Chapter 6

Experiment 4: Wave Interference

WARNING

This experiment will employ Class III(a) lasers as a coherent, monochromatic light source. The student must read and understand the laser safety instructions on page 90 before attending this week’s laboratory.

Helpful Tip

Before beginning this experiment, it is probably a good idea to review the first several pages of the Sound Experiment on page 25. Waves of all kinds are quite similar in most respects so that the same interference between an incident and reflected sound wave that caused standing waves in the sound tube leads to the effects that we will observe today using light waves.

6.1 Introduction

In this and the following lab the wavelike nature of light is explored. We have learned that waves usually exhibit some periodic or oscillatory nature such as the water level at a particular point on a pond steadily rising and falling as waves pass. For light (electromagnetic radiation) it is the electric and magnetic fields which oscillate in strength and direction. At visible frequencies these oscillations occur so fast that they are exceedingly difficult to observe directly \((10^{14}\text{ Hz.})\)

Another important characteristic of waves is that they may interfere with each other. In simple terms interference is the result of oscillating dips and rises adding together or canceling out. A good example of interference is seen when the circular wave patterns from two stones thrown into a pond meet. High points from the separate waves add to make even
Figure 6.1: In each illustration the top two waveforms add point by point to yield the bottom waveform. The resulting effect can add as in the first case, cancel as in the second case, or anywhere in between. The difference between the three cases is the amount of constant phase that has been added to the second waveform. This phase is illustrated as a fraction of the total period. Waves have exactly the same dependence on position where the constant phase is in that case a fraction of wavelength instead of period.

higher points, dips from the two waves also add making lower dips, and when a dip from one wave passes the same point as a high from the other they cancel.

Waves are partly characterized by surfaces of constant phase. If time is frozen, then the argument of the wave’s function varies only in space and the function $\omega t = \varphi = k \cdot r$ defines a surface in space. In the specific case where $\omega t = \varphi = k \cdot r = kx$, $x$ is constant but $y$ and $z$ are arbitrary. Each $x$ defines the function of a plane parallel to the $y$-$z$ plane. This is a plane wave. Alternately, $\varphi = kr$ defines surfaces of constant radius or spheres centered at $r = 0$. This is a spherical wave. Many other surfaces are possible, but these two are the most important ones. Plane waves are by far the easiest to treat mathematically and spherical waves are generated by point sources like atoms, molecules, and surface structures.

Checkpoint

What is a plane wave?

Figure 6.1 shows how two sinusoidal waves add together to produce a new wave. In this lab we will see how interference affects the intensity pattern of light that has passed through two narrow adjacent slits, many parallel slits (diffraction grating), and a special bulls-eye pattern called a Fresnel zone plate.
Figure 6.2: A plane wave incident upon two slits is transmitted as an interference pattern. In the center of the pattern \((m = 0)\), the distance \(L_1\) from the top slit to the screen equals the distance \(L_2\) between the bottom slit and the screen. The time needed to transit this distance is exactly the same for the two paths. The two waves are in phase as in Figure 6.1(a). As we move down off the optical axis, \(L_2\) gets shorter and \(L_1\) gets longer until their difference is half a wavelength and they cancel as in Figure 6.1(b). Further off axis the path difference is a full wavelength and the peaks and valleys are aligned again \((m = -1)\) as in Figure 6.1(a).

### 6.1.1 Two Slit Interference – Young’s Double Slit Experiment

**Historical Aside**

In 1801 Thomas Young devised a classic experiment for demonstrating the wave nature of light similar to the one diagrammed in Figure 6.2.

By passing a wide plane wave through two slits he effectively created two separate sources that could interfere with each other. (A plane wave is a wave consisting of parallel wavefronts all traveling in the same direction.) For a given direction, \(\theta\), the intensity observed at point \(P\) on the screen will be a superposition of the light originating from every slit. Whether the two waves add or cancel at a particular point on the screen depends upon the relative phase between the waves at that point. (The phase of a wave is the stage it is at along its periodic cycle; the argument of the sine or cosine function.) The relative phase of the two waves at the screen varies in an oscillatory manner as a function of direction causing periodic intensity minima and maxima called fringes (see Figure 6.2). The spacing of the fringes depends on the distance between the two slits and the wavelength of the light. If the width of the slits is narrow compared to a wavelength, the intensity of the two-slit
At an angle $\theta$ above the optical axis, the distance between the screen and the bottom slit is longer than the distance between the screen and the top slit by $\delta = d \sin \theta$ as shown by the geometry of the right triangle indicated.

The interference pattern is given by

$$E(\theta) = E_0 \cos \beta \quad \text{and}$$

$$I(\theta) = I_0 \cos^2 \beta \propto |E(\theta)|^2 \quad \text{for}$$

$$\beta = \frac{\pi d}{\lambda} \sin \theta$$

where $\theta$ is the angle between the initial direction, the location of the slits, and the observation. $d$ is the distance between the two slits, $\lambda$ is the wavelength of the light, $E$ is the electric field, and $I$ is the light intensity with units of energy per area per second.

Equation (6.1) is easily derived. The amplitude of the electromagnetic wave emanating from either slit may be written as $E = E_0 \sin(kx - \omega t)$. After the wave emanating from one of the slits travels a distance $l$ the wave undergoes a phase shift $\varphi = kl = \frac{2\pi l}{\lambda}$ so that the electric field is of the form $E = E_0 \sin(\omega t + \varphi)$. The light arrives at the two slits at the
same time, so the phase difference is due to the difference in distances between the point of observation, $P$, and the respective slit. This difference in distance is shown in Figure 6.3 as a function of angle. It is seen that the wave front emanating from one slit lags behind the other by $\Delta l = d \sin \theta$. This gives a relative phase difference of

$$\Delta \phi = \frac{2 \pi \Delta l}{\lambda} = \frac{2 \pi d}{\lambda} \sin \theta.$$  

Since we are interested only in relative differences we can measure from the midpoint between the slits and write the fields due to each slit as

$$E_1 = E_0 \sin \left(\omega t + \frac{1}{2} \Delta \phi\right) \quad \text{and} \quad E_2 = E_0 \sin \left(\omega t - \frac{1}{2} \Delta \phi\right).$$

Using the trigonometric identity, $\sin A + \sin B = 2 \sin \frac{1}{2}(A + B) \cos \frac{1}{2}(A - B)$ gives

$$E = E_1 + E_2 = 2E_0 \sin \omega t \cos \frac{1}{2} \Delta \phi = 2E_0 \cos \beta \sin \omega t.$$  \hspace{1cm} (6.2)

When squared and averaged over time, this produces Equation (6.1). Maximum intensity peaks occur at angles such that $|\cos \beta| = 1$ and the difference between the two path lengths is an integral multiple of wavelengths $\Delta l = d \sin \theta = m\lambda$ (see Figure 6.3). This translates to $\beta = m\pi \ (m = 0, \pm 1, \pm 2, \ldots)$ and thus the angles of maximum intensity satisfy,

$$m\lambda = d \sin \theta \quad \text{and} \quad m = 0, \pm 1, \pm 2, \ldots$$  \hspace{1cm} (6.3)

If the width of the slits is greater than a wavelength the interference fringes are seen to oscillate in brightness as seen in Figure 6.4. This is due to diffraction and will be studied in a later experiment.

**Checkpoint**

Sketch two sinusoidal waves out of phase by 45°.

**Checkpoint**

Two waves in phase and each one with an amplitude $A$ and intensity $I$ meet at the same point. What is the intensity of the signal observed at that point?

### 6.2 The Experiment

A helium-neon (HeNe) laser will be used to illuminate various patterned slides at normal incidence. A two-meter stick and a ruler will be used to quantify the locations of the slides,
Figure 6.4: The intensity profile of light transmitted through two actual slits. By changing only the slit width $a$, we can verify that the intensity modulation indicated by the yellow trace is due to the slit width $a$ and NOT to the slit separation $d$. Presently, we are interested in the closely-spaced, bright spots (the RED trace). The distance $x$ between the slits and the screen and the distance $y$ between the optical axis and each spot can be measured with a meter stick to determine $\theta$ using $\tan \theta = \frac{y}{x}$. 

the original beam path, and the observation points that have bright spots. This information will be used to determine the angle $\theta$ needed to relate the laser’s wavelength to the distance between two slits’ centers.

6.2.1 Two-Slit Interference

You are provided a slide that has several sets of slits on it with separations and widths as shown in Figure 6.5. You use the HeNe laser as your plane wave light source. The experimental set up is as shown in Figure 6.6. If the laser beam is pointed onto a pair of slits, an interference pattern similar to Figure 6.7 may be observed on a screen placed nearby. The steel plates will hold magnets where we may install paper strip(s) to record the spot locations. This allows the distance between fringes to be marked and measured easily and conveniently. Alternately, stripped cm scales may be used as a screen and the interference pattern and scale may be photographed and studied later.

Observe the interference patterns on a white paper strip. Draw what you see for each pattern on the slide in your Data Notebook. You may photograph them also if you prefer; however, practice drawing them anyway. Be sure also to note the slit dimensions for each pattern. How does the slit separation affect the interference pattern? Can you detect any intensity variations in the different fringes? Why does this occur? For each set of double
slits identify the intensity distribution caused by the width, $a$, of each slit as well as the interference fringes whose spacing is related to the distance, $d$, between the two slits. Two of the double-slits in Figure 6.5 are particularly interesting to observe because they have the same slit width, $a$, but different slit separations, $d$. Consequently the minimum of the diffraction pattern must occur at the same $\theta$ value but the spacing of the interference fringe pattern varies. Also, two of the double-slits in Figure 6.5 have the same slit separation, $d$, but different slit widths, $a$. This combination of four double-slit patterns allows you to isolate the effects of these two independent parameters.

You will now make measurements using one double-slit pattern. When the slit mask is perpendicular to the beam, its reflection will re-enter the laser. Orient the screen perpendicular to the beam as well and mark the locations of the interference fringes (the finely spaced bright spots) on a fresh strip. Mark all visible spots; if fewer than five spots are visible on each side of the beam axis, choose a different double-slit pattern. Also measure the distance from the double-slit slide to the screen, $x$, taking care not to disturb the setup until the measurement is obtained. Remove the double-slits and circle the undeflected laser spot; this is the zero order ($m = 0$) fringe. Make a table in your Data, and measure the distances from the beam axis to the bright spots, $y_m$. It will be more accurate to measure $2y_m$ (the distance between the pair of dots on opposite sides of the beam axis) and then divide by two. Be sure to note your error estimates and units for these measurements. Using $\lambda = 632.8\text{ nm}$ as the wavelength of the laser-light and Equation (6.3), calculate the slit separation. (Recall that nano (n) means $\times 10^{-9}$.) Do the distances, $R_m$, between the different bright spots and the slits vary from $x$ significantly more than the total of your measurement errors? If not then we can make an approximation that

$$\sin \theta_m = \frac{y_m}{R_m} \approx \frac{y_m}{x} = \tan \theta_m$$

so that

$$d = \frac{m\lambda}{\sin \theta_m} \approx \frac{m\lambda x}{y_m}.$$  

You can download a suitable setup file for Ga3 from the lab’s website at

http://groups.physics.northwestern.edu/lab/interference.html

Otherwise, load one column in Vernier Software’s Graphical Analysis 3.4 (Ga3) program with the order numbers, load your measurements of $y_m$ (or $2y_m$ if this is what you measured) into another column, and create a calculated column to hold your slit width results. Double-click the column headers to correct the labels and to add the correct units. What will be the units of $d$? (If $x$ and $y_m$ have the same units, then these units will cancel in the ratio;
Figure 6.6: A sketch of our apparatus showing how the HeNe laser can be directed through the slits and projected onto a view screen.

otherwise, it will be more difficult for you to determine the units of \( d \).) Use the “Column” button to choose column “m” and column “\( y_m \)” to enter the appropriate formula for \( d \) into the “Equation” edit control. (You will see your column names instead of mine.) If you measured \( 2y_m \), you will need to divide this column (the denominator) by two or to multiply the numerator by two. You will also need to type in appropriate numbers for \( \lambda \) and \( x \); Ga3 only works with numbers. You will also need explicitly to tell Ga3 to multiply \( \ast \).

Figure 6.7: One example of an interference pattern due to HeNe laser light passing through double slits. Large spots indicate large intensity. If the slit dimensions satisfy \( d = na \) for integer \( n \), there will be missing orders at every \( n^{th} \) spot as indicated by the black circles containing no spot. The positions of the spots are measured as indicated for \( m = 3 \) on the right and for \( m = -7 \) on the left. We MUST count the black circles despite the missing spot.

Plot the slit separation column on the \( y \)-axis and the order, \( m \), on the \( x \)-axis. Click on the smallest \( y \)-axis number and change the range beginning to 0. Do your calculated values of \( d \) vary more than you expect your errors to explain? If not, draw a box around your data points with the mouse and Analyze/Statistics. Move the parameters box away from your data points. Use the strategy in Section 2.6.1 to specify your best measurement, its correct tolerance, and units. Keeping in mind that their tolerances are 1%, note Pasco’s mask specification as well.

6.2.2 Diffraction Gratings

A system consisting of a large number of equally spaced, parallel slits is called a “grating”. The diffraction pattern expected from a grating can be calculated in a way similar to the two slit system. A simple way to do this calculation is to represent the amplitude and phase of
CHAPTER 6: EXPERIMENT 4

each wave originating from one of the grating slits and impinging on a point on the screen by a vector called a phasor (see textbook or Appendix of Experiment 8 from the second quarter manual). Summing the phasors corresponding to each slit gives the correct amplitude and phase of the total wave. The intensity distribution of the light scattered off a grating with \( N \) very narrow slits is

\[
I(\theta) = I(0) \left( \frac{\sin N\beta}{N \sin \beta} \right)^2
\]

where \( \beta \) is defined in Equation 6.3. This relation is very similar to the one for the interference of two slits. In fact if \( N = 2 \), then substituting the identity \( \sin 2\beta = 2 \sin \beta \cos \beta \) in the last term of Equation 6.5) gives

\[
\left( \frac{\sin N\beta}{N \sin \beta} \right)^2 = \left( \frac{\sin 2\beta}{2 \sin \beta} \right)^2 = \left( \frac{2 \sin \beta \cos \beta}{2 \sin \beta} \right)^2 = (\cos \beta)^2
\]

and it is seen to reduce to the two slit formula Equation (6.2).

The angular dependence of the interference maxima are described by the function \( \sin \frac{N\beta}{N \sin \beta} \). With a large number of slits, this function has a numerator rapidly varying with \( \theta \) and a denominator slowly varying. As the number of illuminated slits increases, the intensity peaks will narrow. The position of the fringe maxima generated by a grating is given by Equation 6.3. Gratings are commonly used to measure the wavelength of light, \( \lambda \). This may be determined using Equation (6.5) and by measuring the angle between the maxima. These measurements are called spectroscopy and will be very important in our last experiment.

Place the diffraction grating in the path of the green laser and perpendicular to the beam. The number of slits per inch should be written on one side of the slide. Because the spacing between slits is so small the angles of the maxima are quite large. You will need to keep the screen close so that the 1\(^{st}\) and 2\(^{nd}\) order maxima can be seen. Measure the distance between each of these spots and the central maximum. Don’t forget your units and error. Also measure the distance between the screen and the grating slide. How much does the intensity vary for the central, the 1\(^{st}\), and the 2\(^{nd}\) order maxima?

For both measurements (or all four if you did not measure 2 \( y_m \)) made in Section 6.2.1 calculate values of the wavelength of the light. Note that these angles are not small so that \( \sin \theta \approx \tan \theta \) does NOT apply to this case. You must use

\[
\lambda = \frac{d}{m \sin \tan^{-1} \left( \frac{y_m}{x} \right)}.
\]

Compare these values to the given wavelength of the laser, 532.0 nm. Since we used \( y_m \), \( x \), and \( d \) to calculate \( \lambda \), we must expect that their errors will contribute to an experimental error in \( \lambda \).
Figure 6.8: Two Fresnel Zone Plates. Dark rings block light that would be out of phase and would interfere destructively on the optical axis. Much smaller Fresnel Zone Plates can focus X-rays without a lens. These are suitable for microwaves.

**Checkpoint**

Calculate the maximum number of fringes that you are expected to observe with the grating. (Hint \( \sin \theta \leq 1 \)). Do you observe all of the expected fringes?

**Checkpoint**

What are gratings useful for?

### 6.2.3 The Fresnel Zone Plate

In the *Fraunhofer* formalism light is represented by plane waves, and the distance between light-source and scattering object or scattering object and observer is assumed to be very large compared to the dimensions of the obstacle in the light path. A point source at a lens’ focus also yields plane waves. If you drop these conditions, the light propagates as spherical waves and the phenomena observed are described by the *Fresnel* formalism.

Figure 6.9 shows the spherical surface corresponding to the primary wave front at some arbitrary time \( t \) after it has been emitted from \( S \) at \( t = 0 \). As illustrated the wave front is divided into a number of annular regions. The boundaries between these regions correspond to the intersections of the wave front with a series of spheres centered at \( P \) and having radius \( r_0 + \frac{1}{2}\lambda, r_0 + \lambda, r_0 + \frac{3}{2}\lambda, \) etc.
Figure 6.9: The Fresnel Zones. The surface of the spherical wavefront generated at point \( S \) in the figure above has been divided into several ‘Fresnel Zones’. Each area is comprised of points that are close to the same distance from point \( P \) and thus all the secondary wavelets emanating from within the same Fresnel Zone will add constructively at \( P \).

These are Fresnel zones or half-period zones. The sum of the optical disturbances from all \( m \) zones at \( P \) is

\[
E = E_1 + E_2 + \ldots + E_m
\]  

(6.7)

Checkpoint

What is the difference between Fraunhofer and Fresnel scattering?

The path length for the light passing through each consecutive zone increases by \( \lambda/2 \) so the phase changes by \( \pi \) and the amplitude, \( E_\zeta \), of zone \( \zeta \) alternates between positive and negative values depending on whether \( \zeta \) is odd or even. As a result contributions from adjacent zones are out of phase and tend to cancel.

The surface of the spherical wavefront generated at point \( S \) in Figure 6.9 has been divided into several ‘Fresnel Zones’. Each area is comprised of points that are close to the same distance from point \( P \) and thus all the secondary wavelets emanating from within the same Fresnel Zone will add constructively at \( P \).

This suggests that we would observe a tremendous increase in irradiance (intensity) at \( P \) if we remove all of either the even or the odd zones. A screen which alters the light, either in amplitude or phase, coming from every other half-period is called a zone plate. Examples are shown in Figure 6.8. Suppose that we construct a zone plate which passes only the first 20 odd zones and obstructs the even ones,

\[
E = E_1 + E_3 + \ldots + E_{39}
\]  

(6.8)
and that each of these terms is approximately equal. For a wavefront passing through a circular aperture the size of the fortieth zone, the disturbance at $P$ (demonstrated by Fresnel but not obvious) would be $E_1/2$ with corresponding intensity $I \propto (E_1/2)^2$. However, with the zone plate in place $E \approx 20E_1$ at $P$. The intensity $I$ of the light at $P$ has been increased by a factor of 1600 times. The zone plate acts as a lens with the focusing being done by interference rather than by refraction!

To calculate the radii of the zones shown in Figure 6.8, refer to Figure 6.10. The outer edge of the $m$th zone has radius $R_m$. By definition, a wave which travels the path $S-R_m-P$ must arrive out of phase by $m\lambda/2$ with a wave which travels the path $S-O-P$, that is,

$$\text{Path difference} = (\rho_m + r_m) - (\rho_0 + r_0) = \frac{m\lambda}{2}.$$  \hspace{1cm} (6.9)

According to the Pythagorean theorem $\rho_m = \sqrt{R_m^2 + \rho_0^2}$ and $r_m = \sqrt{R_m^2 + r_0^2}$. Expand both these expressions using the binomial series and retain only the first two terms, $\rho_m \approx \rho_0 + \frac{R_m^2}{2\rho_0}$ and $r_m \approx r_0 + \frac{R_m^2}{2r_0}$. Substituting into Equation (6.9) gives the criterion the zone radii must satisfy to maintain the alternating $\lambda/2$ phase shift between zones,

$$\frac{m\lambda}{2} = (\rho_m + r_m) - (\rho_0 + r_0) \approx \frac{R_m^2}{2\rho_0} + \frac{R_m^2}{2r_0} \quad \text{and} \quad R_m^2 \approx \frac{m\lambda}{\rho_0 + \frac{1}{r_0}}.$$

(6.10)

The width of zone $m$ is proportional to $\frac{1}{\sqrt{m}}$. Since the circumference (length) of the ring is $2\pi R_m \sim \sqrt{m}$, the area and transmitted intensity is the same for all $m$. For $m = 1$ we may rewrite Equation (6.10) as

$$\frac{1}{\rho_0} + \frac{1}{r_0} = \frac{\lambda}{R_1^2} = \frac{1}{f_1}.$$  \hspace{1cm} (6.11)
to put it in a form identical to the thin lens equation \( \frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f} \) with primary focal length \( f_1 = \frac{R^2}{\lambda} \). Once the zone plate is made, the ring radii are fixed so that the focal length is different for every wavelength. This property might be utilized in a spectrometer.

The spacing of the zones on the Fresnel Zone Plate are very fine. If a narrow beam is passed through the plate it will act as a diffraction grating. Put the Fresnel Zone Plate in the red laser path and observe the diffraction pattern. Move the plate back and forth so that you observe the results when light hits the center and the outside of the plate. Is the slit separation smaller on the outside or center of the plate? What have you observed that supports this answer? Can you sketch the zones in your Data? This kind of experiment used with X-rays allows us to quantify the properties of crystalline solids.

Measure the primary focal length of your setup’s zone plate. Do this by centering the laser beam on the double concave lens which will diffuse the beam and generate on the screen a uniformly illuminated circle. The divergent beam emanating from the double concave lens now acts as a point source or ‘object’. Now place the zone plate between the lens and the screen, and change the position of the screen along the direction of the beam to observe the refocusing of the laser beam. Measure the two quantities \( \rho_0 \) and \( r_0 \) (Figure 6.10) for at least 2 different initial values of \( r_0 \) and then use Equation (6.11) to calculate \( f_1 \). Don’t forget your units. How far can you move the zone plate before the spot is out of focus? You must consider all of these locations when estimating your errors.

Often zone plates are made of plastic material or metal with adjacent rings supported by spokes so that the transparent regions are devoid of any material. These will function as lenses in the range from ultraviolet to soft (low energy) X-rays (1 to 1000 Å) where ordinary glass is opaque. Zone plates have also been used to focus low energy neutron beams.

**Checkpoint**

What is a Fresnel half-period zone? Sketch a zone plate. What are the zone plates used to do?

### 6.3 Analysis

Use the strategy in Section 2.9.1 to compare your measured slit separation to that of the Pasco specification; recall that Pasco specifies their dimensions to 1%. Repeat for the measured green laser wavelength versus (532.0 ± 0.1) nm.

Can you think of anything in the three experiments that might reasonably affect the measurements that we have not already included in our measurement errors? Discuss each of these using complete sentences and paragraphs. In your view is it more likely that the theory that has survived for a hundred years of professional tests is false or that you have made some mistake or overlooked some significant sources of error that might explain your nonzero difference(s)?
6.4 Conclusions

What physics have your experiments supported? Contradicted? What have you investigated and failed to resolve? Is $m\lambda = d\sin \theta$? What do these symbols mean? Does light exhibit wave properties? If so, which properties? Do Fresnel zone plates focus light? Don’t forget that we communicate using complete sentences.

Have you measured anything that might be of value in the future? If so, state what they are, give your measured values, their errors, and their units. How might we apply our discoveries? What improvements might you suggest?

6.4.1 References

Sections of this write up were taken from:

- Bruno Gobi, *Physics Laboratory: Third Quarter*, Northwestern University
- E. Hecht/A. Zajac, *Optics*, Addison-Wesley Publishing
6.5 Appendix

A discussion of relevant error propagation and gas laser operation.

6.5.1 Error Propagation

We might estimate the error in $\lambda$ using

$$\delta \lambda = \lambda \sqrt{\left(\frac{\delta d}{d}\right)^2 + \left(\frac{\delta \theta}{\tan \theta}\right)^2} \quad \text{tan} \theta \neq 0. \quad (6.12)$$

In this case $\delta \theta$ must be expressed in rad. (Radian measure is $\delta \theta = \frac{\pi}{180^\circ} \delta \theta$ in degrees.) This experiment does not yield the angle directly but the needed uncertainty is mostly due to the uncertainty in $y_m$

$$\delta \theta_m \approx \frac{\delta y_m}{x} \cos^2 \theta; \quad (6.13)$$

this already has radian units.

6.5.2 The Helium-Neon Laser

The laser (Light Amplification by Stimulated Emission of Radiation) is a source of radiation capable of delivering intense coherent electromagnetic fields. The laser beam coherence is manifest in two ways:

1) it is highly monochromatic, and
2) it is spatially coherent;

the emitted light has a nearly constant wavefront and directionality. The special characteristics of laser radiation are directly attributable to the phenomenon of stimulated emission and to the positive feedback mechanism provided by the cavity structure. Figure 6.11, a schematic diagram of a typical gas laser, highlights the major elements of a gas system. It consists of an active medium (lasing material), which in this case is a gas, but may consist of a solid or a liquid. A power supply pumps energy into the active medium, exciting the atoms. To achieve oscillations mirrors are placed perpendicular to the axis of the active medium to form an optical resonant cavity. Stimulated emission in the active medium results in the required amplification, whereas the mirrors supply the feedback required for regenerative action and oscillation. The rest of this appendix explains these concepts in greater detail.

Checkpoint

Which ones are the unique properties of light emitted by a laser?
CHAPTER 6: EXPERIMENT 4

Figure 6.11: Two energy diagrams illustrating photo-emission and photo-absorption. In the first case an atom gives its energy of excitation to the creation of a photon. In the second case a photon collides with an electron and promotes it to a more energetic atomic orbit.

6.5.3 Atomic Levels

Figure 6.12: Two energy diagrams illustrating photo-absorption and stimulated emission. If the atom is initially not excited, the photon excites the atom. If the atom is initially excited, the incident photon stimulates the emission and both photons are identical to the incident photon.

The atom (see 8th lab) consists of a small heavy nucleus surrounded by orbiting electrons. Quantum Mechanics describes the orbits of the electrons as standing waves. Consequently, the electrons belonging to a specific orbit have a discrete energy value. Of the infinite number of allowed electron orbits (energy levels), only the lowest energy levels are filled when the atom is in the ground state. When the atom is excited, electrons are transferred from the low energy stable orbits to higher states, leaving vacancies in the original orbits. The excited atom will then decay (the electron in the excited state jumps back down to lower levels) with a time constant characteristic of the level. (The time constant of an atomic level has exactly the same meaning as the time constant of an $RC$ circuit.) This time is usually around $10^{-8}$ s, but in some cases it is on the order of a second (it is then called metastable). Figure 10.2 in
the 8th lab shows, as an example, the energy levels of hydrogen. For the sake of simplicity, we shall consider an atom having only two possible energy states: an upper state \( E_2 \) and a lower state \( E_1 \) as shown in Figure 6.12. If the atom is in the upper state (excited state) and makes a transition to the lower state, then \( \Delta E \) energy can be emitted in the form of radiation with the frequency

\[
\nu = \frac{E_2 - E_1}{h}
\]

(\( h \) is the Planck constant: \( h = 6.626 \times 10^{-34} \text{ J} \cdot \text{s} \)). On the other hand, if the atom is initially in the lower energy state, \( E_1 \), and makes a transition to the higher state, \( E_2 \), then energy and hence radiation with a frequency given by Equation 6.14, must be absorbed. Albert Einstein showed that emission (transition from \( E_2 \rightarrow E_1 \)) can occur in two ways:

1) by the atom randomly changing to the lower state with lifetime \( \tau \). This is called spontaneous emission and is contrasted with absorption in Figure 6.11.

2) by a photon having an energy equal to the energy difference between the two levels interacting with the atom in the upper state and causing it to change to the lower state with the creation of a second photon. This process can be thought of as the converse of absorption and is known as stimulated emission. See Figure 6.12.

There are two very important points about stimulated emission upon which the properties of laser light depend. First, the photon produced by stimulated emission is of almost equal energy to that which caused stimulated emission and hence the light waves associated with them must be of nearly equal frequency. Second, the light waves associated with the two photons are in phase — they are said to be coherent. In the case of spontaneous emission the random emission of photons results in waves of random phase and the light is said to be incoherent.

**Checkpoint**

Sketch an energy level diagram of an atom. Indicate one transition and the frequency of the transition. How many possible energy levels are present in an atom?

**Checkpoint**

What does it mean that a level of an atom has a lifetime of \( 10^{-8} \text{ s} \)? What is a metastable level (or state) of an atom?

**Checkpoint**

What is the difference between spontaneous emission and stimulated emission?
6.5.4 Population Inversion

Consider two two-level energy systems representing an atom in an upper and lower state. Suppose that a photon of energy equal to the energy difference between the two levels approaches the two atoms, then which event is more likely to take place, absorption or stimulated emission? Einstein showed that under normal circumstances, both processes are equally probable. A larger population (number of atoms) in the upper level will result in stimulated emission dominating, while if there are more atoms in the lower level, there will be more absorption than stimulated emission. Normally, there are more atoms in the lower level, since in equilibrium such a system tends to minimize its energy (remember your chemistry!) Hence, under the normal thermodynamic circumstances absorption will dominate. However, if we can somehow put energy into the system in such a way that there are more atoms in the upper level, then stimulated emission will dominate. When we have such a situation, where we have put more atoms into the upper level than there are in the lower level, we have achieved population inversion.
6.5.5 Excitation of Atoms

Atoms can be excited by electrical discharge, radio frequency signals (RF), heat, chemical reactions, collisions with electrons, or collisions with other atoms. The excitation by collision (two different kinds) is the most relevant process in a gas laser.

1. Collisions of the first kind involve the interaction of an energetic electron with an atom in the ground state. The impact of the electron causes it to exchange some of its energy with the atom and the latter becomes excited. This process can be represented by an equation similar to the one used in chemical reactions, i.e.

\[
\text{He} + e_1 \rightarrow \text{He}^* + e_2
\]

(6.15)

the * signifies that the He is in an excited state. The kinetic energy of the electron before and after the collision is designated by \(e_1\) and \(e_2\) respectively. Since all electrons are identical, we cannot know whether two electrons exchange places in the collision; indeed, many argue that the distinction is meaningless.

2. A collision of the second kind involves the collision of an excited atom in a metastable state with another atom of a different element in an unexcited state, the final result being a transfer of energy from the metastable state to the unexcited atom, thus exciting it. The metastable atom therefore loses energy and reverts to a lower state. An example of a collision of the second kind is represented by the following equation,

\[
\text{He}^* + \text{Ne} \rightarrow \text{He} + \text{Ne}^*.
\]

(6.16)

The excited Ne atom is then ready to emit a photon and to return to its ground-state.

Checkpoint

How do you excite an atom?

6.6 The Helium-Neon Laser

Figure 6.13 shows juxtaposed the energy levels of He and Ne. Only the levels relevant to the operation of this laser are indicated. The boxes in the figure indicate groups of tightly spaced levels (energy bands) rather than sharp single ones, and the letter M indicates that the level is metastable (long lifetime).

The beam of light of the Helium-Neon Laser is produced by the stimulated emission of radiation produced in the decay of excited electrons in the 3S and 2S levels of neon. These levels of the neon atom, however, are not easily excited because the outer orbitals have a much greater probability of being if excited before them. Instead, energy supplied by an
electrical current (in the form of energetic electrons) is stored in the metastable levels of helium and later transferred by collisions of the second kind to neon. The two $^2\text{S}$ and $^2\text{S}$ metastable levels of helium are very close in energy to the 2S and 3S levels of neon which we wish to populate with excited electrons. Therefore, collisions between helium and neon (which are ensured by selecting a ratio of He to Ne gas of 10 to 1) leave the 2S and 3S levels of neon well populated. The observed light of the HeNe laser is generated only by the decay of 3S electrons to the 2P state (632.8 nm). Sustained operation of the laser is ensured by the He/Ne ratio and by the fact that the lower level (2P) decays 10 times faster (decay time $t = 10^{-8}$ s) than the upper (3S) one ($t = 10^{-7}$ s). (So that basically all electrons that decay to the 2P level undergo further decay). Light generated in any other forms of decay is suppressed (for example by inserting a prism into the cavity, which diverges the unwanted frequencies away from the axis of the laser tube).

![Figure 6.14: A schematic diagram of a resonant cavity filled with a plasma tube. The light gain medium in the cavity replaces cavity losses to sustain the laser oscillator.](image)

We have so far described how energy is pumped into the plasma to achieve population inversion. To have an operating laser we still need signal amplification and the proper feedback to sustain the oscillation. This will be discussed next.

**Checkpoint**

In a He-Ne laser, which gas emits the observed transition of 632.8 nm?

**Checkpoint**

How are the Ne atoms excited in a He-Ne laser?

**Checkpoint**

What is the meaning of “population inversion” when explaining the operation of a laser?
6.6.1 Amplification and feedback

We have indicated above that the observed light of the HeNe laser is generated only by the transition of 3S electrons to the 2P state (632.8 nm). In this process of spontaneous emission, a great part of the radiation propagates away from the cavity in arbitrary directions and is lost. (Though some will be absorbed by a 2P state of neon and will repopulate the 3S level.) A small portion, however, will strike other neon atoms (excited to the 3S level) and cause stimulated emission. Of the photons emitted here, however, only the small portion emitted in a direction perpendicular to the mirrors may stay in the cavity and contribute to the amplification process.

Upon successive reflections by the mirrors, light traveling through the laser medium is further coherently amplified. Since the HeNe laser light amplification is only 1.02 on each pass of the beam from one mirror to the other, all losses must be kept below 2%. Losses in the cavity formed by plane-parallel mirrors will be extremely sensitive to the parallelism of the two mirrors. Slight angular misalignment might result in the reflected beam missing the opposite mirror and lead to excessive loss and difficulties in attaining oscillation. Even under perfect parallelism, it can be shown that a resonator formed by plane-parallel mirrors has higher diffraction losses than a cavity formed by various combinations of plane and curved mirrors. Since the laser tube and the mirrors form an optical resonant cavity, the optical path length between successive reflections of the mirror must be an integral number of wavelengths to produce reinforcement of the wave. This arrangement of mirrors was familiar long before the advent of lasers and is known as the Fabry-Perot interferometer.

6.6.2 The Laser’s Mirrors

To produce an external laser beam, the mirror at the front of the laser tube is a partial reflector which reflects 99% of the light and transmits approximately 1%. The mirror at the
back end of the laser tube has a higher reflectivity and reflects about 99.9% of the light while transmitting less than 0.1%.

![Figure 6.15: A sketch of light propagating through an excited active gain medium. Spontaneous emissions with a radial component quickly leave the cavity after very little amplification; contrarily, the axial mode gets amplified continuously and repetitively.](image)

Laser mirrors are always made by successive deposition of materials with high and low refractive indices onto substrates of the highest optical quality. By making the optical thickness of each layer a quarter of a wavelength thick, very high reflectivities can be achieved. The mirror is the most expensive part of a laser – often good quality mirrors are made of 20 or more layers of dielectric material.

### Checkpoint

Why is the plasma tube of a gas laser enclosed by two parallel mirrors normal to the axis of the tube? How do you extract the light from a laser?

#### 6.6.3 Mode Configuration

A beam of light propagating down the optical axis of the laser will be amplified, and providing sufficient gain is available, will emerge from the laser output mirror. Generally speaking, a photon traveling at right angles to the optical axis will not be amplified to form a laser beam. Although it may cause many stimulated emissions, there will not be a large enough number to overcome all losses or, in other words, the gain will be inadequate. However, in the situation where a wave is traveling only slightly off axis, the wave may be able to zig-zag or “walk” between the mirrors a sufficient number of times to produce enough gain to overcome the losses. This produces the laser output, which consists of a variety of complicated patterns.
These patterns are the result of the laser operating in what is referred to as transverse modes. (Some light rays are not aligned on the laser’s axis.) Such modes are designated by the notation TEM$_{pq}$ modes (TEM standing for transverse electromagnetic).

The most commonly used mode is the TEM$_{00}$ mode. A laser operating in this mode has a Gaussian flux density pattern over the beam cross section. This pattern – a single disk area – produces a single spot of light.

**Checkpoint**

What does the beam spot look like for a laser operating in the TEM$_{00}$ mode?

### 6.6.4 References

Sections of this write up were taken from:
