11 SUPERPOSITION

H2 Physics 9478

Content **Page**

Learning Outcomes

Candidates should be able to:

(a) explain and use the principle of superposition in simple applications

(b) show an understanding of experiments which demonstrate standing (stationary) waves using microwaves, stretched strings and air columns.

(c) explain the formation of a standing (stationary) wave using a graphical method, and identify nodes and antinodes, differentiating between pressure and displacement nodes and antinodes for sound waves.

(d) determine the wavelength of sound using standing (stationary) waves.

(e) show an understanding of the terms interference, coherence, phase difference and path difference.

(f) show an understanding of phenomena which demonstrate two-source interference using water waves, sound waves, light waves and microwaves.

(g) show an understanding of the conditions required for two-source interference fringes to be observed.

(h) recall and use the equation $ax/D = \lambda$ to solve problems for double-slit interference, where *a* is the slit separation and *x* is the fringe separation.

(i) recall and use the equation $a\sin\theta = n\lambda$ to solve problems involving the principal maxima of a diffraction grating, where *a* is the slit separation.

(j) describe the use of a diffraction grating to determine the wavelength of light (knowledge of the structure and use of a spectrometer is not required).

(k) show an understanding of phenomena which demonstrate diffraction through a single slit or aperture, or across an edge, such as the diffraction of water waves in a ripple tank with both a wide gap and a narrow gap, or the diffraction of sound waves from loudspeakers or around corners.

(l) recall and use the equation $b\sin\theta = \lambda$ to solve problems involving the positions of the first minima for diffraction through a single slit of width *b*.

(m) recall and use the Rayleigh criterion $\theta \approx \lambda/b$ for the resolving power of a single aperture.

11.1 Principle of Superposition

❖ Introduction

A wave is a disturbance that travels through a medium or vacuum. For mechanical waves, like sound and water waves, the disturbance refers to the displacement of the particles of the media from their equilibrium position. For electromagnetic waves, the disturbance refers to the varying electric and magnetic field.

What happens when two waves of the same type (e.g., sound waves from two sources) meet at a point in space?

Principle of Superposition

The **principle of superposition** states that when two or more waves of the same type meet at a point in space, the **resultant displacement** at that point is equal to the **vector sum of the displacements of the individual waves** at that point.

Fig. 11.1 illustrates two different sets of two waves meeting at an instant of time.

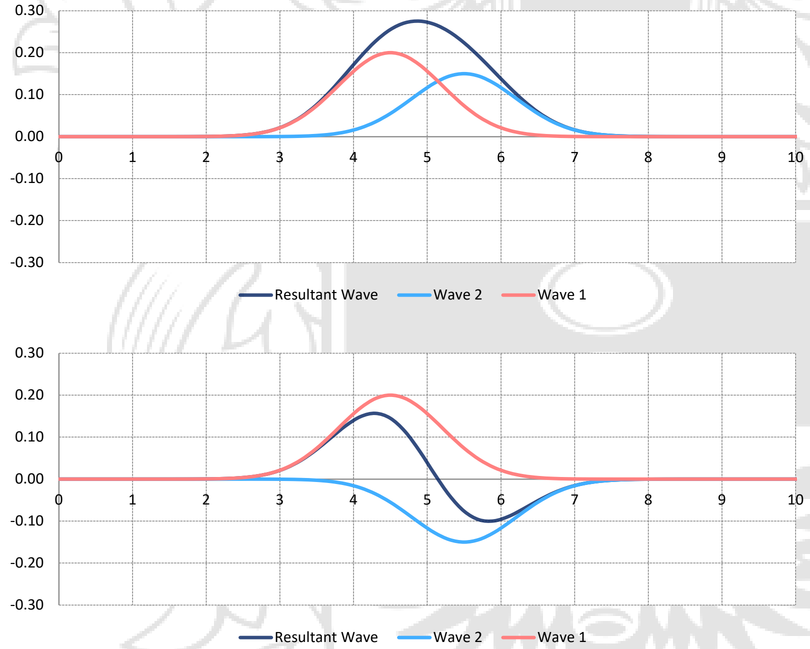


Fig. 11.1

Note that displacement of the resultant wave at any position is the vector sum of the displacements due to the two waves.

11.2 Standing Waves

❖ Formation of a Standing Wave

Standing Waves

A **standing wave** is formed by the superposition of two progressive waves of the **same type**, **amplitude**, **frequency** and **speed**, travelling along the same path but in **opposite directions**.

Note:

Since the waves have the same frequency and speed, they also have the same wavelength.

A standing wave is also known as a **stationary wave**.

Fig. 11.2 shows a set of graphs representing two progressive waves of equal amplitude and frequency travelling in opposite directions.

The superposition of the two waves is observed along the line of propagation, giving rise to a resultant waveform.

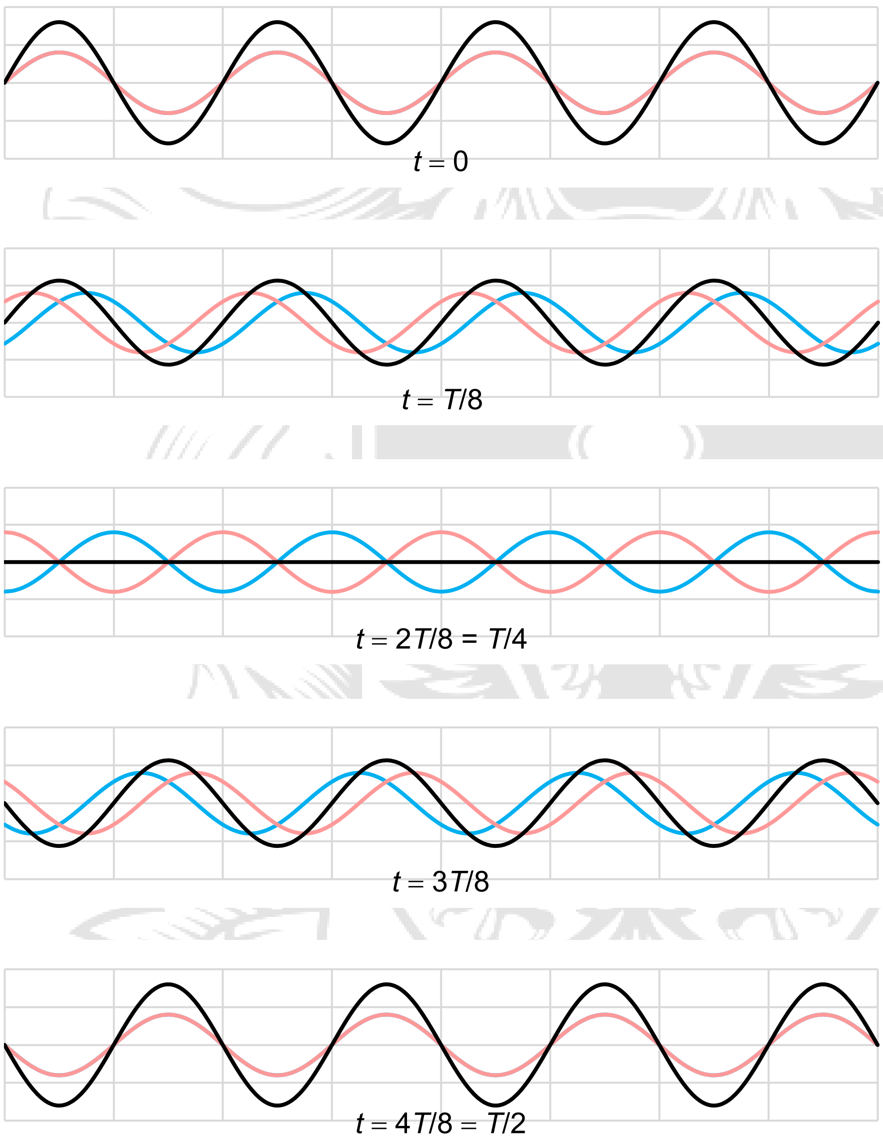


Fig. 11.2

At $t = 0$, superposition of the waves gives rise to a resultant wave which has twice the amplitude of either progressive wave

At $t = \frac{1}{8}T$, the waves have moved $\lambda/8$ in opposite directions.

Resultant wave has lower maximum amplitude.

At $t = \frac{1}{4}T$, the waves have moved $2\lambda/8 = \lambda/4$ in opposite directions.

The resultant amplitude is zero everywhere.

At $t = \frac{3}{8}T$, the waves have moved $3\lambda/8$ in opposite directions.

Resultant wave has lower maximum amplitude.

At $t = \frac{1}{2}T$, the resultant wave has twice the amplitude of either progressive wave once again.

This continues to $t = \frac{3}{4}T$, and resultant displacement is zero everywhere. Finally, at exactly one period, when $t = T$, the resultant wave has twice the amplitude of either progressive wave once again.

The graphical representation of a standing wave is shown in Fig. 11.3.

The solid and dashed lines represent both extremes when the constituent waves are in phase.

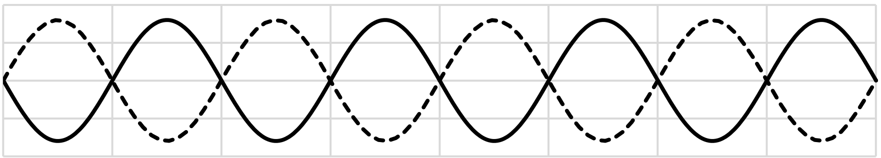


Fig. 11.3

Each loop is described as an **envelope** where the resultant displacement varies rapidly and is outlined by curves that represents the amplitudes of the wave at each position.

❖ **Properties of a Standing Wave**

1. The wave profile does not propagate. As such, the resultant wave is known as a standing wave or stationary wave.
2. The particles of the wave oscillate (except those at the nodes) about their respective equilibrium positions with the **same frequency**, but **different amplitudes**.

   The frequency is the same as that of the two component waves.
3. An **antinode** is a point in a standing wave where the amplitude is the maximum.

   The component waves always arrive in phase at the antinodes.
4. A **node** is a point in a standing wave where the amplitude is zero.

   The component waves always arrive anti-phase at the nodes.
5. Between two adjacent nodes, all particles oscillate **in phase**, i.e., they reach their respective maxima, minima and equilibrium positions at the same instant.

   Note that these particles do not have the same amplitude.
6. Distance between two adjacent nodes (or antinodes) is ½λ.

   Particles in neighbouring segments vibrate 180° (or π rad) out of phase with each other.

Antinode

An **antinode** is a point in a standing wave where the amplitude is the **maximum**.

Node

A **node** is a point in a standing wave where the amplitude is **zero**.

❖ Comparisons of Standing vs Progressive Waves

	Standing Wave	Progressive Wave
Wave profile	The wave profile **does not move**.	The wave profile **moves** in the direction of propagation with the velocity of the wave.
Energy	**No net energy** is transported.	**Energy** is transported in the direction of propagation.
Amplitude	Amplitude varies from **zero at the nodes** to **maximum at the antinodes**.	Every particle along the wave oscillates with the **same amplitude**.
Frequency	All particles **except the nodes** oscillate in simple harmonic motion with the same frequency as the progressive wave.	**All** particles oscillate in simple harmonic motion with the **same frequency** as the progressive wave.
Phase	All particles between adjacent nodes oscillate **in phase**. Particles between adjacent segments oscillate **in anti-phase**.	Particles within one wavelength oscillate with **different phases**.
Wavelength	Distance between **adjacent nodes** or **adjacent antinodes** is $\frac{1}{2}\lambda$.	Distance between two **adjacent particles** oscillating **in phase** is λ.

Example 1 [J90/1/13]

Progressive waves of frequency 300 Hz are superposed to produce a system of stationary waves in which adjacent nodes are 1.5 m apart.

Calculate the speed of the progressive waves.

[Solution]

Since wavelength of progressive waves is twice the internodal distance,

$\lambda =$

Speed

$v =$

❖ Standing Waves along Strings and in Pipes

Standing waves can be set up along a stretched wire or in a pipe.

When a string on a guitar is plucked, two wave trains are sent to the opposite fixed ends and reflected. The reflected wave trains superpose to form standing waves along the string.

At these frequencies, the amplitude of vibration is large, and the string is said to be undergoing **resonance**.

The vibrations of the string energize the soundboard and in turn the air inside the body. The sound is then projected through the cavity in the body.

In the case of a pipe, air blown across the opening of a tube sends longitudinal sound wave along the tube. The sound is then reflected from the other end. The superposition of the reflected sound wave and incident sound wave produces standing waves along the pipe. This is how music is produced in wind instruments.

The vibrations set up on a string or in a tube typically comprise of standing waves of different frequencies.

The standing wave with the lowest frequency is said to be vibrating with the **fundamental frequency** or the **1st harmonic**. Higher modes of vibration are known as the **overtones**.

Music & Resonance

❖ Standing Waves along a String

Consider a string that is stretched and fixed at the two ends. The **fixed ends** ensure that **nodes** will always be produced at these points.

This is a boundary condition that is necessary for the formation of a standing wave in a string.

The simplest way which the string can vibrate is shown in Fig. 11.4, where the wave pattern is a single loop. This is called the fundamental mode of vibration and the frequency at which the standing wave is vibrating at is known as the fundamental frequency.

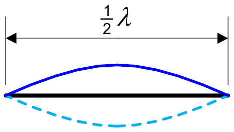


Fig. 11.4

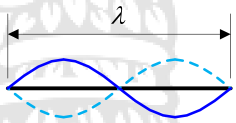


Fig. 11.5

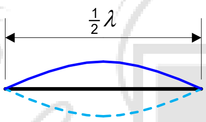
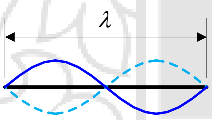
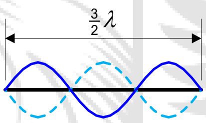
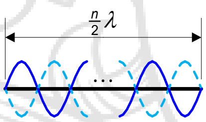
Standing wave along a string demonstration

If the frequency of vibration is gradually increase, it will next resonate at a particular frequency, with its **second mode of vibration**, whereby the wave pattern has two loops as shown in Fig. 11.5.

This is also known as the 1st overtone or 2nd harmonic.

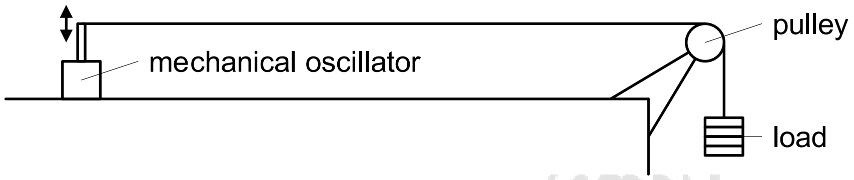
As the frequency of the vibration continues to increase, the string will continue to exhibit higher modes of vibration.

In reality, the presence of different harmonics gives the timbre or characteristics of an instrument.

Mode of vibration	Graphical representation	Wavelength	Frequency	Also known as
Fundamental frequency		$\lambda = 2L$	$f = \dfrac{v}{2L}$	1st harmonic
1st overtone		$\lambda = L$	$f = \mathbf{2}\left(\dfrac{v}{2L}\right)$	2nd harmonic
2nd overtone		$\lambda = \dfrac{2L}{3}$	$f = \mathbf{3}\left(\dfrac{v}{2L}\right)$	3rd harmonic
$(n-1)^{\text{th}}$ overtone		$\lambda = \dfrac{2L}{n}$	$f = \boldsymbol{n}\left(\dfrac{v}{2L}\right)$	n^{th} harmonic

Example 2 [N95/3/3(b)]

In order to investigate stationary waves on a stretched string, a student sets up the apparatus shown below.



(a) State what is meant by a node. Explain why a node must exist at the pulley and at the oscillator.

(b) Explain why it is necessary to adjust either the length of the string or the frequency of the oscillator in order to obtain observable stationary waves on the string.

The end of the string that is attached to the oscillator is also considered as a node because its amplitude of oscillations is small compared to that of the antinode.

[Solution]

(a) A node is a point on the stationary wave where the particle is always at rest.

A node must exist at the pulley because the string at the pulley ____________________.

A node must exist (approximately) at the oscillator because the amplitude of vibration of the oscillator is small. So, the ____________________ of the string at the oscillator is ________ compared to the antinode.

(b) The speed of the wave on the string is __________ as the tension in the string is constant. To form observable stationary waves, the length of the string must be ____ of the wave (i.e., when resonance occurs) as ______________________________ of the string.

$$L = n \times \tfrac{1}{2}\lambda = n \times \tfrac{1}{2}\frac{v}{f} \qquad \text{where } n = 1, 2, 3\ldots$$

We can achieve this condition for a fixed frequency f by ____________________, or for a fixed length L by ____________________.

The speed v of a wave in a string is given by

$$v = \sqrt{T/\mu}$$

where T is the tension of the string and μ is the mass per unit length of the string.

Example 3 [J83/2/12]

A taut wire is clamped at two points 1.0 m apart and plucked near one end.

Calculate the three longest wavelengths present on the vibrating wire.

[Solution]

The 3 longest wavelengths are obtained in the fundamental mode (1st harmonic), 2nd harmonic and the 3rd harmonic.

1st harmonic: $\lambda_1 =$

2nd harmonic: $\lambda_2 =$

3rd harmonic: $\lambda_3 =$

Notice that the factor $2/n$ where n is the order of harmonics.

❖ Standing Waves in an Open Pipe

A pipe is termed **open** (Fig. 11.6) if **both ends of the pipe are open**.

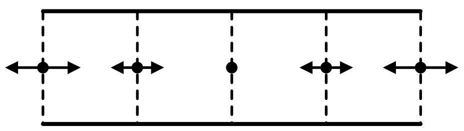


Fig. 11.6 Open pipe

When a sound wave is sent into an **open pipe**, the wave propagates to the end of the pipe. Part of the wave is reflected off the surrounding air outside the pipe. The reflected wave superposes with the incident wave and a standing wave is formed.

As the air molecules can move freely at both ends, displacement antinodes are formed at both ends.

The dashed lines indicate the equilibrium positions of the air molecules, and the lengths of the arrows indicate the amplitude of their vibrations.

Graphically, the standing wave in an open pipe can also be represented by an amplitude-position graph. Fig. 11.7 shows the variation with position x, of amplitude y of the air molecules. The dotted line represents the other amplitude positions of the air molecules. These two lines form the envelope of the standing wave.

A pressure node is formed at the open ends so that the pressure there is atmospheric.

This results in displacement antinodes at the open ends.

This is a boundary condition that is necessary for the formation of a standing wave in an open pipe.

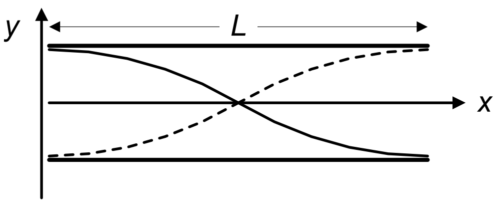
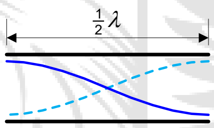
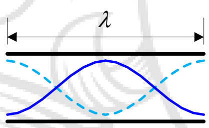
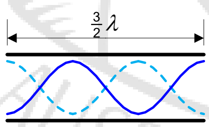
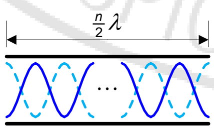


Fig. 11.7

Fig. 11.7 only shows the amplitude of vibration of the air molecules, but it does not mean that sound wave is a transverse wave!

Mode of vibration	Graphical representation	Wavelength	Frequency	Also known as
Fundamental frequency		$\lambda = 2L$	$f = \dfrac{v}{2L}$	1st harmonic
1st overtone		$\lambda = L$	$f = 2\left(\dfrac{v}{2L}\right)$	2nd harmonic
2nd overtone		$\lambda = \dfrac{2L}{3}$	$f = 3\left(\dfrac{v}{2L}\right)$	3rd harmonic
$(n-1)^{th}$ overtone		$\lambda = \dfrac{2L}{n}$	$f = n\left(\dfrac{v}{2L}\right)$	n^{th} harmonic

❖ Standing Waves in a Closed Pipe

A pipe is termed **closed** (Fig. 11.8) if **one end of the pipe is closed** while the other is open.

Sound Resonance Tube (calculation of the speed of sound)

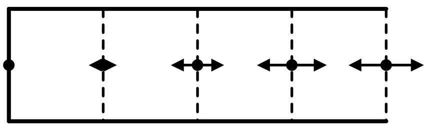


Fig. 11.8 Closed pipe

When a sound wave is sent into a **closed pipe**, the wave propagates to the end of the pipe and is reflected. The reflected wave superposes with the incident wave and a standing wave is formed. A displacement node is formed at the closed end of the pipe, while a displacement antinode is formed at the open end.

This is a boundary condition that is necessary for the formation of a standing wave in a closed pipe.

The dashed lines indicate the equilibrium positions of the air molecules, and the lengths of the arrows indicate the amplitude of their vibrations.

Graphically, the standing wave in a closed pipe can be represented by an amplitude-position graph. Fig. 11.9 shows the variation with position *x*, of **amplitude** *y* of the air molecules. The dotted line represents the other amplitude positions of the air molecules. These two lines form the envelope of the standing wave.

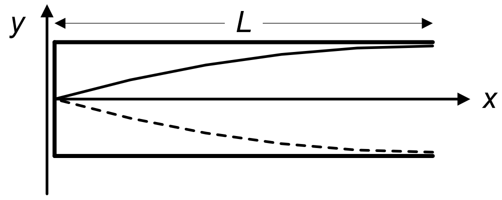
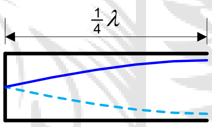
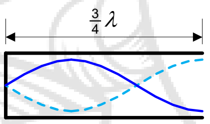
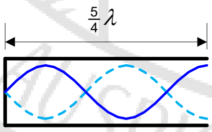
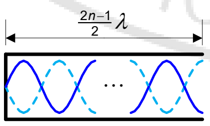


Fig. 11.9

Mode of vibration	Graphical representation	Wavelength	Frequency	Also known as
Fundamental frequency		$\lambda = 4L$	$f = \dfrac{v}{4L}$	1st harmonic
1st overtone		$\lambda = \dfrac{4L}{3}$	$f = 3\left(\dfrac{v}{4L}\right)$	3rd harmonic
2nd overtone		$\lambda = \dfrac{4L}{5}$	$f = 5\left(\dfrac{v}{4L}\right)$	5th harmonic
$(n-1)^{th}$ overtone		$\lambda = \dfrac{4L}{(2n-1)}$	$f = (\mathbf{2n-1})\left(\dfrac{v}{4L}\right)$	$(2n-1)^{th}$ harmonic

❖ End Correction

The displacement antinodes at the open ends of pipes are actually located slightly outside the pipes. This results in a small **end correction** c to be included in the calculation of the resonance frequencies.

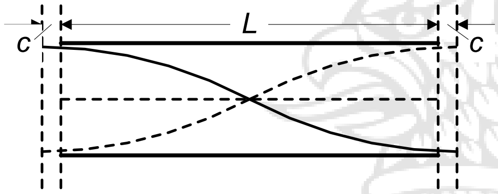


Fig. 11.10

For open pipes (Fig. 11.10), there are two open ends and the effective length is

$$L_{\text{eff}} = L + 2c$$

The frequencies of the higher modes of vibrations are

$$f_n = \frac{nv}{2(L+2c)}$$

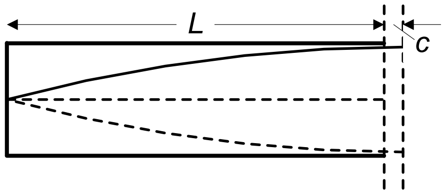


Fig. 11.11

For closed pipes (Fig. 11.11), there is one open end and the effective length is

$$L_{\text{eff}} = L + c$$

The frequencies of the higher modes of vibrations are

$$f_n = \frac{(2n-1)v}{4(L+c)}$$

Example 4 [N76/2/3]

An organ pipe, 0.33 m long, is open at one end and closed at the other. The speed of sound in air is 330 m s^{-1}.

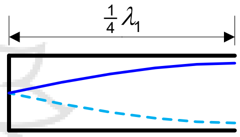
Assuming that end corrections are negligible, calculate

(a) the frequencies of the fundamental and the first overtone,

(b) the length of a pipe which is open at both ends and which has a fundamental frequency equal to the difference of those calculated in **(a)**.

[Solution]

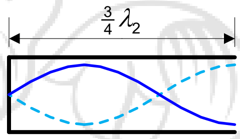
(a) For fundamental mode (1st harmonic) of closed pipe



$\lambda_1 =$

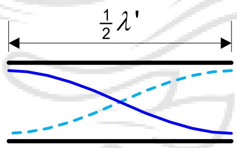
$f_1 = \dfrac{v}{\lambda} =$

For 1st overtone (3rd harmonic) of closed pipe



$f_2 =$

(b) For fundamental frequency of open pipe



$f =$

Length of open pipe

$L =$

Example 5

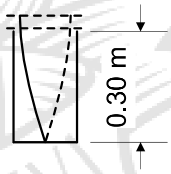
A source of sound of frequency 250 Hz is used with a resonance tube, closed at one end, to measure the speed of sound in air. Strong resonance is first obtained at tube lengths of 0.31 m and then 0.96 m. Calculate

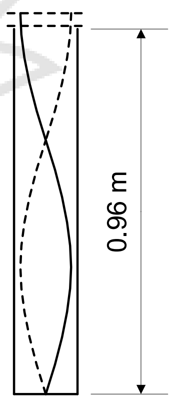
(a) the speed of sound in air,

(b) the end correction of the tube.

[Solution]

(a) From diagrams, internodal distance is the difference in tube lengths





$\frac{1}{2}\lambda =$

$v = f\lambda =$

(b) At first position of resonance

$\frac{1}{4}\lambda =$

$c =$

❖ Visualising Longitudinal Standing Waves

Fig. 11.12 shows the actual positions of the particles in a longitudinal standing wave. Dark regions represent areas of **compression (high pressure)**, while light regions represent areas of **rarefaction (low pressure)**.

Notice that both the **maximum compressions** and **maximum rarefactions** occur at the displacement nodes. Hence, **displacement nodes** are also **pressure antinodes**, and vice versa.

The positions of particles at the displacement antinodes are represented by light grey dots.

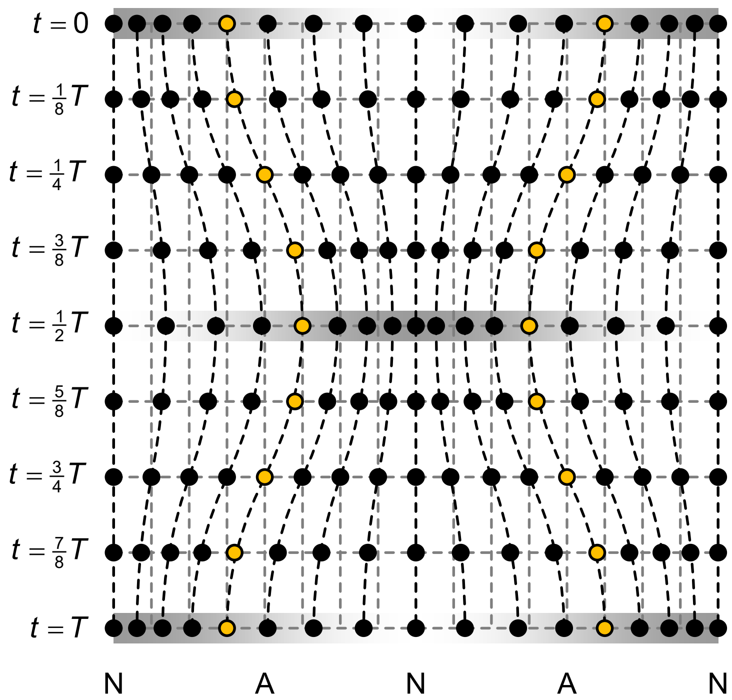


Fig. 11.12 Positions of particles at various times.

Fig. 11.13 represents the displacement of the particles and the excess pressure at $t = 0$.

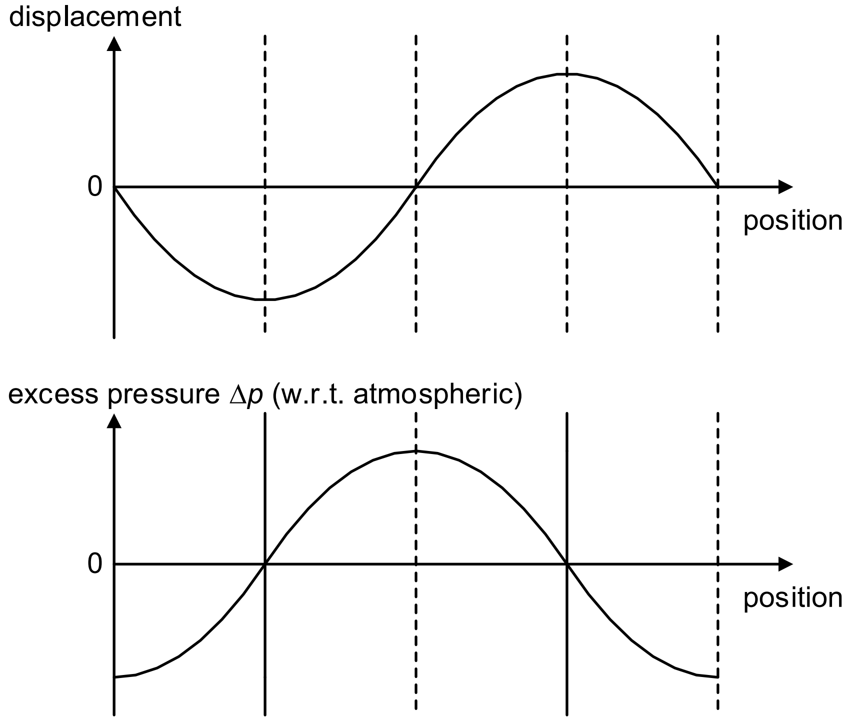


Fig. 11.13

Pressure nodes occurs at displacement antinodes while pressure antinodes occurs at displacement nodes.

11.3 Diffraction Through a Finite-size Gap

Diffraction

Diffraction is the bending of waves after passing through an aperture or round an obstacle.

An example of diffraction is the ability to hear sounds around a corner from where they were generated.

Diffraction causes waves to deviate from a straight trajectory and reach areas that would otherwise remain obstructed.

❖ **Relative size of Wavelength and Aperture**

The degree of diffraction depends on the relative size of the wavelength and the aperture (Fig. 11.14 and Fig. 11.15).

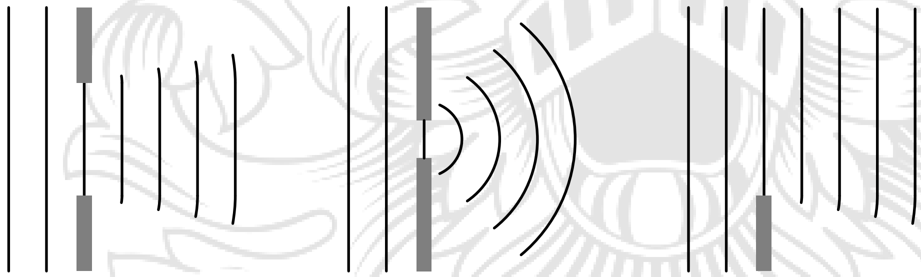


Fig. 11.14 constant wavelength with differently sized apertures

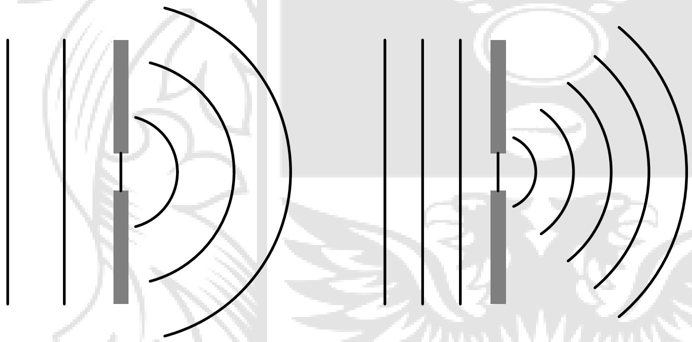


Fig. 11.15 differing wavelengths with constant aperture

Generally, diffraction is **pronounced** when the **wavelength** of the wave is of the **same order of magnitude** as the width of the **aperture**.

This is the reason why under normal circumstances, we do not observe any diffraction of light because the holes and apertures that we come across everyday are much larger than the wavelengths of light.

11.4 Interference of Two or More Coherent Sources

❖ Coherent Waves or Sources

Coherent waves are waves that have **constant phase difference**.

Coherent waves must have the same frequency and wavelength but waves that have the same frequency and wavelength **may not be coherent**.

Coherent waves must have the same speed since $v = f\lambda$.

Sources that emit coherent waves are known as **coherent sources**.

Fig. 11.16 shows two sound waves emitted by two speakers driven by the same signal generator. These sounds waves have a constant phase difference of $\pi/2$ and are **coherent** despite having unequal amplitudes.

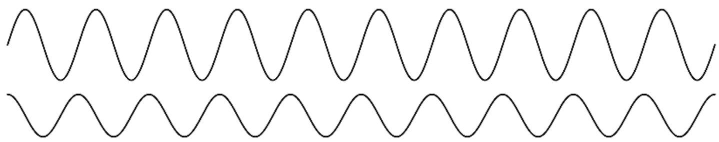


Fig. 11.16 Two sound waves (different amplitudes) with a phase difference of $\pi/2$ is coherent.

Fig. 11.17 shows two light waves emitted by two lasers. The light wave emitted by a laser suffers frequent random phase change at different instants. As such, these light waves are **not coherent**.

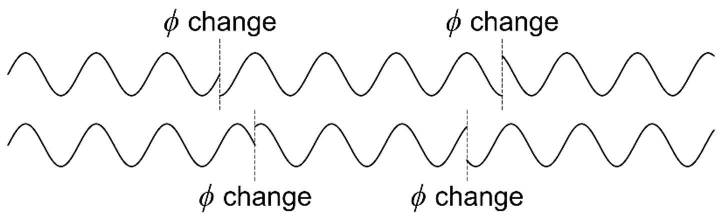


Fig. 11.17 Two monochromatic light waves (with frequent phase change) emitted by two lasers are not coherent

In general, there are two ways to produce two coherent light waves:

- splitting a light wave into two using partial reflecting mirrors (division of amplitude).
- splitting the wavefront of a light wave into two using multiple slits (division of wavefront).

Examples of Coherent Sources	Examples of Incoherent Sources
Two speakers driven by the same signal generator	A single light bulb (due to large filament size)
Diffracted light waves after a laser beam passes through two slits	Light waves from two lasers

Coherent (*adjective*)

Waves are **coherent** when they have a **constant phase difference**.

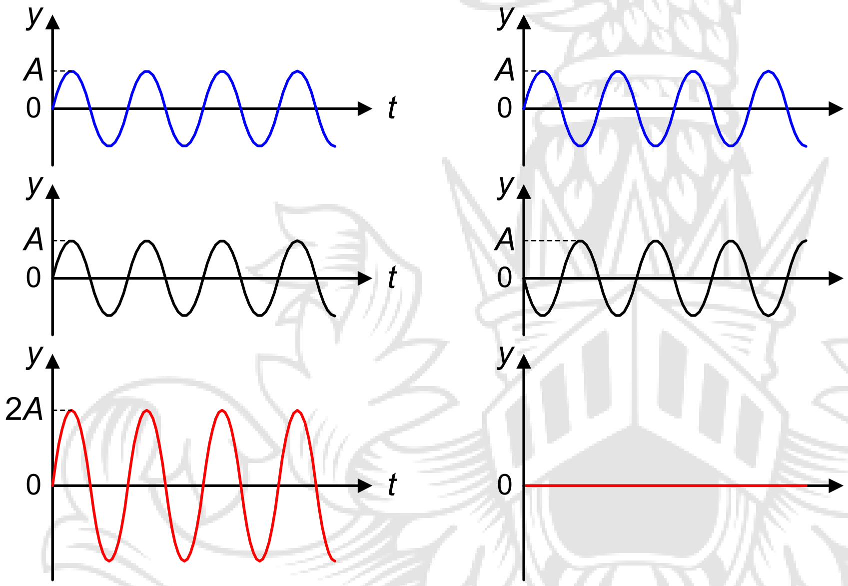
Coherence (*noun*)

Waves exhibit **coherence** when they have a **constant phase difference.**

❖ Interference

Interference is the **superposition** of two or more coherent waves to give a resultant wave whose amplitude is given by the **Principle of Superposition**.

The amplitude at any point is the largest when the waves interfere **constructively**, and the smallest when they interfere **destructively**.



- **Constructive interference** occurs at a point when two waves arrived at the same point **in phase**.
- Resultant amplitude is $2A$.

- **Destructive interference** occurs at a point when two waves arrived at the same point with a **phase difference of π rad**.
- Resultant amplitude is **zero**.

Fig. 11.18

❖ Conditions for Observable Interference

1. The waves or sources must be **coherent** (i.e., constant phase difference).
2. The waves must have **similar amplitude** (for a better contrast).
3. For electromagnetic waves, they must be **unpolarised** or **polarised** in the same plane.
4. The waves of the **same type** must **overlap** (to produce regions of constructive and destructive interference).

When two or more waves of the same type meet, superposition will happen. Whether or not a pattern is easily observed will depend on these factors.

> **Interference**
>
> Interference is the **superposition** of two or more coherent waves to give a resultant wave whose **amplitude** is given by the **Principle of Superposition**.

❖ The Ripple Tank Experiment

Fig. 11.19 shows two ball-ended dippers S_1 and S_2, attached to a mechanical oscillator (and hence are coherent sources), send out two sets of circular wavefronts (let's say crests). These waves interfere when they overlap.

Two-Source Interference

Falstad Ripple Tank

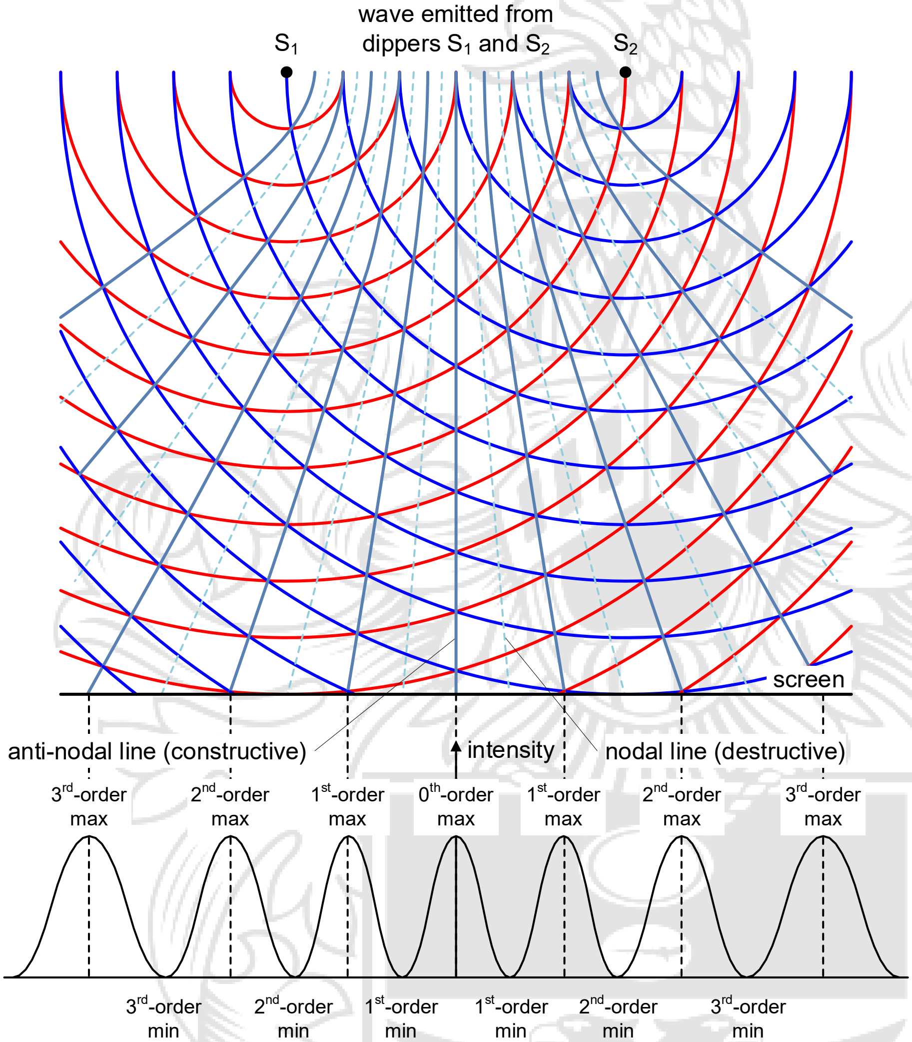


Fig. 11.19

By the principle of superposition, when two waves arrive **in phase** at points along the **anti-nodal lines** (solid lines), and **constructive interference** takes place, and the intensity is **maximum**.

In between the anti-nodal lines, the two waves arrive π rad **out of phase** along the **nodal lines** (dotted lines), and **destructive interference** takes place, and the intensity is **minimum**.

Note that the distance between adjacent wavefronts is λ.

❖ Path Difference

The path difference between two waves is the difference in the distance that each wave travels from its sources to the point where they meet.

Consider waves that are emitted **in phase** by sources S_1 and S_2.

The points P, Q, R, S indicate positions where two waves from S_1 and S_2 can possibly meet (Fig. 11.20).

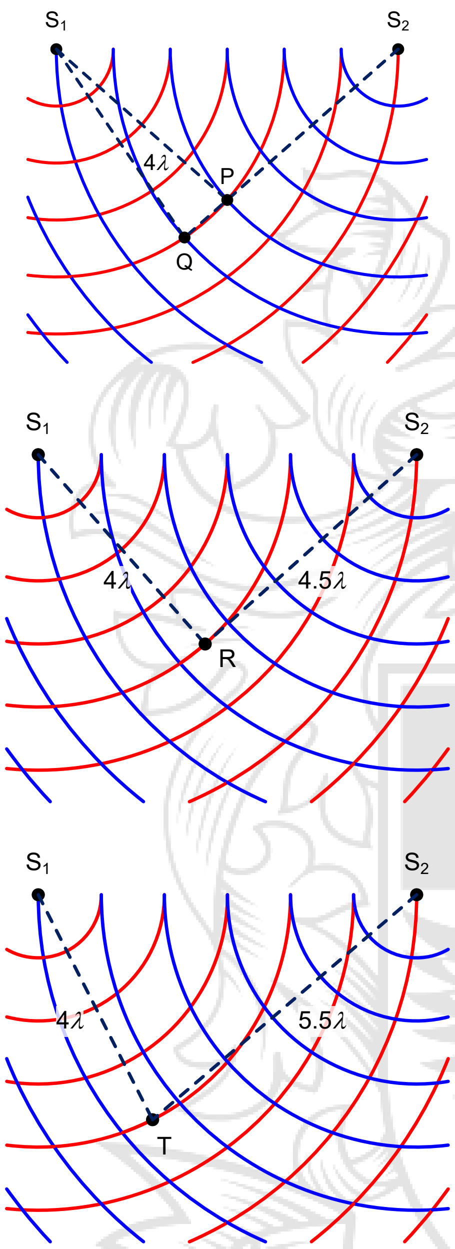


Fig. 11.20

Path difference at P

$$\Delta = |S_2P - S_1P| = 0$$

Phase difference: 0 rad

Path difference at Q

$$\Delta = |S_2Q - S_1Q| = \lambda$$

Phase difference: 2π rad $\equiv 0$ rad

Path difference at R

$$\Delta = |S_2R - S_1R| = 0.5\lambda$$

Phase difference: π rad

Path difference at T

$$\Delta = |S_2T - S_1T| = 1.5\lambda$$

Phase difference: 3π rad $\equiv \pi$ rad

Phase difference is proportional to path difference.

As such, path difference can be used to analyse whether two waves meet in phase or out of phase and subsequently, whether constructive or destructive interference results.

The phase difference between the waves arriving at a point depends on

1. the phase difference of the sources, and
2. the path difference between the waves arriving at that point.

Path Difference

The path difference between two waves is the difference in the distance that each wave travels from its source to the point where they meet.

Recall: **Phase difference** between two particles in a wave or between two waves at a point is a measure of the fraction of a cycle which one is ahead of the other.

For two sources that are **in phase**,

Constructive interference occurs when

path difference $\Delta = n\lambda$ (where $n = 0,\ 1,\ 2, \ldots$)

The n^{th} order maximum (e.g., bright fringe or loud sound) occurs at positions where the path difference is $n\lambda$.

Destructive interference occurs when

path difference $\Delta = \left(n + \frac{1}{2}\right)\lambda$ (where $n = 0,\ 1,\ 2, \ldots$)

The $(n+1)^{\text{th}}$ order minimum (e.g., dark fringe or soft sound) occurs at positions where the path difference is $\left(n - \frac{1}{2}\right)\lambda$.

In Fig. 11.20, constructive interference occurs at P and Q while destructive interference occurs at point R and T.

For two sources that are in **antiphase**,

Constructive interference occurs when

path difference $\Delta = \left(n + \frac{1}{2}\right)\lambda$ (where $n = 0,\ 1,\ 2, \ldots$)

Destructive interference occurs when

path difference $\Delta = n\lambda$ (where $n = 0,\ 1,\ 2, \ldots$)

❖ Interference with One Source

Interference can also occur with a single source in the presence of a reflector.

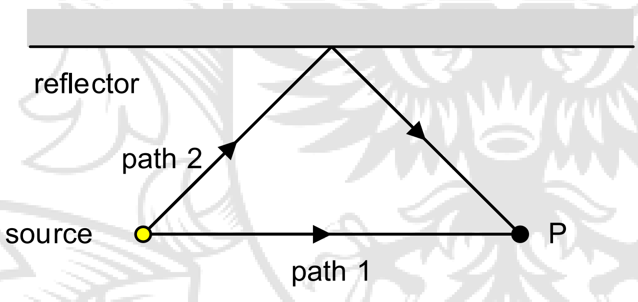


Fig. 11.21

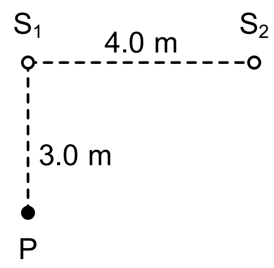
If two waves (e.g. incident wave and reflected wave) from the same sources arrive at a point via two different paths such that the path difference is **integer multiple of one wavelength**, then these two waves meet **in phase** and **constructive interference** occurs.

Assume there is no phase change along any path

If two waves from the same sources arrive at a point via two different paths such that the path difference is **half-integer multiple of one wavelength**, then these two waves meet in **antiphase** and **destructive interference** occurs (assuming there is no phase change along any path).

Example 6 [N88/1/8]

Two wave-generators S_1 and S_2 produce water waves of wavelength 1.0 m. They are placed 4.0 m apart in a water tank and a detector P is placed on the water surface 3.0 m from S_1 as shown in the diagram.



When operated alone, each generator produces a wave at P which has amplitude *A*.

When the generators are operating together and in phase, determine the resultant amplitude at P.

[Solution]

Path length

$S_2P =$

Path difference

$\Delta = S_2P - S_1P =$

Since $\Delta = 2\lambda$ and that the waves are ________________, _____________ interference occurs at P.

Hence, the resultant amplitude is ____.

❖ Young's Double-Slit Experiment

The Young's double-slit experiment (Fig. 11.22) is an experimental setup for viewing two-source interference pattern with light.

A monochromatic light source is placed behind a single slit to create a small, well-defined point source of light. Because light from the two slits originate from the same single slit, they are **coherent** and create a sustained and observable interference pattern.

At points of constructive interference (maxima), bright fringes are observed. At points of destructive interference (minima), dark fringes are observed.

This is conceptually similar to the ripple tank experiment.

Young's Double-Slit Interference

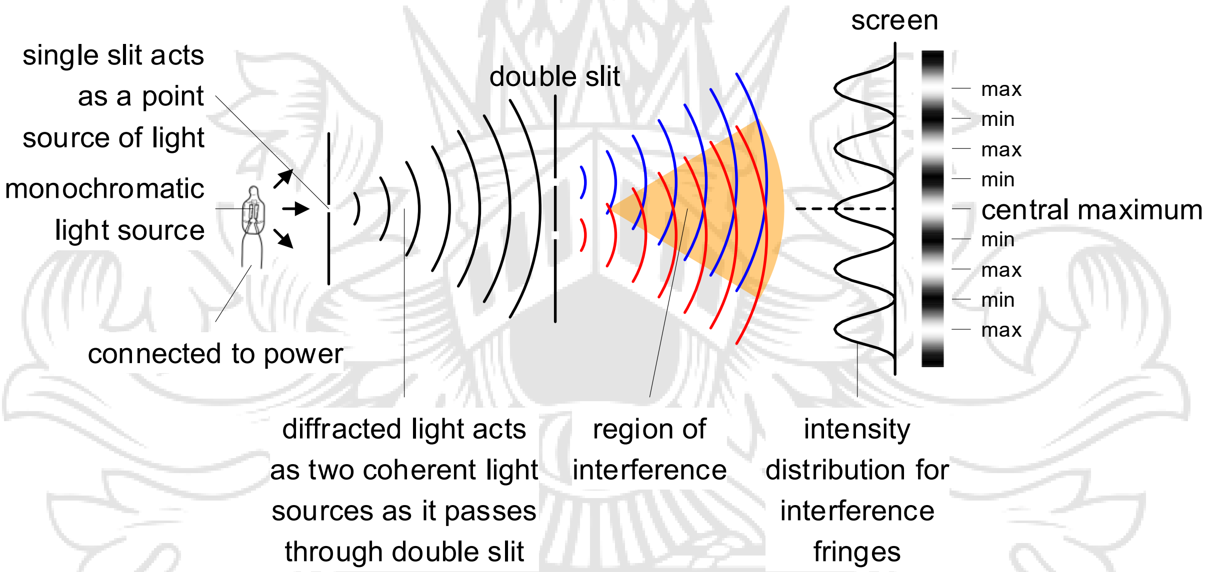


Fig. 11.22

❖ Derivation of Equation for Fringe Separation

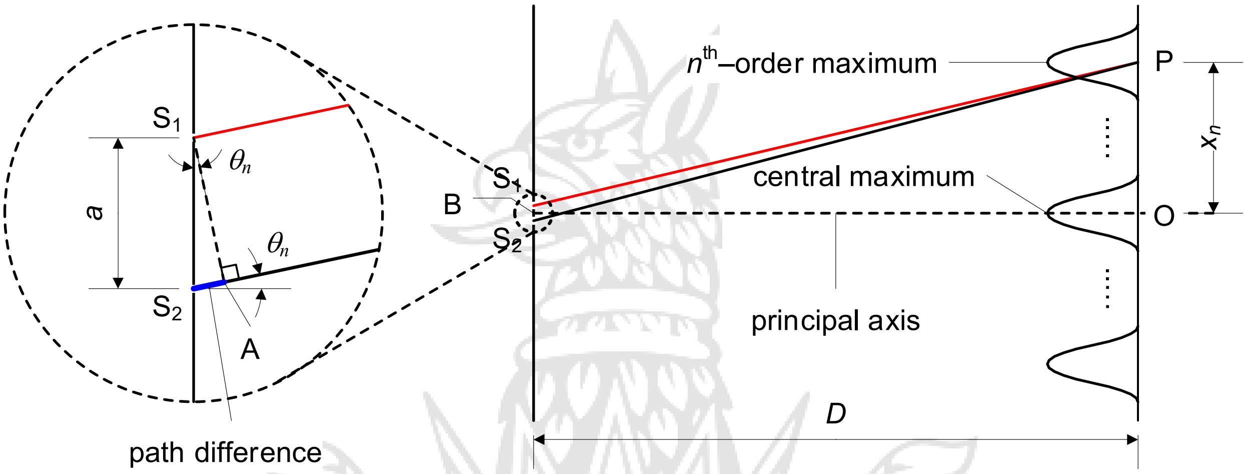


Fig. 11.23

Consider at P where n^{th}-order bright fringe (maximum) is obtained.

The distance between the two sources and the screen should be a lot larger than the slit separation ($D \gg a$), the rays from S_1 and S_2 will be parallel and $\angle A \simeq 90°$.

$$\text{path difference at T} = S_2P - S_1P = S_2A = a\sin\theta_n$$

Since the waves passing through slits S_1 and S_2 are in phase, for constructive interference,

$$a\sin\theta_n = n\lambda \qquad \text{where } n = 0, 1, 2, \ldots$$

From $\triangle OBP$, $\tan\theta_n = \dfrac{x_n}{D}$

For $a \gg \lambda$ (typically $a = 0.5$ mm and $\lambda = 600$ nm), θ_n is small.

$$\sin\theta_n \simeq \tan\theta_n$$

That is,

$$\frac{n\lambda}{a} = \frac{x_n}{D}$$

Notations for Fig. 11.23

x_n is the distance of n^{th} bright fringe from central fringe.

a is the slit separation.

D is the distance between slits and screen.

For $a = 0.5$ mm and $\lambda = 500$ nm,

$$\sin\theta_n = 0.001n$$

For **Two-Source Interference**,

Constructive interference takes place at these positions

$$x_n = \frac{n\lambda D}{a} \qquad \text{where } n = 0,\ 1,\ 2, \ldots$$

Fringe separation (spacing between adjacent bright fringes)

$$x = \frac{\lambda D}{a}$$

Proof:

$$\begin{aligned} x &= x_n - x_{n-1} \\ &= \frac{n\lambda D}{a} - \frac{(n-1)\lambda D}{a} \\ &= \frac{\lambda D}{a} \end{aligned}$$

❖ Intensity Distribution of Interference Fringes

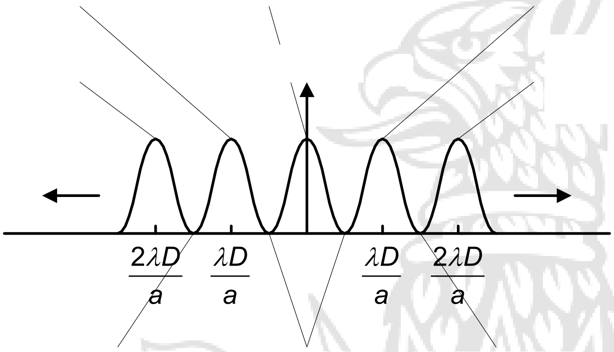


Fig. 11.24

The intensity of the interference fringes is assumed to be constant as the waves from each slit have the same amplitude *A*. Hence, all maxima have a resultant amplitude of 2*A*.

In reality, the intensity is **modulated** by the **single slit diffraction envelope** which decreases with increasing order.

This will be explained in more detail on page 34.

Points to note:

- The formula for fringe separation $x = \frac{\lambda D}{a}$ is applicable only if
  - the rays are almost parallel, i.e., $D \gg a$
  - the angle θ subtended by the fringes from the principal axis is small, i.e., $a \gg \lambda$
- For larger θ, the fringe separation *x* increases.
- Typical values for double-slit interference:

slit width	$b \simeq 0.2$ mm
slit separation	$a \simeq 0.5$ mm
slit-screen distance	$D \simeq 1$ m
wavelength of light	$\lambda \simeq 500$ nm

$D \gg a$

$a \gg \lambda$

- The sources must be separated by more than one wavelength ($a > \lambda$) to produce an observable interference pattern.
- The **single slit** acts as a point source of light to ensure that waves incident on and exiting the double-slits are **coherent**.
- If the two coherent sources have a phase difference of π rad, the positions of constructive and destructive interference would be interchanged.

Example 7

In a Young's double-slit experiment, the separation between the first and the fifth bright fringe is 2.5 mm when the wavelength used is 620 nm. The distance from the double-slit to the screen is 0.80 m.

Calculate the separation of the two slits.

[Solution]

Fringe separation

$x =$

Slit separation

$a =$

Example 8 [N99/2/3]

A student sets up the apparatus shown below to observe two-source interference fringes.

State and explain what change, if any, occurs in the separation of the fringes and in the contrast between bright and dark fringes observed on the screen, when each of the following changes is made separately:

(a) increasing the intensity of light incident on the double-slit.

(b) increasing the distance between the double-slit and the screen.

(c) reducing the intensity of light incident on one of the double-slit.

[Solution]

Fringe separation $x = \dfrac{\lambda D}{a}$

(a) __________ to fringe separation x as λ, D and a remain the same.

Contrast __________ because intensity of bright fringes increases.

(b) Fringe separation x __________ as D increases.

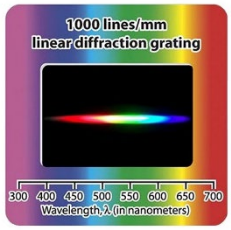
Contrast __________ because intensity of bright fringes decreases due to increased slit to screen distance.

(c) __________ to fringe separation x as λ, D and a remain the same.

Contrast __________ because the bright fringes are not as ________ while the dark fringes are ____________________ due to incomplete destructive interference.

❖ Diffraction Grating (Multiple Sources)

A diffraction grating is an optical component that consists of a large number of parallel, equally spaced lines (or slits) of equal width. A diffraction grating typically consists of 100 lines to 1000 lines per mm.



When a narrow beam of monochromatic light is incident on a diffraction grating, sharp maxima are obtained at various angular positions θ_n as shown.

Diffraction gratings are used in spectrometers (for investigating spectra).

By measuring the angle at which a diffracted image appears, the wavelength of the light producing that image may be determined.

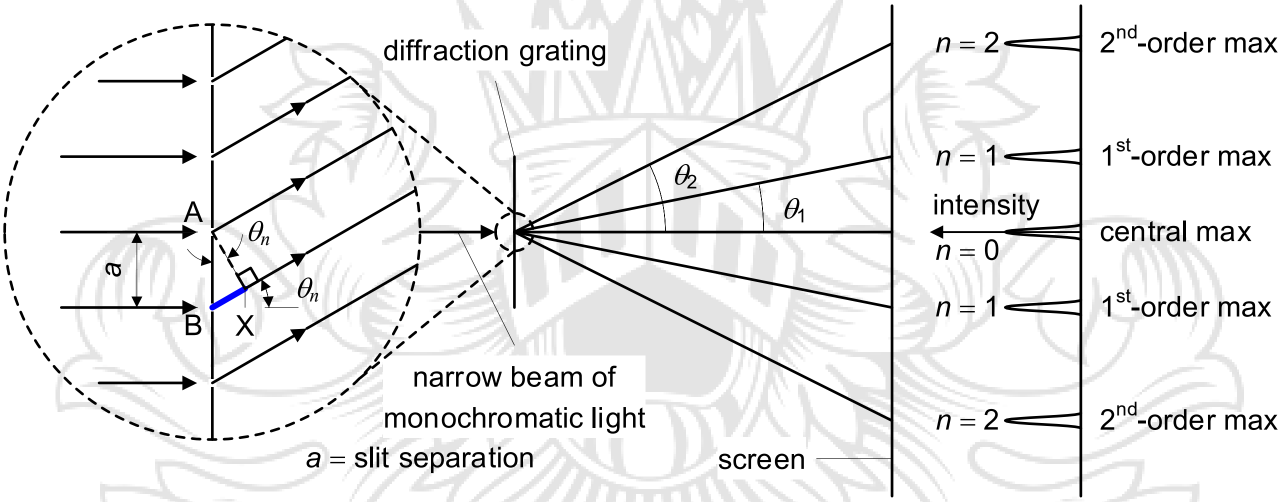


Fig. 11.25

Consider a narrow beam of monochromatic light, incident on a diffraction grating of N lines per metre.

The slit separation

$$a = \frac{1}{N}$$

For a grating with 500 lines per mm, the slit separation

$$a = \frac{1}{500000} = 2.0\ \mu\text{m}$$

At angular position of θ_n where n^{th}-order bright fringe (maxima) is obtained.

path difference between adjacent slits $= \text{BX} = a\sin\theta_n$

For **Diffraction Grating,**

Since the waves passing through all the slits are **in phase**, for **constructive** interference,

$$a\sin\theta_n = n\lambda$$

where $n = 0,\ 1,\ 2, \ldots$

$a =$ slit separation $= \dfrac{1}{N}$ (in metre)

The slit separation a (order of 10^{-6} m or larger) is comparable to the wavelength of visible light 0.4 to 0.7×10^{-6} m.

Hence, the small angle approximation used in Young's double-slit cannot be used for diffraction grating

❖ Maximum Observable Order of Diffraction

Higher order bright fringes occur at larger θ. Since θ cannot exceed $90°$, there is a highest order that can be observed.

The highest order of bright fringes is derived as follows:

Since $\theta_n \le 90°$, $\sin\theta \le 1$.

$$\sin\theta_n = \frac{n\lambda}{a} \le 1$$

$$n \le \frac{a}{\lambda}$$

The highest order of bright fringes is the largest integer value of *n*.

Example 9

Light of wavelength 656 nm is incident normally on a diffraction grating which has 400 lines per mm.

Determine the angular positions of the first, second and third-order maxima.

[Solution]

Slit separation

$$d =$$

For $n = 1$,

$$\sin\theta_1 = \frac{\lambda}{d} = \quad = 0.2624$$

$$\theta_1 =$$

For $n = 2$,

$$\sin\theta_2 = \frac{2\lambda}{d} =$$

$$\theta_2 =$$

For $n = 3$,

$$\sin\theta_3 = \frac{3\lambda}{d} =$$

$$\theta_3 =$$

Note:

The angle subtended between adjacent maxima $\Delta\theta$ increases, with increasing orders:

$$\Delta\theta_{10} = \theta_1 - \theta_0 = 15.2 - 0.0 = 15.2°$$

$$\Delta\theta_{21} = \theta_2 - \theta_1 = 31.7 - 15.2 = 16.5°$$

$$\Delta\theta_{32} = \theta_3 - \theta_2 = 51.9 - 31.7 = 20.2°$$

Example 10

A diffraction grating of 600 lines per mm is illuminated normally with light of wavelength 633 nm.

Determine the number of bright fringes that can be observed.

[Solution]

Slit separation

$d =$

Maximum order of bright fringes

n

Since n must be an __________, the maximum order of bright fringes observable is ____ order.

Hence, the total number of bright fringes is ___ (central fringe and 2 orders of bright fringes on either side of principal axis).

❖ Diffraction of White Light

When a beam of white light is incident on a diffraction grating, each wavelength is diffracted by a different amount, according to the equation $a\sin\theta_n = n\lambda$.

Hence, the beam of white light is split into its component colours to produce a **continuous spectrum** for $n \geq 1$.

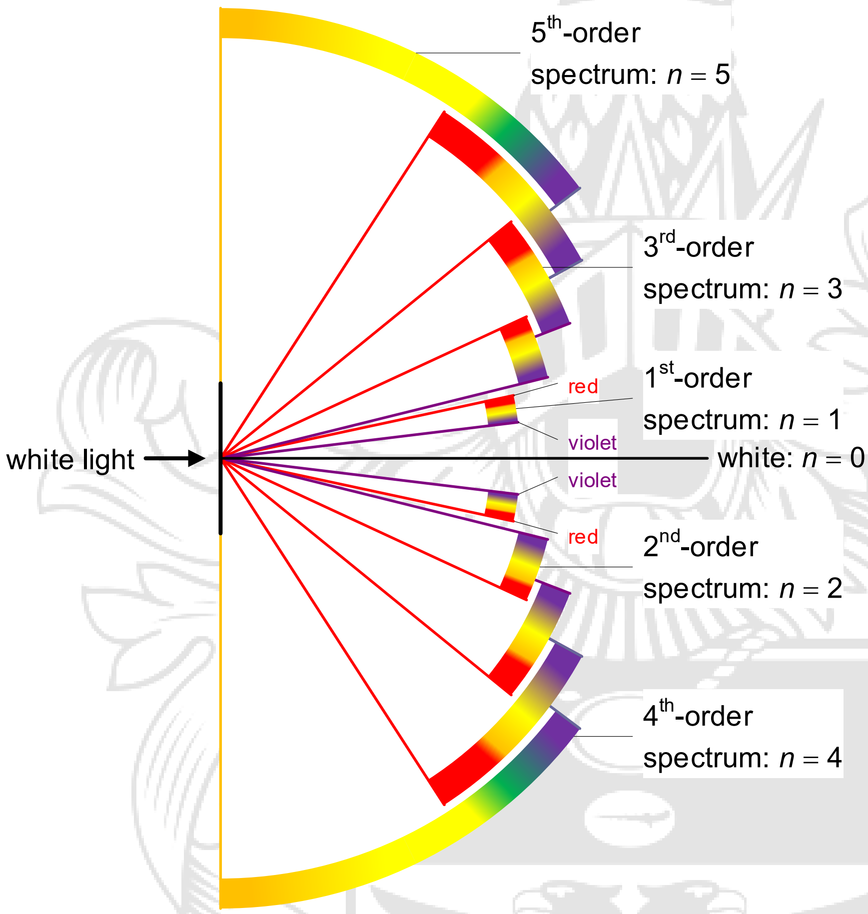


Fig. 11.26 higher order diffraction spectra (violet to red)

The central (0th order) maximum is white because all component colours converge with zero path difference at this point.

For each subsequent higher order, the longest wavelength deviates the most from the principal axis.

There may be some overlapping of different orders. For example, the red component of the 2nd order spectrum will overlap with the blue component of the 3rd order spectrum.

Example 11

White light having wavelengths from 380 nm (violet) to 780 nm (red) is incident normally on a diffraction grating of 500 lines per mm.

Describe, with the aid of a labelled diagram, the appearance of the 1st order spectrum.

[Solution]

Slit separation

$$d =$$

For violet light,

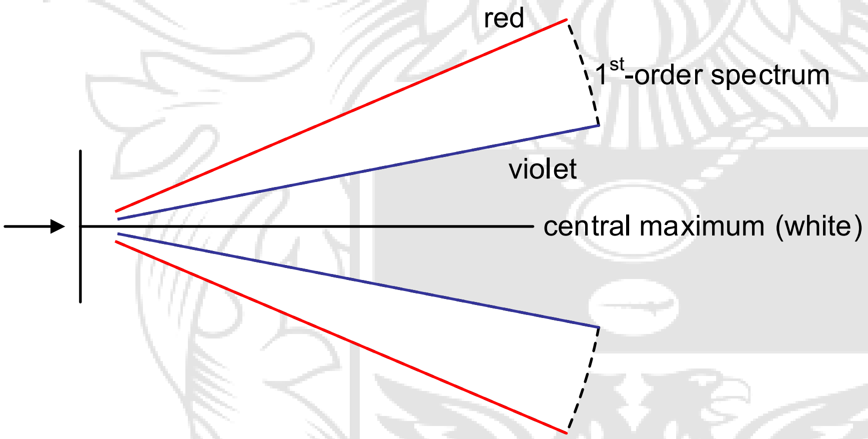
$$\sin\theta_V = \frac{\lambda_V}{d} =$$

$$\theta_V =$$

For red light,

$$\sin\theta_R = \frac{\lambda_R}{d} =$$

$$\theta_R =$$



- Within each order, red (longest wavelength) deviates the most and violet, the least, from the principal axis.
- The central maximum remains white as all component colours are present.

11.5 Single-Slit Diffraction

❖ Huygens' Principle

If a single drop of water falls into a pond, it will create a circular wavefront that spreads outwards. Dutch physicist Christian Huygens, a contemporary of Newton, put forward a wave theory of light which was based on the way in which circular wavefronts advance.

Huygens' principle is also called **Huygens-Fresnel** principle.

He proposed that at any instant, all points on a wavefront could be regarded as secondary sources giving rise to their own outward spreading circular wavelets. The envelope (or tangent curve) of the wavefronts produced by these secondary sources gives the new position of the original wavefront (Fig. 11.27).

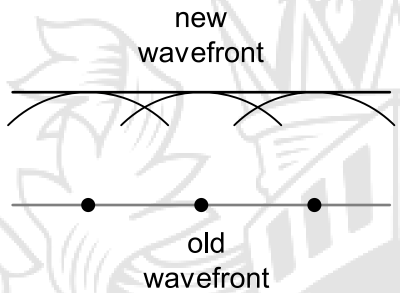


Fig. 11.27 Huygens' construction for a plane wave and spherical wave.

If a plane (or straight) wavefront is restricted by an aperture as shown in Fig. 11.28, some of the wavelets making up the wavefront are removed, causing the edges of the wavefront to be curved.

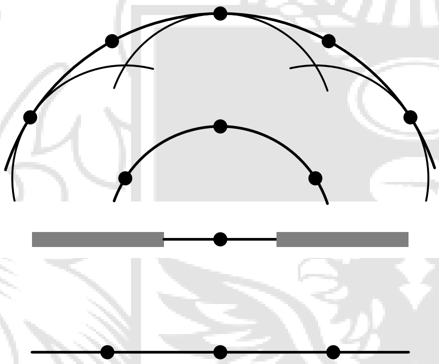


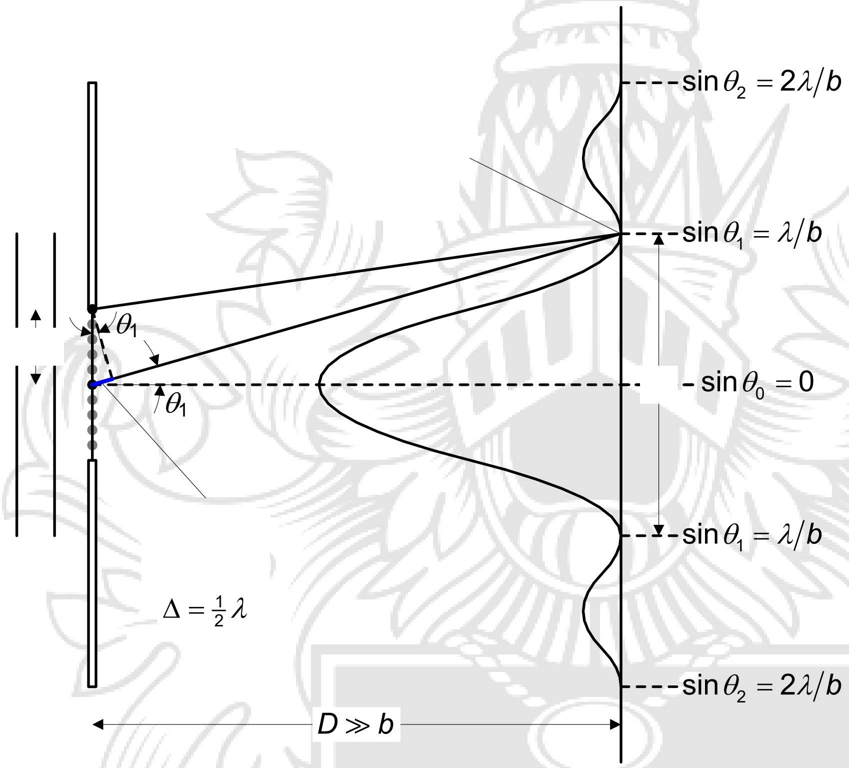
Fig. 11.28 Bending of wave through an aperture.

If the size of the aperture is comparable to the wavelength of the incident wave, the bending becomes pronounced, and the transmitted wavefront looks circular.

❖ Derivation of Single Slit Diffraction Equation

When monochromatic light passes through a single slit, an interference pattern is observed.

According to Huygen's principle, each point on the wavefront passing through the slit acts as a secondary source of wavelets. All wavelets that originate from the slit are in phase. The wavelets spreading out from these points interfere to create an interference pattern.



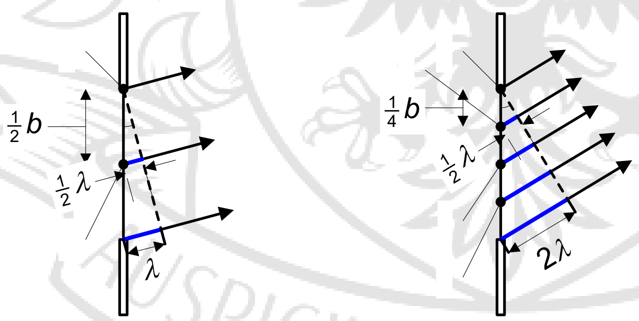
Note that the angles subtended by the waves from the principal axis are equal to θ_1 since $D \gg b$.

The intensities of the secondary maxima are much lower than the intensity at the central maximum.

Fig. 11.29 wavelets from first secondary sources of each half wavefront interfere destructively

The position of the **1st order diffraction minimum** is determined as follows:

To determine the condition for all wavelets emitted by the secondary sources on the wavefront to produce the first minimum, we divide the wavefront into two as shown in Fig. 11.30(a).



If the wavelets from secondary sources 1 and 2 of each half wavefront have a path difference of $\frac{1}{2}\lambda$, destructive interference occur.

Wavelets from subsequent pairs of secondary sources $\frac{1}{2}b$ apart also have a path difference of $\frac{1}{2}\lambda$ to ensure destructive interference.

As such, all wavelets arriving at the screen undergoes destructive interference. This is the position of the 1st order diffraction minimum.

From Fig. 11.30(a), the path difference

$$\tfrac{1}{2}b\sin\theta_1 = \tfrac{1}{2}\lambda$$

Simplifying,

$$\sin\theta_1 = \frac{\lambda}{b}$$

For the position of the **2nd order diffraction minimum**, we divide the wavefront into four as shown on Fig. 11.30(b).

If the wavelets from secondary sources 1 and 2 of the two adjacent quarter-wavefronts have a path difference of $\frac{1}{2}\lambda$, destructive interference occur.

Wavelets from subsequent pairs of secondary sources $\frac{1}{4}b$ apart also have a path difference of $\frac{1}{2}\lambda$ to ensure destructive interference.

The same argument applies to the next two adjacent quarter-wavefronts.

As such, all wavelets arriving at the screen undergoes destructive interference. This is the position of the 2nd order diffraction minimum.

From Fig. 11.30(b), the path difference

$$\tfrac{1}{4}b\sin\theta_2 = \tfrac{1}{2}\lambda$$

Simplifying,

$$\sin\theta_2 = \frac{2\lambda}{b}$$

For **Single-Slit Diffraction,**

The **1st order diffraction minimum** (dark fringe) occurs at angle θ according to

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

The ***m*th order diffraction minimum** (dark fringe) occurs at angle θ_m according to

$$\sin\theta_m = \frac{m\lambda}{b}$$

where *m* is an integer.

Example 12

Light of wavelength 580 nm is incident on a slit of width 0.310 mm. A screen is 2.00 m away from the slit.

Determine the width of the central bright fringe on the screen.

[Solution]

Let the width of the central bright fringe be y.

For the first minimum,

$$\sin\theta_1 = \frac{\lambda}{b} =$$

Since the angle is small,

$$\sin\theta_1 \approx \tan\theta_1$$

Small angle approximation.

$$\frac{\lambda}{b} \approx$$

$$y \approx$$

Alternative Method

$$y \approx 2D\theta_1 \approx 2D \times \sin\theta_1 = 2 \times 2.00 \times 1.93 \times 10^{-3} = 7.72 \times 10^{-3}\ \text{m}$$

Use arc length formula $s = r\phi$.

The position of the first dark fringes from the principal axis

$$\tfrac{1}{2}y = 3.86 \times 10^{-3}\ \text{m}$$

❖ Single Slit Diffraction & Interference of 2 or more sources

On page 23, it was mentioned that the intensity of the double-slit fringes is not constant. This is because the intensity of the light on the screen due to each slit is not constant.

Considering that the double-slit is in fact two single slits, the intensity of the resulting fringes should also vary similarly to a single slit diffraction pattern as shown in Fig. 11.31.

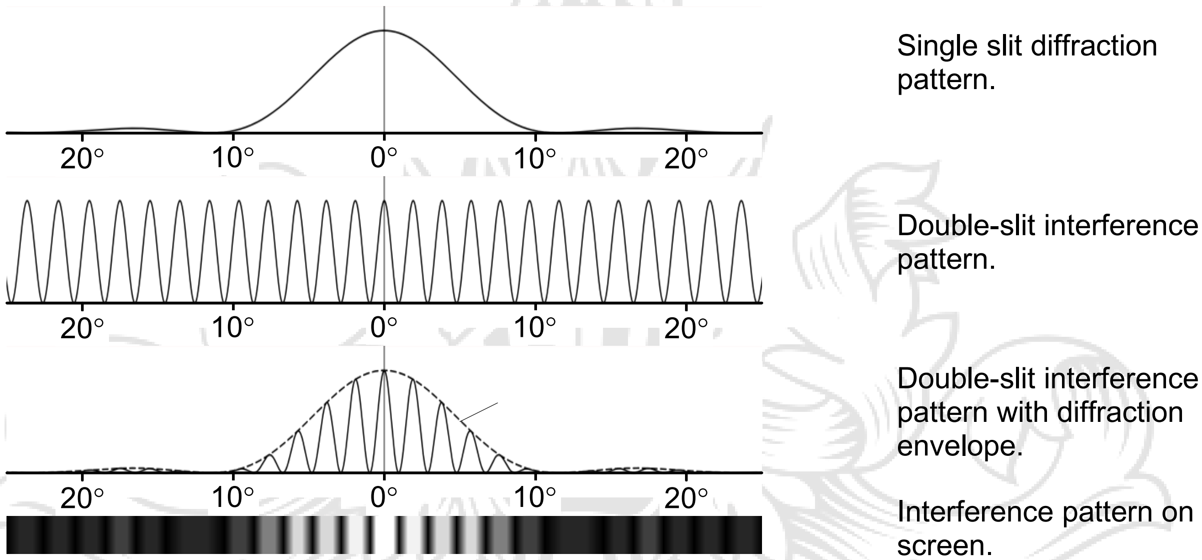


Fig. 11.31

Fig. 11.32 shows the intensity variation of *N*-slits side by side with what is seen on the screen.

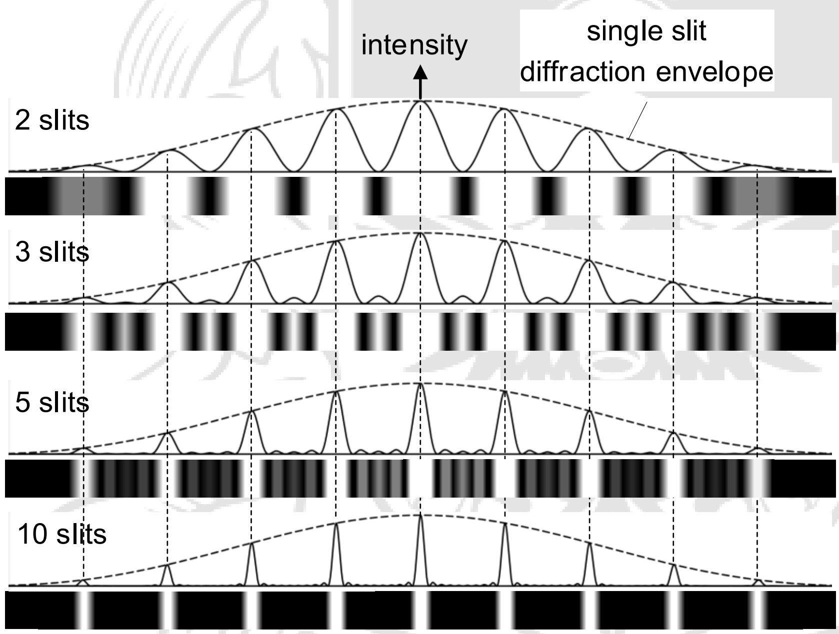


Fig. 11.32

The following are some observations when comparing the interference patterns of single-slit, double-slit and diffraction gratings.

Type	**Intensity**	**Appearance**
Single Slit	The intensity peaks in the centre of the central bright fringe, while gradually decreases to zero towards the first minimum. Successive secondary bright fringes are significantly dimmer and narrower than the central one. The **single-slit diffraction envelope** sets the **limiting intensities for multiple-slit** arrangements.	Broad central bright fringe with narrow, dark fringes on either side. Beyond these is a further succession of bright and dark fringes
Double-Slit	Intensity of the fringes is modulated by the broader diffraction envelope of the single slit.	Fringes are evenly spaced (near the central region).
Diffraction Grating	Intensity of the fringes is limited by the envelope of the single slit diffraction pattern.	Maxima are not evenly spaced (angular separation increases with order).

Example 13

(a) Explain why the higher-order maxima are dimmer in the Young's double-slit interference pattern.

(b) Explain why some orders are missing from the Young's double-slit interference pattern.

(c) In Fig. 11.31 on page 34, the 6th order bright fringe is missing from the double-slit interference pattern.

Determine the ratio of the slit separation *a* to the slit width *b*.

[Solution]

(a) The intensity of the double-slit interference pattern is ___________ by that of the _________________________________, which has a lower intensity at higher orders.

(b) The positions of ________________ of the double-slit interference pattern coincide with the ___________ of the single-slit diffraction pattern. Hence, these maxima are missing from the interference pattern.

(c) The 6th-order bright fringe is missing because it falls at the ___________ as the ______________________ of single-slit pattern.

6th order interference maximum:

1st order diffraction minimum:

Since the angle θ and the wavelength λ are the same in both equations, we have

$$\frac{a}{b} =$$

❖ Resolving of Two Point Sources

As light inevitably undergoes diffraction as it passes through an aperture, the resolving power of an optical instrument such as a microscope or telescope is limited by the size of its objective lens.

Two point sources of light S_1 and S_2 are positioned in front of a small aperture of width *b*. Due to diffraction, light diffracts after passing through the aperture and the resultant intensity on the screen is the sum of the individual diffraction patterns as shown in Fig. 11.33.

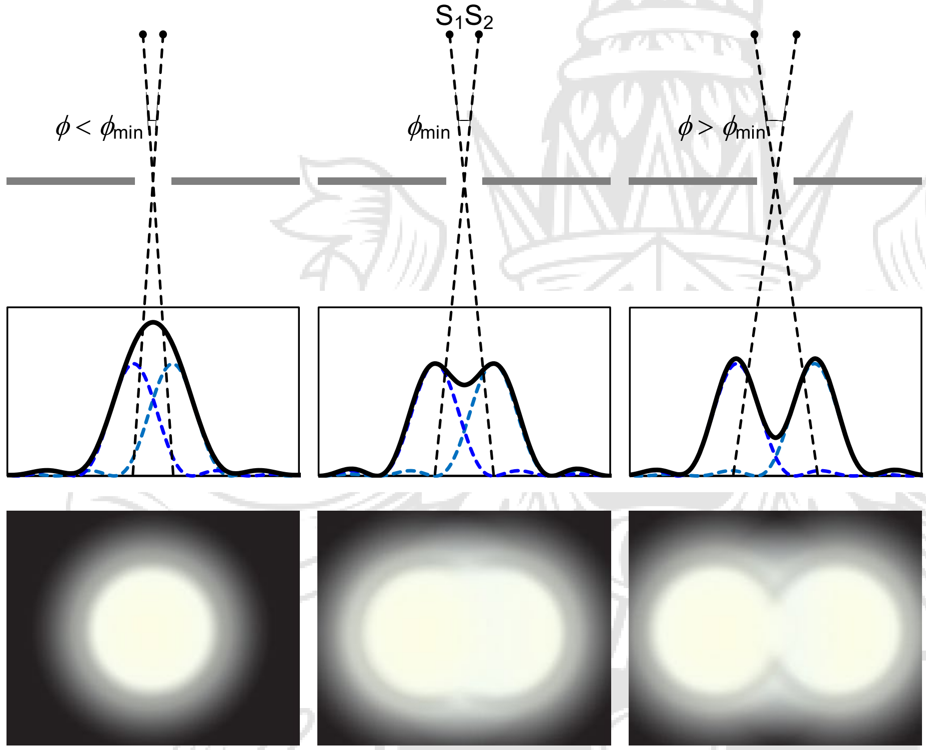


Fig. 11.33(a) **Fig. 11.33(b)** **Fig. 11.33(c)**

When S_1 and S_2 are close together as shown in Fig. 11.33(a), the individual diffraction patterns overlap significantly to produce an **indistinguishable** image. The sources are **unresolved**.

When S_1 and S_2 are further apart in Fig. 11.33(c), the individual diffraction patterns overlap to a lesser degree. The sources are **well-resolved** as the individual images are clearly discernible.

In the case of Fig. 11.33(b), S_1 and S_2 are separated just enough that the individual diffraction pattern is just distinguishable. The sources are **just resolved**. This is known as the **Rayleigh criterion**.

The Rayleigh Criterion

The Rayleigh Criterion states that two images are **just resolved** by an aperture when the **central maximum** of the diffraction pattern of one image **coincides** with the **first minimum** of the diffraction pattern of the other image.

From the study of the single-slit diffraction, the angular position of the **1st minimum** for single slit diffraction is given by

$$\sin\theta_1 = \frac{\lambda}{b}$$

Assuming that θ_1 is small, then $\sin\theta_1 \approx \theta_1$. Hence

$$\theta_1 \approx \frac{\lambda}{b}$$

If the angular separation ϕ of the two sources is greater than θ_1, then the sources are well-resolved.

$$\phi \geq \theta_1 = \frac{\lambda}{b}$$

Rayleigh criterion is satisfied if the angular separation of the two sources is equal to θ_1.

$$\phi_{\text{min}} = \frac{\lambda}{b} \qquad \text{(limiting angle of resolution)}$$

where λ is the wavelength of the light and b is the width of the **rectangular** slit.

For optical systems that use ***circular*** apertures

$$\phi_{\text{min}} = 1.22 \times \frac{\lambda}{D}$$

where D is the diameter of the aperture.

In the A-levels, all slits are assumed to be rectangular, hence

$$\phi_{\text{min}} = \frac{\lambda}{b}$$

Angular separation of images

For two images to be **well resolved**, the angular separation of the images, and hence the objects, must be **greater** than ϕ_{min} or λ/b.

A smaller limiting angle of resolution ϕ_{min} allows finer details to be observed. This increases the resolving power of the optical system.

This can be achieved either with

- a **larger** slit width b
- a **shorter** wavelength λ

Note that we can only observe features that is comparable to the wavelength of light. It is therefore impossible to see an object as small as an atom using visible light with an optical microscope.

Hence, electron microscopes that uses electrons are used to observe very small structures.

Example 14

Assume that the resolution of a human eye of pupil diameter 2.0 mm is only limited by diffraction. For a wavelength of 500 nm, calculate the

(a) limiting angle of resolution for the eye, and

(b) minimum separation between two point sources of light that the eye can distinguish, when the sources are 25 cm from the eye.

[Solution]

(a) Aperture size

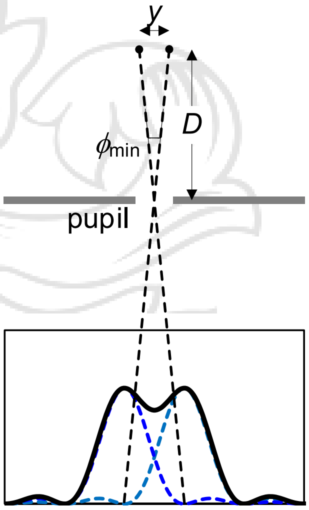
$$b = 2.0 \times 10^{-3} \text{ m}$$

Limiting angle of resolution

$$\phi_{\text{min}} =$$

(b) Since ϕ_{min} is small, we can approximate the separation y using the arc length formula $s = r\theta$.

$$y \geq$$



Alternative Method

Using trigonometry

$$\tan\tfrac{1}{2}\phi_{\text{min}} = \frac{\tfrac{1}{2}y}{D}$$

$$y = 2D\tan\tfrac{1}{2}\phi_{\text{min}} = 2 \times 0.25 \times \tan 1.25 \times 10^{-4} = 6.25 \times 10^{-5} \text{ m}$$

For second method, ensure that your calculator is in radian mode.

11.6 Appendix

❖ Standing Waves – a mathematical approach

An analytical treatment of the production of standing waves from the superposition of two progressive waves having the same frequency and amplitude is given below.

The two progressive waves may be represented by the equations

$$y_1 = a\sin\left(\frac{2\pi x}{\lambda} - \omega t\right) \text{ and } y_2 = a\sin\left(\frac{2\pi x}{\lambda} + \omega t\right)$$

where y_1 and y_2 are travelling toward the right and left respectively.

Using the principle of superposition of waves, the resultant wave can be represented by the equation

$$\begin{aligned} y &= y_1 + y_2 \\ &= a\sin\left(\frac{2\pi x}{\lambda} - \omega t\right) + a\sin\left(\frac{2\pi x}{\lambda} + \omega t\right) \\ &= 2a\sin\frac{2\pi x}{\lambda}\cos\omega t \\ &= \left(2a\sin\frac{2\pi x}{\lambda}\right)\cos\omega t \\ &= A\cos\omega t \end{aligned}$$

where $A = 2a\sin\left(\frac{2\pi x}{\lambda}\right)$ is the amplitude of the resultant standing wave for a point at a distance *x* from a reference point.

At the nodes, the amplitude is always zero and

$$\frac{2\pi x}{\lambda} = 0, \pi, 2\pi, \ldots$$

$$x = 0, \frac{\lambda}{2}, \lambda, \ldots$$

At the antinodes, the amplitude is maximum and equals 2*a*. This occurs when

$$\frac{2\pi x}{\lambda} = \frac{\pi}{2}, \frac{3\pi}{2}, \frac{5\pi}{2}, \ldots$$

$$x = \frac{\lambda}{4}, \frac{3\lambda}{4}, \frac{5\lambda}{4}, \ldots$$

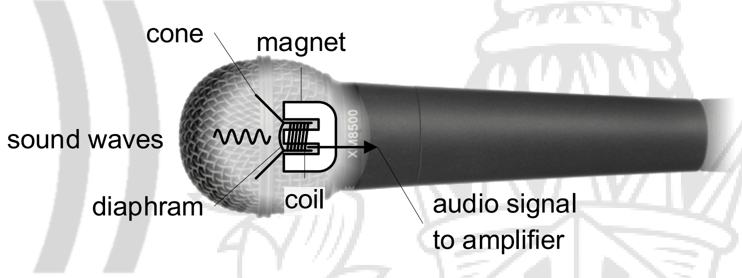
We find that the distance between two successive nodes or antinodes is always $\lambda/2$.

Note: $\sin P + \sin Q = 2\sin\left(\frac{P+Q}{2}\right)\cos\left(\frac{P-Q}{2}\right)$

❖ How Do Microphones Work?

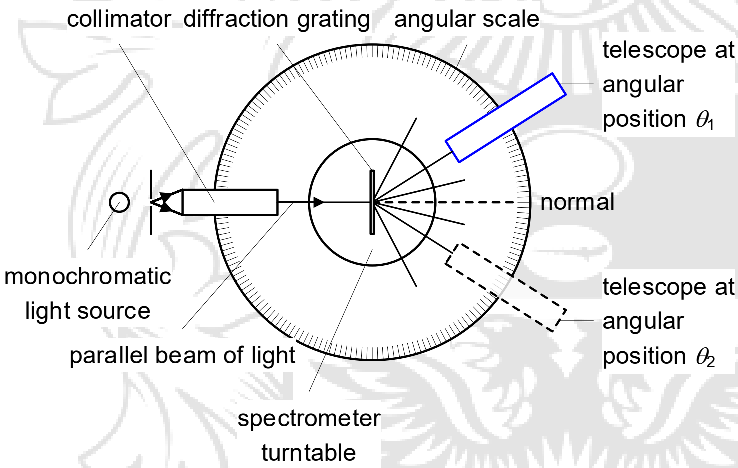
Microphones are a type of transducer – a device which converts energy from one form to another. Microphones convert acoustical energy (sound waves) into electrical energy (audio signal).

Different types of microphones have different ways of converting energy, but they all share one thing in common: the diaphragm. This is a thin piece of material (such as aluminium) which vibrates when it is struck by sound waves. In a typical hand-held microphone like the one below, the diaphragm is located in the head of the microphone.



When the diaphragm vibrates, it causes other components in the microphone to vibrate. These vibrations produce an induced e.m.f. which is amplified and sent to a loudspeaker.

❖ Spectrometer



A parallel, narrow beam of monochromatic light, after emerging from the collimator, is incident normally on the diffraction grating. As a result, bright fringes are produced at various angular positions.

The telescope is first rotated such that the n^{th} order bright fringe is at the centre of the crosswire in the telescope and its angular position θ_1 is noted. It is then positioned on the opposite side of the normal to the grating and the angular position θ_2 of the same order bright fringe is noted.

The angular position θ_n of the n^{th}-order bright fringe from the normal is

$$\theta_n = \tfrac{1}{2}(\theta_2 - \theta_1)$$

Hence, $\lambda = \dfrac{d}{n}\sin\theta_n$.