundamental $f = \frac{v}{2L}$, exactly twice that of a closed pipe of the same length, which is one octave higher. The closed pipe produces only odd harmonics because every allowed mode must

keep a node at the closed end and an antinode at the open end, and only odd multiples of the quarter-wavelength satisfy both conditions at once. An open pipe, with an antinode required at each end, admits every multiple of the half-wavelength, both odd and even, so it produces the complete set of harmonics.

(b) The speed of sound in an ideal gas is $v = \sqrt{\frac{\gamma RT}{M}}$. Solve for γ :

$$\gamma = \frac{Mv^2}{RT}$$

Convert the data: $M = 28.0\text{g/mol} = 0.0280\text{kg/mol}$ and $T = (27 + 273)\text{K} = 300\text{K}$.

$$\gamma = \frac{0.0280\text{kg/mol} \times (353\text{m/s})^2}{8.314\text{J/(mol} \cdot \text{K)} \times 300\text{K}} = 1.40$$

The ratio of the specific heat capacities is 1.40. A monatomic gas has $\gamma = 1.67$ and a diatomic gas has $\gamma = 1.40$, so the gas is **diatomic**.

Example 33

- (a) Explain why the siren of an ambulance racing along United Nations Road sounds higher in pitch while the ambulance approaches a listener and lower in pitch after it has passed, and explain why the change in pitch is sudden rather than gradual.
- (b) A single diesel generator produces a sound level of 88dB at a distance of 5.0m. Treating the generators as incoherent point sources, calculate the distance from a cluster of five such generators at which the sound level falls to 85dB.

Solution

(a) The siren emits wavefronts at a fixed rate. While the ambulance approaches the listener, it moves a little closer between one emission and the next, so the wavefronts ahead of it are crowded into a shorter distance; the listener meets shorter wavelengths, and since the speed of sound in the air is unchanged, a shorter wavelength means a higher frequency. After the ambulance has passed, it moves away between emissions, the wavefronts behind it are stretched, the wavelength is longer and the frequency lower. The change is sudden because, at the instant the ambulance passes the listener, the motion along the line joining them switches abruptly from approach to recession, so the apparent frequency jumps straight from its raised value to its lowered value with nothing in between.

(b) First combine the five generators. Five incoherent equal sources raise the level above that of one source by:

$$\Delta L = 10\log 5 = 10 \times 0.699 \approx 7.0\text{dB}$$

So the cluster produces $88\text{dB} + 7.0\text{dB} = 95\text{dB}$ at 5.0m.

The level must now fall from 95dB to 85dB, a drop of:

$$\Delta L = 85\text{dB} - 95\text{dB} = -10\text{dB}$$

For a point source the level changes with distance as $\Delta L = -20\log\left(\frac{r_2}{r_1}\right)$:

$$-10\text{dB} = -20\log\left(\frac{r_2}{5.0\text{m}}\right)$$

$$r_2 = 15.8\text{m}$$

A worker must stand at least 15.8m from the cluster of five generators for the sound level to fall to 85dB.

Example 34

- (a) A workshop installs more identical machines. Explain why doubling the number of machines raises the measured sound level by only about 3 decibels, and explain why making the workshop sound half as loud to the ear cannot be achieved simply by removing half the machines.
- (b) Calculate the speed of sound in a steel railway rail, for which the Young's modulus is $2.0 \times 10^{11}\text{Pa}$ and the density is 7800kg/m^3 , and in seawater, for which the bulk modulus is $2.2 \times 10^9\text{Pa}$ and the density is 1030kg/m^3 . Hence find the ratio of the two speeds.

Solution

(a) Incoherent sources add their intensities, so doubling the number of machines doubles the total intensity. On the decibel scale a doubling of intensity raises the level by $10\log 2 \approx 3\text{dB}$, regardless of how many machines there were to begin with. The decibel scale is logarithmic and roughly matches the response of the ear, but the ear perceives a doubling of loudness only for a level increase of about 10dB, which corresponds to a tenfold increase in intensity. To make the workshop sound only half as loud, the intensity must be cut to about one tenth, which means removing about nine machines out of every ten. Removing only half the machines lowers the level by just 3dB, far short of the 10dB the ear needs to register a halving of loudness.

(b) In a solid the elastic property is the Young's modulus and the wave speed is $v = \sqrt{\frac{Y}{\rho}}$. For the steel rail:

$$v_{\text{steel}} = \sqrt{\frac{Y}{\rho}} = \sqrt{\frac{2.0 \times 10^{11} \text{Pa}}{7800 \text{kg/m}^3}} = 5064 \text{m/s}$$

In a liquid the elastic property is the bulk modulus and the wave speed is $v = \sqrt{\frac{K}{\rho}}$. For seawater:

$$v_{\text{water}} = \sqrt{\frac{K}{\rho}} = \sqrt{\frac{2.2 \times 10^9 \text{Pa}}{1030 \text{kg/m}^3}} = 1461 \text{m/s}$$

The ratio of the two speeds is:

$$\frac{v_{\text{steel}}}{v_{\text{water}}} = \frac{5064 \text{m/s}}{1461 \text{m/s}} = 3.47$$

Sound travels about 3.5 times faster in the steel rail than in seawater.

Example 35

- (a) Radio waves, visible light, X-rays and gamma rays are all members of the electromagnetic family. Explain why all of them travel at the same speed in a vacuum, and yet behave very differently when they meet matter.
- (b) An open pipe of length 36.0cm and a closed pipe of length 27.5cm, both of the same diameter, are sounded so that the first overtone of the open pipe is in unison with the first overtone of the closed pipe. Calculate the end correction at one open end, and hence the internal radius of the pipes.

Solution

(a) Every electromagnetic wave in a vacuum travels at the speed $c = \frac{1}{\sqrt{\mu_0 \epsilon_0}}$, where μ_0 and ϵ_0 are properties of empty space alone. These two constants do not depend on the frequency or wavelength of the wave, so every member of the family, from the longest radio wave to the shortest gamma ray, travels at exactly the same speed in a vacuum; the members differ only in wavelength and frequency. How a wave interacts with matter, however, depends strongly on its wavelength compared with the size of the objects and structures it meets, and on the energy each oscillation can deliver. A long radio wave diffracts around hills and buildings, visible light is absorbed and reflected by ordinary surfaces, and a short X-ray passes through flesh but is stopped by bone. Same speed in a vacuum, different wavelength, different behaviour in matter.

(b) Let e be the end correction at one open end, so $e = 0.6r$. The effective lengths are:

$$L_{\text{eff}}(\text{open}) = L_o + 2e = 0.360\text{m} + 2e$$

$$L_{\text{eff}}(\text{closed}) = L_c + e = 0.275\text{m} + e$$

The first overtone of the open pipe is its second harmonic, of frequency $\frac{2v}{2L_{\text{eff}}(\text{open})} = \frac{v}{L_{\text{eff}}(\text{open})}$.

The first overtone of the closed pipe is its third harmonic, of frequency $\frac{3v}{4L_{\text{eff}}(\text{closed})}$.

Setting the two equal for unison:

$$\frac{v}{L_{\text{eff}}(\text{open})} = \frac{3v}{4L_{\text{eff}}(\text{closed})}$$

$$4L_{\text{eff}}(\text{closed}) = 3L_{\text{eff}}(\text{open})$$

$$4(L_c + e) = 3(L_o + 2e)$$

$$2e = 4L_c - 3L_o$$

$$e = \frac{4L_c - 3L_o}{2} = \frac{4 \times 0.275\text{m} - 3 \times 0.360\text{m}}{2} = 0.010\text{m}$$

The end correction at one open end is 0.010m, or 1.0cm.

The radius follows from $e = 0.6r$:

$$r = \frac{e}{0.6} = \frac{0.010\text{m}}{0.6} = 0.0167\text{m} = 1.67\text{cm}$$

Example 36

- Explain why an astronaut standing on the surface of the Moon can see a colleague strike two rocks together but cannot hear the strike, and use the same physics to explain why sound is classified as a mechanical wave while light is not.
- A spectral line whose laboratory wavelength is $5000\text{\AA}$ is observed in the light of a star belonging to a binary system to oscillate between $4999.6\text{\AA}$ and $5000.4\text{\AA}$ over one orbital period. Taking the speed of light as $3.0 \times 10^8\text{m/s}$, calculate the orbital speed of the star.

Solution

(a) Sound is a mechanical wave, which means it is a disturbance of a material medium and cannot exist without one. Light is an electromagnetic wave, an oscillation of electric and magnetic fields that needs no medium at all. The Moon has no atmosphere, so on its surface there is no material medium for a sound wave to disturb, and the strike of the rocks produces no sound that can travel. The light reflected from the rocks, needing no medium, crosses the vacuum freely, so the colleague's action is seen but not heard. The same contrast fixes the classification: a wave that requires a medium to carry it is mechanical, and sound, which dies the instant its medium is removed, belongs to that class, while light, which crosses empty space unaided, does not.

(b) The line is most blueshifted when the star moves directly towards the Earth and most redshifted when it moves directly away, both at the full orbital speed v . The shift in each direction has magnitude $\Delta\lambda = \frac{v\lambda}{c}$, so the full swing from the shortest to the longest observed wavelength is:

$$\lambda_{\max} - \lambda_{\min} = 2\Delta\lambda = \frac{2v\lambda}{c}$$

Solve for v :

$$v = \frac{c(\lambda_{\max} - \lambda_{\min})}{2\lambda}$$

$$v = \frac{3.0 \times 10^8\text{m/s} \times (5000.4\text{\AA} - 4999.6\text{\AA})}{2 \times 5000\text{\AA}} = 2.4 \times 10^4\text{m/s}$$

The star moves in its orbit at $2.4 \times 10^4\text{m/s}$, or 24km/s.

Example 37

- Motion does not always produce a Doppler shift. State and explain the three situations in which a source, an observer, or both, are in motion and yet no change of frequency is heard.
- An air-traffic control radar at Julius Nyerere International Airport transmits microwave pulses at a frequency of 9.0GHz towards an aircraft 30km away. Calculate the wavelength of the pulses and the time that elapses between the transmission of a pulse and the return of its reflection.

Solution

(a) There are three such situations.

- When both source and observer are at rest in the medium, the wavefronts are emitted and met at the same rate, and the frequency is unchanged.
- When source and observer move with the same velocity through the medium, they are at rest relative to each other; a passenger seated in a daladala that is sounding its horn hears the horn at its true frequency however fast the daladala travels, because the horn and the passenger share the same motion through the air.

- 3) When the velocity of the source or observer is directed exactly at right angles to the line joining the two, as when a source moves along a circle with the observer at its centre, the distance between them is momentarily not changing, the component of velocity along that line is zero, and no shift is produced.
- (b) Treat the air as effectively a vacuum, so the wave speed is $c = 3.0 \times 10^8 \text{ m/s}$.

Wavelength:

$$\lambda = \frac{c}{f} = \frac{3.0 \times 10^8 \text{ m/s}}{9.0 \times 10^9 \text{ Hz}} = 0.033 \text{ m} = 3.3 \text{ cm}$$

The pulse travels to the aircraft and back, a total distance of $2 \times 30 \text{ km} = 6.0 \times 10^4 \text{ m}$, at the speed of light:

$$t = \frac{2d}{c} = \frac{6.0 \times 10^4 \text{ m}}{3.0 \times 10^8 \text{ m/s}} = 2.0 \times 10^{-4} \text{ s} = 0.20 \text{ ms}$$

The pulses have a wavelength of 3.3cm, and each reflection returns 0.20 milliseconds after the pulse is sent.

Example 38

- (a) Explain why a listener with a portable radio in Dar es Salaam can pick up distant stations from Nairobi or Maputo at midnight but hears only local stations at noon.
- (b) On a cold morning at 13°C , two organ pipes of slightly different lengths produce 6.0 beats per second when sounded together. Calculate the number of beats per second the same two pipes produce on a hot afternoon when the air temperature has risen to 38°C .

Solution

(a) Radio waves of the medium-wave broadcasting band can travel far beyond the horizon only by reflecting off the ionosphere, the electrically charged upper layers of the atmosphere. During the day the Sun's radiation creates a low absorbing layer in the ionosphere that soaks up these radio waves before they can reach the higher reflecting layer, so the listener hears only nearby stations whose signals arrive directly. At night this low absorbing layer disappears, the radio waves rise unhindered to the high reflecting layer, and they are bounced back to the ground hundreds of kilometres away. The listener in Dar es Salaam at midnight is therefore receiving signals that have reflected off the ionosphere from transmitters in distant cities, a path that the daytime absorbing layer had blocked.

(b) The frequency of an organ pipe is proportional to the speed of sound in the air inside it, and the speed of sound is proportional to the square root of the absolute temperature, $v \propto \sqrt{T}$.

When the air warms, the frequency of each pipe rises in the proportion $\sqrt{T}$, and since the beat frequency is the difference of the two pipe frequencies, it is multiplied by the same factor:

$$f_{\text{beat}}(T_2) = f_{\text{beat}}(T_1) \times \sqrt{\frac{T_2}{T_1}}$$

Convert the temperatures: $T_1 = (13 + 273)\text{K} = 286\text{K}$ and $T_2 = (38 + 273)\text{K} = 311\text{K}$.

$$f_{\text{beat}}(T_2) = 6.0 \text{ beats/s} \times \sqrt{\frac{311\text{K}}{286\text{K}}} = 6.3 \text{ beats/s}$$

The two pipes produce about 6.3 beats per second on the hot afternoon.

Example 39

- (a) Explain why the speed of sound in air depends strongly on the temperature of the air but not at all on its pressure.
- (b) A train sounds a whistle of true frequency 510Hz. As it travels along a straight track at constant speed, the frequency heard by a stationary observer while the train approaches is to the frequency heard after the train has passed in the ratio 9:8. Taking the speed of sound in still air as 340m/s, calculate the speed of the train and the two apparent frequencies.

Solution

(a) The speed of sound in an ideal gas can be written $v = \sqrt{\frac{\gamma RT}{M}}$, in which the density has cancelled out, leaving only the absolute temperature, the molar mass, and the constant γ . The speed therefore rises with temperature, in proportion to the square root of the absolute temperature, because warmer molecules move faster and pass the disturbance along more quickly. Pressure does not appear in this form at all. Although the alternative form $v = \sqrt{\frac{\gamma P}{\rho}}$ seems to show a dependence on pressure, at a fixed temperature the density of a gas is itself proportional to its pressure, so the ratio $\frac{P}{\rho}$ stays constant; doubling the pressure doubles the density, and the speed is unchanged. Temperature alters the speed of sound, pressure does not.

(b) For the approaching train the apparent frequency is $f' = \frac{fv}{v-u}$; for the receding train it is $f' = \frac{fv}{v+u}$. The ratio of the two is:

$$\frac{f'(\text{approach})}{f'(\text{recede})} = \frac{\frac{fv}{v-u}}{\frac{fv}{v+u}} = \frac{v+u}{v-u}$$

Set this equal to $\frac{9}{8}$:

$$\frac{v+u}{v-u} = \frac{9}{8}$$

$$8(v+u) = 9(v-u)$$

$$17u = v$$

$$u = \frac{v}{17} = \frac{340\text{m/s}}{17} = 20\text{m/s}$$

The train is travelling at 20m/s.

The two apparent frequencies are:

$$f'(\text{approach}) = \frac{fv}{v-u} = \frac{510\text{Hz} \times 340\text{m/s}}{340\text{m/s} - 20\text{m/s}} = 541.9\text{Hz}$$

$$f'(\text{recede}) = \frac{fv}{v+u} = \frac{510\text{Hz} \times 340\text{m/s}}{340\text{m/s} + 20\text{m/s}} = 481.7\text{Hz}$$

The observer hears 541.9Hz on approach and 481.7Hz on recession.

Example 40

- Explain why an end correction must be applied to the open end of a pipe, and explain why a student measuring the speed of sound with a resonance tube is advised to use two consecutive resonance lengths rather than a single one.
- A radar speed gun on the Morogoro Road operates at a frequency of 24GHz. It detects a Doppler shift of 3.6kHz in the wave reflected from an approaching car. Taking the speed of light as $3.0 \times 10^8\text{m/s}$, calculate the speed of the car.

Solution

(a) The displacement antinode at the open end of a pipe does not sit exactly at the geometric end. The air just inside the open end is free to move, and its motion extends a little way into the air beyond the pipe, so the antinode lies a small distance outside the geometric end. This extra distance is the end correction, and it must be added to the measured length to give the true effective length of the air column. A student measuring the speed of sound with a resonance tube cannot easily measure this end correction directly. If, however, two consecutive resonance lengths are recorded, the difference between them is exactly half a wavelength and the unknown end correction cancels out of the subtraction. Using two lengths rather than one therefore removes the systematic error the end correction would otherwise introduce.

(b) For a radar gun the wave is shifted twice, once when the moving car receives it and once when the car reflects it, so the detected shift for an approaching target is $\Delta f = \frac{2fu}{c}$. Solve for u :

$$u = \frac{\Delta f \times c}{2f}$$

$$u = \frac{3.6 \times 10^3 \text{ Hz} \times 3.0 \times 10^8 \text{ m/s}}{2 \times 24 \times 10^9 \text{ Hz}} = 22.5 \text{ m/s} = 81 \text{ km/h}$$

The car is travelling at 22.5m/s, or about 81km/h.

Example 41

- (a) Explain why the spectral lines of almost every distant galaxy are observed to be shifted towards the red end of the spectrum, and state what an astronomer learns about a galaxy from the size and the direction of the shift.
- (b) For nitrogen gas at 300K, with $\gamma = 1.40$, molar mass 28.0g/mol and $R = 8.314\text{J}/(\text{mol} \cdot \text{K})$, calculate the speed of sound and the root-mean-square speed of the molecules, and show that their ratio equals $\sqrt{\frac{\gamma}{3}}$.

Solution

(a) The light from almost every distant galaxy arrives with its spectral lines shifted towards longer wavelengths, the red end of the spectrum. This redshift is a Doppler effect: the galaxy is moving away from the Earth, so the wavelengths of the light it emits are stretched before they reach us. Almost all distant galaxies show this shift because the universe itself is expanding, carrying every galaxy away from every other. From the direction of the shift the astronomer learns whether the galaxy is receding, shown by a shift to the red, or approaching, shown by a shift to the blue. From the size of the shift, expressed as the fractional change in wavelength, the astronomer calculates the speed of the galaxy along the line of sight, since that fractional change equals the ratio of the recession speed to the speed of light.

(b) Convert the molar mass: $M = 28.0\text{g/mol} = 0.0280\text{kg/mol}$.

Speed of sound:

$$v_{\text{sound}} = \sqrt{\frac{\gamma RT}{M}} = \sqrt{\frac{1.40 \times 8.314\text{J}/(\text{mol} \cdot \text{K}) \times 300\text{K}}{0.0280\text{kg/mol}}} = 353\text{m/s}$$

Root-mean-square speed of the molecules:

$$v_{\text{rms}} = \sqrt{\frac{3RT}{M}} = \sqrt{\frac{3 \times 8.314\text{J}/(\text{mol} \cdot \text{K}) \times 300\text{K}}{0.0280\text{kg/mol}}} = 517\text{m/s}$$

Their ratio:

$$\frac{v_{\text{sound}}}{v_{\text{rms}}} = \frac{353\text{m/s}}{517\text{m/s}} = 0.683$$

Compare with $\sqrt{\frac{\gamma}{3}} = \sqrt{\frac{1.40}{3}} = \sqrt{0.467} = 0.683$.

Hence the ratio of the two speeds equals $\sqrt{\frac{\gamma}{3}}$, as required.

Example 42

- (a) Explain why a person hearing a recording of his own voice for the first time finds that it does not sound like the voice he hears when he speaks, even though the recording is faithful.
- (b) A car travels at 20m/s directly towards a tall vertical cliff while sounding a horn of true frequency 400Hz. Taking the speed of sound in air as 340m/s, calculate the beat frequency the driver hears.

Solution

(a) The quality, or timbre, of a voice is fixed by the particular mixture of harmonics it contains, not by its fundamental frequency alone. When a person speaks, the sound reaches his own ears by two routes. One route is through the air to the outer ear; the other is directly through the bones of his skull, which conduct

the lower harmonics especially well and so make the voice in his head sound deeper and fuller. A recording captures only the sound that travelled through the air, which carries a larger proportion of the higher harmonics. The recording is faithful to that airborne sound, but it lacks the bone-conducted low harmonics the speaker is used to hearing, so the harmonic mixture is different and the voice sounds unfamiliar to him, even though everyone else hears him exactly as the recording does.

(b) The driver hears two tones at once. The first is the sound coming directly from the horn; because the horn and the driver move together with the car, they are at rest relative to each other and the driver hears this direct sound at the true frequency, 400Hz.

The second tone is the echo from the cliff, shifted twice. First the cliff receives the sound from the approaching car:

$$f_1 = \frac{fv}{v - u}$$

The cliff then reflects f_1 , and the driver approaches the cliff as an observer:

$$f_{\text{echo}} = \frac{f_1(v + u)}{v} = \frac{fv}{v - u} \times \frac{v + u}{v} = \frac{f(v + u)}{v - u}$$

$$f_{\text{echo}} = \frac{400\text{Hz} \times (340\text{m/s} + 20\text{m/s})}{340\text{m/s} - 20\text{m/s}} = 450\text{Hz}$$

The beat frequency is the difference between the echo and the direct sound:

$$f_{\text{beat}} = f_{\text{echo}} - f = 450\text{Hz} - 400\text{Hz} = 50\text{Hz}$$

The driver hears 50 beats per second.

Example 43

- (a) Explain why an X-ray can be used to photograph a broken bone inside the body while ordinary visible light cannot, even though both are electromagnetic waves.
- (b) A tuning fork sounded with a sonometer wire of vibrating length 25.0cm produces 3 beats per second. When the vibrating length is shortened to 24.0cm, with the tension unchanged, the wire still produces 3 beats per second. Calculate the frequency of the tuning fork.

Solution

(a) Whether an electromagnetic wave passes through a material or is stopped by it depends on the wave and on the material. Visible light is strongly absorbed and reflected by the skin and soft tissue of the body, so it cannot reach a bone inside and carries back no information about it. X-rays have a much shorter wavelength and carry far more energy in each oscillation; they pass readily through soft tissue, which absorbs them only weakly, but they are absorbed much more strongly by the denser, heavier material of bone. An X-ray beam sent through a limb therefore emerges strongly weakened where it has crossed bone and barely weakened where it has crossed only flesh, and this difference, recorded on a detector, forms an image of the bone. Visible light cannot form such an image because it never reaches the bone in the first place.

(b) The fundamental frequency of a stretched wire at fixed tension is inversely proportional to its vibrating length, $f \propto \frac{1}{L}$. Shortening the wire from 25.0cm to 24.0cm therefore raises its frequency. The beat rate is the same magnitude before and after, so the wire must have been below the fork at 25.0cm and above the fork at 24.0cm. Writing f_0 for the fork frequency:

At 25.0cm the wire frequency is $f_0 - 3$

At 24.0cm the wire frequency is $f_0 + 3$

Since the wire frequency is inversely proportional to length:

$$\frac{f_0 - 3}{f_0 + 3} = \frac{L_2}{L_1} = \frac{24.0\text{cm}}{25.0\text{cm}}$$

$$25(f_0 - 3) = 24(f_0 + 3)$$

$$f_0 = 147\text{Hz}$$

The frequency of the tuning fork is 147Hz.

Example 44

- (a) A daladala sounds its horn while travelling along the Bibi Titi road. Explain why a pedestrian standing on the verge hears the pitch of the horn change as the daladala passes, while a passenger seated inside the daladala hears no change of pitch at all, however fast the daladala travels.
- (b) A steel railway rail on the TAZARA line is 1700m long. A single blow is struck on the rail at one end. Taking the speed of sound in steel as 5000m/s and in air as 340m/s, calculate the time interval between the two sounds heard by a worker with an ear pressed to the rail at the other end.

Solution

(a) The Doppler effect arises only when the source and the observer move differently through the air. The pedestrian on the verge is at rest in the air while the daladala carrying the horn moves; source and observer move differently, so the pedestrian hears the pitch rise as the horn approaches and fall as it recedes. The passenger seated inside the daladala, on the other hand, moves with the daladala, and so does the horn; the horn and the passenger share exactly the same motion through the air and are at rest relative to each other. With no difference in their motion through the medium there is no Doppler shift, and the passenger hears the horn at its true frequency however fast the daladala travels.

(b) The single blow sends a sound disturbance into the steel rail and also into the surrounding air. The sound through the steel covers the length of the rail at the speed of sound in steel:

$$t_{\text{steel}} = \frac{d}{v_{\text{steel}}} = \frac{1700\text{m}}{5000\text{m/s}} = 0.34\text{s}$$

The sound through the air covers the same distance at the speed of sound in air:

$$t_{\text{air}} = \frac{d}{v_{\text{air}}} = \frac{1700\text{m}}{340\text{m/s}} = 5.0\text{s}$$

The time interval between the two sounds is the difference:

$$\Delta t = t_{\text{air}} - t_{\text{steel}} = 5.0\text{s} - 0.34\text{s} = 4.66\text{s}$$

The worker hears the second sound, carried by the air, 4.66s after the first sound, carried by the rail.

These worked examples have carried the physics of sound, electromagnetic waves, and the Doppler effect through one situation after another, drawing the separate ideas of the chapter into a single connected picture. With that picture now clear, we are ready to put it to work. The Digging Deeper Exercise on the next page is where these ideas are tested, stretched, and made firmly our own.

DIGGING DEEPER EXERCISE 9

EXERCISE 9A: BINDER QUESTIONS

Question 1

A student states that an air particle in a sound wave is surrounded by the greatest pressure at the moment its displacement is greatest. Explain why this statement is incorrect.

Question 2

Explain why sound travels slightly faster through humid air than through dry air at the same temperature.

Question 3

Music played by a band reaches a distant listener in time and in tune, with no note arriving ahead of any other. Account for this observation.

Question 4

Explain why blowing harder across the mouth of a pipe eventually makes its note jump abruptly to a higher pitch, instead of rising smoothly.

Question 5

A student reads "0 dB" and concludes there is no sound at all. Account for what 0 dB represents and what a negative decibel level would mean.

Question 6

Two pure tones, of frequencies 50Hz and 3kHz, are produced at the same intensity. Explain why a listener does not hear them as equally loud.

Question 7

Explain why a sound wave cannot be polarised, whereas a light wave can.

Question 8

State and explain the phase relationship between the electric and magnetic fields of an electromagnetic wave travelling through a vacuum.

Question 9

Explain why the ultraviolet part of sunlight damages the skin, while the more intense visible light does not.

Question 10

In the Doppler effect for sound, explain why the motion of the source and the motion of the observer affect the observed frequency in different ways, even when their speeds are equal.

Question 11

A distant galaxy shows a large redshift. Explain whether this is better understood as motion of the galaxy through space, or as an expansion of the space between the galaxy and the Earth.

Question 12

A galaxy moves at high speed, but directly across the line of sight from the Earth. Explain why its spectral lines show no Doppler shift in wavelength.

EXERCISE 9B: REAL QUESTIONS

Question 13

Elephants in the Serengeti communicate with one another over distances of several kilometres using calls whose frequency lies below 20Hz. Explain why such low-frequency sound carries much further than ordinary speech.

Question 14

A wall between two rooms muffles the sound of a conversation next door, yet the deep bass of the music played there passes through clearly. Explain this difference.

Question 15

Kipanga: *“I can hear Mr. Akilikubwa clearly on the far side of that building, yet I cannot see him at all.”*

Kipute: *“The same building stands in the way of both his voice and his image.”*

Explain why the sound reaches Kipanga but the light does not.

Question 16

An echo from a cliff returns noticeably later on a cold morning than on a hot afternoon. Account for this difference.

Question 17

As a bucket is filled with water from a tap, the sound it makes rises steadily in pitch. Account for this rise in pitch.

Question 18

The whine of a mosquito is clearly heard in a quiet bedroom at night, but the same mosquito cannot be heard in a busy Kariakoo market by day. Account for this difference.

Question 19

A tall barrier built along the Morogoro Road reduces the traffic noise reaching the houses behind it, but never makes them completely silent. Explain why complete silence is not achieved.

Question 20

A mobile phone often loses its signal the moment its user steps inside a metal lift. Explain why the signal is lost.

Question 21

An AM radio crackles and pops during a thunderstorm, even when it is tuned to a strong station. Account for the crackling.

Question 22

A thermal-imaging camera can produce a clear picture of a person standing in complete darkness. Explain how this is possible.

Question 23

A television remote control must be pointed at the set and will not work from another room, yet a radio receiver plays normally far from the transmitter and behind walls. Explain this difference.

Question 24

Sealed medical instruments at a hospital are sterilised by exposing them to gamma rays. Explain why gamma rays are suitable for this purpose.

Question 25

Of the whole electromagnetic spectrum, the human eye is sensitive only to the narrow band of visible light. Explain why it is this band that the eye evolved to detect.

Question 26

A Doppler monitor in a clinic detects the heartbeat of an unborn baby. Explain how the movement of the baby's heart is detected.

Question 27

Kipanga: *“Since the universe is expanding, every galaxy must be moving away from us and must show a redshift.”*

Kipute: *“Then explain the Andromeda galaxy, whose light is blueshifted.”*

Account for the blueshift of the Andromeda galaxy.

EXERCISE 9C: HOT QUESTIONS**Question 28**

On a cold night in Iringa the air temperature is 10°C , and on a hot afternoon it is 40°C . Calculate the percentage change in the speed of sound in air between the two.

Question 29

Carbon dioxide has $\gamma = 1.30$ and molar mass 44g/mol ; hydrogen has $\gamma = 1.41$ and molar mass 2.0g/mol . Calculate the ratio of the speed of sound in carbon dioxide to that in hydrogen at the same temperature.

Question 30

One machine produces a sound level of 74dB at a certain point. Treating the machines as incoherent sources of equal intensity, calculate how many identical machines, running together, would raise the level at that point to 86dB .

Question 31

A point source radiates sound uniformly in the open air, producing a sound level of 90dB at a distance of 6.0m . Taking the threshold of hearing as $I_0 = 1.0 \times 10^{-12}\text{W/m}^2$, calculate the acoustic power output of the source.

Question 32

At a certain point in a sound wave the displacement amplitude is halved, the frequency remaining unchanged. Calculate the resulting change in the sound level at that point.

Question 33

A pulse of light crosses a glass block 0.30m thick, of refractive index 1.50 . Taking the speed of light in vacuum as $c = 3.0 \times 10^8\text{m/s}$, calculate the extra time the pulse takes compared with crossing the same thickness of vacuum.

Question 34

A hydrogen spectral line of laboratory wavelength $6563\text{\AA}$ is observed from a distant galaxy at $6647\text{\AA}$. Taking $c = 3.0 \times 10^8\text{m/s}$ and the Hubble constant as 70km/s per Mpc , calculate the recession velocity of the galaxy and its distance from the Earth.

Question 35

Taking the Hubble constant as $H_0 = 70\text{km/s per Mpc}$, and $1\text{Mpc} = 3.09 \times 10^{19}\text{km}$, estimate the Hubble time in years and state what it represents.

Question 36

A whistle mounted on a truck is an open pipe of length 0.20m sounding its fundamental note. The truck travels at 30m/s directly towards a stationary roadside listener. Taking the speed of sound in air as 340m/s and neglecting end correction, calculate the frequency the listener hears.

Question 37

A train approaches a stationary observer at 25m/s , sounding a whistle of frequency 480Hz , on a day when the air temperature is 35°C . Taking the speed of sound in air at 0°C as 331m/s , calculate the frequency the observer hears.

Question 38

The cone of a loudspeaker, and the air in contact with it, executes simple harmonic motion of amplitude $1.0 \times 10^{-5}\text{m}$ at a frequency of 400Hz . Take the density of air as 1.2kg/m^3 , the speed of sound as 340m/s , and $I_0 = 1.0 \times 10^{-12}\text{W/m}^2$. Calculate (a) the maximum speed of the air particles, (b) the pressure amplitude of the sound wave, (c) its intensity, and (d) the sound level produced.

Question 39

A closed organ pipe of length 0.25m is sounded together with a sonometer wire of vibrating length 0.50m and linear mass density $1.2 \times 10^{-3}\text{kg/m}$, the two being in unison at their fundamentals. Taking the speed of sound in air as 340m/s and neglecting end correction, calculate the tension in the wire.

Question 40

An ambulance siren of true frequency 700Hz approaches a stationary observer at 30m/s. Taking the speed of sound as 340m/s and treating the siren as a point source, calculate (a) the frequency the observer hears, and (b) the increase in the sound level the observer measures as the ambulance moves from 80m away to 20m away.

Question 41

In a resonance-tube experiment a tuning fork of frequency 512Hz gives consecutive resonances at air-column lengths of 16.0cm and 49.2cm. Taking the speed of sound in air at 0°C as 331m/s, calculate the speed of sound in the laboratory air, and hence deduce the air temperature.

Question 42

A satellite in low orbit transmits a radio beacon at a steady frequency of 150MHz, and approaches a ground station at 7.6km/s along the line of sight. Taking $c = 3.0 \times 10^8$ m/s, calculate the change in the received frequency, and explain why this shift falls to zero when the satellite passes directly overhead.

EXAMINATION QUESTIONS FOR PART TWO

Question 1

- (a) A transverse pulse travelling along a stretched string is inverted when it reflects from a rigidly fixed end, but returns the same way up when it reflects from a free end. Account for this difference.
- (b) In a 3D cinema the two lenses of the viewing glasses are polarising filters with their transmission axes set at right angles to one another. Account for how this lets each eye receive a separate image.
- (c) A radio station in Dar es Salaam broadcasts at a frequency of 90.0MHz. Taking the speed of electromagnetic waves as $3.00 \times 10^8 \text{m/s}$, calculate the wavelength of the broadcast and the time taken for the signal to reach a receiver 75km away.

Question 2

- (a) Light passing through a wide doorway shows no noticeable spreading, while light passing through a slit a fraction of a millimetre wide spreads markedly beyond the straight-through direction. Account for this.
- (b) A wave rolling in from deep water towards a gently shelving beach near Bagamoyo slows down and grows steeper as the water becomes shallower. Account for this.
- (c) At a workshop one machine alone produces a sound level of 82dB at a worker's position, and a second machine alone produces 78dB at the same position. Calculate the sound level at that position when both machines run together. Take the threshold of hearing as $I_0 = 1.0 \times 10^{-12} \text{W/m}^2$.

Question 3

- (a) The loudness of sound is measured on the decibel scale, which is logarithmic rather than linear. Account for the use of a logarithmic scale.
- (b) Waves rolling into the harbour at Tanga strike the vertical harbour wall and travel back out, crossing the incoming waves and producing a choppy criss-cross pattern on the water. Account for this.
- (c) Monochromatic light of wavelength 600nm falls normally on a diffraction grating ruled with 400 lines per millimetre. Calculate the highest order of diffraction maximum that can be observed.

Question 4

- (a) Two transverse pulses travelling in opposite directions along a stretched string meet, overlap, and then continue on their way, each emerging with its original shape, size and speed. Account for this.
- (b) A traffic police officer points a radar gun at a car approaching along the road. Account for how the gun is able to determine the car's speed.
- (c) Plane-polarised light is incident on an analysing polaroid. Find the angle between the transmission axis of the polaroid and the plane of polarisation of the light for which the transmitted intensity is one quarter of the incident intensity.

Question 5

- (a) An astronaut standing on the airless surface of the Moon would not hear a rock fall and shatter a few metres away. Account for this.
- (b) The feathers of a peacock and the wing-cases of certain beetles show brilliant, shifting colours, yet contain no coloured pigment. Account for this.
- (c) A transverse wave of amplitude 8.0mm and frequency 50Hz travels along a stretched string of linear mass density 0.020kg/m held under a tension of 80N. Calculate the speed of the wave and the average power it carries along the string.

Question 6

- (a) A guitarist can change the frequency of the wave he sends along a string, but not the speed at which it travels. Account for the fact that the wave speed is fixed by the string while the frequency is fixed by the source.
- (b) Looking at a distant streetlamp at night through the fine wire mesh of a window screen, an observer sees the single lamp spread into a regular pattern of bright spots. Account for this.
- (c) An organ pipe closed at one end has a length of 0.55m. Taking the speed of sound in air as 340m/s, calculate the three lowest frequencies at which the pipe resonates.

Question 7

- (a) A diffraction grating ruled with many thousands of lines produces much sharper and narrower bright lines than a double slit illuminated with the same light. Explain this.
- (b) A bat hunting at dusk emits ultrasonic pulses and listens for the echoes. Explain how the bat can tell whether a moth ahead of it is flying towards it or away from it.
- (c) Two waves of the same frequency travel in the same direction along a stretched rope. Their amplitudes are 3.0cm and 5.0cm and there is a constant phase difference of 60° between them. Calculate the amplitude of the resultant wave.

Question 8

- (a) In a stationary wave on a stretched string, all the particles in one loop reach their extreme displacements at the same instant, while a particle in the next loop reaches its extreme on the opposite side of the string at that same instant. Explain.
- (b) A military jet flying low and very fast over Dar es Salaam is followed by a sudden, loud bang heard at the ground. Account for this bang.
- (c) Monochromatic light of wavelength 600nm passes through a single slit of width 0.20mm, and the diffraction pattern is observed on a screen 2.0m away. Calculate the angular position of the first dark fringe and the width of the central bright band on the screen.

Question 9

- (a) In a two-slit interference pattern the bright fringes are all of very nearly the same brightness, but in a single-slit diffraction pattern the bright bands grow steadily fainter on either side of the centre. Account for this difference.
- (b) Two children make a telephone from two empty tins joined by a length of string. When the string is pulled taut a child speaking into one tin is heard clearly at the other; when the string goes slack, nothing is heard. Account for this.
- (c) A car travelling at 30m/s towards a vertical cliff sounds a horn of frequency 480Hz. Taking the speed of sound in air as 340m/s, calculate the frequency of the echo heard by the driver.

Question 10

- (a) During a storm the flash of lightning is seen well before the thunder is heard, although both are produced at the same instant by the same lightning stroke. Account for the delay.
- (b) A blacksmith heats a steel tool; as it cools in the air its polished surface passes through a sequence of colours, from pale straw through brown and purple to blue. Account for these colours.
- (c) A wire fixed at both ends sounds a particular fundamental note, its length and mass per unit length kept constant. Determine the factor by which the tension must be increased to raise the frequency of the fundamental in the ratio 3:2.

Question 11

- (a) A mass oscillating on a spring and a swinging pendulum bob appear to be very different systems, yet both execute simple harmonic motion. Account for the feature they share that makes this so.
- (b) When a glass screen-protector is pressed onto a phone screen, patches of coloured rings and bands appear in the regions where the two surfaces almost, but not quite, touch. Account for these patterns.
- (c) A motorcyclist rides at 20m/s directly towards a stationary siren that emits a steady note of frequency 660Hz. Taking the speed of sound in air as 340m/s, calculate the frequency of the note heard by the motorcyclist.

Question 12

- (a) In a Newton's-rings experiment the dark and bright fringes are seen as a series of concentric circles centred on the point of contact. Account for the circular shape of the fringes.
- (b) A diving board continues to move up and down for a short while after the diver has sprung off it. Account for this continued motion.
- (c) A machine treated as a point source produces a sound level of 88dB at a distance of 2.0m in the open air. Calculate the sound level at a distance of 8.0m from the machine.

Question 13

- (a) An organ pipe closed at one end sounds a note an octave lower than an open pipe of the same length. Account for this.
- (b) A baby placed in a bouncer, a seat hung from a doorframe by an elastic strap, bounces up and down with a steady, regular rhythm. Account for this rhythmic bouncing.
- (c) In a Young's double-slit experiment the bright fringes on the screen are 1.8mm apart when the apparatus is in air. The whole apparatus is then immersed in water of refractive index 1.33. Calculate the new spacing of the bright fringes.

Question 14

- (a) If the amplitude of a simple harmonic oscillator is doubled, its maximum speed and its maximum acceleration both double, yet its period is unchanged. Account for this.
- (b) In the great domed hall of a certain building, a person whispering quietly at one particular point can be heard clearly by someone far away at another particular point, though listeners between the two hear nothing. Account for this.
- (c) A soap film of refractive index 1.33 and thickness 1.10×10^{-7} m is illuminated normally by white light. Calculate the wavelength in the visible range that is most strongly reflected.

Question 15

- (a) Sound travels faster through warm air than through cold air. Account for this.
- (b) Seen through a windscreen covered with fine scratches left by the wipers, the headlights of oncoming cars at night appear drawn out into bright streaks and spikes of light. Account for this.
- (c) A 0.60kg mass hung from a single light spring oscillates vertically with a period of 0.80s. The mass is then hung from two such springs side by side, both supporting it together. Calculate the period of the vertical oscillations of the mass on the pair of springs.

Question 16

- (a) Account for the fact that the total energy of a simple harmonic oscillator is proportional to the square of its amplitude.
- (b) When a thin straight wire is held a short distance in front of one eye and a distant bright vertical lamp is viewed past its edge, a series of fine light and dark bands is seen. Account for these bands.
- (c) An organ pipe open at both ends is to sound a fundamental note of frequency 256Hz. Taking the speed of sound in air as 340m/s, calculate the length of the pipe and the frequency of its second harmonic.

Question 17

- (a) In Young's double-slit experiment a single narrow slit is placed between the lamp and the double slit. Account for the need for this single slit.
- (b) A person who hears a recording of their own voice for the first time usually finds that it sounds unfamiliar and somewhat thinner than the voice they are used to hearing themselves speak in. Account for this.
- (c) A 0.30kg mass hangs from a vertical spring of force constant 30N/m and is set oscillating with an amplitude of 0.080m. Calculate the period of the oscillation and the maximum and minimum tension in the spring during the motion. Take $g = 9.8\text{m/s}^2$.

Question 18

- (a) Account for the fact that the kinetic energy of a simple harmonic oscillator varies with time at twice the frequency of its displacement.
- (b) A doctor's stethoscope makes the faint sounds of a patient's heartbeat loud enough to hear clearly. Account for how it does this.
- (c) Monochromatic light falls normally on a diffraction grating ruled with 500 lines per millimetre. The second-order maximum is observed at an angle of 36.0° to the straight-through direction. Calculate the wavelength of the light.

Question 19

- (a) Sunlight reflected from a smooth water surface is found to be partially polarised, and at one particular angle of incidence it is completely polarised. Account for these observations.
- (b) A heavy load hangs from the long cable of a crane. When the load is set swinging, the operator notices that paying out more cable makes the swinging slower and lazier. Account for this.

- (c) A radar station sends out a pulse of microwaves and receives the echo reflected from an aircraft 1.20×10^{-4} s later. Taking the speed of electromagnetic waves as 3.00×10^8 m/s, calculate the distance of the aircraft from the station.

Question 20

- (a) Radio waves, visible light and X-rays are put to very different uses and have very different effects, yet they are all regarded as the same kind of wave. Account for this.
- (b) Looking through polarised sunglasses at the toughened glass rear window of a car in daylight, a passenger sees a pattern of dark spots and blotches spread across the glass. Account for this.
- (c) Calculate the length of a simple pendulum that beats seconds, that is, has a period of 2.0s, at a place where $g = 9.8\text{m/s}^2$. If this pendulum is then lengthened by 1.0%, calculate the resulting percentage change in its period.

Question 21

- (a) Explain what is meant by the natural frequency of an oscillating system, and state the physical quantities that fix it for a mass on a spring and for a simple pendulum.
- (b) At a football match the spectators rise and sit in turn, sending a 'wave' that travels right round the stadium, although no spectator leaves their seat. Account for the fact that the wave travels round the ground while the spectators do not.
- (c) A girl standing in front of a high cliff claps her hands once and hears the echo 1.5s later. Taking the speed of sound in air as 340m/s, calculate her distance from the cliff. She then walks 60m directly towards the cliff and claps again; calculate the time after which she now hears the echo.

Question 22

- (a) Explain why sound travels faster through a solid such as steel, and through a liquid such as water, than it does through air.
- (b) A bungee jumper, after the first plunge, bounces up and down several times, each bounce smaller than the one before, until at last she hangs almost at rest. Explain this behaviour in terms of oscillations.
- (c) Water waves travelling across a pond move at 1.8m/s and pass a fixed post 2.5 times every second. Calculate the wavelength of the waves, and determine the phase difference between two points lying 0.30m apart along the direction in which the waves travel.

Question 23

- (a) A wave travelling along a light rope reaches a join where the light rope is tied to a much heavier rope. Explain why part of the wave is reflected at the join and part is transmitted into the heavier rope.
- (b) A person standing near the corner of a building can hear someone speaking just around the corner, even though the speaker cannot be seen. Explain how the sound reaches the listener.
- (c) A particle executing simple harmonic motion has a maximum speed of 0.30m/s and a maximum acceleration of 4.5m/s^2 . Calculate the angular frequency, the period and the amplitude of the motion.

Question 24

- (a) During simple harmonic motion the displacement, the velocity and the acceleration of the body all vary with time. Explain how the velocity and the acceleration are each related, in phase, to the displacement.
- (b) A survey ship finds the depth of water beneath it by sending a pulse of sound down to the sea-bed and detecting the pulse that returns. Account for the way this method reveals the depth.
- (c) A long string has a mass per unit length of $2.0 \times 10^{-3}\text{kg/m}$ and is held under a tension of 50N. A vibrator at one end sends transverse waves of frequency 100Hz along it. Calculate the speed of the waves on the string and their wavelength.

Question 25

- (a) The note of the siren of an ambulance is heard to fall in pitch at the moment the ambulance passes a stationary listener. Explain why the pitch falls.
- (b) When a stone is dropped into a still pond, circular ripples spread outwards and grow steadily weaker the further they travel from the point where the stone entered. Account for the weakening of the ripples as they spread.

- (c) A 0.25kg mass hung from a light spring oscillates vertically with a period of 0.50s. Calculate the force constant of the spring, and determine the further mass that must be added to the 0.25kg so that the period becomes 1.0s.

Question 26

- (a) Use Huygens' principle to explain how a wavefront advances through a medium.
- (b) A musician's metronome marks time with a steady tick, and the musician changes the tempo by sliding the small weight up or down its swinging arm. Explain, in terms of oscillations, why moving the weight changes the rate of ticking.
- (c) The speed of sound in air is 340m/s when the air temperature is 15°C, that is 288K. Calculate the speed of sound when the air warms to 35°C. Hence calculate the wavelength in air of the note of a 480Hz tuning fork at each of these two temperatures.

Question 27

- (a) Distinguish between the loudness and the pitch of a musical note, and state the physical property of the sound wave on which each of them depends.
- (b) A heavy adult and a light child sit in turn on the same playground swing. Explain why each of them completes one full to-and-fro swing in the same time.
- (c) Transverse wave pulses travel at 40m/s along a long cable stretched between two fixed posts 100m apart. A pulse is sent from one post; calculate the time after which the pulse, reflected from the far post, returns to its starting point. The tension in the cable is then increased to four times its original value; calculate the new return time.

Question 28

- (a) Explain why the wavefronts produced by a small source of waves are very nearly circular close to the source, but appear practically straight to an observer far away from it.
- (b) In a hospital, an image of an unborn baby is formed using ultrasound rather than ordinary audible sound. Suggest why ultrasound is used for this purpose.
- (c) A body of mass 0.50kg executes simple harmonic motion with a total energy of 0.16J and an amplitude of 0.080m. Calculate the force constant of the system and the maximum speed of the body.

Question 29

- (a) Explain why the amplitude of a forced oscillation becomes very large when the frequency of the driving force is close to the natural frequency of the system.
- (b) A voice in an empty room with bare walls and floor rings on and sounds hollow, but once the room is furnished with curtains, carpets and soft chairs the hollow ringing disappears. Account for this difference.
- (c) A stretched string vibrating as a stationary wave has a distance of 0.12m between a node and the antinode next to it, and the string vibrates 25 times each second. Calculate the wavelength of the waves on the string and the speed at which they travel along it.

Question 30

- (a) Explain why an echo can be heard as a sound separate from the original only when the reflecting surface is sufficiently far from the listener.
- (b) When two children turn a skipping rope steadily, the rope settles into a single broad loop, with the ends almost still and the middle sweeping out the widest path. Explain why this pattern forms.
- (c) A particle moving with simple harmonic motion has a speed of 8.0cm/s when it is 6.0cm from the centre of its path, and a speed of 6.0cm/s when it is 8.0cm from the centre. Calculate the amplitude and the period of the motion.

Question 31

- (a) Explain why a body executing simple harmonic motion spends, over each complete cycle, more of its time near the extremes of its path than near the centre.
- (b) Boats moored in the sheltered water behind a harbour breakwater are still seen to rock gently, although the breakwater stands directly between them and the open sea. Account for the gentle rocking of the boats.

- (c) Monochromatic light of wavelength 590nm is incident normally on a diffraction grating, and the first-order bright image is seen at an angle of 18.0° to the straight-through direction. Calculate the spacing of the grating lines and the number of lines per millimetre of the grating.

Question 32

- (a) In a two-slit interference pattern there are dark fringes at which light arrives from both slits and yet no light is seen. Explain how light from the two slits can combine to give darkness, and account for what becomes of the light energy at a dark fringe.
- (b) A hacksaw blade clamped firmly to the edge of a bench and twanged vibrates with a steady buzz; sliding more of the blade out beyond the edge of the bench makes the buzz deepen. Explain this behaviour in terms of oscillations.
- (c) A wire 0.90m long is fixed at both ends and is made to vibrate in its third harmonic, sounding a note of frequency 200Hz. Calculate the wavelength of the waves on the wire and the speed at which they travel along it.

Question 33

- (a) Explain why a stationary wave has points, the nodes, at which the medium never moves at all, even though each of the two waves that combine to form it would, on its own, move every point of the medium.
- (b) A tuning fork sounds the very same musical note whether it is struck hard or struck gently. Explain why striking it harder changes the loudness of the note but not the note itself.
- (c) In a Young's double-slit experiment the two slits are 0.40mm apart and the fringes are observed on a screen 1.5m away. The bright fringes are found to be 2.1mm apart. Calculate the wavelength of the light used.

Question 34

- (a) A diffraction grating and a glass prism both spread white light into a spectrum, but the grating bends red light most and violet least, while the prism does the reverse. Account for the fact that the grating spreads the colours in the opposite order to the prism.
- (b) A child pushes and pulls the end of a long springy coil lying on the floor, sending a pulse of compression running along its length to the far end. Account for the way the pulse travels along the coil.
- (c) A spring of force constant 18N/m is cut into two equal pieces. Calculate the force constant of each piece, and the period of vertical oscillation of a 0.40kg mass hung from one of the pieces.

Question 35

- (a) Explain why a mass suspended from a vertical spring oscillates with exactly the same period as it would if the same spring and mass were laid horizontally, even though gravity pulls on the hanging mass.
- (b) On a misty night a bright street lamp is seen surrounded by a faint set of coloured rings. Suggest why these coloured rings appear.
- (c) A continuous train of waves of frequency 20Hz travels along a stretched cord at a speed of 8.0m/s. Calculate the wavelength of the waves, and the number of complete waves contained in a 12m length of the cord.

Question 36

- (a) Explain why the energy carried past a point each second by a progressive wave depends on the square of the amplitude of the wave.
- (b) When a car approaches from far away at night, its two headlamps appear at first as a single point of light, and only when the car is much nearer do they become distinguishable as two separate lamps. Suggest why the two lamps cannot be told apart until the car is close.
- (c) A particle executes simple harmonic motion with an amplitude of 0.10m. Calculate the displacement at which its kinetic energy is three times its potential energy, and the fraction of the total energy that is kinetic when the displacement of the particle is 0.080m.

Question 37

- (a) A mass oscillating on a spring is replaced by a mass twice as large on the same spring. Explain why the period of oscillation increases by a factor of $\sqrt{2}$ rather than doubling.
- (b) When a whip is cracked, the wave sent travelling down it moves faster, and with greater amplitude, as it nears the thin tip. Account for the speeding-up of the wave as it travels towards the tip.

- (c) Unpolarised light of intensity 16W/m^2 is incident on a pair of polaroids whose transmission axes are set at 60° to each other. Calculate the intensity of the light transmitted through the first polaroid, and the intensity of the light emerging from the second.

Question 38

- (a) Account for the equal spacing of the bright fringes across the screen in a Young's double-slit interference pattern.
- (b) A child bounces steadily up and down on a pogo stick, a vertical toy with a strong spring for hopping, each bounce taking the same time as the one before. Explain, in terms of oscillations, what makes the bouncing so regular.
- (c) A transverse wave of amplitude 1.5cm and frequency 12Hz travels along a stretched string at a speed of 9.0m/s . Calculate the maximum speed of a particle of the string, and state how it compares with the speed of the wave itself.

Question 39

- (a) Explain why a train of water waves changes its direction of travel when it crosses, at an angle, into a region where the water is shallower and the waves move more slowly.
- (b) A small boat at sea, rocked to one side by a passing wave, continues to roll to and fro with a steady rhythm of its own after the wave has gone. Account for this rolling.
- (c) In a Newton's-rings experiment the plano-convex lens has a radius of curvature of 1.2m . The radius of the eighth dark ring, counted from the centre, is measured to be 2.4mm . Calculate the wavelength of the light used.

Question 40

- (a) Explain why the sky appears blue to an observer on the ground during the day.
- (b) Waves approaching a straight beach are often seen to arrive very nearly parallel to the shoreline, even when, far out, they were travelling towards the shore at a noticeable angle. Explain why the waves arrive nearly parallel to the shore.
- (c) A particle moving with simple harmonic motion has a period of 0.40s and travels a total distance of 24cm in each complete oscillation. Calculate the amplitude of the motion, its maximum speed and its maximum acceleration.

ANSWERS TO DIGGING DEEPER EXERCISES

EXERCISE 6A

1. The condition for SHM is that the acceleration is directed towards a fixed point and proportional in magnitude to the displacement from that point, $a = -\omega^2 x$. A particle in uniform circular motion has a centripetal acceleration of **constant** magnitude $\omega^2 A$ directed towards the centre, so the magnitude of the acceleration does not vary with the displacement; the motion is therefore not SHM. The projection of the position vector on a diameter, however, has displacement $x = A \cos(\omega t)$ and acceleration $\frac{d^2 x}{dt^2} = -\omega^2 A \cos(\omega t) = -\omega^2 x$, which does satisfy the SHM condition. Kipute is correct.

2. The error is to confuse a free choice with a vacuous one. The experimenter is free to choose where $t = 0$ lies, but once that choice is made, ε is fixed and is no longer at the experimenter's disposal. It encodes where the oscillator is, and which way it is moving, at that chosen instant: $y(0) = A \sin \varepsilon$ and $v(0) = A\omega \cos \varepsilon$. When two oscillators of the same A and ω are observed together, their phase difference $\Delta \varepsilon = \varepsilon_1 - \varepsilon_2$ is the physical relationship between them, telling whether they are in step ($\Delta \varepsilon = 0$), out of step ($\Delta \varepsilon = \pi$), or in quadrature ($\Delta \varepsilon = \pi/2$).

3. The two formulas describe the same situation written in two languages. Starting from $y = A \sin(\omega t + \varepsilon)$ and $v = A\omega \cos(\omega t + \varepsilon)$, squaring and adding gives:

$$y^2 + \frac{v^2}{\omega^2} = A^2$$

so that $v = \omega \sqrt{A^2 - y^2}$: Kipanga's formula is in y and Kipute's in t . The speed is maximum at $y = 0$ because that is where the oscillator has been falling "downhill" in the energy diagram for the longest time without yet having to climb; it is zero at $y = \pm A$ because all the kinetic energy has there been converted to potential energy. The acceleration is unrelated to the speed; it depends on the displacement only, and so is greatest where the speed is least.

4. The error is the assumption that zero velocity implies zero acceleration. The acceleration in SHM is $a = -\omega^2 x$; it depends on the displacement, not on the velocity. At the positive extreme, $x = +A$, so $a = -\omega^2 A$ has the largest magnitude it ever has, directed back towards the centre. The body is momentarily stationary at the extreme precisely because it has been decelerating throughout the previous quarter cycle and is now turning round; the acceleration is therefore at its largest, not its smallest.

5. The error is the conflation of an average with an instantaneous equality. With $x = A \sin(\omega t)$ and $v = A\omega \cos(\omega t)$, the kinetic energy is $K = \frac{1}{2} k A^2 \cos^2(\omega t)$ and the potential energy is $U = \frac{1}{2} k A^2 \sin^2(\omega t)$. They are nowhere equal at every instant; on the contrary, when one is at its maximum the other is zero. They are equal only on the time-average, since the mean of $\sin^2(\omega t)$ and of $\cos^2(\omega t)$ over a complete cycle is each $\frac{1}{2}$, and so each averages to $\frac{1}{4} k A^2$, together accounting for the total energy $\frac{1}{2} k A^2$. The equipartition of average energies between the two stores reflects the symmetric way in which the oscillator passes between them.

6. The error is to confuse "plays no part" with "does not appear in the period formula". Gravity does act on the mass, and it is responsible for the new equilibrium position being lower than the natural length of the spring by $e = mg/k$. When the mass is displaced by x from this new equilibrium, the net force is $F = -k(e + x) + mg = -kx$, because the constant gravitational pull mg is exactly cancelled by the constant extra spring tension ke . Gravity has therefore been absorbed into the choice of equilibrium and disappears from the equation of motion, leaving $T = 2\pi \sqrt{\frac{m}{k}}$. A weaker g simply moves the equilibrium up; the oscillation about it is unchanged.

7. The exact restoring acceleration is $a = -g \sin \theta$, which is not proportional to θ ; the motion is therefore not strictly SHM. The formula $T = 2\pi \sqrt{\frac{l}{g}}$ is obtained by replacing $\sin \theta$ with θ , an approximation which is only valid when θ is small. For larger amplitudes $\sin \theta < \theta$, the restoring acceleration is weaker than the SHM model predicts, and the period is correspondingly longer; it grows with amplitude. Kipute is therefore right. Kipanga's claim is acceptable for amplitudes up to about 10° , where $\sin \theta$ differs from θ by less than 0.5%, and the resulting error in T is well below 1%.

8. The squared period rearranges as:

$$T^2 = \frac{4\pi^2(k^2 + h^2)}{gh} = \frac{4\pi^2}{g} \left(\frac{k^2}{h} + h \right),$$

exposing two competing terms that depend on h in opposite ways. As $h \rightarrow 0$, the term $\frac{k^2}{h}$ diverges and $T \rightarrow \infty$: pivoting through the centre of mass leaves no gravitational restoring torque. As h approaches the end of the stick, the term h dominates and the period rises towards $2\pi \sqrt{h/g}$. A minimum therefore lies between.

9. The shock absorber is designed close to critical damping, where the damping coefficient γ is approximately equal to ω_0 ; the suspension returns to equilibrium in the shortest possible time without overshooting, and a passenger feels no oscillation after the bump. The swing door is under-damped, with $\gamma < \omega_0$; the door oscillates about its closed position with a slowly decaying amplitude before friction in the hinge brings it to rest. An over-damped system would correspond to a swing door that drags shut so slowly that it never visibly oscillates and takes a long time to settle, like a heavy fire door fitted with an over-stiff closer. Each engineering choice serves a different goal: passenger comfort calls for critical damping; gentle automatic closure tolerates light under-damping; safety doors rely on heavy over-damping.

EXERCISE 6B

10. A new shock absorber is critically damped: the damping is just sufficient to bring the suspension back to its rest height in the shortest possible time, without any overshoot. As the seal wears and the absorber loses oil, the damping weakens, and the suspension enters the under-damped regime, in which it oscillates about its rest height with a slowly decaying amplitude before coming to rest. The two or three observed bobs after each bump are these under-damped oscillations.

11. After being struck, the bell oscillates at its natural frequency in the under-damped regime. Energy is steadily lost to internal friction in the metal and to the surrounding air, so the amplitude (and hence the loudness of the note) decays steadily until the sound is no longer audible. A heavier bell of the same shape and material stores more vibrational energy at any given amplitude (more mass is moving), while the rate at which energy leaves the bell is set mainly by the surface in contact with the air and is roughly the same. More energy stored at a comparable rate of loss means a longer audible decay, so the heavier bell rings for longer.

12. The floor-machine system has a particular natural frequency at which it would oscillate if struck once and left alone; this frequency is set by the mass of the machine and the stiffness of the floor's response. The slightly unbalanced rotating drum applies a small periodic force at a forcing frequency equal to the drum's rotation rate. As the spin-up sweeps the drum's frequency through the natural frequency of the floor-machine system, the small driving force arrives in step with the system's preferred rhythm, and the response builds up into resonance: the floor shakes violently. At every other speed the driving frequency is well off the natural frequency, the drive and the response fall out of step, and the shaking is small.

13. Let the expansion joints be a regular distance s apart along the road. A vehicle travelling at speed v crosses joints at the rate $f = v/s$ jolts per second, and this is the forcing frequency applied to the suspension. At $v \approx 80\text{km/h}$ this rate coincides with the natural frequency of the suspension; the small periodic jolts arrive in step with the suspension's preferred rhythm and the response builds up into resonance, giving the worst shaking. At 70 and 90km/h the forcing rate falls off the natural frequency, jolts and swing fall out of step, and the response is small.

14. Soldiers marching in step apply a periodic vertical force to the bridge at the frequency of their cadence. If this forcing frequency happens to coincide with the natural frequency of the bridge, the bridge is driven into resonance: each footfall arrives in step with the bridge's own swing, and the response builds up into a large amplitude that may damage the structure. Breaking step destroys the periodicity of the force; the bridge then sees only a series of independent (and on average, cancelling) impulses, never a sustained drive at its natural frequency, so the resonance cannot build up.

15. From $T = 2\pi\sqrt{\frac{l}{g}}$, the period grows as g decreases. At 3000m altitude g is smaller than at sea level (the body is farther from the Earth's centre), so T is larger and each "tick" takes longer; the clock therefore loses time. To restore correct timekeeping, the watchmaker must shorten the pendulum (raise the bob along the rod), since T depends on $\sqrt{l}$, and a smaller l exactly compensates a smaller g .

16. From $T = 2\pi\sqrt{\frac{l}{g}}$, a clock that runs fast has too short a period and therefore too short an effective length. The bob must be moved **down** along the rod, increasing l (and hence T) back to its correct value. g is fixed at the laboratory's location and is not under the technician's control.

17. The pan, together with the tomatoes on it, behaves as a mass m suspended from a spring of stiffness k (the scale's internal spring). When jostled, the pan is displaced from its equilibrium reading; the spring pulls it back, the pan overshoots, the spring pushes it forward, and the system executes SHM about the equilibrium reading with period,

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

The pointer's brief oscillation is exactly this SHM, and its period depends only on the combined pan-plus-tomato mass and on the scale's spring constant.

18. From $T = 2\pi\sqrt{\frac{m}{k}}$, a softer mattress (smaller k) has a longer natural period: each oscillation takes a long time, so the guest perceives the slow rise and fall plainly. A firmer mattress (larger k) has a much shorter period: the oscillation is so rapid that the guest barely notices it before the motion fades. The firmer mattress therefore has the higher spring constant k .

19. The mass per unit length μ of the line is the same in both cases (it depends on the line's material and thickness). When the line is tight, T is large and $v = \sqrt{T/\mu}$ is large, so the pulse crosses the line in a fraction of a second and is barely visible. When the line sags, T is small and v is correspondingly small, so the pulse travels slowly enough to be seen with the naked eye.

20. For small displacements from rest, a slender elastic branch supplies a restoring force on the mango that is approximately proportional to the displacement, $F \approx -kx$, where k depends on the stiffness of the branch. By Newton's second law $a = F/m = -(k/m)x = -\omega^2 x$ with $\omega^2 = k/m$, which is the SHM condition. The motion is therefore approximately SHM at small amplitude. The approximation fails when the displacement becomes large enough that the branch no longer behaves elastically: the wood may bend irreversibly, or the restoring force may cease to be linear in x , and the motion departs from SHM.

EXERCISE 6C

21. The static extension is $e = 0.65\text{m} - 0.50\text{m} = 0.15\text{m}$, giving $k = mg/e = (2.5\text{kg})(9.8\text{m/s}^2)/(0.15\text{m}) = 163\text{N/m}$ and $\omega = \sqrt{k/m} = \sqrt{163/2.5}\text{rad/s} = 8.08\text{rad/s}$. Take downward positive. The block is released from rest at $x(0) = +0.05\text{m}$, so the amplitude is $A = 0.05\text{m}$, and the general SHM equation is:

$$x(t) = A\sin(\omega t + \phi)$$

Imposing the initial conditions $x(0) = +A$ and $v(0) = 0$:

$$A\sin(\phi) = A \Rightarrow \sin(\phi) = 1 \Rightarrow \phi = \pi/2$$

so $x(t) = 0.05\sin(\omega t + \pi/2)\text{m}$. At $t = 1.0\text{s}$, $\omega t = 8.08\text{rad}$, and

$$x(1.0) = 0.05\sin(8.08 + \pi/2)\text{m} = 0.05\sin(9.65\text{rad})\text{m} = 0.05 \times (-0.227)\text{m} = -0.0114\text{m}$$

That is, about **1.14cm above the equilibrium position**.

Alternative solution. Recognising that the block starts at a turning point, where sine reaches its peak with phase $\pi/2$, one may write the motion directly in cosine form,

$$x(t) = A\cos(\omega t) = 0.05\cos(\omega t)\text{m}$$

which automatically satisfies $x(0) = +A$ and $v(0) = 0$. At $t = 1.0\text{s}$:

$$x(1.0) = 0.05\cos(8.08\text{rad})\text{m} = 0.05 \times (-0.227)\text{m} = -0.0114\text{m}$$

The same answer; cos and sin describe the same physical motion, with the phase constant absorbing the difference.

$$K_{\max} = \frac{1}{2}kA^2 = \frac{1}{2}(163\text{N/m})(0.05\text{m})^2 = \mathbf{0.204J}$$

22. 9.79m/s^2

23. With $\omega = 2\pi f = 4\pi\text{rad/s}$, the energy condition $K = U$ requires $U = E/2$, i.e. $\frac{1}{2}kx^2 = \frac{1}{2} \times \frac{1}{2}kA^2$, so

$$x = \pm A/\sqrt{2} = \pm 0.04\text{m}/\sqrt{2} = \pm 0.0283\text{m}$$

Taking $x = A\sin(\omega t)$ with the particle moving in the positive direction at $x = 0$, the time to reach $x = A/2$ satisfies $\sin(\omega t) = 1/2$, i.e. $\omega t = \pi/6$:

$$t = (\pi/6)/\omega = (\pi/6)/(4\pi\text{rad/s}) = \frac{1}{24}\text{s} = 0.0417\text{s}$$

24. 0.568s

25. It loses about 7.8s per day. The clock runs slow.

26. Treating the wire as a spring of stiffness $k = AY/L$:

$$k = \frac{(0.50 \times 10^{-6}\text{m}^2)(2.0 \times 10^{11}\text{N/m}^2)}{2.0\text{m}} = 5.0 \times 10^4\text{N/m}$$

Period of vertical oscillations:

$$T = 2\pi\sqrt{\frac{m}{k}} = 2\pi\sqrt{\frac{4.0\text{kg}}{5.0 \times 10^4\text{N/m}}} = 0.0562\text{s}$$

27. 4.5N

28. From $x = 0.05\sin(4\pi t + \pi/3)$, the angular frequency is $\omega = 4\pi\text{rad/s}$ and the amplitude is $A = 0.05\text{m}$. The first positive time at which $x = +0.025\text{m}$ satisfies $\sin(4\pi t + \pi/3) = 0.5$. The first positive root is $4\pi t + \pi/3 = 5\pi/6$, so $4\pi t = \pi/2$ and

$$t = 1/8\text{s} = 0.125\text{s}$$

The velocity at this instant is

$$v = A\omega\cos(5\pi/6) = (0.05\text{m})(4\pi\text{rad/s})(-\frac{\sqrt{3}}{2}) = -0.544\text{m/s}$$

The total mechanical energy is

$$E = \frac{1}{2}m\omega^2 A^2 = \frac{1}{2}(0.20\text{kg})(4\pi\text{rad/s})^2(0.05\text{m})^2 = 0.0395\text{J}$$

29. 22.1s

30. $\text{KE} = 0.0800\text{J}$, $x = \pm 0.0577\text{m}$

31. The amplitude is half the peak-to-peak range, $A = 0.60\text{m}$, and the angular frequency is $\omega = 2\pi/T = \pi/2\text{rad/s} = 1.57\text{rad/s}$. The maximum vertical acceleration is

$$a_{\text{max}} = \omega^2 A = (1.57\text{rad/s})^2(0.60\text{m}) = 1.48\text{m/s}^2$$

Taking $x = -A\cos(\omega t)$ so the boat is at its lowest point at $t = 0$, the position 0.30m above the mean is reached when $\cos(\omega t) = -1/2$, i.e. $\omega t = 2\pi/3$:

$$t = (2\pi/3)/(\pi/2)\text{s} = 4/3\text{s} = 1.33\text{s}$$

32. Mass per unit length: $\mu = M/L = (1.2 \times 10^{-3}\text{kg})/(0.60\text{m}) = 2.0 \times 10^{-3}\text{kg/m}$.

Wave speed: $v = \sqrt{T/\mu} = \sqrt{80/0.002}\text{m/s} = 200\text{m/s}$. The round trip is twice the wire length:

$$t_{\text{rt}} = 2L/v = (2 \times 0.60\text{m})/(200\text{m/s}) = 6.0 \times 10^{-3}\text{s} = 6.0\text{ms}$$

With μ doubled at fixed tension, v falls by $\sqrt{2}$ to $v_{\text{new}} = v/\sqrt{2} = 141\text{m/s}$, and the new round-trip time is

$$t_{\text{rt,new}} = 2L/v_{\text{new}} = (1.20\text{m})/(141\text{m/s}) = 8.49 \times 10^{-3}\text{s} = 8.5\text{ms}$$

EXERCISE 7A

1. A progressive wave is a propagating disturbance whose speed is determined entirely by the medium's elastic and inertial properties: the elastic stiffness of the medium provides the restoring force that passes the disturbance from one particle to the next, and the inertia of those particles fixes how quickly the restoring force can accelerate them. The amplitude of the wave is set independently by the source. Increasing the amplitude makes each particle oscillate over a wider range, but it does not change the medium's stiffness or its inertia, so the rate at which the disturbance is handed on from particle to particle, and therefore the wave speed, is unaffected.

2. The two formulas are equivalent because the angular frequency and the linear frequency are connected by the definition $\omega = 2\pi f$. Substituting into Kipanga's expression gives:

$$v_{\text{p,max}} = A\omega = A(2\pi f) = 2\pi fA,$$

which is identical to Kipute's expression. Both quantities describe the maximum speed at which a single particle of the medium oscillates as the wave passes, and both are written purely in terms of the source amplitude and the rate of oscillation.

3. The source on one side of the boundary is producing oscillations at a fixed rate, equal to its frequency. Continuity of the medium at the boundary requires the same rate of oscillations to emerge on the far side: oscillations cannot appear from nowhere, and they cannot disappear without trace. The number of crests crossing the boundary per second is the same on both sides. The frequency is therefore identical on both sides, even though the wave speed and the wavelength generally differ between the two media.

4. The acoustic power P of the source is constant, and the wavefronts spread outward over expanding spheres of surface area $4\pi r^2$. The intensity is therefore:

$$I = \frac{P}{4\pi r^2},$$

so $I \propto 1/r^2$. The intensity of a progressive wave is also proportional to the square of the displacement amplitude:

$$I \propto A^2$$

Combining the two:

$$A^2 \propto \frac{1}{r^2}, \quad \text{hence} \quad A \propto \frac{1}{r}$$

The displacement amplitude therefore falls as $1/r$.

5. Energy is not destroyed at a silent point; it is redistributed away from it to the regions of constructive interference. Across any small region of the field around the two sources, silent zones (destructive interference) are accompanied by louder zones (constructive interference) a fraction of a wavelength away. Integrating the power flowing through any closed surface enclosing the two sources gives a total equal to the power emitted by both sources together, as required by conservation of energy. The “missing” energy at a silent point is exactly accounted for by the “extra” energy at a nearby loud point.

6. The formula $A_R = 2A \cos(\Delta\phi/2)$ comes from adding $y_1 = A \sin(\omega t)$ and $y_2 = A \sin(\omega t + \Delta\phi)$ by the sum-to-product identity:

$$\sin\alpha + \sin\beta = 2 \sin \frac{\alpha + \beta}{2} \cos \frac{\alpha - \beta}{2}$$

The pre-factor of $2A$ arises only because the two waves share the same amplitude A ; the identity collapses to a single sine of amplitude $2A \cos(\Delta\phi/2)$. When the amplitudes are unequal, the same identity does not produce a single-sine form. The correct general expression is:

$$A_R = \sqrt{A_1^2 + A_2^2 + 2A_1A_2 \cos(\Delta\phi)},$$

which reduces to $2A \cos(\Delta\phi/2)$ exactly when $A_1 = A_2 = A$.

7. A travelling wave has its space and time variables in a single combined argument $\omega t \pm kx$, so that the entire pattern slides rigidly through the medium.

$y_1 = A \sin(\omega t - kx)$: the argument is $\omega t - kx$, so the wave is travelling in the $+x$ direction.

$y_3 = A \sin(\omega t + kx)$: the argument is $\omega t + kx$, so the wave is travelling in the $-x$ direction.

A stationary wave has the space and time variables factorised into separate trigonometric functions, so that the spatial pattern is fixed in space while the whole pattern oscillates uniformly in time. $y_2 = 2A \sin(kx) \cos(\omega t)$ has $\sin(kx)$ multiplied by $\cos(\omega t)$, so it is a stationary wave.

8. Both students are partly right. A stationary wave does not transport any net energy along the string, as Kipanga says. But the medium itself is in motion: every particle of the string between two consecutive nodes oscillates vertically with simple harmonic motion, with an amplitude that varies sinusoidally with position (zero at the nodes, maximum at the antinodes). Each oscillating particle carries kinetic energy, with a maximum at the antinodes and zero at the nodes, exactly as Kipute observes. The total kinetic energy of the string oscillates twice each period, exchanged with the elastic potential energy of the strained string. The two energies sum to a constant total, stored in the standing wave but not transported.

9. The fundamental frequency of a stretched string fixed at both ends is:

$$f_1 = \frac{1}{2L} \sqrt{\frac{T}{\mu}}$$

At constant length and tension, $f_1 \propto 1/\sqrt{\mu}$, so f_1 falls as μ rises. Physically, a string with greater linear mass density has greater inertia for the same tension-driven restoring force; the disturbance is passed more slowly from one element of the string to the next, the transverse wave speed $v = \sqrt{T/\mu}$ is reduced, and the time taken for one half-cycle round-trip ($L = \lambda/2$ in the fundamental) is therefore lengthened, giving a lower fundamental frequency.

10. The claim is **incorrect**. For a stretched string vibrating in its third harmonic, the angular frequency is $\omega_3 = 3\omega_1$, so at any antinode the maximum particle speed is:

$$v_{p,\max} = A_{\text{ant}} \omega_3 = 3A_{\text{ant}} \omega_1,$$

three times the corresponding maximum at an antinode of the fundamental of the same antinode amplitude. But this applies only to particles at the antinodes. At the nodes of the third harmonic, located at $x = 0, L/3, 2L/3, L$, the particles do not move at all. Particles between a node and an antinode have amplitudes (and therefore peak speeds) smaller than the antinode value. The blanket statement that “every particle” moves three times as fast is therefore wrong.

11. The speed of a longitudinal wave through a uniform metal rod is:

$$v = \sqrt{\frac{Y}{\rho}},$$

where Y is Young's modulus and ρ is the density of the rod's material. The cross-sectional area enters the rod's mass per unit length ($\mu = \rho A$) and also enters the elastic restoring force per unit strain (proportional to YA). The two factors of A cancel exactly when the wave equation is written for a thin rod, leaving the speed dependent only on the ratio Y/ρ . A thicker rod has more inertia per metre but is also proportionally stiffer, and the two effects exactly compensate.

EXERCISE 7B

12. A wave is a propagating disturbance, not a flow of matter. The water beneath the leaf executes a small simple-harmonic up-and-down motion as each ripple passes; the molecules return to their original positions once the ripple has gone, having merely transmitted the disturbance to their neighbours. The leaf, riding on the water, shares only this local up-and-down motion. What moves outward is the energy of the disturbance and its geometric pattern; what does not move is the bulk of the water, and therefore the leaf does not travel with the ripple.

13. When the ultrasound pulse crosses the fat-muscle boundary, the source frequency is preserved at 5MHz: continuity of oscillation count requires the same number of cycles per second on both sides of the boundary, otherwise oscillations would have to appear from nowhere or disappear without trace. The wave speed, however, changes because muscle has different elastic and inertial properties from fat. Since $v = f\lambda$ with f fixed, the wavelength scales in proportion to the wave speed: the wavelength in muscle is longer than in fat by the ratio of the wave speeds. The amplitude is also reduced on transmission, because part of the wave energy is reflected at the boundary owing to the acoustic-impedance mismatch.

14. Sound travels as a longitudinal wave whose speed is set by $v = \sqrt{E/\rho}$, where E is the relevant elastic modulus of the medium. Solid soil has a much greater elastic modulus than air, while its density is only about a thousand times greater. The ratio E/ρ is therefore much larger in soil than in air, and the sound speed in soil (typically several thousand metres per second) is many times greater than in air (340m/s). A pulse generated by the cattle's hooves reaches the listener through the soil long before the same disturbance reaches him through the air, giving the earlier warning.

15. The two coherent loudspeakers, driven in phase at 680Hz, radiate sound of wavelength

$$\lambda = \frac{v}{f} = \frac{340\text{m/s}}{680\text{Hz}} = 0.50\text{m}.$$

As the patron walks along the line parallel to the stage, the path difference Δx from the two loudspeakers changes continuously. At positions where $\Delta x = n\lambda$ (with integer n), the two waves arrive in phase and superpose constructively, producing a loud note. At positions where $\Delta x = (n + 1/2)\lambda$, the two waves arrive in antiphase and cancel, producing a faint or silent zone. The patron walks through alternating regions of constructive and destructive interference, hearing the loudness rise and fall at regular intervals.

16. The initial beat frequency is $f_b = 6\text{Hz}$, so the string's frequency lies at $440\text{Hz} \pm 6\text{Hz}$, i.e., either 434Hz or 446Hz. Tightening a string raises its tension and therefore its frequency, by the second law of vibration $f \propto \sqrt{T}$. After tightening, the beat rate has fallen to 2Hz, which means the string's frequency is now closer to that of the fork. If the string had originally been at 446Hz, tightening would push it further above the fork and increase the beat rate (inconsistent). If it had originally been at 434Hz, tightening raises it to 438Hz and the beat rate falls to 2Hz (consistent). The original frequency of the fifth string was therefore 434Hz.

17. Pressing the string against a fret reduces its vibrating length to the section of string between the fret and the bridge. The fundamental frequency of a stretched string at constant tension and mass per unit length varies as $f_1 \propto 1/L$ (first law of vibration). To raise the pitch by a full octave, the frequency must double, so the vibrating length must be halved:

$$L_{\text{new}} = \frac{L}{2}.$$

The twelfth fret therefore lies exactly halfway along the open-string vibrating length, i.e., midway between the nut and the bridge.

18. A cable suspended between two fixed supports is mechanically equivalent to a stretched string and supports a discrete set of natural frequencies of transverse vibration, $f_n = (n/2L)\sqrt{T/\mu}$, with integer n . The wind, blowing past the cable, sheds vortices alternately on either side at a frequency proportional to the wind speed (the Strouhal frequency). At most wind speeds, this forcing frequency is detuned from every natural frequency of the cable; the cable's response is a small forced oscillation, barely visible. At one particular wind speed, the forcing frequency matches one of the cable's natural frequencies. The resonance condition is met: each shed vortex delivers energy in step with the cable's own oscillation, the amplitude builds up cycle by cycle, and the cable swings vigorously.

19. The speed of a transverse wave on a stretched string is

$$v = \sqrt{\frac{T}{\mu}}$$

where T is the tension in the string and μ is the linear mass density. Tightening the washing-line gives it a large tension and therefore a large wave speed; the pulse crosses the line in a fraction of a second. Reducing the tension to almost zero (a slack droop) leaves μ essentially unchanged but cuts v by a large factor. The same disturbance now crosses the line slowly enough that the eye can resolve the wave's progress.

20. Two tuning forks of nominally identical frequency are never exactly identical in practice. Small manufacturing differences in mass distribution, prong length, or mounting shift each fork's natural frequency slightly. When sounded together, the small frequency difference Δf produces beats: the two waves drift slowly into and out of phase at the listener's ear, giving a resultant whose amplitude rises and falls at the beat frequency

$$f_b = |f_1 - f_2|.$$

A beat rate of about 2Hz, observed by the class, implies that the two forks differ in frequency by about 2Hz despite their nominally identical 440Hz rating.

21. A long stay-cable behaves as a stretched string fixed at both ends, supporting a discrete set of natural transverse frequencies $f_n = (n/2L)\sqrt{T/\mu}$. A steady wind blowing past the cable sheds vortices alternately on either side at a frequency proportional to the wind speed. At most wind speeds this vortex-shedding frequency is detuned from every natural frequency of the cable, and the cable's response amplitude is too small to radiate audible sound. At one particular wind speed, however, the shedding frequency matches one of the cable's natural frequencies. The resonance condition is met, the amplitude of the cable's vibration builds up, and the resulting standing wave radiates an audible note. Above and below this speed the resonance is lost and the cable returns to silence.

EXERCISE 7C

22. Wavelength: 1.0m, Period: 0.040s, Wave speed: 25m/s, Maximum particle speed: 9.42m/s, Maximum particle acceleration: 1480m/s².

23. $I_1 = 1.27 \times 10^{-2} \text{W/m}^2$, $I_2 = 3.18 \times 10^{-3} \text{W/m}^2$, $\frac{A_1}{A_2} = 2$.

24. The resultant has amplitude 4.0cm and lags y_1 by $\pi/3$ rad.

25. Wavelength: 0.50m, Path difference: 1.50m.

Since $\Delta\phi = 6\pi = 3(2\pi)$, the path difference is exactly three full wavelengths ($\Delta x = 3\lambda$). The two waves arrive at P in phase and superpose constructively, so P lies in a region of constructive interference.

26. Increasing the tension raises the frequency of a stretched wire (second law of vibration, $f \propto \sqrt{T}$). After increasing the tension the beats vanish, so the wire is now at 512Hz. The original frequency must therefore have been below 512Hz, namely: $f_{\text{old}} = 504\text{Hz}$. $T_{\text{new}} = 92.9\text{N}$.

27. Speed: 192m/s, Tension: 184N.

28. Wave speed: 200m/s, Fundamental frequency: 83.33Hz, Highest audible harmonic: 240th

29. The wave speed on a stretched wire is $v = \sqrt{T/\mu}$, where $\mu = \rho A$ is the linear mass density of a uniform wire of cross-section A. The maximum tension that the wire can sustain is $T_{\text{max}} = \sigma_{\text{max}} A$. Substituting:

$$v_{\text{max}} = \sqrt{\frac{T_{\text{max}}}{\mu}} = \sqrt{\frac{\sigma_{\text{max}} A}{\rho A}} = \sqrt{\frac{\sigma_{\text{max}}}{\rho}}$$

The cross-sectional area cancels exactly, leaving v_{max} a function only of the material constants σ_{max} and ρ .

For copper:

$$v_{\text{max}} = \sqrt{\frac{2.2 \times 10^8 \text{N/m}^2}{8900 \text{kg/m}^3}} \approx 157 \text{m/s}$$

Maximum tension on a wire of cross-section $A = 1.5 \times 10^{-6} \text{m}^2$:

$$T_{\text{max}} = \sigma_{\text{max}} A = (2.2 \times 10^8 \text{N/m}^2)(1.5 \times 10^{-6} \text{m}^2) = 330 \text{N}.$$

30. Speed: 5092m/s, Time: $9.82 \times 10^{-3} \text{s}$.

31. Amplitude: 0.015m, Wavelength: 0.50m, Wave speed: 20m/s,

Antinodes occur where $\sin(kx) = \pm 1$, i.e., $kx = (2n + 1)\pi/2$ for integer $n \geq 0$. The nearest to $x = 0$ corresponds to $n = 0$:

$$x = \frac{\pi/2}{k} = \frac{\pi/2}{4\pi\text{rad/m}} = 0.125\text{m}.$$

Displacement of that antinode at $t = 0.020\text{s}$: $9.27 \times 10^{-3}\text{m}$.

EXERCISE 8A

1. A stable interference pattern requires the two illuminating sources to maintain a fixed phase relationship over a time long compared with the detector's integration time. Two physically separate tungsten lamps emit independent sequences of wavetrains, each lasting only a few nanoseconds before being replaced by the next emission. The phase relationship between the two lamps therefore shifts randomly on this nanosecond timescale, while the eye and the screen integrate over much longer times. The pattern of fringes is averaged out to a uniform illumination, because no coherence exists between the two independent sources. To make the experiment work, both beams must be derived from a single primary source (for example, by passing light from one narrow slit through two further slits).

2. For an air wedge of small angle θ between two glass plates, a dark fringe of order n in reflected light occurs where the air-gap thickness t satisfies $2t = n\lambda$ (allowing for the π phase change on reflection at the lower glass surface). The position x_n along the wedge at which this thickness is reached is:

$$x_n = \frac{t_n}{\theta} = \frac{n\lambda}{2\theta}$$

The spacing between consecutive dark fringes is therefore:

$$\Delta x = x_{n+1} - x_n = \frac{\lambda}{2\theta},$$

which depends only on the wavelength and the wedge angle and not on the position along the wedge. Consecutive dark fringes are therefore equally spaced everywhere along the wedge.

3. The fine periodic structure of the data tracks on the reflective surface of a compact disc makes the disc act as a reflection diffraction grating. For light of wavelength λ incident at angle i and diffracted at angle θ relative to the disc's normal, the principal maxima satisfy

$$d(\sin i + \sin \theta) = n\lambda,$$

with d the track spacing. For a given disc and a given source, different wavelengths satisfy this condition at different angles θ . As the disc is tilted, i changes; the diffracted-maximum angle θ for each wavelength changes in step, so the colour seen at any given viewing direction shifts continuously with the disc's orientation.

4. The student's claim is incorrect. The bright fringes occur where the path difference between the two slits equals an integer number of wavelengths, $\delta = n\lambda$, giving positions $y_{n,\text{br}} = n\lambda D/d$. The dark fringes occur where the path difference is an odd half-integer of wavelengths, $\delta = (n + 1/2)\lambda$, giving positions $y_{n,\text{dk}} = (n + 1/2)\lambda D/d$. Each dark fringe therefore lies exactly midway between the bright fringe of order n and the bright fringe of order $n + 1$, at a precise and predictable position fixed by the path-difference condition.

5. Kipute is correct. In reflected light, the two interfering beams reflect from the lower surface of the lens (no phase change, glass-to-air) and from the upper surface of the plate (a π phase change, air-to-glass). At the centre, where the geometric path difference vanishes, the two beams differ in phase by π and interfere destructively, giving a dark centre.

In transmitted light, the two interfering beams are the directly transmitted beam (with no internal reflections) and the beam that has undergone one reflection at the upper plate surface followed by one reflection at the lower lens surface. These two internal reflections occur at the same kind of interface (air into glass on each side) and contribute equal phase shifts, so the net relative phase between the directly transmitted and the doubly reflected beams is zero. At the centre, the path difference also vanishes, so the two beams arrive in phase and interfere constructively. The centre is therefore bright. The reflected and transmitted patterns are complementary, consistent with conservation of energy.

6. The student's distinction is not physically meaningful. Diffraction and interference are both manifestations of the same underlying principle: the superposition of coherent secondary wavelets according to Huygens' principle. When a wavefront passes through any aperture, each point on the wavefront acts as a secondary source, and the field at any observation point is the coherent sum of the contributions from all those secondary sources. When the aperture is small (a single slit), the secondary sources form a continuum and the resulting pattern is conventionally called a diffraction pattern. When the aperture is two separate slits, the secondary sources cluster around two locations and the resulting fine fringe structure is conventionally called an interference pattern. Both are interference patterns; the labels are historical.

7. The first student is correct. The dark fringes of a single-slit pattern occur where $a \sin \theta = n\lambda$ for non-zero integers n . The angular separation between consecutive dark fringes is λ/a , the same for every pair. The central maximum, however, is bounded on both sides by the first dark fringe at $\sin \theta = \pm \lambda/a$, so its angular full-width is $2\lambda/a$. Every secondary maximum is bounded by two consecutive dark fringes a distance λ/a apart, so it is half the width of the central maximum. The factor of two arises because the central maximum is bounded on both sides by the same diffraction order (the first), while every other maximum is bounded by two successive orders.

8. An unpolarised beam consists of light whose electric vector has no preferred direction in the plane perpendicular to the direction of propagation. The instantaneous electric vector can be decomposed into two perpendicular components: one parallel to the Polaroid's transmission axis and one perpendicular to it. Because the unpolarised light's electric vector points equally probably in any direction in the transverse plane, the two components carry, on average, equal intensity. The Polaroid transmits the component along its transmission axis and absorbs the component perpendicular to it. Exactly one half of the incident intensity is therefore transmitted. The result is independent of the orientation of the Polaroid's transmission axis in the transverse plane, because any orientation defines two perpendicular components that share the intensity equally for unpolarised light.

9. The error in Kipanga's claim is the implicit assumption that the refractive index of a transparent medium is independent of wavelength. In reality every real transparent medium is dispersive: its refractive index varies with the wavelength of the light, with shorter wavelengths typically refracting more strongly than longer ones. Brewster's angle is $\tan i_p = n$. Since n depends on λ , so does i_p . Different colours of light therefore have slightly different Brewster angles, and at any single fixed angle of incidence the polarisation is exact only for one wavelength at a time. The variation across the visible spectrum is small (typically a fraction of a degree), but it is real and is the reason that the polarisation of reflected white light is never perfect at any single angle.

10. The bright fringes in a Young's double-slit pattern occur at positions on the screen where the path difference from the two slits equals an integer number of wavelengths,

$$\delta = n\lambda \quad \Rightarrow \quad y_{n,br} = \frac{n\lambda D}{d}$$

The dark fringes occur at positions where the path difference equals an odd half-integer of wavelengths,

$$\delta = (n + 1/2)\lambda \quad \Rightarrow \quad y_{n,dk} = \frac{(n + 1/2)\lambda D}{d}$$

The position of the n th dark fringe relative to the n th bright fringe is:

$$y_{n,dk} - y_{n,br} = \frac{\lambda D}{2d} = \frac{\beta}{2}$$

where $\beta = \lambda D/d$ is the bright-fringe spacing. The dark fringe therefore lies exactly halfway between the bright fringe of order n and the bright fringe of order $n + 1$.

11. The student's error is to assume that a change in wavelength shifts the central fringe. The position of the central fringe is determined by zero path difference between the two beams, not by the wavelength. Immersing the entire apparatus in water does not break the symmetry of the two paths from the slits to the centre of the screen, because both paths travel through water of the same refractive index. The path difference at the centre therefore remains zero, and the central fringe stays where it was.

What actually happens is that the wavelength of the light inside water is reduced by a factor equal to the refractive index of water, so the fringe spacing $\beta = \lambda D/d$ shrinks by the same factor. All non-central fringes therefore move closer to the centre while the central fringe is anchored. The pattern is compressed about its centre, but it is not shifted.

12. The dark ring of order n in a Newton's-rings pattern occurs where the air-gap thickness t satisfies $2t = n\lambda$. The geometric relation between the ring radius r and the gap thickness t for a spherical lens of radius of curvature R resting on a flat plate, valid for small t , is $r^2 = 2tR$ (from the chord-of-circle approximation). Combining the two relations:

$$r^2 = 2\left(\frac{n\lambda}{2}\right)R = n\lambda R \quad \Rightarrow \quad r = \sqrt{n\lambda R}$$

The ring radius therefore grows as $\sqrt{n}$, not in direct proportion to n . The geometric origin is the curvature of the lens: as one moves outward from the centre, the air-gap thickness grows quadratically with radius rather than linearly, so a unit increase in ring order (corresponding to a fixed linear increase in thickness) corresponds to a smaller and smaller incremental radial step as the rings move outward. The rings therefore crowd together as n grows.

13. Both the soap film and the glass pane are thin transparent layers bounded by two reflecting surfaces, and in principle both should produce interference between the front-reflected and back-reflected beams. The difference lies in the comparison between the coherence length of the white-light illumination and the optical path difference $2nt$ across each layer.

White light from an ordinary source has a very short coherence length, typically a few micrometres. The two reflected beams interfere only if the path difference $2nt$ is smaller than this coherence length. A soap film is typically a few hundred nanometres to

a few micrometres thick, so $2nt$ fits comfortably within the coherence length and the two reflections interfere to give visible colour bands.

A glass window-pane is several millimetres thick, so $2nt$ is many thousands of times the coherence length of white light. The two reflections from the front and back surfaces of the pane are then essentially independent (incoherent), and no interference fringes are produced. The pane simply transmits the white light unchanged, with two faint uncoloured reflections.

EXERCISE 8B

14. The Newton's-rings pattern in reflected light arises from interference between two beams: one reflecting from the lower surface of the lens (a glass-to-air interface, with no phase change) and one reflecting from the upper surface of the flat plate (an air-to-glass interface, with a π phase change). At the centre of the pattern, where the air-gap thickness is essentially zero, the two beams have zero geometric path difference but they differ in phase by π , so they cancel destructively and the centre is dark.

Silvering the lower surface of the lens replaces the glass-to-air interface with a perfect mirror. All incident light is now reflected back at the lens itself; no light enters the air gap to reach the plate below. The interference between the two air-gap reflections, which is the entire mechanism by which the Newton's rings are produced, is therefore destroyed. The field of view no longer shows any ring pattern, only a uniform bright reflection from the silvered surface. The previously dark centre is replaced by a uniform bright field.

15. The light emerging from a smartphone screen is plane-polarised, because the LCD's outer polariser fixes the direction of the electric vector along a single axis (typically at 45° to the screen edges). When this polarisation direction is parallel to the transmission axis of the wearer's polarised sunglasses, the light is transmitted and the screen looks bright. When the phone is rotated through 90° about the viewing axis, the LCD's polarisation direction also rotates through 90° and becomes perpendicular to the sunglasses' axis. By Malus's law,

$$I = I_0 \cos^2 \theta = I_0 \cos^2(90^\circ) = 0,$$

so no light reaches the eye through the sunglasses, and the screen appears black. The observation is a direct consequence of the LCD emitting already-polarised light and the sunglasses acting as an analyser whose transmission depends on the relative orientation of the two axes.

16. A soap bubble is a thin film of soapy water held together by surface tension. Gravity drains the liquid downward inside the film, so that the film becomes thicker at the bottom and thinner at the top. The thickness t varies continuously and smoothly with height on the bubble's surface.

Light reflected from the front and back surfaces of the film interferes. The condition for a bright reflected fringe at wavelength λ is

$$2nt = (m + 1/2)\lambda,$$

with the extra half-wavelength accounting for the π phase change on reflection at the front surface. At each thickness t , this condition picks out a specific wavelength for bright reflection, and so a specific colour appears at that height. As the film continues to drain, t changes continuously with time at every height; the height at which any given colour satisfies the bright condition therefore moves downward, and the colour bands flow downward across the bubble's surface.

17. A thin layer of petrol on water acts as a thin film bounded by two reflecting surfaces, the air-petrol interface and the petrol-water interface. The two reflections interfere, with optical path difference $2nt \cos r$ between them, where n is the refractive index of petrol, t is the film thickness and r is the angle of refraction inside the film. The condition for a bright reflected fringe at wavelength λ is:

$$2nt \cos r = (m + 1/2)\lambda.$$

When the motorcyclist changes his viewing angle, the angle of refraction r changes; the effective path difference $2nt \cos r$ changes correspondingly, and the wavelength that satisfies the constructive condition at any given point on the puddle shifts to a different value. Different colours therefore appear at the same point on the petrol layer when viewed from different angles. The shifting colours are the angle-dependence of the same interference mechanism that produces them.

18. Light reflected from the surface of a body of water at large angles of incidence is partially polarised in the horizontal direction (the component of the electric vector parallel to the lake's surface is preferentially reflected). At the Brewster angle i_p for water, given by

$$\tan i_p = n_{\text{water}} = 1.33 \Rightarrow i_p \approx 53^\circ,$$

the reflected light is completely plane-polarised in the horizontal direction. Polaroid sunglasses with a vertical transmission axis block almost all of this horizontally polarised glare. The sunglasses are therefore most effective when the fisherman views the lake

along a line of sight close to Brewster's angle from the vertical, i.e., his line of sight makes about 53° with the normal to the lake (about 37° above the horizontal). At other angles of view, the reflected light is only partially polarised and the suppression is less complete.

19. On a hot day, the tarmac surface heats the air immediately above it to a temperature much higher than the air higher up. Hot air is less dense than cool air, and so has a slightly smaller refractive index. The refractive index of the air therefore decreases continuously as one approaches the road surface.

A wavefront of light from the sky, approaching the road at a glancing angle, passes through layers of progressively smaller refractive index. By Snell's law, the wavefront bends further from the normal at each layer interface, and eventually undergoes total internal reflection at a layer immediately above the tarmac, curving back upward toward the driver's eye. The driver sees light from the sky as if it had been reflected from a pool of water on the road. As he approaches, the geometry of the bending changes, and the "pool" appears to recede; he can never quite reach it.

20. Light incident on the coated lens reflects partially at the air-coating interface and partially at the coating-glass interface. Both reflections occur at boundaries where the index increases (air to coating, and coating to glass), so each reflection undergoes a π phase change. The two π shifts cancel, and the relative phase of the two reflected beams is set entirely by the optical path difference $2nt$ accumulated by the wave inside the coating.

Choosing the thickness so that $2nt = \lambda/2$, that is,

$$t = \frac{\lambda}{4n},$$

makes the two reflected beams meet 180° out of phase at the air-coating boundary, and they cancel destructively. The intensity reflected at the chosen wavelength of 550nm is therefore strongly reduced. The two reflections have nearly equal amplitude because the refractive index of magnesium fluoride lies between that of air and that of glass, allowing nearly complete cancellation.

21. Sunlight scattered by air molecules into a direction making a large angle with the original beam is strongly linearly polarised, a consequence of Rayleigh scattering. The scattered electric vector must be perpendicular both to the incoming direction of polarisation and to the direction of scattering, so light scattered through approximately 90° from the original beam is most strongly polarised.

At midday, light reaching the camera from directly overhead has been scattered through approximately 90° from the incoming sunlight, and is therefore strongly polarised. Light from a direction close to the sun has been scattered through only small angles, and is weakly polarised. When the photographer rotates the polarising filter, she varies the angle between the filter's transmission axis and the polarisation direction of the overhead scattered light; by Malus's law $I = I_0 \cos^2 \theta$, the transmitted intensity ranges from full at $\theta = 0$ down to zero at $\theta = 90^\circ$. The overhead sky therefore darkens visibly at one particular orientation and brightens again at the perpendicular orientation. The sky close to the sun, being only weakly polarised, hardly changes its brightness regardless of the filter's orientation.

22. A hologram records the interference pattern between a reference beam and the light scattered from an object, in a photosensitive emulsion. Reconstruction of the original object wavefronts requires illumination of the hologram with a beam of the same wavelength as the reference beam used during recording. The diffraction angles produced by the recorded fringe pattern follow the grating equation:

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

If the illuminating wavelength matches the recording wavelength, the diffracted wavefronts reconstruct the original object wavefronts exactly, and the image appears sharp. If the illuminating wavelength is different, the diffracted angles shift, and the reconstructed wavefronts no longer align with the original object geometry; the image becomes distorted and blurred.

The security hologram was recorded with red laser light. Under the red LED inspection lamp (wavelength close to the recording wavelength), the reconstruction is faithful and the image appears sharp and three-dimensional. Under the yellow sodium street-lamps (wavelength 589nm, very different from the red recording wavelength), the diffraction angles shift, the wavefronts misalign, and the image looks blurred and washed-out.

23. Most clear plastics are optically isotropic at rest: the refractive index inside the plastic is the same for any direction of polarisation, so plane-polarised light passing through the plastic emerges still plane-polarised, and is blocked by a crossed analyser. The view between the crossed Polaroids therefore looks uniformly dark (or shows only a faint colour pattern from residual stresses left in manufacture).

When the plastic is mechanically stressed, its optical isotropy is broken: the refractive index for light polarised along the local stress direction becomes slightly different from that for light polarised perpendicular to it. The material is then said to be stress-induced birefringent. Plane-polarised light entering the stressed region emerges elliptically polarised, with the ellipse parameters depending on the local stress. When this elliptical light passes through the crossed analyser, a fraction of the incident intensity is transmitted, the fraction depending on the local stress. Under white light, different wavelengths transmit different fractions, producing a

coloured pattern that maps the stress field. Squeezing the protractor increases the stress throughout the plastic, strengthens the birefringence, and brightens the pattern; releasing the squeeze removes the stress, restores isotropy, and the pattern fades.

24. A prism separates colours by refraction: the refractive index of glass depends on the wavelength of light, so different colours bend by different angles when passing through the prism. There is a single, smoothly-varying angle of deviation that depends on wavelength.

A CD has no glass and there is no refraction involved. The reflective surface of the CD is ruled with a fine periodic structure (the data tracks), and acts as a reflection diffraction grating. White light incident on the disc is reflected at angles satisfying the grating equation

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

with d the track spacing and n an integer order. Different wavelengths satisfy this equation at different angles, separating into a spectrum at each diffraction order. The CD therefore spreads colour by wavelength-dependent diffraction off a periodic surface structure, while the prism spreads colour by wavelength-dependent refraction through a transparent medium. The two mechanisms are physically distinct, and produce slightly different orderings of the colours in the spectrum.

25. The source of the ultrasound pulse emits oscillations at a fixed rate (here 5MHz). Continuity at the fat-muscle boundary requires the same rate of oscillations on both sides of the boundary: cycles cannot be created from nothing or be destroyed without trace as the wave crosses the interface. The frequency f is therefore preserved across the boundary, in both fat and muscle.

The wave speed v , by contrast, depends on the elastic and inertial properties of the medium, which are different in fat and in muscle. Muscle has different bulk elasticity and density from fat, and the longitudinal wave speed in muscle is larger. Since $v = f \lambda$ and f is fixed, the wavelength must scale in proportion to v :

$$\lambda_{\text{muscle}} = \frac{v_{\text{muscle}}}{f} > \lambda_{\text{fat}} = \frac{v_{\text{fat}}}{f}.$$

The wavelength inside muscle is therefore larger than the wavelength inside fat, while the frequency stays the same.

EXERCISE 8C

26. 30.7° , 400nm , $2.00 \times 10^8\text{m/s}$.

27. $2.0 \times 10^{-10}\text{s}$, 6.0cm .

28. 2.64mm , 0.126° .

29. Slit separation: 0.60mm , New fringe width: 0.775mm .

30. $3.25 \times 10^{-6}\text{m}$ or $3.25 \mu\text{m}$.

31. Wedge angle:

$$\theta = \frac{t}{L} = \frac{0.040 \times 10^{-3}\text{m}}{8.0 \times 10^{-2}\text{m}} = 5.0 \times 10^{-4}\text{rad}.$$

Spacing between consecutive dark fringes:

$$\Delta x = \frac{\lambda}{2\theta} = \frac{589 \times 10^{-9}\text{m}}{2(5.0 \times 10^{-4})} = 5.89 \times 10^{-4}\text{m} = 0.589\text{mm}.$$

Total number of dark fringes from the apex to the paper end (where the air-gap thickness is 0.040mm):

$$N = \frac{2t_{\text{max}}}{\lambda} = \frac{2(0.040 \times 10^{-3}\text{m})}{589 \times 10^{-9}\text{m}} \approx 135.8.$$

Rounded to the nearest integer, approximately 136 dark fringes are observed along the wedge (orders $n = 0$ at the apex up to $n = 135$ near the paper end).

32. Smallest thickness: 85.7nm , Next-smallest thickness: 257nm .

33. $r_{10} = 2.32\text{mm}$, $R' = 1.08\text{m}$.

34. 1.35.

35. 0.245mm , 0.413° .

36. $\theta_{\text{red}} = 25.9^\circ$, $\theta_{\text{blue}} = 16.3^\circ$.

EXERCISE 9A

1. The statement is incorrect because, in a sound wave, the displacement of a particle and the pressure variation are a quarter of a cycle out of phase. At its displacement extreme the particle is momentarily at rest, and the air around it is then neither compressed nor rarefied, so the pressure equals the normal atmospheric value. The pressure is greatest where the displacement is zero: as the particle passes through its equilibrium position, neighbouring particles crowd in on one side while those on the other side move away, producing the maximum compression or rarefaction. Greatest pressure therefore coincides with zero displacement, not with greatest displacement.
2. Humid air contains water vapour. A water molecule has a molar mass of 18g/mol, which is less than that of the nitrogen (28g/mol) and oxygen (32g/mol) molecules it replaces, so humid air is less dense than dry air at the same temperature and pressure. The speed of sound in a gas is $v = \sqrt{(\gamma RT/M)}$, so a smaller average molar mass M gives a larger speed. Hence sound travels slightly faster through humid air.
3. All audible frequencies travel through air at the same speed. The speed of sound $v = \sqrt{(\gamma RT/M)}$ depends only on the nature of the gas and its temperature, not on the frequency of the wave; air is a non-dispersive medium for sound. The high notes and the low notes therefore cover the distance to the listener in exactly the same time, so the chords stay together and the melody arrives undistorted, however far away the listener stands.
4. A pipe can sustain a standing wave only at its discrete set of natural frequencies, the fundamental and its harmonics, because only these fit the boundary conditions of the pipe. Gentle blowing excites the fundamental. Blowing much harder supplies enough energy to set up the next allowed mode, and the air column locks onto that higher harmonic. Since no standing wave can exist at the frequencies that lie between two harmonics, the pitch cannot rise smoothly; it jumps abruptly from one allowed mode to the next.
5. A sound level of 0dB does not mean the absence of sound. The decibel scale measures the intensity of a sound relative to a fixed reference, the threshold of hearing, $I_0 = 10^{-12} \text{ W/m}^2$; the level is ten times the logarithm, to base ten, of the ratio of the sound's intensity to this reference. A level of 0dB therefore means that the sound's intensity is equal to I_0 , the faintest intensity a normal human ear can just detect. A negative decibel level describes a sound whose intensity is below I_0 : a real sound that exists physically but is too faint for the human ear to detect.
6. Loudness is not the same quantity as intensity. Intensity is the physical power carried by the wave per unit area; loudness is the sensation the sound produces, and the sensitivity of the ear depends strongly on frequency. The ear is most sensitive to frequencies of about 2–5kHz and much less sensitive at low frequencies. The 3kHz tone lies near the peak of the ear's sensitivity, while the 50Hz tone lies where the ear responds poorly. At equal intensity, the 3kHz tone is therefore heard as the louder of the two.
7. Polarisation is the restriction of a wave's oscillation to a single transverse direction, and it is possible only for a transverse wave, whose oscillation is already perpendicular to the direction of travel and so offers a choice of plane. Light is a transverse wave and can therefore be polarised. Sound in air is a longitudinal wave: the air oscillates along the direction in which the wave travels, so there is no transverse direction for a polariser to select. A sound wave therefore cannot be polarised.
8. The electric and magnetic fields are in phase: they reach their maximum values at the same instant and fall to zero together. This is because the two fields are not independent. A changing electric field generates the magnetic field and a changing magnetic field generates the electric field, so the two are produced together and travel together as a single wave, with no continual exchange of energy between them of the kind that occurs between the kinetic and potential energy in simple harmonic motion. Their magnitudes are linked at every instant by $E = cB$.
9. Whether radiation harms the skin depends not on the total power it delivers but on its position in the electromagnetic spectrum, that is, on its frequency. The spectrum is ordered by frequency, and the higher the frequency, the more energetically the radiation interacts with the matter it meets. Ultraviolet lies at a higher frequency than visible light, high enough to disrupt the chemistry of the cells in the skin. Visible light, of lower frequency, cannot do this however intense it is. It is therefore the ultraviolet, and not the brighter visible light, that burns the skin.
10. A moving source and a moving observer do physically different things. A moving source travels through the air as it emits, chasing the wavefronts ahead of it and drawing away from those behind, so it crowds the wavefronts closer together in front and stretches them out behind: it changes the wavelength of the sound in the air itself. A moving observer leaves the wave in the air unchanged, the wavelength is unaltered, but by moving meets the wavefronts at a faster or slower rate than a stationary observer would. Because the source changes the wave while the observer changes only the rate at which the wavefronts are met, the two motions affect the observed frequency in different ways.
11. It is better understood as an expansion of the space between the galaxy and the Earth. The distant galaxies are very nearly at rest with respect to their own surroundings; they are not driving themselves through space. They are carried apart because the space lying between them steadily expands. This is supported by Hubble's law, which shows that the recession speed of a galaxy is proportional to its distance, exactly what a uniform expansion of space produces, since the more space lies between two galaxies, the more of it stretches and the faster they separate. Objects merely flying apart through a fixed space would not give this simple proportionality.
12. The Doppler shift of light responds only to the component of a galaxy's velocity along the line of sight, the radial component, because only that component changes the distance between the source and the observer and so stretches or compresses the

wavelength. A galaxy moving purely across the line of sight keeps its distance from the Earth momentarily unchanged, so it produces no shift in wavelength. A measured redshift therefore reveals only the radial part of a galaxy's motion; its motion across the sky goes undetected by the shift.

EXERCISE 9B

13. Sound of very low frequency has a very long wavelength. A wave is strongly diffracted by an obstacle only when its wavelength is comparable with, or larger than, the obstacle, so the long waves of infrasound bend around hills, trees and buildings instead of being stopped by them. Low-frequency sound is also absorbed by the air far more slowly than high-frequency sound. For both reasons, infrasound travels much greater distances than ordinary speech, whose shorter waves are more easily blocked and absorbed.

14. The deep bass of the music is low-frequency sound of long wavelength, while the consonants that carry speech are high-frequency sound of short wavelength. Long-wavelength sound passes through the wall, and diffracts around its edges and any gaps, far more readily than short-wavelength sound, which the wall reflects and absorbs much more effectively. The wall therefore lets the bass through while muffling the conversation.

15. Sound and light are both waves, but their wavelengths are vastly different. The wavelength of ordinary speech is of the order of one metre, comparable with the size of the building, so the sound wave diffracts strongly around the corner and reaches Kipanga. The wavelength of light is less than one millionth of a metre, far smaller than the building, so light diffracts only negligibly and travels in practically straight lines; the building blocks it. The same obstacle therefore stops the light while the sound bends around it.

16. The speed of sound in air increases with temperature. On a cold morning the air is cooler, so sound travels more slowly and takes longer to cover the distance to the cliff and back; the echo returns later. On a hot afternoon the warmer air carries the sound faster, so the same round trip takes less time and the echo returns sooner.

17. The air column above the water in the bucket acts as a closed pipe and resonates as the bucket fills. As the water rises, the length of this air column steadily decreases. The resonant frequency of a closed air column is inversely proportional to its length, so a shorter column resonates at a higher frequency. The pitch of the sound therefore rises steadily as the bucket fills.

18. The whine of a mosquito is a faint sound of low intensity. In a quiet bedroom the background sound is fainter still, so the whine stands well above the level at which the ear can detect it. In a busy market the loud background noise raises the level a sound must reach before the ear can pick it out, and the faint whine is masked by the much louder surrounding noise. The mosquito produces the same sound in both places; what changes is whether that sound stands out against the background.

19. The barrier blocks the straight-line path of the sound from the road to the houses, casting a sound shadow behind it. But sound has a long wavelength and diffracts: at the top edge of the barrier the waves bend over and spread downwards into the shadow region behind it. Some sound therefore always reaches the houses by diffracting over the barrier, so the noise is reduced but never completely removed.

20. A mobile phone communicates by radio waves, which are electromagnetic. The walls of the lift are made of metal, which is a conductor. Radio waves cannot pass through a conductor: the electric field of the wave drives currents in the metal, and these currents reflect and absorb the wave so that very little of it enters the enclosed space. The metal lift therefore screens the phone from the signal, and the connection is lost.

21. A lightning flash is a sudden, violent surge of electric charge. Accelerating charge radiates electromagnetic waves, and the abrupt lightning discharge sends out a brief, powerful burst of radio waves spread over a wide range of frequencies. The radio receiver picks up these bursts along with the station it is tuned to, and they are heard as the crackling and popping.

22. Every object emits electromagnetic radiation because of the motion of its particles, and a warm object such as the human body emits this radiation mainly in the infrared part of the spectrum. A thermal-imaging camera is built to detect infrared rather than visible light. It forms its picture from the infrared radiation the body itself gives out, so it needs no visible light at all and works just as well in complete darkness.

23. A television remote control sends infrared radiation, which has a very short wavelength and behaves much like visible light: it travels in straight lines and is stopped by walls, so it needs an unobstructed path to the set. The radio waves used for broadcasting have a much longer wavelength; they diffract around obstacles and pass through walls, so a radio receiver works even when it is far from the transmitter and separated from it by walls. The difference follows from the great difference in wavelength between the two kinds of wave.

24. Gamma rays lie at the highest-frequency end of the electromagnetic spectrum. They are extremely penetrating, so they pass straight through sealed packaging and through the whole body of the instrument, reaching every part of it. They also interact very energetically with matter, energetically enough to destroy the cells of bacteria and other micro-organisms. A gamma source can therefore sterilise an instrument completely without the package ever being opened.

25. The eye evolved to detect the radiation that is both most plentiful and most usable at the Earth's surface. The Sun radiates most strongly in the visible band, and the Earth's atmosphere is highly transparent to visible light while absorbing much of the radiation in the neighbouring bands. The visible band is therefore the part of the spectrum that reaches the ground in the greatest amount, so an eye sensitive to that band gives the clearest view of the surroundings. Natural selection accordingly favoured eyes tuned to visible light.

26. The monitor sends a beam of ultrasound into the body. When the ultrasound meets the moving wall of the baby's heart it is reflected, and because the reflecting surface is moving, the reflected wave returns at a slightly changed frequency: this is the Doppler effect. The monitor compares the frequency of the reflected wave with the frequency it sent out. The reflected frequency is raised while the heart wall moves towards the probe and lowered while it moves away, so the rhythmic change in frequency reveals and follows the beating of the heart.

27. A redshift means a galaxy is receding from us and a blueshift means it is approaching. The expansion of space tends to carry every galaxy away from us, but each galaxy also has its own local motion through space. The Andromeda galaxy is one of our nearest neighbours, close enough that the gravitational attraction between it and the Milky Way is drawing the two together. Over so short a distance the small recession due to the expansion is outweighed by this local motion of approach, so Andromeda is on the whole moving towards us and its light is blueshifted. The general redshift appears only for galaxies far enough away that the expansion outweighs any local motion.

EXERCISE 9C

28. The speed of sound increases by about 5.2% between the cold night and the hot afternoon.

29. 0.20

30. About 16 machines are required.

31. 0.45W

32. The sound level falls by 6.0dB.

33. $5.0 \times 10^{-10}\text{s}$ or 0.50ns.

34. 3840km/s, 55Mpc

35. The Hubble time is $\frac{1}{H_0}$. Expressing H_0 in basic units:

$$H_0 = \frac{70\text{km/s}}{3.09 \times 10^{19}\text{km}} = 2.27 \times 10^{-18}\text{s}^{-1}$$

Hence the Hubble time is:

$$\frac{1}{H_0} = \frac{1}{2.27 \times 10^{-18}\text{s}^{-1}} = 4.4 \times 10^{17}\text{s}$$

Dividing by the number of seconds in a year:

$$\frac{4.4 \times 10^{17}\text{s}}{3.16 \times 10^7\text{s}} = 1.4 \times 10^{10} \text{ years}$$

This is a rough estimate of the age of the universe, the time the present rate of expansion would have taken to produce the universe's present size.

36. 932Hz

37. 517Hz

38. (a) The angular frequency is $\omega = 2\pi f$:

$$\omega = 2\pi \times 400\text{Hz} = 2513\text{rad/s}$$

The maximum speed of the air particles is:

$$v_{\text{max}} = \omega s_0 = 2513\text{rad/s} \times 1.0 \times 10^{-5}\text{m} = 0.025\text{m/s}$$

(b) The pressure amplitude is:

$$p_0 = \rho v \omega s_0 = 1.2\text{kg/m}^3 \times 340\text{m/s} \times 2513\text{rad/s} \times 1.0 \times 10^{-5}\text{m} = 10.3\text{Pa}$$

(c) The intensity is:

$$I = \frac{p_0^2}{2\rho v} = \frac{(10.3\text{Pa})^2}{2 \times 1.2\text{kg/m}^3 \times 340\text{m/s}} = 0.13\text{W/m}^2$$

(d) The sound level is:

$$L = 10 \log \frac{I}{I_0} = 10 \log \frac{0.13\text{W/m}^2}{1.0 \times 10^{-12}\text{W/m}^2} = 111\text{dB}$$

39. 139N

40. (a) 768Hz (b) 12dB

41. 340m/s, 15°C.

42. For a radio wave with the source speed much less than c , the fractional Doppler shift is $\frac{\Delta f}{f} = \frac{v}{c}$. Hence:

$$\Delta f = \frac{fv}{c} = \frac{150 \times 10^6\text{Hz} \times 7.6 \times 10^3\text{m/s}}{3.0 \times 10^8\text{m/s}} = 3.8 \times 10^3\text{Hz}$$

The received frequency is raised by about 3.8kHz while the satellite approaches.

When the satellite passes directly overhead, its velocity is momentarily perpendicular to the line of sight. The radial component of the velocity is then zero, and since only the radial component produces a first-order Doppler shift, the shift falls to zero at that instant.

PART TWO

Question 1

(a) At a fixed end the string cannot move. The arriving pulse pulls upward on the support, and by Newton's third law the support pulls downward on the string with an equal force, which launches a reflected pulse on the opposite side, an inversion. At a free end no such reaction force acts: the end is free to overshoot, and the reflected pulse keeps its original orientation. Reflection at a fixed boundary introduces a phase change of π ; reflection at a free boundary introduces none.

(b) The projector sends two images to the screen, each carried by light polarised in one of two perpendicular planes. Each lens transmits only the light whose plane of polarisation matches its own transmission axis and blocks the perpendicular component almost completely. So each eye receives only its intended image, and the brain fuses the two slightly different views into one three-dimensional impression.

(c) The wavelength follows from $c = f\lambda$:

$$\lambda = \frac{c}{f} = \frac{3.00 \times 10^8\text{m/s}}{90.0 \times 10^6\text{Hz}} = 3.33\text{m}$$

The signal travels at the speed of light, so the time to cover 75km is:

$$t = \frac{d}{c} = \frac{75 \times 10^3\text{m}}{3.00 \times 10^8\text{m/s}} = 2.5 \times 10^{-4}\text{s}$$

Question 2

(a) Diffraction, the spreading of a wave as it passes an aperture, becomes appreciable only when the width of the aperture is comparable with the wavelength. The wavelength of light is around $5 \times 10^{-7}\text{m}$, vastly smaller than a doorway, so the spreading is utterly negligible and the light appears to travel in straight lines. A slit a fraction of a millimetre wide is far closer to the scale of the wavelength, so the spreading becomes noticeable.

(b) The speed of a water wave in shallow water depends on the depth, falling as the water becomes shallower. As the wave advances up the beach its speed therefore decreases. Its frequency is fixed by the distant source, so since $v = f\lambda$ the wavelength must shorten. The same energy is now packed into a shorter, slower wave, so the crest rises and the wave steepens.

(c) The intensity for a level L is $I = I_0 \times \log^{-1}\left(\frac{L}{10}\right)$. For the two machines:

$$I_1 = (1.0 \times 10^{-12}\text{W/m}^2) \times \log^{-1}(8.2) = 1.58 \times 10^{-4}\text{W/m}^2$$

$$I_2 = (1.0 \times 10^{-12}\text{W/m}^2) \times \log^{-1}(7.8) = 0.63 \times 10^{-4}\text{W/m}^2$$

Intensities from independent sources add:

$$I = I_1 + I_2 = 2.21 \times 10^{-4}\text{W/m}^2$$

Converting the combined intensity back to a sound level:

$$L = 10 \log(2.21 \times 10^8) = 83.5 \text{ dB}$$

Question 3

(a) The range of intensities the ear responds to is enormous: from the faintest audible sound to the threshold of pain spans a factor of about 10^{12} . A linear scale would be hopelessly unwieldy over so vast a range. A logarithmic scale compresses it into a convenient span of roughly 0 to 120. It also matches the ear itself, whose response to intensity is roughly logarithmic, so that equal multiplicative steps in intensity are heard as roughly equal steps in loudness.

(b) The vertical wall reflects the incoming waves with little loss, sending a reflected train back across the path of the incoming train. At every point the water carries both trains at once, and by the principle of superposition the resultant displacement is the sum of the two. Because the trains travel in different directions, their crests and troughs overlap in a continually shifting pattern, giving the choppy criss-cross appearance.

(c) The grating spacing is the reciprocal of the line density:

$$d = \frac{1 \text{ mm}}{400} = 2.5 \times 10^{-6} \text{ m}$$

The grating equation is $d \sin \theta = m\lambda$. The largest order corresponds to $\sin \theta = 1$:

$$m_{\text{max}} = \frac{d}{\lambda} = \frac{2.5 \times 10^{-6} \text{ m}}{600 \times 10^{-9} \text{ m}} = 4.17$$

The order must be a whole number, so the highest maximum that can be observed is the **fourth** order.

Question 4

(a) While the pulses overlap, the displacement of each point of the string is the algebraic sum of the displacements the two pulses would separately produce, which is the principle of superposition. The medium responds to each pulse independently, so neither is altered by the other. Once they have passed through each other, each pulse continues exactly as before, since no energy has been exchanged between them.

(b) The gun emits microwaves of known frequency, which reflect from the moving car and return. Because the car is moving, the reflected waves are Doppler-shifted: the car first acts as a moving receiver and then as a moving re-emitter, so the echo is shifted twice. The gun measures the frequency difference between the emitted and returned waves; this difference is proportional to the car's speed, from which the speed is found.

(c) By Malus's law the transmitted intensity is $I = I_0(\cos \theta)^2$. Setting $I = \frac{1}{4}I_0$:

$$\begin{aligned} (\cos \theta)^2 &= \frac{1}{4} \\ \cos \theta &= \frac{1}{2} \end{aligned}$$

The required angle is therefore:

$$\theta = 60^\circ$$

Question 5

(a) Sound is a mechanical wave: it travels only by the successive disturbance of the particles of a material medium, each layer pushing on the next. The Moon has no atmosphere, so between the shattering rock and the astronaut there is no medium to carry the compressions and rarefactions. With nothing to vibrate, no sound wave can form, and the event reaches the astronaut as silence.

(b) The colours come from interference, not pigment. The surfaces carry extremely fine, regularly spaced structures whose spacing is comparable with the wavelength of light. Light reflected from successive layers interferes; for a given viewing angle one wavelength is reinforced while others are cancelled, so a colour is seen. As the viewing angle changes, the path difference changes, the reinforced wavelength changes, and the colour shifts. This is structural colour.

(c) The wave speed is set by the tension and the linear density:

$$v = \sqrt{\frac{T}{\mu}} = \sqrt{\frac{80 \text{ N}}{0.020 \text{ kg/m}}} = 63.2 \text{ m/s}$$

The average power carried by the wave is $P = \frac{1}{2} \mu \omega^2 A^2 v$, with $\omega = 2\pi f = 314 \text{ rad/s}$. Substituting:

$$P = \frac{1}{2}(0.020\text{kg/m})(314\text{rad/s})^2(0.0080\text{m})^2(63.2\text{m/s}) = 3.99\text{W}$$

Question 6

(a) The speed of a transverse wave on a string is $v = \sqrt{T/\mu}$: it is fixed entirely by the tension and the mass per unit length, properties of the string itself, and is independent of how the wave is made. The frequency is the rate at which the source shakes the end of the string, each oscillation launching one cycle, so the source alone fixes it. The wavelength then adjusts to satisfy $\lambda = v/f$.

(b) The regularly spaced openings of the mesh act as a two-dimensional diffraction grating. Light from the lamp passing through the many openings is diffracted, and the diffracted beams interfere. In directions where the path difference between neighbouring openings is a whole number of wavelengths the interference is constructive, giving a bright spot; the regular spacing of the mesh produces a regular pattern of such spots.

(c) A pipe closed at one end has a node at the closed end and an antinode at the open end, so the fundamental wavelength is four times the length:

$$\lambda_1 = 4L = 4(0.55\text{m}) = 2.20\text{m}$$

$$f_1 = \frac{v}{\lambda_1} = \frac{340\text{m/s}}{2.20\text{m}} = 154.5\text{Hz}$$

A closed pipe resonates only in the odd harmonics, so the next two resonances are the third and fifth harmonics:

$$f_3 = 3f_1 = 463.6\text{Hz}$$

$$f_5 = 5f_1 = 772.7\text{Hz}$$

Question 7

(a) In a double slit only two beams interfere, so the intensity varies gradually with angle and the bright fringes are broad. In a grating, light from many thousands of slits interferes. At an exact maximum every beam arrives in phase. A very slight turn away from that direction is enough for the many beams, spread across the whole grating, to fall progressively out of step and cancel completely. The maxima are therefore confined to very narrow ranges of angle, and the more slits there are, the sharper the line.

(b) The echo returning from the moth is Doppler-shifted. If the moth is flying towards the bat it acts as an approaching reflector and the echo returns at a higher frequency than the emitted pulse; if it is flying away, the echo returns at a lower frequency. By comparing the pitch of the echo with that of its own pulse, the bat senses whether the moth is closing in or escaping.

(c) For two waves of the same frequency the resultant amplitude is:

$$A_R = \sqrt{A_1^2 + A_2^2 + 2A_1A_2\cos\varphi}$$

Substituting $A_1 = 3.0\text{cm}$, $A_2 = 5.0\text{cm}$ and $\varphi = 60^\circ$:

$$A_R = \sqrt{(3.0\text{cm})^2 + (5.0\text{cm})^2 + 2(3.0\text{cm})(5.0\text{cm})(0.50)}$$

$$A_R = 7.0\text{cm}$$

Question 8

(a) In a stationary wave every particle oscillates with the same frequency, but the amplitude varies with position, being zero at the nodes and greatest at the antinodes. Within one loop all the particles pass through equilibrium together and reach their extremes together: they oscillate exactly in phase, differing only in amplitude. Across a node the sign of the displacement reverses, so particles in adjacent loops are always half a cycle out of phase, reaching their extremes on opposite sides at the same instant.

(b) When the jet moves faster than the speed of sound it outruns the waves it produces. These waves cannot move ahead of it; instead they pile up and overlap along a cone-shaped shock front trailing the aircraft, across which there is a sharp, intense pressure change. When this shock front sweeps over an observer on the ground, the abrupt pressure change is heard as the single loud bang.

(c) The first dark fringe of a single slit occurs where a $\sin\theta = \lambda$:

$$\sin\theta = \frac{\lambda}{a} = \frac{600 \times 10^{-9}\text{m}}{0.20 \times 10^{-3}\text{m}} = 3.0 \times 10^{-3}$$

So $\theta = 3.0 \times 10^{-3}\text{rad}$.

The central band runs from this fringe on one side to the matching fringe on the other, so on a screen at distance D its width is:

$$w = 2D \sin\theta = 2(2.0\text{m})(3.0 \times 10^{-3}) = 1.2 \times 10^{-2}\text{m}$$

The central bright band is therefore 12mm wide.

Question 9

(a) In two-slit interference the two beams have effectively equal amplitude wherever they meet, so wherever they arrive in phase they reinforce to the same maximum brightness, and all the bright fringes are alike. In single-slit diffraction the slit acts as very many tiny sources across its width. At the centre all their contributions arrive in phase and add fully; away from the centre they are increasingly out of step, so only a partial reinforcement survives, and the bright bands fade with distance from the centre.

(b) Speech sets the base of the first tin vibrating; the vibration passes into the string and travels along it as a wave to the second tin, whose base then re-radiates the sound. A string can carry such a wave only while it is taut, because the wave needs tension to supply the restoring force, its speed being $v = \sqrt{T/\mu}$. When the string goes slack the tension falls almost to zero, the string can no longer transmit the vibration, and the sound fails to cross.

(c) The sound is Doppler-shifted twice: once as the moving car sends sound to the cliff, and again as the moving driver receives the reflected sound. For a source and observer moving towards a stationary reflector at the same speed u , the two shifts combine to give:

$$f' = f \frac{v+u}{v-u}$$

$$f' = 480\text{Hz} \times \frac{340\text{m/s} + 30\text{m/s}}{340\text{m/s} - 30\text{m/s}} = 573\text{Hz}$$

The driver therefore hears the echo at:

$$f' = 573\text{Hz}$$

Question 10

(a) The flash and the thunder are produced together, but they reach the observer at very different speeds. Light travels at about $3 \times 10^8\text{m/s}$, so the flash arrives almost instantly. Sound travels through air at only about 340m/s, nearly a million times slower, so the thunder takes an appreciable time to cover the same distance. The further away the stroke, the longer the gap between the flash and the thunder.

(b) Heating in air grows a very thin, transparent film of oxide on the steel, and the film thickens the longer and hotter the heating. Light reflected from the top of the oxide film and light reflected from the steel beneath it interfere. The film thickness fixes the path difference, and so fixes which wavelength is reinforced and which cancelled, so the surface shows an interference colour. As the film thickens the reinforced wavelength changes, carrying the surface through the observed sequence of colours.

(c) The fundamental frequency of a wire fixed at both ends is $f = \frac{1}{2L} \sqrt{T/\mu}$. With the length and the linear density unchanged, $f \propto \sqrt{T}$, and therefore $T \propto f^2$.

The frequency is to rise in the ratio $f_2/f_1 = 3/2$, so:

$$\frac{T_2}{T_1} = \left(\frac{f_2}{f_1}\right)^2 = \left(\frac{3}{2}\right)^2 = 2.25$$

The tension must be increased to 2.25 times its original value.

Question 11

(a) Both systems, when displaced from equilibrium, experience a restoring force directed back towards the equilibrium position and proportional to the displacement from it. For the spring it is the elastic force $F = -kx$; for the pendulum at small angles it is the tangential component of gravity, which is likewise proportional to the displacement. Whenever the restoring force is proportional to the displacement and oppositely directed, the acceleration obeys $a = -\omega^2 x$, which is the defining condition of simple harmonic motion. The two systems differ in their physical make-up but share this one essential property.

(b) Where the protector and the screen nearly touch there is a very thin film of trapped air whose thickness varies from point to point. Light reflected from the lower surface of the protector and light reflected from the screen beneath it interfere. Where the air gap makes the path difference a whole number of wavelengths the interference is constructive for that colour, and where it makes an odd number of half-wavelengths it is destructive; since the gap thickness varies across the contact region, different colours are reinforced at different places, producing the coloured rings and bands. This is the same effect as Newton's rings.

(c) The observer moves towards a stationary source, so the observed frequency is raised. With the observer's speed u_o added to the speed of sound:

$$f' = f \frac{v + u_o}{v}$$

$$f' = 660\text{Hz} \times \frac{(340 + 20)\text{m/s}}{340\text{m/s}} = 660\text{Hz} \times \frac{360\text{m/s}}{340\text{m/s}}$$

The motorcyclist therefore hears:

$$f' = 699\text{Hz}$$

Question 12

(a) The fringes form a thin-film interference pattern, and a given fringe is the locus of all points at which the air gap between the curved lens and the flat plate has one particular thickness. The lens surface is part of a sphere resting on a plane, so the air gap has the same thickness at every point lying at a given distance from the point of contact. Points of equal gap thickness therefore lie on circles centred on the contact point, and so each fringe, bright or dark, is itself a circle.

(b) Springing off leaves the board both displaced and moving, so it has been given elastic potential energy and kinetic energy. With the diver gone, no further force drives it; the board simply oscillates freely at its own natural frequency, energy passing back and forth between the elastic strain of the bent board and the kinetic energy of its motion. This is a free oscillation, and it dies away only as the energy is gradually lost to air resistance and internal friction.

(c) For a point source the intensity falls with the inverse square of the distance:

$$\frac{I_2}{I_1} = \left(\frac{r_1}{r_2}\right)^2 = \left(\frac{2.0\text{m}}{8.0\text{m}}\right)^2 = \frac{1}{16}$$

The corresponding change in sound level is:

$$\Delta L = 10 \log\left(\frac{I_2}{I_1}\right) = 10 \log\left(\frac{1}{16}\right) = -12.0\text{dB}$$

The level at 8.0m is therefore:

$$L_2 = 88\text{dB} - 12.0\text{dB} = 76.0\text{dB}$$

Question 13

(a) An open pipe has a displacement antinode at each end, so its fundamental fits half a wavelength into the pipe: the fundamental wavelength is twice the pipe length. A pipe closed at one end has a node at the closed end and an antinode at the open end, so its fundamental fits only a quarter of a wavelength into the same length: its fundamental wavelength is four times the pipe length, twice that of the open pipe. Since frequency is inversely proportional to wavelength, the closed pipe sounds at half the frequency, which is one octave lower.

(b) The baby and the elastic strap together form a mass-on-a-spring system. The strap stretches under the baby's weight to an equilibrium position; if the baby is displaced below it the extra tension provides a restoring force proportional to the displacement, and above it the reduced tension does the same. The restoring force being proportional to displacement, the baby executes simple harmonic motion, with a regular period $T = 2\pi\sqrt{m/k}$ set by the baby's mass and the stiffness of the strap, hence the steady rhythm.

(c) The fringe spacing is $y = \frac{\lambda D}{d}$, so it is proportional to the wavelength of the light. On immersion in water the wavelength shortens to $\lambda_w = \lambda/n$, the frequency being unchanged, so the fringe spacing shrinks in the same ratio:

$$y_w = \frac{y}{n} = \frac{1.8\text{mm}}{1.33} = 1.35\text{mm}$$

Question 14

(a) In simple harmonic motion the maximum speed is $v_{\max} = \omega A$ and the maximum acceleration is $a_{\max} = \omega^2 A$; both are directly proportional to the amplitude, so doubling the amplitude doubles each of them. The period $T = 2\pi/\omega$ is fixed by the angular frequency alone, and ω is set by the properties of the system, the stiffness and mass or the length and gravity, not by how far the oscillator is displaced. Doubling the amplitude does not change ω , so the period is unchanged: the oscillator simply covers twice the distance each cycle, moving twice as fast to do so in the same time.

(b) The smooth curved wall of the dome reflects sound well. Sound spreading out from the whisper at one point strikes the curved surface and is reflected, and the curvature is such that the reflected sound is brought together again, or focused, at the second point. The whisper's energy, instead of spreading thinly in every direction, is concentrated by these reflections onto that one listening point, so it arrives there loud enough to hear, while points away from the focus receive only the faint direct sound.

(c) Light reflected from the top surface of the film undergoes a phase change of π , while light reflected from the bottom surface does not. Constructive interference in the reflected light, giving a strongly reflected colour, therefore requires:

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

The optical path is $2nt = 2(1.33)(1.10 \times 10^{-7}\text{m}) = 2.93 \times 10^{-7}\text{m}$. Taking the lowest order, $m = 0$:

$$\lambda = 2 \times (2.93 \times 10^{-7}\text{m}) = 5.85 \times 10^{-7}\text{m}$$

The most strongly reflected visible wavelength is 585nm.

(The next order, $m = 1$, would give 195nm, which lies in the ultraviolet.)

Question 15

(a) The speed of sound in a gas depends on how rapidly the molecules pass the disturbance from one to the next. In warmer air the molecules move with greater average speed, so they collide more frequently and transmit the compressions and rarefactions more quickly. The speed of sound in a gas is in fact proportional to the square root of the absolute temperature, so a rise in temperature raises the speed.

(b) Each fine scratch is a narrow transparent track of a width not enormously greater than the wavelength of light, and it diffracts the light passing through it, spreading that light out in the direction across the scratch. The bright point of a distant headlight is therefore smeared into a streak of light running perpendicular to the scratch. Because the wiper scratches lie along many arcs at different orientations, the diffracted light spreads in a range of directions, producing the spikes and streaks seen around each headlight.

(c) For a mass on a spring $T = 2\pi\sqrt{m/k}$, so the period is inversely proportional to the square root of the stiffness. Two identical springs side by side, sharing the load, act as a single spring of twice the stiffness:

$$k_p = 2k$$

The period therefore shortens by a factor of $\sqrt{2}$:

$$T_2 = \frac{T_1}{\sqrt{2}} = \frac{0.80\text{s}}{\sqrt{2}} = 0.57\text{s}$$

Question 16

(a) At the extremes of the motion the oscillator is momentarily at rest, so all its energy is potential. For a system with restoring force $F = -kx$, the potential energy stored at displacement x is $\frac{1}{2}kx^2$, and at the extreme, where $x = A$, this is $\frac{1}{2}kA^2$. Since the total energy is constant throughout the motion, in the absence of damping, and equals this maximum potential energy, the total energy is $\frac{1}{2}kA^2$, proportional to the square of the amplitude. Doubling the amplitude therefore quadruples the energy of the oscillation.

(b) The wire is a narrow obstacle of width comparable with the wavelength of light. As light from the distant lamp passes the wire it is diffracted, bending into the region behind the wire. Light arriving at a point in that region by the two paths passing the opposite sides of the wire has a path difference that depends on the direction; where the path difference is a whole number of wavelengths the waves reinforce, and where it is an odd number of half-wavelengths they cancel. This alternation produces the regular pattern of light and dark bands.

(c) An open pipe has a displacement antinode at each end, so its fundamental wavelength is twice the length of the pipe, $\lambda_1 = 2L$. Hence:

$$L = \frac{v}{2f_1} = \frac{340\text{m/s}}{2 \times 256\text{Hz}} = 0.664\text{m}$$

An open pipe sounds all the harmonics, so the second harmonic has twice the fundamental frequency:

$$f_2 = 2f_1 = 2 \times 256\text{Hz} = 512\text{Hz}$$

Question 17

(a) For a steady interference pattern the light reaching the two slits must be coherent: the two slits must always be illuminated by waves with a constant phase relationship. An ordinary lamp is an extended source whose different parts emit independently, so light reaching the two slits from it directly would have no fixed phase relationship. The single narrow slit acts as one small, effectively point-like source: the light spreading from it reaches both slits as parts of the same wavefront, so the two slits are illuminated coherently and a stable pattern is formed.

(b) When a person speaks, the sound reaches their own ears by two routes, through the air and directly through the bones and tissues of the head. Bone conduction carries the lower frequencies particularly well, so the voice a speaker normally hears is enriched in low frequencies and sounds fuller and deeper to them. A recording captures only the air-conducted sound; played back,

it reaches the ears by the air route alone, without the bone-conducted bass, so it sounds thinner and unfamiliar to the speaker, though it is the voice that others always hear.

(c) The period depends only on the mass and the stiffness:

$$T = 2\pi\sqrt{\frac{m}{k}} = 2\pi\sqrt{\frac{0.30\text{kg}}{30\text{N/m}}} = 0.63\text{s}$$

At the equilibrium position the spring tension balances the weight, $mg = 0.30\text{kg} \times 9.8\text{m/s}^2 = 2.94\text{N}$. During the motion the tension varies above and below this by $kA = 30\text{N/m} \times 0.080\text{m} = 2.4\text{N}$. The greatest tension, at the lowest point, is:

$$F_{\text{max}} = mg + kA = 2.94\text{N} + 2.4\text{N} = 5.34\text{N}$$

And the least tension, at the highest point, is:

$$F_{\text{min}} = mg - kA = 2.94\text{N} - 2.4\text{N} = 0.54\text{N}$$

Question 18

(a) The displacement completes one full cycle in each period. The kinetic energy depends on the square of the speed. In one period of the displacement the oscillator passes through the equilibrium position twice, once moving each way, and at each passage the speed, and hence the kinetic energy, reaches its maximum; the kinetic energy falls to zero twice as well, at the two extremes. The kinetic energy therefore rises and falls twice in every single cycle of the displacement, so it varies at twice the frequency.

(b) The chest-piece picks up the heartbeat sounds, and the hollow tubes of the stethoscope carry the sound waves directly to the doctor's ears. Confined within the narrow tubes, the sound energy cannot spread out and weaken as it would in the open air, where it would diminish with the square of the distance. Almost all of the collected sound energy is therefore delivered to the ears instead of being dispersed, so the heartbeat is heard far more loudly than it would be by simply listening near the chest.

(c) The grating spacing is:

$$d = \frac{1\text{mm}}{500} = 2.0 \times 10^{-6}\text{m}$$

From the grating equation $d \sin\theta = m\lambda$, with $m = 2$ and $\theta = 36.0^\circ$:

$$\lambda = \frac{d \sin\theta}{m} = \frac{(2.0 \times 10^{-6}\text{m})(\sin 36.0^\circ)}{2}$$

$$\lambda = 5.88 \times 10^{-7}\text{m}$$

The wavelength of the light is 588nm.

Question 19

(a) When light strikes a surface, the component of its vibration parallel to the surface is reflected more strongly than the component lying in the plane of incidence. The reflected beam is therefore richer in the parallel vibration, that is, partially polarised. At one particular angle of incidence, Brewster's angle, the reflected and refracted rays are at right angles to each other; the vibration in the plane of incidence is then not reflected at all, and the reflected light consists solely of the vibration parallel to the surface, so it is completely polarised.

(b) The hanging load behaves as a simple pendulum, with the cable serving as the pendulum's length. The period of a simple pendulum is $T = 2\pi\sqrt{l/g}$: it increases as the square root of the length. Paying out more cable lengthens the pendulum, so the period grows and each swing takes longer, the load swinging more slowly. The mass of the load does not enter the formula, so it is the cable length alone that the operator alters when the swinging becomes lazier.

(c) The pulse travels out to the aircraft and back, so in the measured time it covers twice the distance to the aircraft:

$$2d = ct$$

$$d = \frac{ct}{2} = \frac{(3.00 \times 10^8\text{m/s})(1.20 \times 10^{-4}\text{s})}{2}$$

The aircraft is therefore:

$$d = 1.80 \times 10^4\text{m} = 18.0\text{km}$$

Question 20

(a) All three are electromagnetic waves: each consists of oscillating electric and magnetic fields, at right angles to each other and to the direction of travel, and each is produced by accelerating electric charges. They all travel through a vacuum at the same speed,

$3 \times 10^8 \text{ m/s}$, and all show the wave behaviours of reflection, refraction, interference, diffraction and polarisation. They differ only in frequency, and so in wavelength; this difference gives them their different uses and effects, but their common nature makes them one single family, the electromagnetic spectrum.

(b) Toughened glass is made by cooling its surfaces rapidly, which leaves the glass with permanent built-in stresses that vary from place to place. Stressed glass is birefringent: it affects the two perpendicular components of a light wave differently, altering the polarisation of light passing through it by an amount that depends on the local stress. The daylight reaching the window is already partly polarised, and the sunglasses act as an analysing polariser. Where the stress in the glass has turned the polarisation away from the sunglasses' axis, less light passes and a dark blotch is seen; the varying stress thus shows itself as a pattern of light and dark patches.

(c) From the period formula $T = 2\pi\sqrt{l/g}$, the length is:

$$l = \frac{gT^2}{4\pi^2} = \frac{(9.8 \text{ m/s}^2)(2.0 \text{ s})^2}{4\pi^2} = 0.993 \text{ m}$$

Since $T \propto \sqrt{l}$, a small fractional change in length produces half that fractional change in period:

$$\frac{\Delta T}{T} = \frac{1}{2} \frac{\Delta l}{l} = \frac{1}{2}(1.0\%) = 0.5\%$$

The period increases by 0.5%.

Question 21

(a) The natural frequency of a system is the frequency at which it oscillates on its own once it has been displaced and released, with no external driving force acting. It is fixed entirely by the properties of the system. For a mass on a spring it is set by the force constant of the spring and the mass, through $f = \frac{1}{2\pi}\sqrt{k/m}$. For a simple pendulum it is set by the length of the pendulum and the acceleration of free fall, through $f = \frac{1}{2\pi}\sqrt{g/l}$.

(b) Each spectator only rises and sits down again about their own seat; the disturbance, the act of standing, is passed on from each spectator to the next, so the pattern of standing sweeps around the ground. What travels round the stadium is the disturbance and the energy carried with it, while the spectators, who form the medium, merely oscillate in place. This is the defining feature of a wave: the wave profile and its energy move forward, the medium itself does not.

(c) The clap travels from the girl to the cliff and back, covering twice her distance from the cliff, so:

$$d = \frac{vt}{2} = \frac{(340 \text{ m/s})(1.5 \text{ s})}{2} = 255 \text{ m}$$

After walking 60m closer, her new distance from the cliff is $255 \text{ m} - 60 \text{ m} = 195 \text{ m}$. The echo now covers twice this distance:

$$t' = \frac{2d'}{v} = \frac{2(195 \text{ m})}{340 \text{ m/s}} = 1.15 \text{ s}$$

Question 22

(a) Sound is carried forward by each particle of the medium disturbing its neighbours, and the speed depends on how stiffly the particles are coupled and on the density of the medium. In a solid or a liquid the particles are much more closely bound, and the elastic restoring forces between them far stronger, than in a gas, so a disturbance is handed on from particle to particle far more rapidly. Solids and liquids are also denser, which on its own would slow the sound, but the great increase in stiffness outweighs the increase in density, so the net result is a much higher speed.

(b) On the stretched cord the jumper oscillates up and down: below the equilibrium extension the cord pulls her up, above it her weight pulls her down, giving a restoring force and so an oscillation that is close to simple harmonic. The motion is damped. At each bounce some mechanical energy is lost, to air resistance and to internal friction within the stretching cord, and is carried away as heat. With less energy in each successive cycle the amplitude steadily falls, bounce by bounce, until the surplus energy is spent and she hangs at rest at the equilibrium extension.

(c) The wavelength follows from $v = f\lambda$:

$$\lambda = \frac{v}{f} = \frac{1.8 \text{ m/s}}{2.5 \text{ Hz}} = 0.72 \text{ m}$$

One whole wavelength corresponds to a phase difference of 2π radians, so two points a distance d apart differ in phase by:

$$\Delta\phi = 2\pi\left(\frac{d}{\lambda}\right) = 2\pi\left(\frac{0.30 \text{ m}}{0.72 \text{ m}}\right)$$

which gives:

$$\Delta\phi = \frac{5\pi}{6} \text{ rad} = 2.6 \text{ rad}$$

Question 23

(a) The join is a boundary between two media along which the wave travels at different speeds, the wave being slower on the heavier rope. A wave meeting such a boundary cannot pass on undisturbed: the two ropes cannot keep their motions matched at the join unless some of the wave is sent back. Part of the energy is therefore carried forward into the heavy rope as a transmitted wave, and the rest is returned along the light rope as a reflected wave. Because it meets a heavier, slower medium, the reflected pulse is also inverted.

(b) Sound spreads out when it passes the edge of an obstacle or an opening, an effect called diffraction, and the spreading is large when the wavelength is comparable with, or larger than, the obstacle. The wavelength of ordinary speech is about a metre, comparable with the size of a wall or a doorway, so the sound diffracts strongly and bends round the corner to reach the listener. Light has a wavelength far smaller than any everyday obstacle, so it diffracts only negligibly and travels on in straight lines; hence the speaker cannot be seen.

(c) For simple harmonic motion the maximum speed is $v_{\max} = \omega A$ and the maximum acceleration is $a_{\max} = \omega^2 A$. Dividing one by the other eliminates the amplitude:

$$\omega = \frac{a_{\max}}{v_{\max}} = \frac{4.5 \text{ m/s}^2}{0.30 \text{ m/s}} = 15 \text{ rad/s}$$

The period follows from the angular frequency:

$$T = \frac{2\pi}{\omega} = \frac{2\pi}{15 \text{ rad/s}} = 0.42 \text{ s}$$

and the amplitude from $v_{\max} = \omega A$:

$$A = \frac{v_{\max}}{\omega} = \frac{0.30 \text{ m/s}}{15 \text{ rad/s}} = 0.020 \text{ m}$$

Question 24

(a) Take the displacement as $x = A \sin \omega t$. The velocity is the rate of change of displacement, $v = A\omega \cos \omega t$, and the acceleration is $a = -A\omega^2 \sin \omega t$. The velocity is a quarter of a cycle, a phase angle of $\frac{\pi}{2}$, ahead of the displacement: it is greatest as the body crosses the centre, where the displacement is zero, and zero at the extremes, where the displacement is greatest. The acceleration is exactly half a cycle, a phase angle of π , out of step with the displacement: it is always directed opposite to the displacement, greatest at the extremes and zero at the centre.

(b) The ship sends a short pulse of sound vertically downwards. The pulse travels down to the sea-bed, is reflected there, and travels back up to a detector on the ship. The instrument measures the time between sending the pulse and receiving its echo. In that time the pulse has covered the depth twice, and since the speed of sound in seawater is known, the depth is half the product of that speed and the measured time. A longer delay therefore means deeper water.

(c) The speed of a transverse wave on a string is set by the tension and the linear mass density:

$$v = \sqrt{\frac{T}{\mu}} = \sqrt{\frac{50 \text{ N}}{2.0 \times 10^{-3} \text{ kg/m}}} = 158 \text{ m/s}$$

The wavelength then follows from $v = f\lambda$:

$$\lambda = \frac{v}{f} = \frac{158 \text{ m/s}}{100 \text{ Hz}} = 1.58 \text{ m}$$

Question 25

(a) While the ambulance approaches, each successive compression of the sound wave is sent out from a point closer to the listener than the one before, so the compressions arrive more closely spaced: the effective wavelength is shortened and the frequency raised. While the ambulance recedes, each compression is sent out from a point further away, so they arrive more widely spaced: the wavelength is lengthened and the frequency lowered. At the instant of passing, the source changes from approaching to receding, so the pitch heard drops from the raised value to the lowered one.

(b) The stone gives the water a fixed amount of energy, and this energy is spread along the expanding circular ripple. As the ripple travels outwards its circumference grows, so the same energy must be shared along an ever-greater length of crest, and the energy

carried by each portion of the ripple steadily decreases. Since the energy of a wave depends on the square of its amplitude, this falling energy shows itself as a falling amplitude, and the ripples grow weaker as they spread. Some energy is also lost in overcoming the viscosity of the water.

(c) For a mass on a spring $T = 2\pi\sqrt{m/k}$, which rearranges to give the force constant:

$$k = \frac{4\pi^2 m}{T^2} = \frac{4\pi^2(0.25\text{kg})}{(0.50\text{s})^2} = 39.5\text{N/m}$$

Since $T \propto \sqrt{m}$, doubling the period to 1.0s requires the mass to become four times as large, that is $4 \times 0.25\text{kg} = 1.0\text{kg}$.

Checking with the period formula:

$$m' = \frac{kT'^2}{4\pi^2} = \frac{(39.5\text{N/m})(1.0\text{s})^2}{4\pi^2} = 1.0\text{kg}$$

The further mass that must be added is therefore:

$$\Delta m = 1.0\text{kg} - 0.25\text{kg} = 0.75\text{kg}$$

Question 26

(a) Huygens' principle states that every point on a wavefront may be regarded as a source of secondary spherical wavelets, which spread out forwards with the speed of the wave, and that the new wavefront a short time later is the surface that just touches, the envelope of, all these secondary wavelets. To find how a wavefront advances, one draws from each point of the present wavefront a small wavelet, every wavelet of the same radius, equal to the wave speed multiplied by the time elapsed, and then draws the surface tangent to them all. That tangent surface is the wavefront in its new position.

(b) The metronome arm carrying its weight is an oscillating system that swings to and fro, and the rate of ticking is simply its natural frequency. That natural frequency depends on how the mass is distributed about the pivot. Sliding the weight outwards, away from the pivot, increases the rotational inertia of the arm, so each swing takes longer and the ticking slows. Sliding the weight inwards, towards the pivot, reduces the rotational inertia, so each swing is quicker and the ticking speeds up. The musician sets the tempo simply by choosing where the weight sits.

(c) The speed of sound in a gas is proportional to the square root of its absolute temperature, $v \propto \sqrt{T}$. Hence the speed at the higher temperature is:

$$v_2 = v_1 \sqrt{\frac{T_2}{T_1}} = (340\text{m/s}) \sqrt{\frac{308\text{K}}{288\text{K}}} = 352\text{m/s}$$

The wavelength of the 480Hz note follows from $\lambda = v/f$. At 15°C:

$$\lambda_1 = \frac{v_1}{f} = \frac{340\text{m/s}}{480\text{Hz}} = 0.708\text{m}$$

and at 35°C:

$$\lambda_2 = \frac{v_2}{f} = \frac{352\text{m/s}}{480\text{Hz}} = 0.733\text{m}$$

Question 27

(a) The pitch of a note, how high or low it sounds, is determined by the frequency of the sound wave: a higher frequency is heard as a higher pitch. The loudness of the note, how strong or faint it sounds, is determined by the intensity of the wave, and so by the amplitude of the vibration, since the intensity depends on the square of the amplitude: a larger amplitude is heard as a louder sound. The two properties are independent: two notes may share the same pitch but differ in loudness, or share the same loudness but differ in pitch, because each depends on a different feature of the wave.

(b) A person sitting on a swing is a simple pendulum, and the period of a simple pendulum, $T = 2\pi\sqrt{l/g}$, depends only on the length of the swing and on the acceleration of free fall, not on the mass of the person. A heavier person is pulled towards the lowest point by a larger gravitational restoring force, but that same person also has correspondingly greater inertia and so is harder to set moving; the two effects cancel exactly. The time for one complete to-and-fro swing is therefore the same for the heavy adult and the light child on a swing of the same length.

(c) The pulse travels to the far post and back, a total distance of $2 \times 100\text{m}$, at 40m/s:

$$t = \frac{2L}{v} = \frac{2(100\text{m})}{40\text{m/s}} = 5.0\text{s}$$

The speed of a wave on the cable is $v = \sqrt{T/\mu}$, so $v \propto \sqrt{T}$. Raising the tension to four times its value multiplies the speed by $\sqrt{4} = 2$, giving a new speed of 80m/s. The new return time is therefore:

$$t' = \frac{2L}{v'} = \frac{2(100\text{m})}{80\text{m/s}} = 2.5\text{s}$$

Question 28

(a) A wavefront is the surface reached by all the parts of the wave that are in step with one another. Close to a small source the wave has spread only a short distance, and the wavefront is a small circle, or sphere, centred on the source, with a pronounced curvature. As the wave travels outwards the radius of this circle grows; far from the source the radius is very large, and any small portion of so large a circle, which is the part an observer actually sees, is almost indistinguishable from a straight line. The wavefront therefore appears practically straight, a plane wavefront, at a great distance from the source.

(b) Ultrasound has a far higher frequency than audible sound, and therefore a much shorter wavelength. A wave can resolve, and so form a clear image of, detail only about as small as its own wavelength; the very short wavelength of ultrasound allows the fine structure of the baby to be made out, which long-wavelength audible sound could never show. Ultrasound is also strongly reflected at the boundaries between the different soft tissues, providing the echoes from which the image is built, and it can be sent in a narrow, well-directed beam. It is, in addition, safe for both mother and baby.

(c) The total energy of a simple harmonic oscillator is $E = \frac{1}{2}kA^2$, which gives the force constant:

$$k = \frac{2E}{A^2} = \frac{2(0.16\text{J})}{(0.080\text{m})^2} = 50\text{N/m}$$

At the centre of the motion all the energy is kinetic, $E = \frac{1}{2}mv_{\text{max}}^2$, which gives the maximum speed:

$$v_{\text{max}} = \sqrt{\frac{2E}{m}} = \sqrt{\frac{2(0.16\text{J})}{0.50\text{kg}}} = 0.80\text{m/s}$$

Question 29

(a) A driven system takes in energy from the driving force most effectively when it is pushed in step with its own natural motion. When the driving frequency matches the natural frequency, each push from the driving force acts in the same direction as the system's velocity, and so does positive work on it cycle after cycle, feeding energy in steadily. The amplitude grows until the rate at which energy is fed in is balanced by the rate at which it is lost to damping; when the damping is light this balance is reached only at a very large amplitude. At driving frequencies away from the natural frequency the force is repeatedly out of step, often opposing the motion, so far less energy is delivered and the amplitude stays small.

(b) Sound striking a surface is partly reflected and partly absorbed. Bare hard walls and floors reflect sound strongly and absorb very little, so a voice is reflected back and forth many times and dies away only slowly; the sound rings on and the room seems hollow. Soft furnishings, such as curtains, carpets and padded chairs, are good absorbers: their soft, porous surfaces convert the energy of the sound into heat at each encounter. Once the room is furnished the reflected sound is quickly absorbed, the ringing dies away rapidly, and the hollow quality disappears.

(c) In a stationary wave the distance from a node to the antinode next to it is one quarter of a wavelength, so:

$$\lambda = 4 \times 0.12\text{m} = 0.48\text{m}$$

The string vibrates 25 times each second, so $f = 25\text{Hz}$, and the speed of the waves follows from $v = f\lambda$:

$$v = f\lambda = (25\text{Hz})(0.48\text{m}) = 12\text{m/s}$$

Question 30

(a) The ear and brain cannot separate two sounds that arrive within about a tenth of a second of each other; such sounds merge into one. For a reflected sound to be heard as a distinct echo it must therefore arrive at least about a tenth of a second after the direct sound. The reflected sound has to travel from the listener to the surface and back again, so the surface must be far enough away for this extra journey to take more than about a tenth of a second; at the speed of sound in air this means a reflecting surface more than roughly 17m away. A nearer surface returns the sound too soon, and it is heard only as a prolonging of the original.

(b) The turned skipping rope is held at its two ends and carries a stationary wave. The held ends cannot move, so they must be nodes, points of zero displacement, while the middle is free to swing out fully and is an antinode. The simplest stationary wave that has a node at each end with one antinode between them is a single loop, half a wavelength long; this is the fundamental mode, the one set up at the steady, moderate rate at which the children naturally turn the rope. The rope therefore settles into one broad loop, its ends almost still and its middle sweeping out the widest path.

(c) For simple harmonic motion the speed at a displacement x is given by $v^2 = \omega^2(A^2 - x^2)$. Applying this at the two stated points:

$$(8.0\text{cm/s})^2 = \omega^2(A^2 - (6.0\text{cm})^2)$$

$$(6.0\text{cm/s})^2 = \omega^2(A^2 - (8.0\text{cm})^2)$$

Dividing the first equation by the second eliminates ω :

$$\frac{64}{36} = \frac{A^2 - 36}{A^2 - 64}$$

Cross-multiplying and solving for the amplitude:

$$64(A^2 - 64) = 36(A^2 - 36)$$

$$A = 10\text{cm}$$

Substituting back into the first equation gives the angular frequency:

$$\omega^2 = \frac{64}{A^2 - 36} = \frac{64}{64} = 1.0$$

$$\omega = 1.0\text{rad/s}$$

And hence the period:

$$T = \frac{2\pi}{\omega} = \frac{2\pi}{1.0\text{rad/s}} = 6.3\text{s}$$

Question 31

(a) In simple harmonic motion the speed is greatest at the centre and falls to zero at the extremes, since $v = \omega\sqrt{A^2 - x^2}$. Near the extremes the body is moving slowly, so it takes a relatively long time to cover any given small stretch of its path there; near the centre it is moving quickly and passes through in a short time. Over one complete cycle, therefore, more of the total time is spent in the slow-moving regions near the two extremes than in the fast-moving region near the centre.

(b) The breakwater has an end, and the open water beyond that end acts as a gap past which the sea waves travel. Waves passing the end of the wall are diffracted: they spread sideways into the region of calm water behind the wall. Because the wavelength of the sea waves is not small compared with the wall, this spreading is appreciable, so some of the wave energy curls round into the sheltered water and reaches the moored boats, setting them gently rocking. The waves there are weaker than in the open sea, since only the diffracted part of the wave arrives.

(c) The grating equation is $d \sin\theta = m\lambda$. For the first-order image, $m = 1$, the line spacing is:

$$d = \frac{m\lambda}{\sin\theta} = \frac{(1)(590 \times 10^{-9}\text{m})}{\sin 18.0^\circ} = 1.91 \times 10^{-6}\text{m}$$

This spacing is $1.91 \times 10^{-3}\text{mm}$, so the number of lines per millimetre of the grating is:

$$N = \frac{1\text{mm}}{1.91 \times 10^{-3}\text{mm}} = 524$$

Question 32

(a) At a dark fringe the light from the two slits arrives exactly out of step, the crest of one wave coinciding with the trough of the other. By the principle of superposition the two displacements cancel, so the resultant disturbance, and with it the intensity, is zero. The light energy is not destroyed but redistributed. Where the two waves arrive in step, at the bright fringes, they reinforce, and the intensity there is greater than the sum of the two separate intensities. Averaged over the whole pattern the total energy is exactly that supplied by the two slits; interference simply moves energy from the dark fringes to the bright ones.

(b) The clamped blade, with its free end, is an oscillating system; once twanged it vibrates freely at its own natural frequency, which is heard as the pitch of the buzz. That natural frequency depends on the stiffness of the vibrating part and on its mass. Sliding more of the blade out beyond the bench lengthens the vibrating portion: a longer projecting blade bends more easily, so it is less stiff, and it also carries more mass. Both changes lower the natural frequency, and so the buzz is heard to deepen.

(c) A string fixed at both ends and vibrating in its third harmonic carries three half-wavelengths along its length, so:

$$\lambda = \frac{2L}{3} = \frac{2(0.90\text{m})}{3} = 0.60\text{m}$$

The speed of the waves on the wire then follows from $v = f\lambda$:

$$v = f\lambda = (200\text{Hz})(0.60\text{m}) = 120\text{m/s}$$

Question 33

(a) A stationary wave is formed by two identical waves travelling through the medium in opposite directions. At a node the two waves always arrive exactly out of step: whatever displacement one wave would give to that point, the other gives an equal and opposite displacement at the same instant, and this holds at every instant. By the principle of superposition the two displacements cancel completely and permanently, so the medium at a node never moves. Each wave on its own would indeed move that point, but their contributions there are equal and opposite at all times, and their sum is always zero.

(b) Striking the fork harder gives its prongs more energy and so a larger amplitude of vibration, which is heard as a louder sound. But the frequency at which the prongs vibrate, and it is the frequency that fixes the musical note, is the natural frequency of the fork, set by the stiffness and the mass of the prongs. This natural frequency does not depend on the amplitude; the oscillation is isochronous. So however hard or gently the fork is struck, it vibrates at the same frequency and sounds the same note, differing only in loudness.

(c) The spacing of the bright fringes in a Young's pattern is $y = \frac{\lambda D}{d}$, which rearranges to give the wavelength:

$$\lambda = \frac{yd}{D} = \frac{(2.1 \times 10^{-3}\text{m})(0.40 \times 10^{-3}\text{m})}{1.5\text{m}}$$

which gives:

$$\lambda = 5.6 \times 10^{-7}\text{m}$$

Question 34

(a) A prism spreads light by refraction. Glass has a slightly greater refractive index for violet light than for red, so violet light is slowed and bent the more; violet is deviated most and red least. A grating spreads light by diffraction and interference: the direction of a bright order is fixed by $d \sin \theta = m\lambda$, so the angle of deviation increases with wavelength. Red light, having the longer wavelength, satisfies the grating equation at a larger angle than violet, so with a grating red is deviated most and violet least. The two instruments order the colours oppositely because they spread the light by entirely different mechanisms.

(b) Pushing and pulling the end of the coil squeezes the nearby turns together, forming a compression, and then draws them apart, forming a rarefaction. Each disturbed turn pushes and pulls on the next, so the compression is handed on from turn to turn and a longitudinal wave travels to the far end. Each individual turn of the coil only moves a short way back and forth about its own position; it is the pattern of compression, together with the energy carried with it, that travels the length of the coil.

(c) The force constant measures the force needed for each unit of extension. A given pull stretches the whole spring twice as far as it stretches either half alone, so each half is twice as stiff as the whole spring:

$$k' = 2k = 2(18\text{N/m}) = 36\text{N/m}$$

The period of the 0.40kg mass on one half-spring is then:

$$T = 2\pi\sqrt{\frac{m}{k'}} = 2\pi\sqrt{\frac{0.40\text{kg}}{36\text{N/m}}} = 0.66\text{s}$$

Question 35

(a) When a mass hangs on a vertical spring, gravity stretches the spring to a new equilibrium position at which the spring's pull exactly balances the weight. If the mass is then displaced from this position, the extra force that restores it is supplied entirely by the change in the spring's tension, just as it is for a horizontal spring displaced from its own equilibrium. The constant weight is already balanced and plays no further part in the motion. The restoring force per unit displacement is therefore the force constant of the spring in both cases, so the period $T = 2\pi\sqrt{m/k}$ is the same; gravity only shifts the equilibrium position, it does not change the period.

(b) The mist consists of a great many tiny water droplets. Light from the street lamp passing the droplets is diffracted, spreading slightly as it passes each small droplet, and the diffracted light from the many droplets interferes. The angle through which the light is bent depends on its wavelength, so each colour is reinforced at a slightly different angle away from the lamp. The eye therefore sees the lamp ringed by faint bands in which the colours are separated, a corona produced by diffraction.

(c) The wavelength follows from $v = f\lambda$:

$$\lambda = \frac{v}{f} = \frac{8.0\text{m/s}}{20\text{Hz}} = 0.40\text{m}$$

The number of complete waves contained in a 12m length is that length divided by one wavelength:

$$n = \frac{12\text{m}}{0.40\text{m}} = 30$$

Question 36

(a) As a wave passes, each particle of the medium is set oscillating, and the energy handed to a particle is the energy of its oscillation. The energy of an oscillating particle is proportional to the square of its amplitude: its greatest kinetic energy is $\frac{1}{2}mv_{\text{max}}^2$, and with $v_{\text{max}} = \omega A$ this is proportional to A^2 . The wave passes this energy on from particle to particle, so the energy carried past any point each second, which is the power of the wave, is likewise proportional to the square of the amplitude. Doubling the amplitude therefore carries four times the energy, although the speed and frequency of the wave are unchanged.

(b) Light from each headlamp, on entering the eye, passes through the pupil, which is a small circular opening, and is diffracted there. Instead of forming a perfect point on the retina, each lamp forms a small spread-out patch of light. When the car is far away the two lamps are very close together in angle, so their two patches overlap and merge into what looks like a single light. As the car comes nearer the angle between the lamps grows; once it is large enough for the two patches to be seen apart, the lamps are made out as two separate lights.

(c) The total energy is $E = \frac{1}{2}kA^2$ and the potential energy at a displacement x is $\frac{1}{2}kx^2$, so $\frac{PE}{E} = \left(\frac{x}{A}\right)^2$. When the kinetic energy is three times the potential energy, the potential energy is one quarter of the total:

$$\left(\frac{x}{A}\right)^2 = \frac{1}{4}$$

$$x = \frac{A}{2} = \frac{0.10\text{m}}{2} = 0.050\text{m}$$

The kinetic energy is whatever part of the total is not potential, so the fraction that is kinetic is $1 - \left(\frac{x}{A}\right)^2$. At a displacement of 0.080m:

$$\frac{KE}{E} = 1 - \left(\frac{0.080\text{m}}{0.10\text{m}}\right)^2 = 0.36$$

Question 37

(a) The period of a mass on a spring is $T = 2\pi\sqrt{m/k}$. The period depends not on the mass itself but on its square root. Doubling the mass therefore multiplies the period not by 2 but by $\sqrt{2}$. The heavier mass does have more inertia and so is accelerated less by the same restoring force, which lengthens the period; but the period responds only to the square root of the mass, because it is $\sqrt{m}$, and not m , that appears in the formula. The period therefore rises by a factor of $\sqrt{2}$, about 1.4, and not by a factor of 2.

(b) A whip is not uniform: it tapers, growing thinner and lighter towards the tip. The speed of a transverse wave along it is $v = \sqrt{T/\mu}$, which is greater where the mass per unit length is smaller. As the wave runs towards the thin tip the mass per unit length steadily decreases, so the wave travels faster and faster. At the same time the energy carried by the wave is being delivered to less and less mass, so the amplitude of the motion grows, until the tip is flicked so violently that it produces the sharp crack.

(c) An ideal polaroid transmits one half of the intensity of unpolarised light, whatever its orientation, so after the first polaroid:

$$I_1 = \frac{I_0}{2} = \frac{16\text{W/m}^2}{2} = 8.0\text{W/m}^2$$

The light is now plane-polarised. By Malus's law the second polaroid, with its axis at 60° to the first, transmits:

$$I_2 = I_1(\cos 60^\circ)^2 = (8.0\text{W/m}^2)(0.50)^2 = 2.0\text{W/m}^2$$

Question 38

(a) A bright fringe is formed wherever the path difference between the light from the two slits is a whole number of wavelengths: $0, \lambda, 2\lambda, 3\lambda$ and so on. In passing from one bright fringe to the next, the path difference increases by exactly one wavelength. For a screen far from the slits the path difference changes in direct proportion to the distance moved along the screen, so equal steps of one wavelength in path difference correspond to equal steps of distance on the screen. The bright fringes are therefore evenly spaced, the spacing being $\frac{\lambda D}{d}$.

(b) The child and the spring of the pogo stick together form a mass-on-a-spring system. When the spring is compressed below the equilibrium position it pushes back harder than the child's weight, and the restoring force is proportional to how far the spring is compressed. A restoring force proportional to the displacement is the condition for simple harmonic motion, whose period $T =$

$2\pi\sqrt{m/k}$ is fixed by the mass of the child and the stiffness of the spring, and does not depend on how hard the child bounces. Each bounce therefore takes the same time, giving the steady rhythm.

(c) A particle of the string performs simple harmonic motion, so its maximum speed is $v_{\max} = \omega A$, with $\omega = 2\pi f$:

$$v_{\max} = 2\pi f A = 2\pi(12\text{Hz})(0.015\text{m}) = 1.1\text{m/s}$$

The wave itself travels along the string at 9.0m/s. The maximum speed of a particle, 1.1m/s, is therefore much smaller than the speed of the wave; the two are quite distinct quantities.

Question 39

(a) When a train of waves crosses obliquely into shallower water it travels more slowly there. The wavefronts meet the boundary at an angle, so one end of a wavefront enters the shallow water and is slowed while the other end, still in deeper water, continues at the higher speed. Each wavefront therefore swings round as it crosses, and the whole train changes its direction of travel. The frequency is unchanged, so since $v = f\lambda$ the slower waves also have a shorter wavelength; the change of direction is the result of the two ends of each wavefront advancing unequally.

(b) When a wave tilts the boat to one side, the upward push of the water, the buoyancy force, and the weight of the boat no longer act along the same line, and together they produce a turning effect that acts to right the boat. This restoring turning effect grows with the angle of tilt, so once the wave has passed the boat rolls to and fro about its upright position in an oscillation that is close to simple harmonic. The rhythm of the roll is the boat's own natural period, fixed by its shape and by the way its mass is distributed, not by the wave that started the rolling.

(c) For Newton's rings the radius of the n -th dark ring is given by $r_n^2 = nR\lambda$, which rearranges to give the wavelength:

$$\lambda = \frac{r_n^2}{nR} = \frac{(2.4 \times 10^{-3}\text{m})^2}{(8)(1.2\text{m})}$$

which gives:

$$\lambda = 6.0 \times 10^{-7}\text{m}$$

Question 40

(a) Sunlight passing through the atmosphere is scattered by the air molecules, which are very much smaller than the wavelength of light. For scattering by such small particles the amount of scattering is far greater for short wavelengths than for long ones, being inversely proportional to the fourth power of the wavelength (Rayleigh scattering). Blue light, having a shorter wavelength than red, is therefore scattered much more strongly. Light reaching the observer's eye from all directions of the sky is this scattered light, in which the blue is dominant, so the sky appears blue.

(b) As waves approach a shelving beach the water grows shallower towards the shore, and waves travel more slowly in shallower water. When a wavefront comes in at an angle, the end of it nearer the shore is in shallower water and moves more slowly than the end still in deeper water. The slow end is held back while the fast end advances, so the wavefront steadily swings round until it lies almost parallel to the shoreline. By the time the waves reach the beach they are travelling nearly straight in, whatever angle they had far out; this turning is the refraction of the waves.

(c) In one complete oscillation the particle travels from one extreme to the other and back again, covering the amplitude four times, so the total path length is $4A$:

$$A = \frac{24\text{cm}}{4} = 6.0\text{cm}$$

The angular frequency is $\omega = 2\pi/T$, and the maximum speed is $v_{\max} = \omega A$:

$$\omega = \frac{2\pi}{T} = \frac{2\pi}{0.40\text{s}} = 15.7\text{rad/s}$$

$$v_{\max} = \omega A = (15.7\text{rad/s})(0.060\text{m}) = 0.94\text{m/s}$$

and the maximum acceleration is $a_{\max} = \omega^2 A$:

$$a_{\max} = \omega^2 A = (15.7\text{rad/s})^2(0.060\text{m}) = 14.8\text{m/s}^2$$