RADICALLY MODERN INTRODUCTORY PHYSICS TEXT I



David J. Raymond
New Mexico Tech



New Mexico Tech

Radically Modern Introductory Physics Text I

David J. Raymond

This text is disseminated via the Open Education Resource (OER) LibreTexts Project (https://LibreTexts.org) and like the thousands of other texts available within this powerful platform, it is freely available for reading, printing, and "consuming."

The LibreTexts mission is to bring together students, faculty, and scholars in a collaborative effort to provide an accessible, and comprehensive platform that empowers our community to develop, curate, adapt, and adopt openly licensed resources and technologies; through these efforts we can reduce the financial burden born from traditional educational resource costs, ensuring education is more accessible for students and communities worldwide.

Most, but not all, pages in the library have licenses that may allow individuals to make changes, save, and print this book. Carefully consult the applicable license(s) before pursuing such effects. Instructors can adopt existing LibreTexts texts or Remix them to quickly build course-specific resources to meet the needs of their students. Unlike traditional textbooks, LibreTexts' web based origins allow powerful integration of advanced features and new technologies to support learning.



LibreTexts is the adaptable, user-friendly non-profit open education resource platform that educators trust for creating, customizing, and sharing accessible, interactive textbooks, adaptive homework, and ancillary materials. We collaborate with individuals and organizations to champion open education initiatives, support institutional publishing programs, drive curriculum development projects, and more.

The LibreTexts libraries are Powered by NICE CXone Expert and was supported by the Department of Education Open Textbook Pilot Project, the California Education Learning Lab, the UC Davis Office of the Provost, the UC Davis Library, the California State University Affordable Learning Solutions Program, and Merlot. This material is based upon work supported by the National Science Foundation under Grant No. 1246120, 1525057, and 1413739.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation nor the US Department of Education.

Have questions or comments? For information about adoptions or adaptions contact info@LibreTexts.org or visit our main website at https://LibreTexts.org.

This text was compiled on 06/09/2026

TABLE OF CONTENTS

Licensing

1: Waves in One Dimension

- 1.1: Transverse and Longitudinal Waves
- 1.2: Sine Waves
- 1.3: Types of Waves
- 1.4: Superposition Principle
- 1.5: Beats
- 1.6: Interferometers
- 1.7: Thin Films
- 1.8: Math Review — Derivatives
- 1.9: Group Velocity
- 1.10: Problems

2: Waves in Two and Three Dimensions

- 2.1: Math Tutorial — Vectors
- 2.2: Plane Waves
- 2.3: Superposition of Plane Waves
- 2.4: Diffraction Through a Single Slit
- 2.5: Two Slits
- 2.6: Diffraction Gratings
- 2.7: Problems

3: Geometrical Optics

- 3.1: Reflection and Refraction
- 3.2: Total Internal Reflection
- 3.3: Anisotropic Media
- 3.4: Thin Lens Equation and Optical Instruments
- 3.5: Fermat's Principle
- 3.6: Problems

4: Special Relativity

- 4.1: Galilean Spacetime Thinking
- 4.2: Spacetime Thinking in Special Relativity
- 4.3: Postulates of Special Relativity
- 4.4: Time Dilation
- 4.5: Lorentz Contraction
- 4.6: Twin Paradox
- 4.7: Problems

5: Applications of Special Relativity

- 5.1: Waves in Spacetime
- 5.2: Math Tutorial – Four-Vectors
- 5.3: Principle of Relativity Applied
- 5.4: Characteristics of Relativistic Waves
- 5.5: The Doppler Shift

- 5.6: Addition of Velocities
- 5.7: Problems

6: Acceleration and General Relativity

- 6.1: Acceleration
- 6.2: Circular Motion
- 6.3: Acceleration, Force, and Mass
- 6.4: Acceleration in Special Relativity
- 6.5: Accelerated Reference Frames
- 6.6: Gravitational Red Shift
- 6.7: Event Horizons
- 6.8: Problems

7: Matter Waves

- 7.1: Bragg's Law
- 7.2: X-Ray Diffraction Techniques
- 7.3: Meaning of Quantum Wave Function
- 7.4: Sense and Nonsense in Quantum Mechanics
- 7.5: Mass, Momentum, and Energy
- 7.6: Heisenberg Uncertainty Principle
- 7.7: Problems

8: Geometrical Optics and Newton's Laws

- 8.1: Fundamental Principles of Dynamics
- 8.2: Potential Energy
- 8.3: Work and Power
- 8.4: Mechanics and Geometrical Optics
- 8.5: Math Tutorial – Partial Derivatives
- 8.6: Motion in Two and Three Dimensions
- 8.7: Kinetic and Total Momentum
- 8.8: Problems

9: Symmetry and Bound States

- 9.1: Math Tutorial — Complex Waves
- 9.2: Symmetry and Quantum Mechanics
- 9.3: Confined Matter Waves
- 9.4: Problems

10: Dynamics of Multiple Particles

- 10.1: Momentum and Newton's Second Law
- 10.2: Newton's Third Law
- 10.3: Conservation of Momentum
- 10.4: Collisions
- 10.5: Rockets and Conveyor Belts
- 10.6: Problems

11: Rotational Dynamics

- 11.1: Math Tutorial — Cross Product
- 11.2: Torque and Angular Momentum

- 11.3: Two Particles
- 11.4: The Uneven Dumbbell
- 11.5: Many Particles
- 11.6: New Page
- 11.7: Statics
- 11.8: Problem

12: Harmonic Oscillator

- 12.1: Energy Analysis
- 12.2: Analysis Using Newton’s Laws
- 12.3: Forced Oscillator
- 12.4: Complex Exponential Solutions
- 12.5: Quantum Mechanical Harmonic Oscillator
- 12.6: Problems

13: Appendix A, Constants, Units, and Conversions

- 13.1: Appendix A Constants, Units, and Conversions

Index

Glossary

Detailed Licensing

Licensing

*A detailed breakdown of this resource's licensing can be found in **Back Matter/Detailed Licensing**.*

CHAPTER OVERVIEW

1: Waves in One Dimension

The wave is a universal phenomenon which occurs in a multitude of physical contexts. The purpose of this section is to describe the *kinematics* of waves, i. e., to provide tools for describing the form and motion of all waves irrespective of their underlying physical mechanisms.

Many examples of waves are well known to you. You undoubtedly know about ocean waves and have probably played with a stretched slinky toy, producing undulations which move rapidly along the slinky. Other examples of waves are sound, vibrations in solids, and light.

In this chapter we learn first about the basic properties of waves and introduce a special type of wave called the sine wave. Examples of waves seen in the real world are presented. We then learn about the superposition principle, which allows us to construct complex wave patterns by superimposing sine waves. Using these ideas, we discuss the related ideas of beats and interferometry. Finally, the ideas of wave packets and group velocity are introduced.

1.1: Transverse and Longitudinal Waves
1.2: Sine Waves
1.3: Types of Waves
1.4: Superposition Principle
1.5: Beats
1.6: Interferometers
1.7: Thin Films
1.8: Math Review — Derivatives
1.9: Group Velocity
1.10: Problems

This page titled 1: Waves in One Dimension is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

1.1: Transverse and Longitudinal Waves

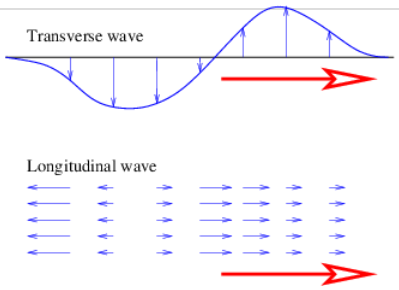


Figure 1.1.1:: Example of displacements in transverse and longitudinal waves. The wave motion is to the right as indicated by the large arrows. The small arrows indicate the displacements at a particular instant.

With the exception of light, waves are undulations in a material medium. For instance, ocean waves are (nearly) vertical undulations in the position of water parcels. The oscillations in neighboring parcels are phased such that a **pattern** moves across the ocean surface. Waves on a slinky are either **transverse** , in that the motion of the material of the slinky is perpendicular to the orientation of the slinky, or they are **longitudinal** , with material motion in the direction of the stretched slinky. (See Figure 1.1.1:.) Some media support only longitudinal waves, others support only transverse waves, while yet others support both types. Light waves are purely transverse, while sound waves are purely longitudinal. Ocean waves are a peculiar mixture of transverse and longitudinal, with parcels of water moving in elliptical trajectories as waves pass.

Light is a form of electromagnetic radiation. The undulations in an electromagnetic wave occur in the electric and magnetic fields. These oscillations are perpendicular to the direction of motion of the wave (in a vacuum), which is why we call light a transverse wave.

Waves

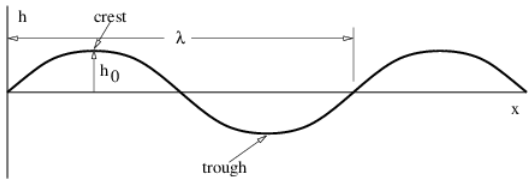


Figure 1.1.2:: Definition sketch for a sine wave, showing the wavelength λ and the amplitude h_0 and the phase ϕ at various points.

A particularly simple kind of wave, the sine wave, is illustrated in Figure 1.1.2:. This has the mathematical form

$$h(x) = h0sin(2\pi x / \lambda), \tag{1.1.1}$$

where h is the **displacement** (which can be either longitudinal or transverse), h_0 is the maximum displacement, also called the **amplitude** of the wave, and λ is the **wavelength** . The oscillatory behavior of the wave is assumed to carry on to infinity in both positive and negative x directions. Notice that the wavelength is the distance through which the sine function completes one full cycle. The **crest** and the **trough** of a wave are the locations of the maximum and minimum displacements, as seen in Figure 1.1.2:.

So far we have only considered a sine wave as it appears at a particular time. All interesting waves move with time. The movement of a sine wave to the right a distance d may be accounted for by replacing x in the above formula by x - d. If this movement occurs in time t, then the wave moves at velocity $c = d / t$. Solving this for d and substituting yields a formula for the displacement of a sine wave as a function of both distance x and time t :

$$h(x,t) = h0sin[2\pi(x - ct) / \lambda]. \tag{1.1.2}$$

The time for a wave to move one wavelength is called the **period** of the wave: $T = \lambda / c$. Thus, we can also write

$$h(x,t) = hsin[2\pi(x / \lambda - t / T)].0 \tag{1.1.3}$$

Physicists actually like to write the equation for a sine wave in a slightly simpler form. Defining the **wavenumber** as $k = 2\pi / \lambda$ and the **angular frequency** as $\omega = 2\pi / T$, we write

$$h(x,t) = h0sin(kx - \omega t). \tag{1.1.4}$$

We normally think of the frequency of oscillatory motion as the number of cycles completed per second. This is called the **rotational frequency** , and is given by $f = 1/T$. It is related to the angular frequency by $\omega = 2\pi f$. The rotational frequency is

usually easier to measure than the angular frequency, but the angular frequency tends to be used more often in theoretical discussions. As shown above, converting between the two is not difficult. Rotational frequency is measured in units of hertz, abbreviated Hz; 1 Hz = 1 cycle s^{-1}. Angular frequency also has the dimensions of inverse time, e. g., radian s^{-1}, but the term “hertz” is generally reserved only for rotational frequency.

The argument of the sine function is by definition an angle. We refer to this angle as the **phase** of the wave, $\phi = kx - \omega t$. The difference in the phase of a wave at fixed time over a distance of one wavelength is 2π, as is the difference in phase at fixed position over a time interval of one wave period.

Since angles are dimensionless, we normally don’t include this in the units for frequency. However, it sometimes clarifies things to refer to the dimensions of rotational frequency as “rotations per second” or angular frequency as “radians per second”.

As previously noted, we call h_0, the maximum displacement of the wave, the amplitude. Often we are interested in the **intensity** of a wave, which is proportional to the square of the amplitude. The intensity is often related to the amount of energy being carried by a wave.

The wave speed we have defined above, $c = \lambda / T$, is actually called the **phase speed** . Since $\lambda = 2\pi / k$ and $T = 2\pi / \omega$, we can write the phase speed in terms of the angular frequency and the wavenumber:

$$c = \omega - (phasespeed).k \tag{1.1.5}$$

Types of Waves

In order to make the above material more concrete, we now examine the characteristics of various types of waves which may be observed in the real world.

Ocean Surface Waves

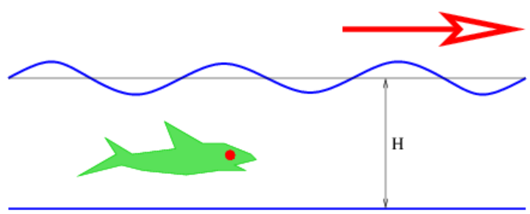


Figure 1.1.3:: Wave on an ocean of depth H. The wave is moving to the right and the particles of water at the surface oscillate in elliptical trajectories as the wave crests and troughs pass.

These waves are manifested as undulations of the ocean surface as seen in Figure 1.1.3:. The speed of ocean waves is given by the formula

$$()1/2gtanh - (kH-) - c = k, \tag{1.1.6}$$

where $g = 9.8$ m s^{-2} is the earth’s gravitational force per unit mass, H is the depth of the ocean, and the hyperbolic tangent is defined as[1]

$$exp(x) - exp(-x)tanh(x) = \frac{------------------}{.exp(x) + exp(-x)} \tag{1.1.7}$$

The equation for the speed of ocean waves comes from the theory for oscillations of a fluid surface in a gravitional field.

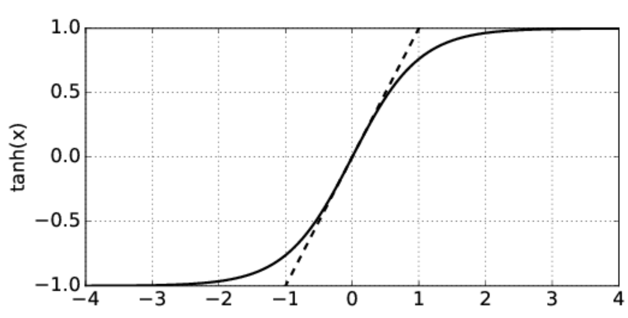


Figure 1.1.4:: Plot of the function tanh(x). The dashed line shows our approximation tanh(x) ≈ x for $|x| \ll 1$.

https://phys.libretexts.org/@go/page/32918

As Figure 1.1.4: shows, for $|x| \ll 1$, we can approximate the hyperbolic tangent by $\tanh(x) \approx x$, while for $|x| \gg 1$ it is +1 for $x > 0$ and -1 for $x < 0$. This leads to two limits: Since $x = kH$, the **shallow water** limit, which occurs when $kH \ll 1$, yields a wave speed of

$$c \approx (gH)1 / 2, (shallowwaterwaves), \tag{1.1.8}$$

while the **deep water** limit, which occurs when $kH \gg 1$, yields

$$c \approx (g / k)1 / 2, (deepwaterwaves). \tag{1.1.9}$$

Notice that the speed of shallow water waves depends only on the depth of the water and on g. In other words, all shallow water waves move at the same speed. On the other hand, deep water waves of longer wavelength (and hence smaller wavenumber) move more rapidly than those with shorter wavelength. Waves for which the wave speed varies with wavelength are called **dispersive** . Thus, deep water waves are dispersive, while shallow water waves are non-dispersive.

For water waves with wavelengths of a few centimeters or less, surface tension becomes important to the dynamics of the waves. In the deep water case, the wave speed at short wavelengths is given by the formula

$$c = (g / k + Ak)1 / 2 \tag{1.1.10}$$

where the constant A is related to an effect called surface tension. For an air-water interface near room temperature, $A \approx 74$ cm^3 s^{-2}.

Sound Waves

Sound is a longitudinal compression-expansion wave in a fluid. The wave speed for sound in an ideal gas is

$$1 / 2c = (\gamma RTabs) \tag{1.1.11}$$

where γ and R are constants and T_{abs} is the **absolute temperature** . The absolute temperature is measured in Kelvins and is numerically given by

$$Tabs = TC + 273\circ \tag{1.1.12}$$

where T_C is the temperature in Celsius degrees. The angular frequency of sound waves is thus given by

$$\omega = ck = (\gamma RTabs)1 / 2k. \tag{1.1.13}$$

The speed of sound in air at normal temperatures is about 340 m s^{-1}.

Light

Light moves in a vacuum at a speed of $c_{vac} = 3 \times 10^8$ m s^{-1}. In transparent materials it moves at a speed less than c_{vac} by a factor n which is called the **refractive index** of the material:

$$c = c / n.vac \tag{1.1.14}$$

Often the refractive index takes the form

$$n2 \approx 1 + - - - - A - - - - - - -, 1 - (k / kR)2 \tag{1.1.15}$$

where k is the wavenumber and k_R and A are positive constants characteristic of the material. The angular frequency of light in a transparent medium is thus

$$\omega = kc = kcvac / n. \tag{1.1.16}$$

Superposition Principle

It is found empirically that as long as the amplitudes of waves in most media are small, two waves in the same physical location don't interact with each other. Thus, for example, two waves moving in the opposite direction simply pass through each other without their shapes or amplitudes being changed. When collocated, the total wave displacement is just the sum of the displacements of the individual waves. This is called the **superposition principle** . At sufficiently large amplitude the superposition principle often breaks down — interacting waves may scatter off of each other, lose amplitude, or change their form.

Interference is a consequence of the superposition principle. When two or more waves are superimposed, the net wave displacement is just the algebraic sum of the displacements of the individual waves. Since these displacements can be positive or negative, the net displacement can either be greater or less than the individual wave displacements. The former case, which occurs

when both displacements are of the same sign, is called **constructive interference** , while **destructive interference** occurs when they are of opposite sign.

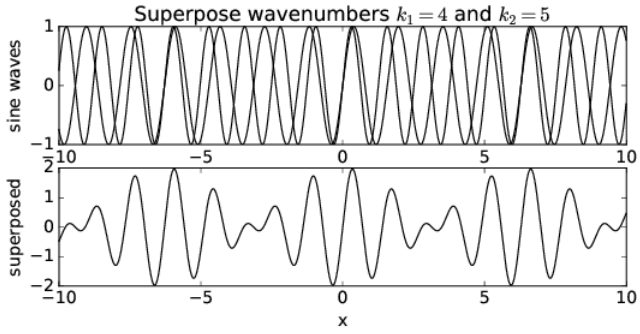


Figure 1.1.5:: Superposition (lower panel) of two sine waves (shown individually in the upper panel) with equal amplitudes and wavenumbers $k_1 = 4$ and $k_2 = 5$.

Let us see what happens when we superimpose two sine waves with different wavenumbers. Figure 1.1.5: shows the superposition of two waves with wavenumbers $k_1 = 4$ and $k_2 = 5$. Notice that the result is a wave with about the same wavelength as the two initial waves, but which varies in amplitude depending on whether the two sine waves are interfering constructively or destructively. We say that the waves are **in phase** if they are interfering constructively, and they are **out of phase** if they are interfering destructively.

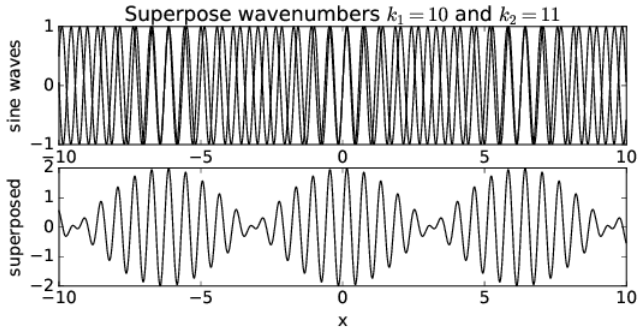


Figure 1.1.6:: Superposition of two sine waves with equal amplitudes and wavenumbers $k_1 = 10$ and $k_2 = 11$.

What happens when the wavenumbers of the two sine waves are changed? Figure 1.1.6: shows the result when $k_1 = 10$ and $k_2 = 11$. Notice that though the wavelength of the resultant wave is decreased, the locations where the amplitude is maximum have the same separation in x as in Figure 1.1.5:.

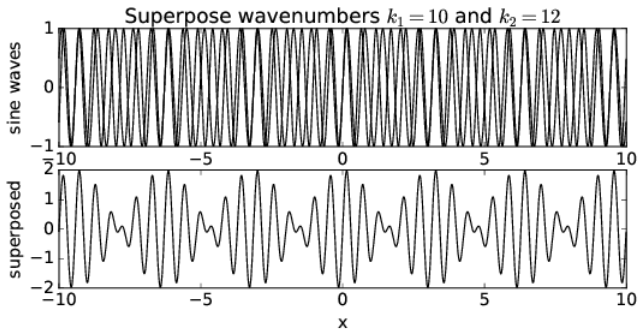


Figure 1.1.7:: Superposition of two sine waves with equal amplitudes and wavenumbers $k_1 = 10$ and $k_2 = 12$.

If we superimpose waves with $k_1 = 10$ and $k_2 = 12$, as is shown in Figure 1.1.7:, we see that the x spacing of the regions of maximum amplitude has decreased by a factor of two. Thus, while the wavenumber of the resultant wave seems to be related to something like the **average** of the wavenumbers of the component waves, the spacing between regions of maximum wave amplitude appears to go inversely with the **difference** of the wavenumbers of the component waves. In other words, if k_1 and k_2 are close together, the amplitude maxima are far apart and vice versa.

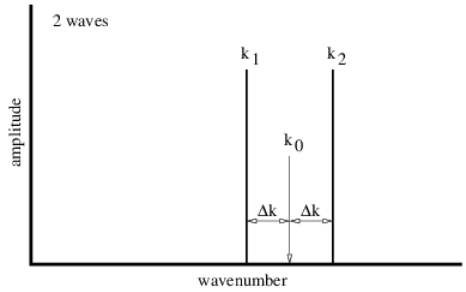


Figure 1.1.8:: Representation of the wavenumbers and amplitudes of two superimposed sine waves.

We can symbolically represent the sine waves that make up figures 1.5, 1.6, and 1.7 by a plot such as that shown in Figure 1.1.8:. The amplitudes and wavenumbers of each of the sine waves are indicated by vertical lines in this figure.

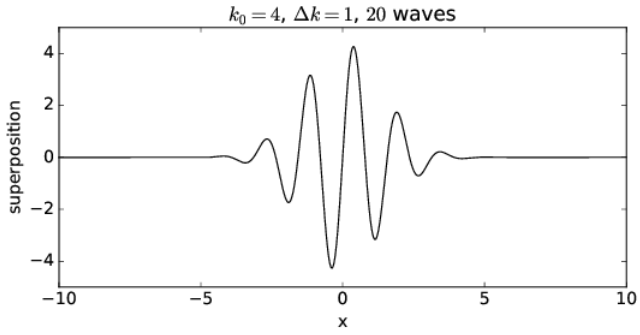


Figure 1.1.9:: Superposition of twenty sine waves with $k_0 = 4$ and $\Delta k = 1$.

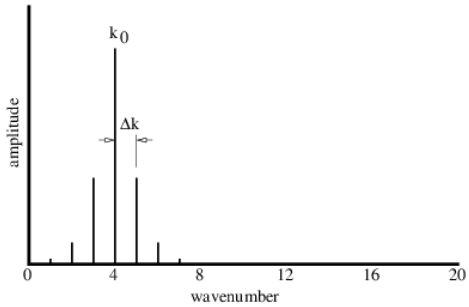


Figure 1.1.10:: Representation of the distribution of wavenumbers and amplitudes of 20 superimposed sine waves with maximum at $k_0 = 4$ and half-width $\Delta k = 1$.

The regions of large wave amplitude are called **wave packets** . Wave packets will play a central role in what is to follow, so it is important that we acquire a good understanding of them. The wave packets produced by only two sine waves are not well separated along the x -axis. However, if we superimpose many waves, we can produce an isolated wave packet. For example, Figure 1.1.9: shows the results of superimposing 20 sine waves with wavenumbers $k = 0.4m$, $m = 1, 2, \ldots, 20$, where the amplitudes of the waves are largest for wavenumbers near $k = 4$. In particular, we assume that the amplitude of each sine wave is proportional to $\exp[-(k - k_0)^2 / \Delta k^2]$, where $k_0 = 4$ defines the maximum of the distribution of wavenumbers and $\Delta k = 1$ defines the half-width of this distribution. The amplitudes of each of the sine waves making up the wave packet in Figure 1.1.9: are shown schematically in Figure 1.1.10:.

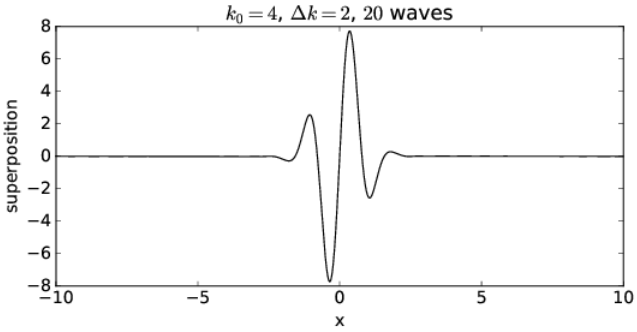


Figure 1.1.11:: Superposition of twenty sine waves with $k_0 = 4$ and $\Delta k = 2$.

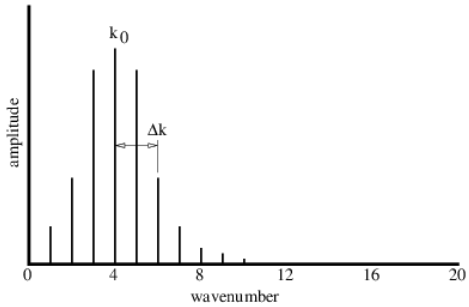


Figure 1.1.12:: Representation of the distribution of wavenumbers and amplitudes of 20 superimposed sine waves with maximum at $k_0 = 4$ and half-width $\Delta k = 2$.

The quantity Δk controls the distribution of the sine waves being superimposed — only those waves with a wavenumber k within approximately Δk of the **central wavenumber** k_0 of the wave packet, i. e., for $3 \le k \le 5$ in this case, contribute significantly to the sum. If Δk is changed to 2, so that wavenumbers in the range $2 \le k \le 6$ contribute significantly, the wavepacket becomes narrower, as is shown in figures 1.11 and 1.12. Δk is called the **wavenumber spread** of the wave packet, and it evidently plays a role similar to the difference in wavenumbers in the superposition of two sine waves — the larger the wavenumber spread, the smaller the physical size of the wave packet. Furthermore, the wavenumber of the oscillations within the wave packet is given approximately by the central wavenumber.

We can better understand how wave packets work by mathematically analyzing the simple case of the superposition of two sine waves. Let us define $k_0 = (k_1 + k_2)/2$ where k_1 and k_2 are the wavenumbers of the component waves. Furthermore let us set $\Delta k = (k_2 - k_1)/2$. The quantities k_0 and Δk are graphically illustrated in Figure 1.1.8:. We can write $k_1 = k_0 - \Delta k$ and $k_2 = k_0 + \Delta k$ and use the trigonometric identity $\sin(a + b) = \sin(a)\cos(b) + \cos(a)\sin(b)$ to find

$$\begin{aligned} sin(k1x) + sin(k2x) = sin[(k0 - \Delta k)x] + sin[(k0 + \Delta k)x] = sin(k0x)cos(\Delta kx) \\ - cos(k0x)sin(\Delta kx) + sin(k0x)cos(\Delta kx) + cos(k0x)sin(\Delta kx) \\ = 2sin(k0x)cos(\Delta kx). \end{aligned} \tag{1.1.17}$$

The sine factor on the bottom line of the above equation produces the oscillations within the wave packet, and as speculated earlier, this oscillation has a wavenumber k_0 equal to the average of the wavenumbers of the component waves. The cosine factor modulates this wave with a spacing between regions of maximum amplitude of

$$\Delta x = \pi / \Delta k. \tag{1.1.18}$$

Thus, as we observed in the earlier examples, the length of the wave packet Δx is inversely related to the spread of the wavenumbers Δk (which in this case is just the difference between the two wavenumbers) of the component waves. This relationship is central to the uncertainty principle of quantum mechanics.

Beats

Suppose two sound waves of different frequency but equal amplitude impinge on your ear at the same time. The displacement perceived by your ear is the superposition of these two waves, with time dependence

$$h(t) = sin(\omega t) + sin(\omega t) = 2sin(\omega t)cos(\Delta\omega t), 120 \tag{1.1.19}$$

where we have used the above math trick, and where $\omega_0 = (\omega_1 + \omega_2)/2$ and $\Delta\omega = (\omega_2 - \omega_1)/2$. What you actually hear is a tone with angular frequency ω_0 which fades in and out with period

$$T = \pi / |\Delta\omega| = 2\pi / |\omega - \omega| = 1 / |f - f|.beat2121 \tag{1.1.20}$$

The **beat frequency** is simply

$$fbeat = 1 / Tbeat = |f2 - f1|. \tag{1.1.21}$$

Note how beats are the time analog of wave packets — the mathematics are the same except that frequency replaces wavenumber and time replaces space.

Interferometers

An interferometer is a device which splits a beam of light (or other wave) into two sub-beams, shifts the phase of one sub-beam with respect to the other, and then superimposes the sub-beams so that they interfere constructively or destructively, depending on the magnitude of the phase shift between them. In this section we study the Michelson interferometer and interferometric effects in thin films.

Michelson Interferometer

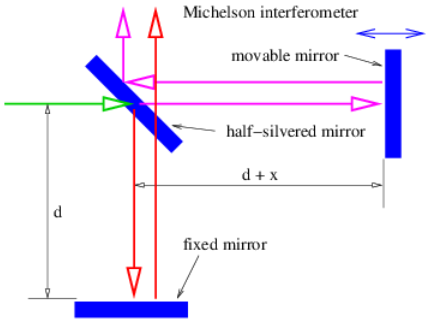


Figure 1.1.13:: Sketch of a Michelson interferometer.

The American physicist Albert Michelson invented the optical interferometer illustrated in Figure 1.1.13:. The incoming beam is split into two beams by the half-silvered mirror. Each sub-beam reflects off of another mirror which returns it to the half-silvered mirror, where the two sub-beams recombine as shown. One of the reflecting mirrors is movable by a sensitive micrometer device, allowing the path length of the corresponding sub-beam, and hence the phase relationship between the two sub-beams, to be altered. As Figure 1.1.13: shows, the difference in path length between the two sub-beams is $2x$ because the horizontal sub-beam traverses the path twice. Thus, constructive interference occurs when this path difference is an integral number of wavelengths, i. e.,

$$2x = m\lambda, m = 0, \pm1, \pm2, ...(Michelsoninterferometer) \tag{1.1.22}$$

where λ is the wavelength of the wave and m is an integer. Note that m is the number of wavelengths that fits evenly into the distance $2x$.

Films

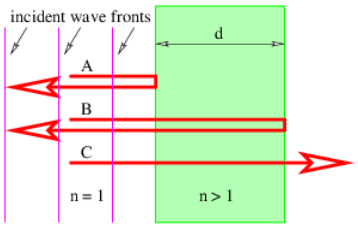


Figure 1.1.14:: Plane light wave normally incident on a transparent thin film of thickness d and index of refraction $n > 1$. Partial reflection occurs at the front surface of the film, resulting in beam A, and at the rear surface, resulting in beam B. Much of the wave passes completely through the film, as with C.

One of the most revealing examples of interference occurs when light interacts with a thin film of transparent material such as a soap bubble. Figure 1.1.14: shows how a plane wave normally incident on the film is partially reflected by the front and rear surfaces. The waves reflected off the front and rear surfaces of the film interfere with each other. The interference can be either constructive or destructive depending on the phase difference between the two reflected waves.

If the wavelength of the incoming wave is λ, one would naively expect constructive interference to occur between the A and B beams if $2d$ were an integral multiple of λ.

Two factors complicate this picture. First, the wavelength inside the film is **not** λ, but λ / n, where n is the index of refraction of the film. Constructive interference would then occur if $2d = m\lambda / n$. Second, it turns out that an additional phase shift of half a wavelength occurs upon reflection when the wave is incident on material with a higher index of refraction than the medium in which the incident beam is immersed. This phase shift doesn't occur when light is reflected from a region with lower index of refraction than felt by the incident beam. Thus beam B doesn't acquire any additional phase shift upon reflection. As a consequence, constructive interference actually occurs when

$$2d = (m + 1 / 2)\lambda / n, m = 0, 1, 2, ...(constructiveinterference) \tag{1.1.23}$$

while destructive interference results when

$$2d = m\lambda / n, m = 0, 1, 2, ...(destructiveinterference). \tag{1.1.24}$$

When we look at a soap bubble, we see bands of colors reflected back from a light source. What is the origin of these bands? Light from ordinary sources is generally a mixture of wavelengths ranging from roughly $\lambda = 4.5 \times 10^{-7}$ m (violet light) to $\lambda = 6.5 \times 10^{-7}$ m (red light). In between violet and red we also have blue, green, and yellow light, in that order. Because of the different wavelengths associated with different colors, it is clear that for a mixed light source we will have some colors interfering constructively while others interfere destructively. Those undergoing constructive interference will be visible in reflection, while those undergoing destructive interference will not.

Another factor enters as well. If the light is not normally incident on the film, the difference in the distances traveled between beams reflected off of the front and rear faces of the film will not be just twice the thickness of the film. To understand this case quantitatively, we need the concept of refraction, which will be developed later in the context of geometrical optics. However, it should be clear that different wavelengths will undergo constructive interference for different angles of incidence of the incoming light. Different portions of the thin film will in general be viewed at different angles, and will therefore exhibit different colors under reflection, resulting in the colorful patterns normally seen in soap bubbles.

Review — Derivatives

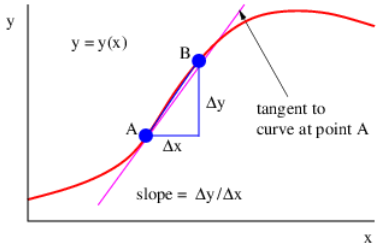


Figure 1.1.15:: Estimation of the derivative, which is the slope of the tangent line. When point B approaches point A, the slope of the line AB approaches the slope of the tangent to the curve at point A.

This section provides a quick review of the idea of the derivative. Often we are interested in the slope of a line tangent to a function y (x) at some value of x. This slope is called the **derivative** and is denoted dy / dx. Since a tangent line to the function can be defined at any point x, the derivative itself is a function of x :

$$dy(x)g(x) = - - - - - .dx \tag{1.1.25}$$

As Figure 1.1.15: illustrates, the slope of the tangent line at some point on the function may be approximated by the slope of a line connecting two points, A and B, set a finite distance apart on the curve:

$$dy - \approx \Delta y - .dx\Delta x \tag{1.1.26}$$

As B is moved closer to A, the approximation becomes better. In the limit when B moves infinitely close to A, it is exact.

Derivatives of some common functions are now given. In each case a is a constant.

$$adx - - = axa - 1dx \tag{1.1.27}$$

(1.27)

$$d - -exp(ax) = aexp(ax)dx \tag{1.1.28}$$

$$d1---log(ax) = --dxx \tag{1.1.29}$$

$$d---sin(ax) = acos(ax)dx \tag{1.1.30}$$

(1.30)

$$ddx - cos(ax) = -asin(ax) \tag{1.1.31}$$

$$daf(x)df(x) --------- = a-------dxdx \tag{1.1.32}$$

$$ddf(x)dg(x)---[f(x)+g(x)] = ------+-------dxdxdx \tag{1.1.33}$$

$$\begin{gathered} ddf(x)dg(x)--f(x)g(x) \\ = -----g(x)+f(x)-----(productrule)dxdxdx \end{gathered} \tag{1.1.34}$$

$$ddfdy--f(y) = -----(chainrule)dxdydx \tag{1.1.35}$$

The product and chain rules are used to compute the derivatives of complex functions. For instance,

$$\begin{gathered} ddsin(x)dcos(x)---(sin(x)cos(x)) \\ = ---------cos(x)+sin(x)--------- \\ = cos2(x)-sin2(x)dxdxdx \end{gathered} \tag{1.1.36}$$

and

$$-d-log(sin(x)) = --1---dsin(x)- = cos(x-).dxsin(x)dxsin(x) \tag{1.1.37}$$

Group Velocity

We now ask the following question: How fast do wave packets move? Surprisingly, we often find that wave packets move at a speed very different from the phase speed, $c = \omega / k$, of the wave composing the wave packet.

We shall find that the speed of motion of wave packets, referred to as the **group velocity** , is given by

$$|d\omega-||u = dk||(groupvelocity).k = k0 \tag{1.1.38}$$

The derivative of ω (k) with respect to k is first computed and then evaluated at $k = k_0$, the central wavenumber of the wave packet of interest.

The relationship between the angular frequency and the wavenumber for a wave, $\omega = \omega$ (k), depends on the type of wave being considered. Whatever this relationship turns out to be in a particular case, it is called the **dispersion relation** for the type of wave in question.

As an example of a group velocity calculation, suppose we want to find the velocity of deep ocean wave packets for a central wavelength of $\lambda_0 = 60$ m. This corresponds to a central wavenumber of $k_0 = 2\pi / \lambda_0 \approx 0.1$ m^{-1}. The phase speed of deep ocean waves is $c = (g / k)^{1/2}$. However, since $c \equiv \omega / k$, we find the frequency of deep ocean waves to be $\omega = (gk)^{1/2}$. The group velocity is therefore $u \equiv d\omega / dk = (g / k)^{1/2} / 2 = c / 2$. For the specified central wavenumber, we find that $u \approx (9.8$ m s$^{-2} / 0.1$ m$^{-1})^{1/2} / 2 \approx 5$ m s^{-1}. By contrast, the phase speed of deep ocean waves with this wavelength is $c \approx 10$ m s^{-1}.

Dispersive waves are waves in which the phase speed varies with wavenumber. It is easy to show that dispersive waves have unequal phase and group velocities, while these velocities are equal for non-dispersive waves.

Derivation of Group Velocity Formula

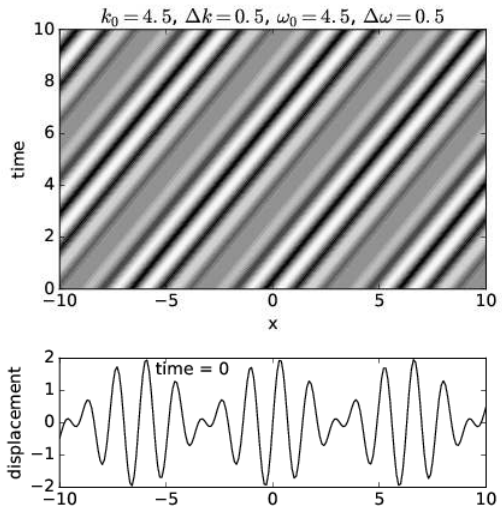


Figure 1.1.16:: Upper panel: Net displacement of the sum of two traveling sine waves plotted in the x - t plane. The white indicates where the displacement is large and positive, while the black indicates where it is large and negative. Non-dispersive case with $u_p = u_g = 1$.. Lower panel: Plot of wave displacement as a function of x at time $t = 0$.

We now derive equation 1.36}. It is easiest to do this for the simplest wave packets, namely those constructed out of the superposition of just two sine waves. We will proceed by adding two waves with full space and time dependence:

$$h = sin(kx - \omega t) + sin(kx - \omega t)1122 \tag{1.1.39}$$

After algebraic and trigonometric manipulations familiar from earlier sections, we find

$$h = 2sin(k0x - \omega 0t)cos(\Delta kx - \Delta\omega t), \tag{1.1.40}$$

where as before we have $k_0 = (k_1 + k_2)/2$, $\omega_0 = (\omega_1 + \omega_2)/2$, $\Delta k = (k_2 - k_1)/2$, and $\Delta\omega = (\omega_2 - \omega_1)/2$.

Again think of this as a sine wave of frequency ω_0 and wavenumber k_0 modulated by a cosine function. In this case the modulation pattern moves with a speed so as to keep the argument of the cosine function constant:

$$\Delta kx - \Delta\omega t = const. \tag{1.1.41}$$

Differentiating this with respect to t while holding Δk and $\Delta\omega$ constant yields

$$dx - \Delta - \omega - u \equiv dt = \Delta k. \tag{1.1.42}$$

In the limit in which the deltas become very small, this reduces to the derivative

$$d\omega u = --, dk \tag{1.1.43}$$

which is the desired result.

Examples

We now illustrate some examples of phase speed and group velocity by showing the displacement resulting from the superposition of two sine waves, as given by equation 1.38}, in the x -t plane. This is an example of a spacetime diagram, of which we will see many examples later on.

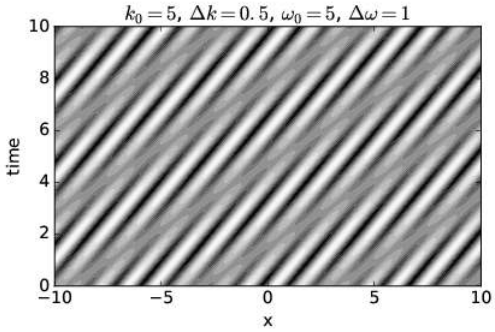


Figure 1.1.17:: As in the upper panel of Figure 1.1.16: except a dispersive case with $u_p = 1$, $u_g = 2$.

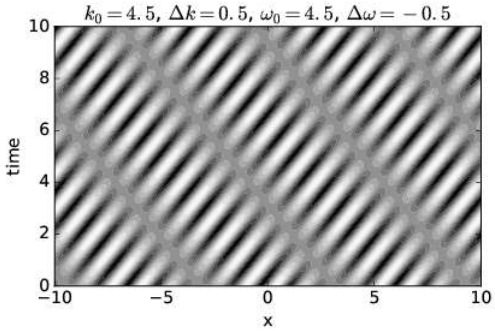


Figure 1.1.18:: As in the upper panel of Figure 1.1.16: except a dispersive case with phase and group velocities in opposite directions, $u_p = 1$, $u_g = -1$.

The upper panel of Figure 1.1.16: shows a non-dispersive case in which the phase speed equals the group velocity. The white and black regions indicate respectively strong wave crests and troughs (i. e., regions of large positive and negative displacements), with grays indicating a displacement near zero. Regions with large displacements indicate the location of wave packets. The positions of waves and wave packets at any given time may therefore be determined by drawing a horizontal line across the graph at the desired time and examining the variations in wave displacement along this line. The lower panel of this figure shows the wave displacement as a function of x at time $t = 0$ as an aid to interpretation of the upper panel.

Notice that as time increases, the crests move to the right. This corresponds to the motion of the waves within the wave packets. Note also that the wave packets, i. e., the broad regions of large positive and negative amplitudes, move to the right with increasing time as well.

Since velocity is distance moved Δx divided by elapsed time Δt, the slope of a line in Figure 1.1.16:, $\Delta t / \Delta x$, is one over the velocity of whatever that line represents. The slopes of lines representing crests are the same as the slopes of lines representing wave packets in this case, which indicates that the two move at the same velocity. Since the speed of movement of wave crests is the phase speed and the speed of movement of wave packets is the group velocity, the two velocities are equal and the non-dispersive nature of this case is confirmed.

Figure 1.1.17: shows a dispersive wave in which the group velocity is twice the phase speed, while Figure 1.1.18: shows a case in which the group velocity is actually opposite in sign to the phase speed. See if you can confirm that the phase and group velocities seen in each figure correspond to the values for these quantities calculated from the specified frequencies and wavenumbers.

Problems

1. Measure your pulse rate. Compute the ordinary frequency of your heart beat in cycles per second. Compute the angular frequency in radians per second. Compute the period.
2. An important wavelength for radio waves in radio astronomy is 21 cm. (This comes from neutral hydrogen.) Compute the wavenumber of this wave. Compute the ordinary and angular frequencies. (The speed of light is 3×10^8 m s^{-1}.)
3. Sketch the resultant wave obtained from superimposing the waves $A = \sin(2x)$ and $B = \sin(3x)$. By using the trigonometric identity given in equation 1.17}, obtain a formula for $A + B$ in terms of $\sin(5x/2)$ and $\cos(x/2)$. Does the wave obtained from sketching this formula agree with your earlier sketch?
4. }

5. Two sine waves with wavelengths λ_1 and λ_2 are superimposed, making wave packets of length L. If we wish to make L larger, should we make λ_1 and λ_2 closer together or farther apart? Explain your reasoning.
6. By examining Figure 1.1.9: versus Figure 1.1.10: and then Figure 1.1.11: versus Figure 1.1.12:, determine whether equation 1.18} works at least in an approximate sense for isolated wave packets.
7. }
8. The frequencies of the chromatic scale in music are given by
9. $$fi = f02i \,/\, 12, i = 0, 1, 2, ..., 11, ''class = ''math - display'' src = ''/book143x.png \tag{1.1.44}$$

where f_0 is a constant equal to the frequency of the lowest note in the scale.

1. Compute f_1 through f_{11} if f_0 = 440 Hz (the "A" note).
2. Using the above results, what is the beat frequency between the "A" (i = 0) and "B" (i = 2) notes? (The frequencies are given here in cycles per second rather than radians per second.)
3. Which pair of the above frequencies f_0 - f_{11} yields the smallest beat frequency? Explain your reasoning.

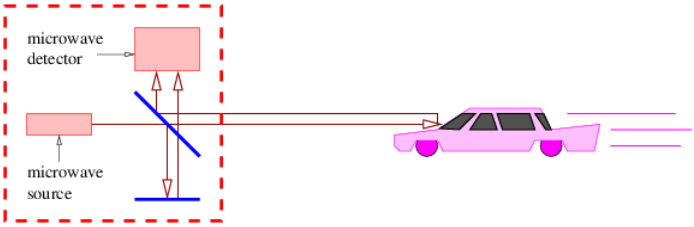


Figure 1.1.19:: Sketch of a police radar.

---

10. Large ships in general cannot move faster than the phase speed of surface waves with a wavelength equal to twice the ship's length. This is because most of the propulsive force goes into making big waves under these conditions rather than accelerating the ship.
    1. How fast can a 300 m long ship move in very deep water?
    2. As the ship moves into shallow water, does its maximum speed increase or decrease? Explain.
11. Given the formula for refractive index of light quoted in this section, for what range of k does the phase speed of light in a transparent material take on real values which exceed the speed of light in a vacuum?
12. A police radar works by splitting a beam of microwaves, part of which is reflected back to the radar from your car where it is made to interfere with the other part which travels a fixed path, as shown in Figure 1.1.19:.
    1. If the wavelength of the microwaves is λ, how far do you have to travel in your car for the interference between the two beams to go from constructive to destructive to constructive?
    2. If you are traveling toward the radar at speed v = 30 m s^{-1}, use the above result to determine the number of times per second constructive interference peaks will occur. Assume that λ = 3 cm.

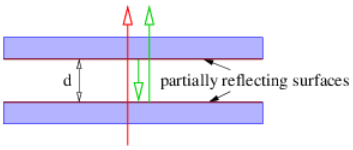


Figure 1.1.20:: Sketch of a Fabry-Perot interferometer.

---

13. Suppose you know the wavelength of light passing through a Michelson interferometer with high accuracy. Describe how you could use the interferometer to measure the length of a small piece of material.
14. A Fabry-Perot interferometer (see Figure 1.1.20:) consists of two parallel half-silvered mirrors placed a distance d from each other as shown. The beam passing straight through interferes with the beam which reflects once off of both of the mirrored surfaces as shown. For wavelength λ, what values of d result in constructive interference?
15. A Fabry-Perot interferometer has spacing d = 2 cm between the glass plates, causing the direct and doubly reflected beams to interfere (see Figure 1.1.20:). As air is pumped out of the gap between the plates, the beams go through 23 cycles of

constructive-destructive-constructive interference. If the wavelength of the light in the interfering beams is 5×10^{-7} m, determine the index of refraction of the air initially in the interferometer.

16. Measurements on a certain kind of wave reveal that the angular frequency of the wave varies with wavenumber as shown in the following table:
ω (s^{-1}) k (m^{-1}) 5 1 20 2 45 3 80 4 125 5

    1. Compute the phase speed of the wave for $k = 3$ m^{-1} and for $k = 4$ m^{-1}.
    2. Estimate the group velocity for $k = 3.5$ m^{-1} using a finite difference approximation to the derivative.

17. Suppose some type of wave has the (admittedly weird) dispersion relation shown in Figure 1.1.21:.

    1. For what values of k is the phase speed of the wave positive?
    2. For what values of k is the group velocity positive?

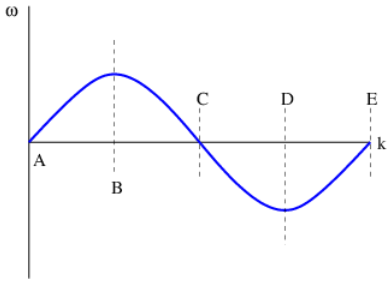


Figure 1.1.21:: Sketch of a weird dispersion relation.

---

18. Group velocities of various waves.

    1. Compute the group velocity for shallow water waves. Compare it with the phase speed of shallow water waves. (Hint: You first need to derive a formula for ω (k) from c (k).)
    2. Repeat the above problem for deep water waves.
    3. Repeat for sound waves. What does this case have in common with shallow water waves?

Chapter 2 Waves in Two and Three Dimensions

In this chapter we extend the ideas of the previous chapter to the case of waves in more than one dimension. The extension of the sine wave to higher dimensions is the **plane wave** . Wave packets in two and three dimensions arise when plane waves moving in different directions are superimposed.

Diffraction results from the disruption of a wave which is impingent upon an object. Those parts of the wave front hitting the object are scattered, modified, or destroyed. The resulting **diffraction pattern** comes from the subsequent interference of the various pieces of the modified wave. A knowledge of diffraction is necessary to understand the behavior and limitations of optical instruments such as telescopes.

Diffraction and interference in two and three dimensions can be manipulated to produce useful devices such as the **diffraction grating** .

Tutorial — Vectors

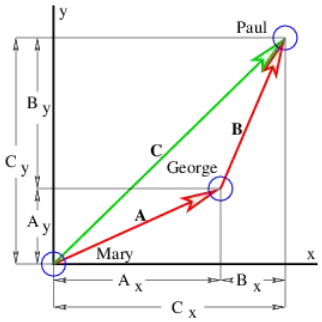


Figure 1.1.1:: Displacement vectors in a plane. Vector A represents the displacement of George from Mary, while vector B represents the displacement of Paul from George. Vector C represents the displacement of Paul from Mary and $C = A + B$. The quantities A_x, A_y, etc., represent the Cartesian components of the vectors.

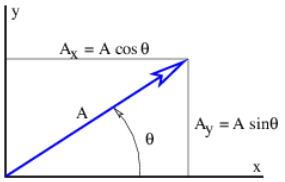


Figure 1.1.2:: Definition sketch for the angle θ representing the orientation of a two dimensional vector.

Before we can proceed further we need to explore the idea of a **vector** . A vector is a quantity which expresses both magnitude and direction. Graphically we represent a vector as an arrow. In typeset notation a vector is represented by a boldface character, while in handwriting an arrow is drawn over the character representing the vector.

Figure 1.1.1: shows some examples of **displacement vectors** , i. e., vectors which represent the displacement of one object from another, and introduces the idea of vector addition. The tail of vector B is collocated with the head of vector A, and the vector which stretches from the tail of A to the head of B is the sum of A and B, called C in Figure 1.1.1:.

The quantities A_x, A_y, etc., represent the **Cartesian components** of the vectors in Figure 1.1.1:. A vector can be represented either by its Cartesian components, which are just the projections of the vector onto the Cartesian coordinate axes, or by its direction and magnitude. The direction of a vector in two dimensions is generally represented by the counterclockwise angle of the vector relative to the x axis, as shown in Figure 1.1.2:. Conversion from one form to the other is given by the equations

$$A = (A2x + A2y)1 / 2\theta = tan - 1(Ay / Ax), \tag{1.1.45}$$

where A is the magnitude of the vector. A vector magnitude is sometimes represented by absolute value notation: $A \equiv |A|$.

Notice that the inverse tangent gives a result which is ambiguous relative to adding or subtracting integer multiples of π. Thus the quadrant in which the angle lies must be resolved by independently examining the signs of A_x and A_y and choosing the appropriate value of θ.

To add two vectors, A and B, it is easiest to convert them to Cartesian component form. The components of the sum $C = A + B$ are then just the sums of the components:

$$Cx = Ax + BxCy = Ay + By. \tag{1.1.46}$$

Subtraction of vectors is done similarly, e. g., if $A = C - B$, then

$$Ax = Cx - BxAy = Cy - By. \tag{1.1.47}$$

A **unit vector** is a vector of unit length. One can always construct a unit vector from an ordinary (non-zero) vector by dividing the vector by its length: $n = A / |A|$. This division operation is carried out by dividing each of the vector components by the number in the denominator. Alternatively, if the vector is expressed in terms of length and direction, the magnitude of the vector is divided by the denominator and the direction is unchanged.

Unit vectors can be used to define a Cartesian coordinate system. Conventionally, i, j, and k indicate the x, y, and z axes of such a system. Note that i, j, and k are mutually perpendicular. Any vector can be represented in terms of unit vectors and its Cartesian components: $A = A_x i + A_y j + A_z k$. An alternate way to represent a vector is as a list of components: $A = (A_x, A_y, A_z)$. We tend to use the latter representation since it is somewhat more economical notation.

There are two ways to multiply two vectors, yielding respectively what are known as the **dot product** and the **cross product** . The cross product yields another vector while the dot product yields a number. Here we will discuss only the dot product. The cross product will be presented later when it is needed.

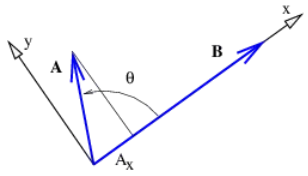


Figure 1.1.3:: Definition sketch for dot product.

Given vectors A and B, the dot product of the two is defined as

$$A \cdot B \equiv |A||B|cos\theta, \tag{1.1.48}$$

where θ is the angle between the two vectors. In two dimensions an alternate expression for the dot product exists in terms of the Cartesian components of the vectors:

$$A \cdot B = AxBx + AyBy. \tag{1.1.49}$$

It is easy to show that this is equivalent to the cosine form of the dot product when the x axis lies along one of the vectors, as in Figure 1.1.3:. Notice in particular that $A_x = |A|\cos\theta$, while $B_x = |B|$ and $B_y = 0$. Thus, $A \cdot B = |A|\cos\theta\,|B|$ in this case, which is identical to the form given in equation 2.5}.

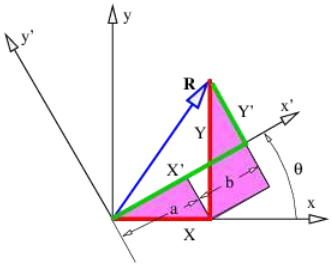


Figure 1.1.4:: Definition figure for rotated coordinate system. The vector R has components X and Y in the unprimed coordinate system and components X' and Y' in the primed coordinate system.

All that remains to be proven for equation 2.6} to hold in general is to show that it yields the same answer regardless of how the Cartesian coordinate system is oriented relative to the vectors. To do this, we must show that $A_x B_x + A_y B_y = A_x' B_x' + A_y' B_y'$, where the primes indicate components in a coordinate system rotated from the original coordinate system.

Figure 1.1.4: shows the vector R resolved in two coordinate systems rotated with respect to each other. From this figure it is clear that $X' = a + b$. Focusing on the shaded triangles, we see that $a = X\cos\theta$ and $b = Y\sin\theta$. Thus, we find $X' = X\cos\theta + Y\sin\theta$. Similar reasoning shows that $Y' = -X\sin\theta + Y\cos\theta$. Substituting these and using the trigonometric identity $\cos^2\theta + \sin^2\theta = 1$ results in

$$\begin{aligned} ''''AxBx + AyBy &= (Axcos\theta + Aysin\theta)(Bxcos\theta + Bysin\theta) \\ &+ (-Axsin\theta + Aycos\theta)(-Bxsin\theta + Bycos\theta) = AxBx + AyBy \end{aligned} \tag{1.1.50}$$

thus proving the complete equivalence of the two forms of the dot product as given by equations (2.5) and (2.6). Multiply out the above expression to verify this.

A numerical quantity that doesn't depend on which coordinate system is being used is called a **scalar** . The dot product of two vectors is a scalar. However, the components of a vector, taken individually, are not scalars, since the components change as the coordinate system changes. Since the laws of physics cannot depend on the choice of coordinate system being used, we insist that physical laws be expressed in terms of scalars and vectors, but not in terms of the components of vectors.

In three dimensions the cosine form of the dot product remains the same, while the component form is

$$A \cdot B = AxBx + AyBy + AzBz. \tag{1.1.51}$$

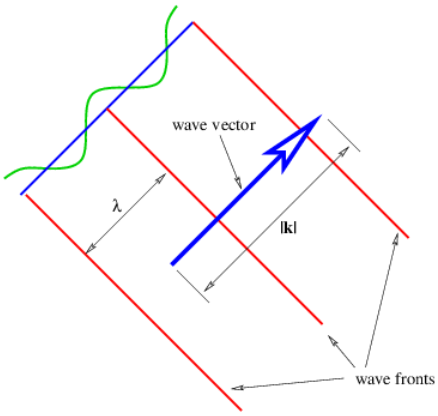


Figure 1.1.5:: Definition sketch for a plane sine wave in two dimensions. The wave fronts are constant phase surfaces separated by one wavelength. The wave vector is normal to the wave fronts and its length is the wavenumber.

A plane wave in two or three dimensions is like a sine wave in one dimension except that crests and troughs aren't points, but form lines (2-D) or planes (3-D) perpendicular to the direction of wave propagation. Figure 1.1.5: shows a plane sine wave in two dimensions. The large arrow is a vector called the **wave vector** , which defines (1) the direction of wave propagation by its orientation perpendicular to the wave fronts, and (2) the wavenumber by its length. We can think of a wave front as a line along the crest of the wave. The equation for the displacement associated with a plane sine wave (of unit amplitude) in three dimensions at some instant in time is

$$h(x,y,z) = sin(k \cdot x) = sin(kxx + kyy + kzz). \tag{1.1.52}$$

Since wave fronts are lines or surfaces of constant phase, the equation defining a wave front is simply $k \cdot x = const$.

In the two dimensional case we simply set $k_z = 0$. Therefore, a wave front, or line of constant phase ϕ in two dimensions is defined by the equation

$$k \cdot x = kxx + kyy = \phi(twodimensions). \tag{1.1.53}$$

This can be easily solved for y to obtain the slope and intercept of the wave front in two dimensions.

As for one dimensional waves, the time evolution of the wave is obtained by adding a term -ωt to the phase of the wave. In three dimensions the wave displacement as a function of both space and time is given by

$$h(x,y,z,t) = sin(kxx + kyy + kzz - \omega t). \tag{1.1.54}$$

The frequency depends in general on all three components of the wave vector. The form of this function, $\omega = \omega\,(k_x, k_y, k_z)$, which as in the one dimensional case is called the **dispersion relation** , contains information about the physical behavior of the wave.

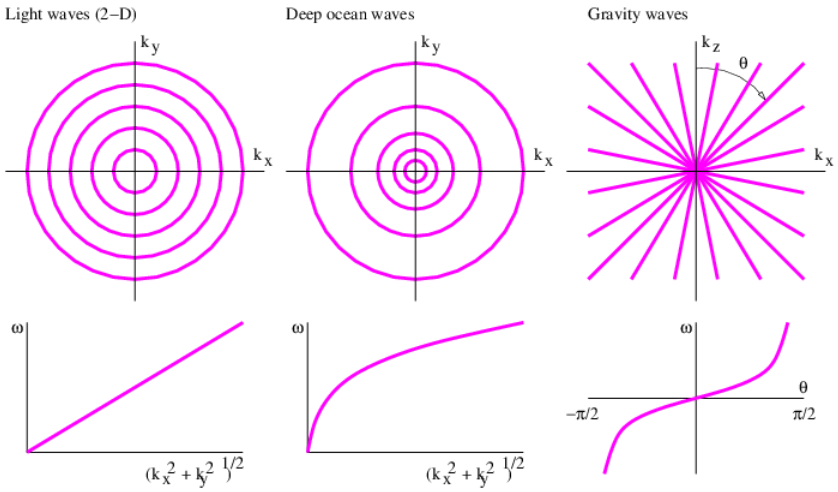


Figure 1.1.6:: Contour plots of the dispersion relations for three kinds of waves in two dimensions. In the upper panels the curves show lines or contours along which the frequency ω takes on constant values. Contours are drawn for equally spaced values of ω. For light and ocean waves the frequency depends only on the magnitude of the wave vector, whereas for gravity waves it depends only on the wave vector's direction, as defined by the angle θ in the upper right panel. These dependences for each wave type are illustrated in the lower panels.

Some examples of dispersion relations for waves in two dimensions are as follows:

- Light waves in a vacuum in two dimensions obey
- $$\omega = c(k2x + k2y)1 / 2(light),''class = ''math - display''src = ''/book155x.png \tag{1.1.55}$$

  where c is the speed of light in a vacuum.
- Deep water ocean waves in two dimensions obey
- $$\omega = g1 / 2(k2x + k2y)1 / 4(oceanwaves),''class = \\ ''math - display''src = ''/book156x.png \tag{1.1.56}$$

  where g is the strength of the Earth's gravitational field as before.
- Certain kinds of atmospheric waves confined to a vertical x -z plane called **gravity waves** (not to be confused with the gravitational waves of general relativity)[1] obey
- $$\omega = N - kx - (gravitywaves), kz \tag{1.1.57}$$

  where N is a constant with the dimensions of inverse time called the Brunt-Väisälä frequency.

Contour plots of these dispersion relations are plotted in the upper panels of Figure 1.1.6:. These plots are to be interpreted like topographic maps, where the lines represent contours of constant elevation. In the case of Figure 1.1.6:, constant values of frequency are represented instead. For simplicity, the actual values of frequency are not labeled on the contour plots, but are represented in the graphs in the lower panels. This is possible because frequency depends only on wave vector magnitude $(k_x^2 + k_y^2)^{1/2}$ for the first two examples, and only on wave vector direction θ for the third.

Superposition of Plane Waves

We now study wave packets in two dimensions by asking what the superposition of two plane sine waves looks like. If the two waves have different wavenumbers, but their wave vectors point in the same direction, the results are identical to those presented in the previous chapter, except that the wave packets are indefinitely elongated without change in form in the direction perpendicular to the wave vector. The wave packets produced in this case move in the direction of the wave vectors and thus appear to a stationary observer like a series of passing **pulses** with broad lateral extent.

Superimposing two plane waves which have the same frequency results in a stationary wave packet through which the individual wave fronts pass. This wave packet is also elongated indefinitely in some direction, but the direction of elongation depends on the dispersion relation for the waves being considered. These wave packets are in the form of steady **beams** , which guide the individual phase waves in some direction, but don't themselves change with time. By superimposing multiple plane waves, all with

the same frequency, one can actually produce a single stationary beam, just as one can produce an isolated pulse by superimposing multiple waves with wave vectors pointing in the same direction.

If the frequency of a wave depends on the magnitude of the wave vector, but not on its direction, the wave's dispersion relation is called **isotropic** ; otherwise it is **anisotropic** . In the isotropic case, two waves have the same frequency only if the lengths of their wave vectors, and hence their wavelengths, are the same. The first two examples in Figure 1.1.6: satisfy this condition, while the last example is anisotropic.

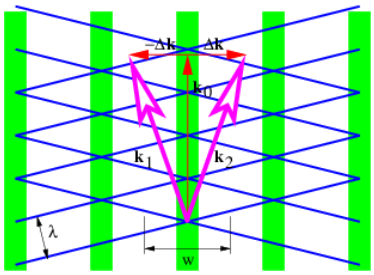


Figure 1.1.7:: Wave fronts and wave vectors (k_1 and k_2) of two plane waves with the same wavelength but oriented in different directions. The vertical bands show regions of constructive interference where wave fronts coincide. The vertical regions in between have destructive interference, and hence define the lateral boundaries of the beams produced by the superposition. The quantities k_0 and Δk are also shown.

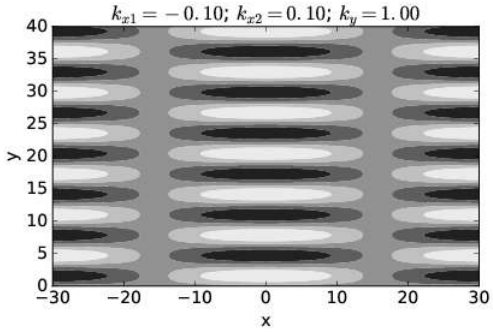


Figure 1.1.8:: Example of beams produced by two plane waves with the same wavelength moving in different directions. The wave vectors of the two waves are $k = (\pm 0.1, 1.0)$. Regions of positive displacement are lighter, while regions of negative displacement are darker.

We now use the language of vectors to investigate the superposition of two plane waves with wave vectors k_1 and k_2:

$$h = sin(k1 \cdot x - \omega t) + sin(k2 \cdot x - \omega t). \tag{1.1.58}$$

Applying the trigonometric identity for the sine of the sum of two angles (as we have done previously), equation 2.15} can be reduced to

$$h = 2sin(k0 \cdot x - \omega t)cos(\Delta k \cdot x) \tag{1.1.59}$$

where

$$k0 = (k1 + k2) \,/\, 2\Delta k = (k2 - k1) \,/\, 2. \tag{1.1.60}$$

This is in the form of a sine wave moving in the k_0 direction with phase speed $c_{\text{phase}} = \omega / |k_0|$ and wavenumber $|k_0|$, modulated in the Δk direction by a cosine function. The lines of destructive interference are normal to Δk. The distance w between lines of destructive interference is the distance between successive zeros of the cosine function in equation 2.16}, implying that $|\Delta k|w = \pi$, which leads to

$$w = \pi \,/\, |\Delta k|. \tag{1.1.61}$$

Thus, the smaller $|\Delta k|$, the greater is the beam diameter.

Waves of Identical Wavelength

In this section we investigate the beams produced by superimposing isotropic waves of the same frequency. Figure 1.1.7: illustrates what happens in such a superposition. Vectors k_1 and k_2 of equal length give rise to a mean wave vector k_0 and half the difference, Δk. As illustrated, the lines of constructive and destructive interference are perpendicular to Δk. Figure 1.1.8: shows a

concrete example of the beams produced by superposition of two plane waves of equal wavelength oriented as in Figure 1.1.7:. The beams are aligned vertically, since Δk is horizontal, with the lines of destructive interference separating the beams located near $x = \pm 16$. The transverse width of the beams of ≈ 32 satisfies equation 2.18} with $|\Delta k| = 0.1$. Each beam is made up of vertically propagating phase waves, with the crests and troughs indicated by the regions of white and black.

Waves of Differing Wavelength

In the third example of Figure 1.1.6:, the frequency of the wave depends only on the direction of the wave vector, independent of its magnitude, which is the reverse of the case for an isotropic dispersion relation. In this highly anisotropic case, different plane waves with the same frequency have wave vectors which point in the same direction, but have different lengths.

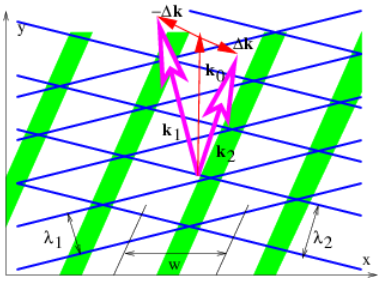


Figure 1.1.9:: Wave fronts and wave vectors (k_1 and k_2) of two plane waves with different wavelengths oriented in different directions. The slanted bands show regions of constructive interference where wave fronts coincide. The slanted regions in between have destructive interference, and as mentioned previously, define the lateral limits of the beams produced by the superposition. The quantities k_0 and Δk are also shown.

More generally, one might have waves for which the frequency depends on **both** the direction and magnitude of the wave vector. In this case, two different plane waves with the same frequency would typically have wave vectors which differ both in direction and magnitude. Such an example is illustrated in figures 2.9 and 2.10.

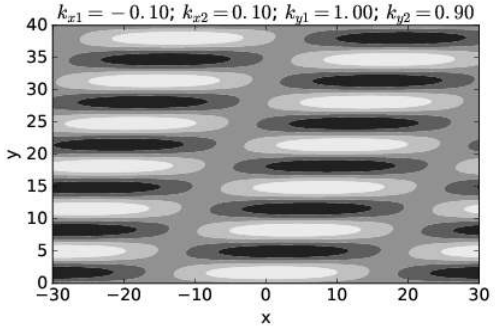


Figure 1.1.10:: Example of beams produced by two plane waves with wave vectors differing in both direction and magnitude. The wave vectors of the two waves are k_1 = (-0.1, 1.0) and k_2 = (0.1, 0.9). Regions of positive displacement are lighter, while regions of negative displacement darker.

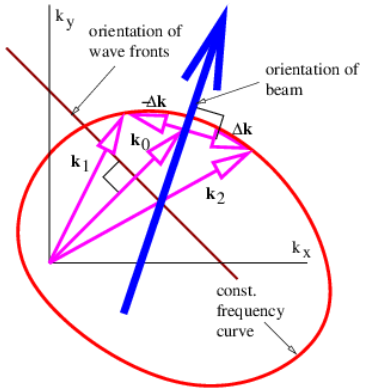


Figure 1.1.11:: Illustration of factors entering the addition of two plane waves with the same frequency. The wave fronts are perpendicular to the vector average of the two wave vectors, $k_0 = (k_1 + k_2)/2$, while the lines of constructive interference, which define the beam orientation, are oriented perpendicular to the difference between these two vectors, $\Delta k = (k_2 - k_1)/2$.

https://phys.libretexts.org/@go/page/32918

Figure 1.1.11: summarizes what we have learned about adding plane waves with the same frequency. In general, the beam orientation (and the lines of constructive interference) are not perpendicular to the wave fronts. This only occurs when the wave frequency is independent of wave vector direction.

Waves with the Same Wavelength

As with wave packets in one dimension, we can add together more than two waves to produce an isolated wave packet. We will confine our attention here to the case of an isotropic dispersion relation in which all the wave vectors for a given frequency are of the same length.

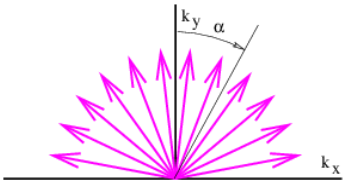


Figure 1.1.12:: Illustration of wave vectors of plane waves which might be added together.

Figure 1.1.12: shows an example of this in which wave vectors of the same wavelength but different directions are added together. Defining α_i as the angle of the i th wave vector clockwise from the vertical, as illustrated in Figure 1.1.12:, we could write the superposition of these waves at time t = 0 as

$$h = \sum hsin(kx + ky)iixiyi \sum = hisin[kxsin(\alpha i) + kycos(\alpha i) \tag{1.1.62}$$

i \]

where we have assumed that $k_{xi} = k \sin(\alpha_i)$ and $k_{yi} = k \cos(\alpha_i)$. The parameter $k = |k|$ is the magnitude of the wave vector and is the same for all the waves. Let us also assume in this example that the amplitude of each wave component decreases with increasing $|\alpha_i|$:

$$hi = exp[-(\alpha i \,/\, \alpha max)2]. \tag{1.1.63}$$

The exponential function decreases rapidly as its argument becomes more negative, and for practical purposes, only wave vectors with $|\alpha_i| \leq \alpha_{max}$ contribute significantly to the sum. We call α_{max} the **spreading angle** .

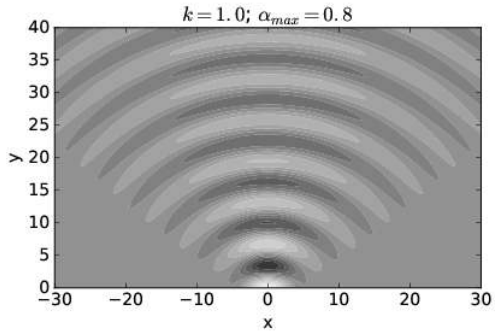


Figure 1.1.13:: Plot of the displacement field $h\,(x, y)$ from equation 2.19} for α_{max} = 0. 8 and k = 1.
}

Figure 1.1.13: shows what $h\,(x, y)$ looks like when α_{max} = 0. 8 radians and k = 1. Notice that for y = 0 the wave amplitude is only large for a small region in the range -4 $< x <$ 4. However, for $y > 0$ the wave spreads into a broad, semicircular pattern.

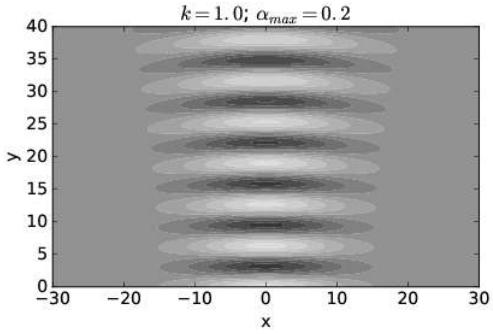


Figure 1.1.14:: Plot of the displacement field $h\,(x, y\,)$ from equation 2.19} for α_{max} = 0. 2 and k = 1.
}

Figure 1.1.14: shows the computed pattern of $h\,(x, y\,)$ when the spreading angle α_{max} = 0. 2 radians. The wave amplitude is large for a much broader range of x at y = 0 in this case, roughly -12 $< x <$ 12. On the other hand, the subsequent spread of the wave is much smaller than in the case of Figure 1.1.13:.

We conclude that a superposition of plane waves with wave vectors spread narrowly about a central wave vector which points in the y direction (as in Figure 1.1.14:) produces a beam which is initially broad in x but for which the breadth increases only slightly with increasing y. However, a superposition of plane waves with wave vectors spread more broadly (as in Figure 1.1.13:) produces a beam which is initially narrow in x but which rapidly increases in width as y increases.

The relationship between the spreading angle α_{max} and the initial breadth of the beam is made more understandable by comparison with the results for the two-wave superposition discussed at the beginning of this section. As indicated by equation 2.18}, large values of k_{x}, and hence α, are associated with small wave packet dimensions in the x direction and vice versa. The superposition of two waves doesn't capture the subsequent spread of the beam which occurs when many waves are superimposed, but it does lead to a rough quantitative relationship between α_{max} (which is just tan $^{-1}(k_{\text{x}}/k_{\text{y}})$ in the two wave case) and the initial breadth of the beam. If we invoke the small angle approximation for $\alpha = \alpha_{\text{max}}$ so that $\alpha_{\text{max}} = \tan^{-1}(k_{\text{x}}/k_{\text{y}}) \approx k_{\text{x}}/k_{\text{y}} \approx k_{\text{x}}/k$, then $k_{\text{x}} \approx k\alpha_{\text{max}}$ and equation (2.18) can be written $w = \pi\,/\,k_{\text{x}} \approx \pi/\,(k\alpha_{\text{max}}) = \lambda/\,(2\alpha_{\text{max}})$. Thus, we can find the approximate spreading angle from the wavelength of the wave λ and the initial breadth of the beam w :

$$\alpha max \approx \lambda\,/\,(2w)(singleslitspreadingangle). \tag{1.1.64}$$

Diffraction Through a Single Slit

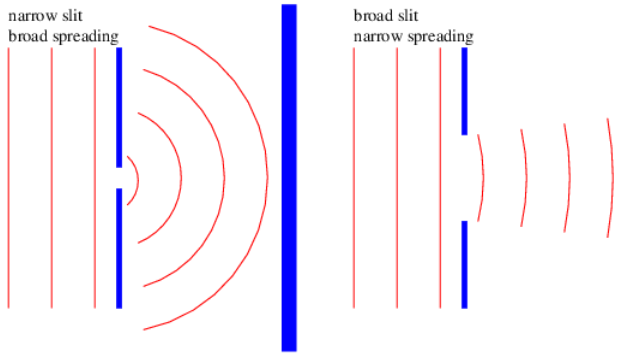


Figure 1.1.15:: Schematic behavior when a plane wave impinges on a narrow slit and a broad slit.

How does all of this apply to the passage of waves through a slit? Imagine a plane wave of wavelength λ impingent on a barrier with a slit. The barrier transforms the plane wave with infinite extent in the lateral direction into a beam with initial transverse dimensions equal to the width of the slit. The subsequent development of the beam is illustrated in figures 2.13 and 2.14, and schematically in Figure 1.1.15:. In particular, if the slit width is comparable to the wavelength, the beam spreads broadly as in Figure 1.1.13:. If the slit width is large compared to the wavelength, the beam doesn't spread as much, as Figure 1.1.14: illustrates. Equation (2.21) gives us an approximate quantitative result for the spreading angle if w is interpreted as the width of the slit.

One use of the above equation is in determining the maximum angular resolution of optical instruments such as telescopes. The primary lens or mirror can be thought of as a rather large "slit". Light from a distant point source is essentially in the form of a plane wave when it arrives at the telescope. However, the light passed by the telescope is no longer a plane wave, but is a beam

with a tendency to spread. The spreading angle α_{max} is given by equation (2.21), and the telescope cannot resolve objects with an angular separation less than α_{max}. Replacing w with the diameter of the lens or mirror in equation (2.21) thus yields the telescope's angular resolution. For instance, a moderate sized telescope with aperture 1 m observing red light with $\lambda \approx 6 \times 10^{-7}$ m has a maximum angular resolution of about 3×10^{-7} radians.

Slits

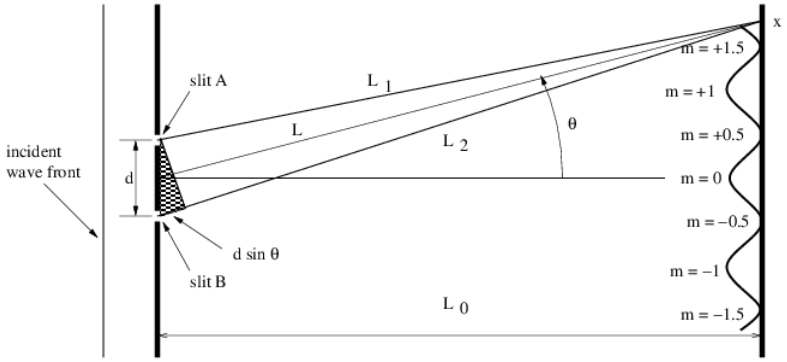


Figure 1.1.16:: Definition sketch for the double slit. Light passing through slit B travels an extra distance to the screen equal to d sin θ compared to light passing through slit A.

Let us now imagine a plane sine wave normally impingent on a screen with two narrow slits spaced by a distance d, as shown in Figure 1.1.16:. Since the slits are narrow relative to the wavelength of the wave impingent on them, the spreading angle of the beams is large and the diffraction pattern from each slit individually is a cylindrical wave spreading out in all directions, as illustrated in Figure 1.1.13:. The cylindrical waves from the two slits interfere, resulting in oscillations in wave intensity at the screen on the right side of Figure 1.1.16:.

Constructive interference occurs when the difference in the paths traveled by the two waves from their originating slits to the screen, L_2 - L_1, is an integer multiple m of the wavelength λ : L_2 - L_1 = $m\lambda$. If $L_0 \gg d$, the lines L_1 and L_2 are nearly parallel, which means that the narrow end of the dark triangle in Figure 1.1.16: has an opening angle of θ. Thus, the path difference between the beams from the two slits is L_2 - L_1 = d sin θ. Substitution of this into the above equation shows that constructive interference occurs when

$$dsin\theta = m\lambda, m = 0, \pm1, \pm2, ...(twoslitinterference). \tag{1.1.65}$$

Destructive interference occurs when m is an integer plus 1/ 2. The integer m is called the **interference order** and is the number of wavelengths by which the two paths differ.

Diffraction Gratings

Since the angular spacing $\Delta\theta$ of interference peaks in the two slit case depends on the wavelength of the incident wave, the two slit system can be used as a crude device to distinguish between the wavelengths of different components of a non-sinusoidal wave impingent on the slits. However, if more slits are added, maintaining a uniform spacing d between slits, we obtain a more sophisticated device for distinguishing beam components. This is called a **diffraction grating** .

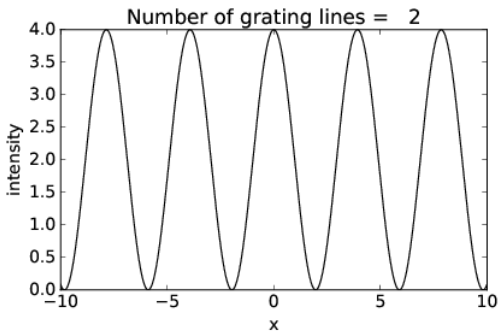


Figure 1.1.17:: Intensity of interference pattern from a diffraction grating with 2 slits on the screen in Figure 1.1.16:. The position x on the screen is proportional to the angle θ in the small angle approximation.

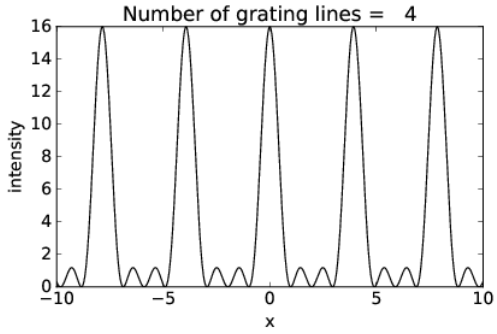


Figure 1.1.18:: Intensity of interference pattern from a diffraction grating with 4 slits.

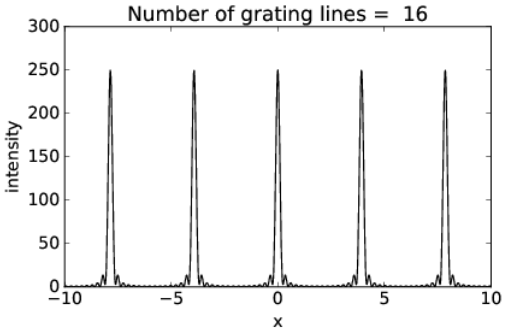


Figure 1.1.19:: Intensity of interference pattern from a diffraction grating with 16 slits.

Figures 2.17-2.19 show the intensity of the diffraction pattern as a function of position x on the display screen (see Figure 1.1.16:) for gratings with 2, 4, and 16 slits respectively, with the same slit spacing. Notice how the interference peaks remain in the same place but increase in sharpness as the number of slits increases.

The width of the peaks is actually related to the overall width of the grating, $w = nd$, where n is the number of slits. Thinking of this width as the dimension of a large single slit, the single slit equation, $\alpha_{\text{max}} = \lambda / (2w)$, tells us the angular width of the peaks.[2]

Whereas the angular width of the interference peaks is governed by the single slit equation, their angular positions are governed by the two slit equation. Let us assume for simplicity that $|\theta| \ll 1$ so that we can make the small angle approximation to the two slit equation, $m\lambda = d \sin\theta \approx d\theta$, and ask the following question: How different do two wavelengths differing by $\Delta\lambda$ have to be in order that the interference peaks from the two waves not overlap? In order for the peaks to be distinguishable, they should be separated in θ by an angle $\Delta\theta = m\,\Delta\lambda / d$, which is greater than the angular width of each peak, α_{max}:

$$\Delta\theta \tag{1.1.66}$$

αmax. " class="math-display" src="/book166x.png"> (2.23)

Substituting in the above expressions for $\Delta\theta$ and α_{max} and solving for $\Delta\lambda$, we get $\Delta\lambda > \lambda / (2mn)$, where λ is the average of the two wavelengths and $n = w / d$ is the number of slits in the diffraction grating. Thus, the fractional difference between wavelengths which can be distinguished by a diffraction grating depends solely on the interference order m and the number of slits n in the grating:

$$\Delta - \lambda - -1 - - - \lambda \tag{1.1.67}$$

2mn . " class="math-display" src="/book167x.png"> (2.24)

Problems

1. Point A is at the origin. Point B is 3 m distant from A at 30˚ counterclockwise from the x axis. Point C is 2 m from point A at 100˚ counterclockwise from the x axis.
    1. Obtain the Cartesian components of the vector D_1 which goes from A to B and the vector D_2 which goes from A to C.
    2. Find the Cartesian components of the vector D_3 which goes from B to C.
    3. Find the direction and magnitude of D_3.

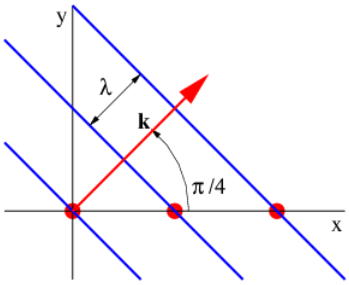
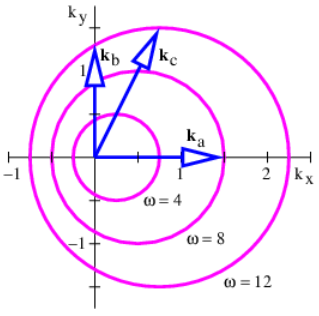


Figure 1.1.20:: Sketch of wave moving at 45° to the x -axis.

---

2. For the vectors in the previous problem, find $D_1 \cdot D_2$ using both the cosine form of the dot product and the Cartesian form. Check to see if the two answers are the same.
3. Show graphically or otherwise that $|A + B| \neq |A| + |B|$ except when the vectors A and B are parallel.
4. A wave in the x -y plane is defined by $h = h_0 \sin(k \cdot x)$ where $k = (1, 2)$ cm^{-1}.
    1. On a piece of graph paper draw x and y axes and then plot a line passing through the origin which is parallel to the vector k.
    2. On the same graph plot the line defined by $k \cdot x = k_x x + k_y y = 0$, $k \cdot x = \pi$, and $k \cdot x = 2\pi$. Check to see if these lines are perpendicular to k.
5. A plane wave in two dimensions in the x - y plane moves in the direction 45° counterclockwise from the x -axis as shown in Figure 1.1.20:. Determine how fast the intersection between a wave front and the x -axis moves to the right in terms of the phase speed c of the wave. Hint: What is the distance between wave fronts along the x -axis compared to the wavelength?
6. Two deep plane ocean waves with the same frequency ω are moving approximately to the east. However, one wave is oriented a small angle β north of east and the other is oriented β south of east.
    1. Determine the orientation of lines of constructive interference between these two waves.
    2. Determine the spacing between lines of constructive interference.
7. An example of waves with a dispersion relation in which the frequency is a function of both wave vector magnitude and direction is shown graphically in Figure 1.1.21:.
    1. What is the phase speed of the waves for each of the three wave vectors? Hint: You may wish to obtain the length of each wave vector graphically.
    2. For each of the wave vectors, what is the orientation of the wave fronts?
    3. For each of the illustrated wave vectors, sketch two other wave vectors whose average value is approximately the illustrated vector, and whose tips lie on the same frequency contour line. Determine the orientation of lines of constructive interference produced by the superimposing pairs of plane waves for which each of the vector pairs are the wave vectors.
8. Two gravity waves have the same frequency, but slightly different wavelengths.
    1. If one wave has an orientation angle $\theta = \pi/4$ radians, what is the orientation angle of the other? (See Figure 1.1.6:.)
    2. Determine the orientation of lines of constructive interference between these two waves.
9. A plane wave impinges on a single slit, spreading out a half-angle α after the slit. If the whole apparatus is submerged in a liquid with index of refraction $n = 1.5$, how does the spreading angle of the light change? (Hint: Recall that the index of refraction in a transparent medium is the ratio of the speed of light in a vacuum to the speed in the medium. Furthermore, when light goes from a vacuum to a transparent medium, the light frequency doesn't change. Therefore, how does the wavelength of the light change?)



https://phys.libretexts.org/@go/page/32918

Figure 1.1.21:: Graphical representation of the dispersion relation for shallow water waves in a river flowing in the x direction. Units of frequency are hertz, units of wavenumber are inverse meters.

---

10. Determine the diameter of the telescope needed to resolve a planet 2×10^8 km from a star which is 6 light years from the earth. (Assume blue light which has a wavelength $\lambda \approx 4 \times 10^{-7}$ m = 400 nm. Also, don't worry about the great difference in brightness between the two for the purposes of this problem.)
11. A laser beam from a laser on the earth is bounced back to the earth by a corner reflector on the moon.
    1. Engineers find that the returned signal is stronger if the laser beam is initially spread out by the beam expander shown in Figure 1.1.22:. Explain why this is so.
    2. The beam has a diameter of 1 m leaving the earth. How broad is it when it reaches the moon, which is 4×10^5 km away? Assume the wavelength of the light to be 5×10^{-7} m.
    3. How broad would the laser beam be at the moon if it weren't initially passed through the beam expander? Assume its initial diameter to be 1 cm.
12. Suppose that a plane wave impinges on two slits in a barrier at an angle, such that the phase of the wave at one slit lags the phase at the other slit by half a wavelength. How does the resulting interference pattern change from the case in which there is no lag?
13. Suppose that a thin piece of glass of index of refraction $n = 1.33$ is placed in front of one slit of a two slit diffraction setup.
    1. How thick does the glass have to be to slow down the incoming wave so that it lags the wave going through the other slit by a phase difference of π ? Take the wavelength of the light to be $\lambda = 6 \times 10^{-7}$ m.
    2. For the above situation, describe qualitatively how the diffraction pattern changes from the case in which there is no glass in front of one of the slits. Explain your results.

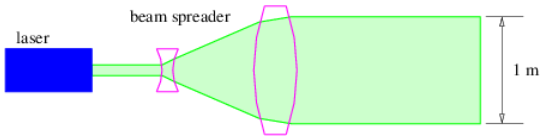


Figure 1.1.22:: Sketch of a beam expander for a laser.

---

14. A light source produces two wavelengths, $\lambda_1 = 400$ nm (blue) and $\lambda_2 = 600$ nm (red).
    1. Qualitatively sketch the two slit diffraction pattern from this source. Sketch the pattern for each wavelength separately.
    2. Qualitatively sketch the 16 slit diffraction pattern from this source, where the slit spacing is the same as in the two slit case.
15. A light source produces two wavelengths, $\lambda_1 = 631$ nm and $\lambda_2 = 635$ nm. What is the minimum number of slits needed in a grating spectrometer to resolve the two wavelengths? (Assume that you are looking at the first order diffraction peak.) Sketch the diffraction peak from each wavelength and indicate how narrow the peaks must be to resolve them.

Chapter 3 Geometrical Optics

As was shown previously, when a plane wave is impingent on an aperture which has dimensions much greater than the wavelength of the wave, diffraction effects are minimal and a segment of the plane wave passes through the aperture essentially unaltered. This plane wave segment can be thought of as a wave packet, called a **beam** or **ray** , consisting of a superposition of wave vectors very close in direction and magnitude to the central wave vector of the wave packet. In most cases the ray simply moves in the direction defined by the central wave vector, i. e., normal to the orientation of the wave fronts. However, this is not true when the medium through which the light propagates is optically anisotropic, i. e., light traveling in different directions moves at different phase speeds. An example of such a medium is a calcite crystal. In the anisotropic case, the orientation of the ray can be determined once the dispersion relation for the waves in question is known, by using the techniques developed in the previous chapter.

If light moves through some apparatus in which all apertures are much greater in dimension than the wavelength of light, then we can use the above rule to follow rays of light through the apparatus. This is called the **geometrical optics** approximation.

Reflection and Refraction

Most of what we need to know about geometrical optics can be summarized in two rules, the laws of reflection and refraction. These rules may both be inferred by considering what happens when a plane wave segment impinges on a flat surface. If the

surface is polished metal, the wave is **reflected** , whereas if the surface is an interface between two transparent media with differing indices of refraction, the wave is partially reflected and partially **refracted** . Reflection means that the wave is turned back into the half-space from which it came, while refraction means that it passes through the interface, acquiring a different direction of motion from that which it had before reaching the interface.

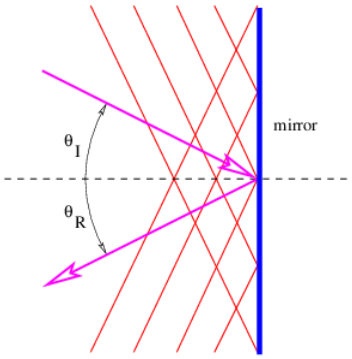


Figure 1.1.1:: Sketch showing the reflection of a wave from a plane mirror. The law of reflection states that $\theta_I = \theta_R$.

Figure 1.1.1: shows the wave vector and wave front of a wave being reflected from a plane mirror. The angles of incidence, θ_I, and reflection, θ_R, are defined to be the angles between the incoming and outgoing wave vectors respectively and the line normal to the mirror. The law of reflection states that $\theta_R = \theta_I$. This is a consequence of the need for the incoming and outgoing wave fronts to be in phase with each other all along the mirror surface. This plus the equality of the incoming and outgoing wavelengths is sufficient to insure the above result.

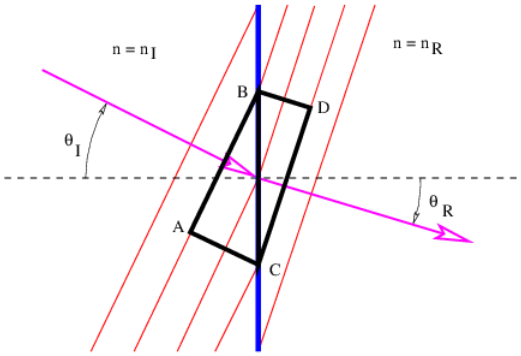


Figure 1.1.2:: Sketch showing the refraction of a wave from an interface between two dielectric media with $n_2 > n_1$.

Refraction, as illustrated in Figure 1.1.2:, is slightly more complicated. Since $n_R > n_I$, the speed of light in the right-hand medium is less than in the left-hand medium. (Recall that the speed of light in a medium with refractive index n is $c_{medium} = c_{vac}/n$.) The frequency of the wave packet doesn't change as it passes through the interface, so the wavelength of the light on the right side is less than the wavelength on the left side.

Let us examine the triangle ABC in Figure 1.1.2:. The side AC is equal to the side BC times sin(θ_I). However, AC is also equal to 2 λ_I, or twice the wavelength of the wave to the left of the interface. Similar reasoning shows that $2\lambda_R$, twice the wavelength to the right of the interface, equals BC times sin(θ_R). Since the interval BC is common to triangles ABC and DBC, we easily see that

$$\lambda I - = -sin(\theta I).\lambda Rsin(\theta R) \tag{1.1.68}$$

Since $\lambda_I = c_I T = c_{vac} T / n_I$ and $\lambda_R = c_R T = c_{vac} T / n_R$ where c_I and c_R are the wave speeds to the left and right of the interface, c_{vac} is the speed of light in a vacuum, and T is the (common) period, we can easily recast the above equation in the form

$$nIsin(\theta I) = nRsin(\theta R). \tag{1.1.69}$$

This is called **Snell's law** , and it governs how a ray of light bends as it passes through a discontinuity in the index of refraction. The angle θ_I is called the incident angle and θ_R is called the refracted angle. Notice that these angles are measured from the normal to the surface, not the tangent.

Total Internal Reflection

When light passes from a medium of lesser index of refraction to one with greater index of refraction, Snell's law indicates that the ray bends **toward** the normal to the interface. The reverse occurs when the passage is in the other direction. In this latter circumstance a special situation arises when Snell's law predicts a value for the sine of the refracted angle greater than one. This is physically untenable. What actually happens is that the incident wave is **reflected** from the interface. This phenomenon is called **total internal reflection** . The minimum incident angle for which total internal reflection occurs is obtained by substituting $\theta_R = \pi / 2$ into equation 3.2}, resulting in

$$sin(\theta I) = nR/nI(totalinternalreflection). \tag{1.1.70}$$

Anisotropic Media

Notice that Snell's law makes the implicit assumption that rays of light move in the direction of the light's wave vector, i. e., normal to the wave fronts. As the analysis in the previous chapter makes clear, this is valid only when the optical medium is isotropic, i. e., the wave frequency depends only on the magnitude of the wave vector, not on its direction.

Certain kinds of crystals, such as those made of calcite, are not isotropic — the speed of light in such crystals, and hence the wave frequency, depends on the orientation of the wave vector. As an example, the angular frequency in an anisotropic medium might take the form

$$\begin{aligned}&[c2(kx+ky)2c2(kx-ky)2]1 / 2\omega \\ &= -1 - - - - - - - - - - - + -2 - - - - - - - - - - - -, 22\end{aligned} \tag{1.1.71}$$

where c_1 is the speed of light for waves in which $k_y = k_x$, and c_2 is its speed when $k_y = -k_x$.

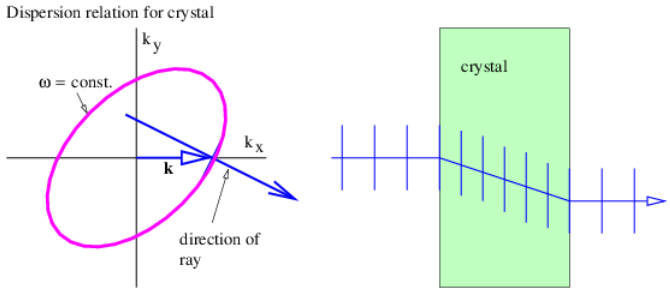


Figure 1.1.3:: The right panel shows the fate of a light ray normally incident on the face of a properly cut calcite crystal. The anisotropic dispersion relation which gives rise to this behavior is shown in the left panel.

Figure 1.1.3: shows an example in which a ray hits a calcite crystal oriented so that constant frequency contours are as specified in equation 3.4}. The wave vector is oriented normal to the surface of the crystal, so that wave fronts are parallel to this surface. Upon entering the crystal, the wave front orientation must stay the same to preserve phase continuity at the surface. However, due to the anisotropy of the dispersion relation for light in the crystal, the ray direction changes as shown in the right panel. This behavior is clearly inconsistent with the usual version of Snell's law!

It is possible to extend Snell's law to the anisotropic case. However, we will not present this here. The following discussions of optical instruments will always assume that isotropic optical media are used.

Lens Equation and Optical Instruments

Given the laws of reflection and refraction, one can see in principle how the passage of light through an optical instrument could be traced. For each of a number of initial rays, the change in the direction of the ray at each mirror surface or refractive index interface can be calculated. Between these points, the ray traces out a straight line.

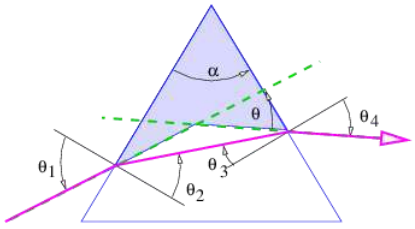


Figure 1.1.4:: Bending of a ray of light as it passes through a prism.

Though simple in conception, this procedure can be quite complex in practice. However, the procedure simplifies if a number of approximations, collectively called the **thin lens approximation** , are valid. We begin with the calculation of the bending of a ray of light as it passes through a prism, as illustrated in Figure 1.1.4:.

The pieces of information needed to find θ, the angle through which the ray is deflected, are as follows: the geometry of the triangle defined by the entry and exit points of the ray and the upper vertex of the prism leads to

$$\alpha + (\pi / 2 - \theta) + (\pi / 2 - \theta) = \pi, 23 \tag{1.1.72}$$

which simplifies to

$$\alpha = \theta 2 + \theta 3. \tag{1.1.73}$$

Snell's law at the entrance and exit points of the ray tell us that

$$sin - (\theta 1) sin - (\theta 4) n = sin(\theta) n = sin(\theta), 23 \tag{1.1.74}$$

where n is the index of refraction of the prism. (The index of refraction of the surroundings is assumed to be unity.) One can also infer that

$$\theta = \theta 1 + \theta 4 - \alpha. \tag{1.1.75}$$

This comes from the fact that the the sum of the internal angles of the shaded quadrangle in Figure 1.1.4: is

$$(\pi / 2 - \theta 1) + \alpha + (\pi / 2 - \theta 4) + (\pi + \theta) = 2\pi. \tag{1.1.76}$$

Combining equations (3.6), (3.7), and (3.8) allows the ray deflection θ to be determined in terms of θ_1 and α, but the resulting expression is very messy. However, great simplification occurs if the following conditions are met:

- The angle $\alpha \ll 1$.
- The angles $\theta_1, \theta_2, \theta_3, \theta_4 \ll 1$.

With these approximations it is easy to show that

$$\theta = \alpha(n - 1)(smallangles). \tag{1.1.77}$$

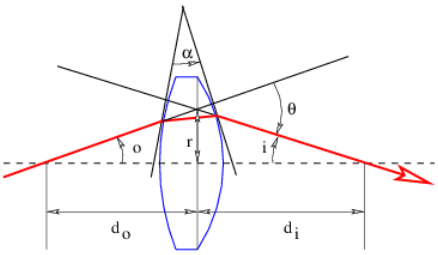


Figure 1.1.5:: Light ray undergoing deflection through an angle θ by a lens. The angle α is the angle between the tangents to the entry and exit points of the ray on the lens.

Generally speaking, lenses and mirrors in optical instruments have curved rather than flat surfaces. However, we can still use the laws for reflection and refraction by plane surfaces as long as the segment of the surface on which the wave packet impinges is not curved very much on the scale of the wave packet dimensions. This condition is easy to satisfy with light impinging on ordinary optical instruments. In this case, the deflection of a ray of light is given by equation 3.10} if α is defined as the intersection of the tangent lines to the entry and exit points of the ray, as illustrated in Figure 1.1.5:.

A **positive lens** is thicker in the center than at the edges. The angle α between the tangent lines to the two surfaces of the lens at a distance r from the central axis takes the form $\alpha = Cr$, where C is a constant. The deflection angle of a beam hitting the lens a distance r from the center is therefore $\theta = Cr$ (n - 1), as indicated in Figure 1.1.5:. The angles o and i sum to the deflection angle: $o + i = \theta = Cr$ (n - 1). However, to the extent that the small angle approximation holds, $o = r / d_o$ and $i = r / d_i$ where d_o is the distance to the object and d_i is the distance to the image of the object. Putting these equations together and cancelling the r results in the **thin lens equation** :

$$1 - -1 - 1 - d + d = C(n - 1) \equiv f.oi \tag{1.1.78}$$

The quantity f is called the **focal length** of the lens. Notice that $f = d_i$ if the object is very far from the lens, i. e., if d_o is extremely large.

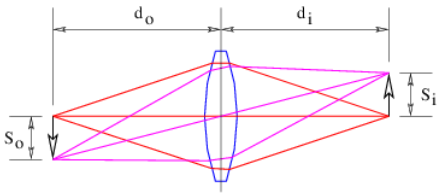


Figure 1.1.6:: A positive lens producing an image on the right of the arrow on the left.

Figure 1.1.6: shows how a positive lens makes an image. The image is produced by all of the light from each point on the object falling on a corresponding point in the image. If the arrow on the left is an illuminated object, an **image** of the arrow will appear at right if the light coming from the lens is allowed to fall on a piece of paper or a ground glass screen. The size of the object S_o and the size of the image S_i are related by simple geometry to the distances of the object and the image from the lens:

$$Si- = di.Sodo \tag{1.1.79}$$

Notice that a positive lens inverts the image.

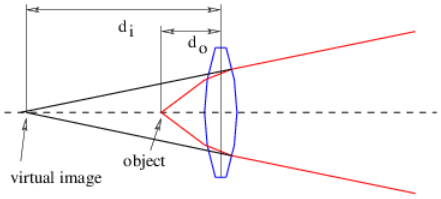


Figure 1.1.7:: Production of a virtual image by a positive lens.

An image will be produced to the right of the lens only if $d_o > f$. If $d_o < f$, the lens is unable to converge the rays from the image to a point, as is seen in Figure 1.1.7:. However, in this case the backward extension of the rays converge at a point called a **virtual image** , which in the case of a positive lens is always farther away from the lens than the object. The image is called virtual because it does not appear on a ground glass screen placed at this point. Unlike the real image seen in Figure 1.1.6:, the virtual image is not inverted. The thin lens equation still applies if the distance from the lens to the image is taken to be negative.

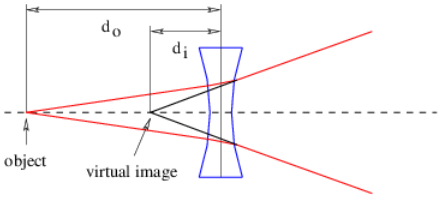


Figure 1.1.8:: Production of a virtual image by a negative lens.

A **negative lens** is thinner in the center than at the edges and produces only virtual images. As seen in Figure 1.1.8:, the virtual image produced by a negative lens is closer to the lens than is the object. Again, the thin lens equation is still valid, but both the distance from the image to the lens and the focal length must be taken as negative. Only the distance to the object remains positive.

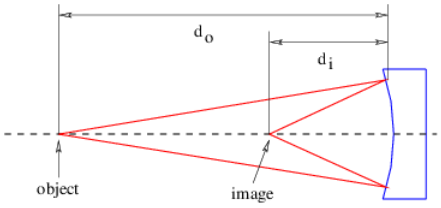


Figure 1.1.9:: Production of a real image by a concave mirror.

Curved mirrors also produce images in a manner similar to a lens, as shown in Figure 1.1.9:. A concave mirror, as seen in this figure, works in analogy to a positive lens, producing a real or a virtual image depending on whether the object is farther from or closer to the mirror than the mirror's focal length. A convex mirror acts like a negative lens, always producing a virtual image. The thin lens equation works in both cases as long as the angles are small.

Fermat's Principle

An alternate approach to geometrical optics can be developed from Fermat's principle. This principle states (in its simplest form) that light waves of a given frequency traverse the path between two points which takes the least time. The most obvious example of

this is the passage of light through a homogeneous medium in which the speed of light doesn't change with position. In this case the shortest time corresponds to the shortest distance between the points, which, as we all know, is a straight line. Thus, Fermat's principle is consistent with light traveling in a straight line in a homogeneous medium.

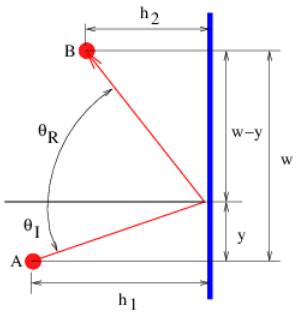


Figure 1.1.10:: Definition sketch for deriving the law of reflection from Fermat's principle. θ_I is the angle of incidence and θ_R the angle of reflection as in Figure 1.1.1:.

Fermat's principle can also be used to derive the laws of reflection and refraction. For instance, Figure 1.1.10: shows a candidate ray for reflection in which the angles of incidence and reflection are not equal. The time required for the light to go from point A to point B is

$$t = ([h2 + y2]1 \,/\, 2 + [h2 + (w - y)2]1 \,/\, 2) \,/\, c12 \tag{1.1.80}$$

where c is the speed of light. We find the minimum time by differentiating t with respect to y and setting the result to zero, with the result that

$$\begin{array}{c} -----y------ = --------w---y------- \\ .[h21 + y2]1 \,/\, 2[h22 + (w - y)2]1 \,/\, 2 \end{array} \tag{1.1.81}$$

However, we note that the left side of this equation is simply sin θ_I, while the right side is sin θ_R, so that the minimum time condition reduces to sin θ_I = sin θ_R or $\theta_I = \theta_R$, which is the law of reflection.

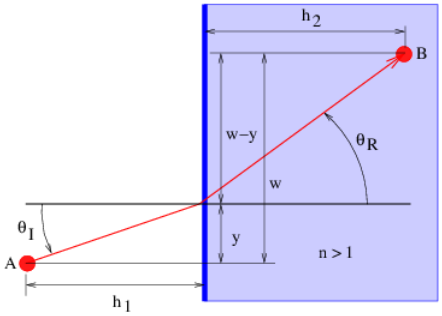


Figure 1.1.11:: Definition sketch for deriving Snell's law of refraction from Fermat's principle. The shaded area has index of refraction $n > 1$.

A similar analysis may be done to derive Snell's law of refraction. The speed of light in a medium with refractive index n is $c \,/\, n$, where c is its speed in a vacuum. Thus, the time required for light to go some distance in such a medium is n times the time light takes to go the same distance in a vacuum. Referring to Figure 1.1.11:, the time required for light to go from A to B becomes

$$221 \,/\, 2221 \,/\, 2t = ([h1 + y] + n[h2 + (w - y)]) \,/\, c. \tag{1.1.82}$$

This results in the condition

$$sin\theta I = nsin\theta R \tag{1.1.83}$$

where θ_R is now the refracted angle. We recognize this result as Snell's law.

Notice that the reflection case illustrates a point about Fermat's principle: The minimum time may actually be a **local** rather than a global minimum — after all, in Figure 1.1.10:, the global minimum distance from A to B is still just a straight line between the two points! In fact, light starting from point A will reach point B by **both** routes — the direct route and the reflected route.

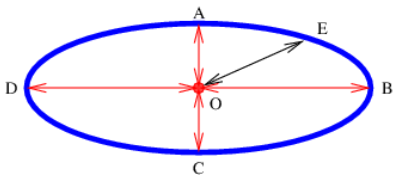


Figure 1.1.12:: Ellipsoidal mirror showing minimum and maximum time rays from the center of the ellipsoid to the mirror surface and back again.

It turns out that trajectories allowed by Fermat's principle don't strictly have to be minimum time trajectories. They can also be **maximum** time trajectories, as illustrated in Figure 1.1.12:. In this case light emitted at point O can be reflected back to point O from four points on the mirror, A, B, C, and D. The trajectories O-A-O and O-C-O are minimum time trajectories while O-B-O and O-D-O are maximum time trajectories.

Fermat's principle seems rather mysterious. However, the American physicist Richard Feynman made sense out of it by invoking an even more fundamental principle, as we now see.

If a light ray originates at point O in Figure 1.1.12:, reflects off of the ellipsoidal mirror surface at point A, and returns to point O, the elapsed time isn't much different from that experienced by a ray which reflects off the mirror a slight distance from point A and returns to O. This is because at point A the beam from point O is perpendicular to the tangent to the surface of the mirror at point A. In contrast, the time experienced by a ray going from point O to point E and back would differ by a much greater amount than the time experienced by a ray reflecting off the mirror a slight distance from point E. This is because the tangent to the mirror surface at point E is not perpendicular to the beam from point O.

Technically, the change in the round trip time varies linearly with the deviation in the reflection point from point E, but quadratically with the deviation from point A. If this deviation is small in the first place, then the change in the round trip time will be much smaller for the quadratic case than for the linear case.

It seems odd that we would speak of a beam reflecting back to point O if it hit the mirror at any point except A, B, C, or D, due to the requirements of the law of reflection. However, recall that the law of reflection itself depends on Fermat's principle, so we cannot assume the validity of that law in this investigation.

Feynman postulated that light rays explore **all** possible paths from one point to another, but that the only paths realized in nature are those for which light taking closely neighboring paths experiences nearly the same elapsed time (or more generally, traverses nearly the same number of wavelengths) as the original path. If this is true, then neighboring rays interfere constructively with each other, resulting in a much brighter beam than would occur in the absence of this constructive interference. Thus, the round-trip paths O-A-O, O-B-O, O-C-O, and O-D-O in Figure 1.1.12: actually occur, but not O-E-O. Feynman explains Fermat's principle by invoking constructive and destructive interference!

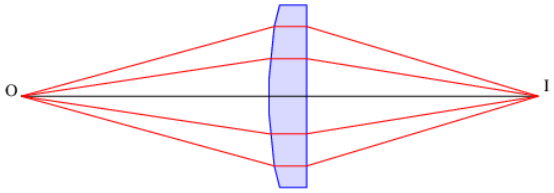


Figure 1.1.13:: Ray trajectories from a point O being focused to another point I by a lens.

Figure 1.1.13: illustrates a rather peculiar situation. Notice that **all** the rays from point O which intercept the lens end up at point I. This would seem to contradict Fermat's principle, in that only the minimum (or maximum) time trajectories should occur. However, a calculation shows that all the illustrated trajectories in this particular case take the **same** time. Thus, the light cannot choose one trajectory over another using Fermat's principle and all of the trajectories are equally favored. Note that this inference applies not to just any set of trajectories, but only those going from an object point to the corresponding image point.

Problems

1. The index of refraction varies as shown in Figure 1.1.14::
    1. Given θ_1, use Snell's law to find θ_2.
    2. Given θ_2, use Snell's law to find θ_3.
    3. From the above results, find θ_3, given θ_1. Do n_2 or θ_2 matter?

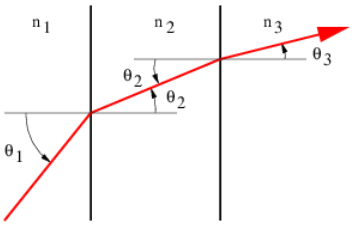


Figure 1.1.14:: Refraction through multiple parallel layers with different refractive indices.

---

2. A 45˚-45˚-90˚ prism is used to totally reflect light through 90˚ as shown in Figure 1.1.15:. What is the minimum index of refraction of the prism needed for this to work?

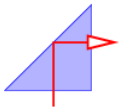


Figure 1.1.15:: Refraction through a 45˚-45˚-90˚ prism.

---

3. Show graphically which way the wave vector must point inside the calcite crystal of Figure 1.1.3: for a light ray to be horizontally oriented. Sketch the orientation of the wave fronts in this case.

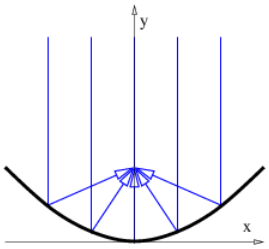


Figure 1.1.16:: Focusing of parallel rays by a parabolic mirror.

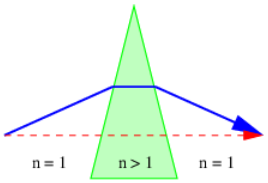


Figure 1.1.17:: Refraction through a wedge-shaped prism.

---

4. The human eye is a lens which focuses images on a screen called the retina. Suppose that the normal focal length of this lens is 4 cm and that this focuses images from far away objects on the retina. Let us assume that the eye is able to focus on nearby objects by changing the shape of the lens, and thus its focal length. (The lens-retina distance remains the same.) If an object is 20 cm from the eye, what must the altered focal length of the eye be in order for the image of this object to be in focus on the retina?
5. An amoeba 0. 01 cm in diameter has its image projected on a screen as shown in Figure 1.1.18: by a positive lens of diameter 0. 1 cm.
    1. How big is the image of the amoeba?
    2. What is the focal length of the lens?
    3. What is the minimum distance between features on the image of the amoeba which can be resolved? Assume that the wavelength of light used is 5×10^{-7} m. (Hint: What is the spreading angle of a beam of light passing through an opening the size of the lens?)

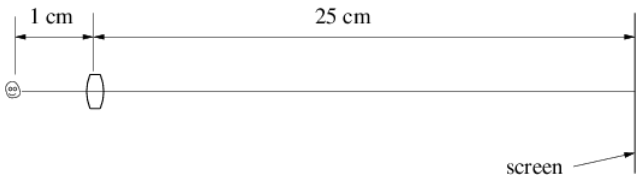


Figure 1.1.18:: An amoeba imaged by a lens.

---

6. The great refractor telescope of Yerkes Observatory in Wisconsin (see Figure 1.1.19:) has primary lens $D = 1.02$ m in diameter with a focal length of $L = 19.4$ m. Use the small angle approximation in all calculations and assume that the light has wavelength 5×10^{-7} m.
    1. Jupiter has a diameter of 1.5×10^5 km and an average distance from the earth of 8×10^8 km. How big is the image of Jupiter (in cm) at the focal plane of the primary lens?
    2. Given perfect atmospheric "seeing" conditions, how far apart must two features be on Jupiter (in km) for the Yerkes telescope to be able to resolve them?
    3. What should the focal length l of the secondary lens or eyepiece be for Jupiter to subtend the same angle as the moon subtends to the naked eye? The moon's diameter is 3.5×10^3 km and its distance from the earth is 3.8×10^5 km. Hint: Imagine that a translucent sheet of ground glass is placed at the focal plane so that the image is seen projected on this ground glass, which scatters light over a broad range of angles. The eyepiece can then be thought of as a magnifying glass with which you can examine the image on the ground glass. Using this artifact, you need consider only light rays that pass through the center of each lens.

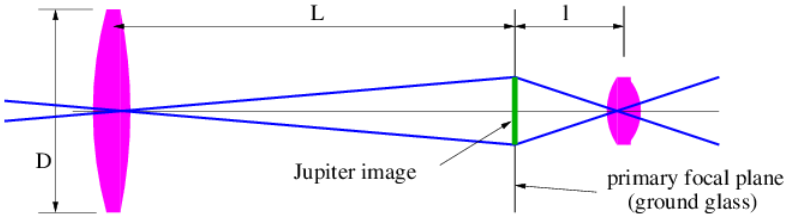


Figure 1.1.19:: The great Yerkes refractor telescope observing Jupiter.

---

7. Show that a concave mirror that focuses incoming rays parallel to the optical axis of the mirror to a point on the optical axis, as illustrated in Figure 1.1.16:, is parabolic in shape. Hint: Since rays following different paths all move from the distant source to the focal point of the mirror, Fermat's principle implies that all of these rays take the same time to do so (why is this?), and therefore all traverse the same distance.
8. Use **Fermat's principle** to explain **qualitatively** why a ray of light follows the solid rather than the dashed line through the wedge of glass shown in Figure 1.1.17:.
9. Test your knowledge of Fermat's principle by using equation 3.15} to derive Snell's law.
10. }

Chapter 4 Special Relativity

Albert Einstein invented the special and general theories of relativity early in the 20th century, though many other people contributed to the intellectual climate which made these discoveries possible. The special theory of relativity arose out of a conflict between the ideas of mechanics as developed by Galileo and Newton, and the theory of electromagnetism. For this reason relativity is often discussed in textbooks after electromagnetism is developed. However, special relativity is actually a valid extension to the Galilean world view which is needed when objects move at very high speeds, and it is only coincidentally related to electromagnetism. For this reason we discuss relativity before electromagnetism.

The only fact from electromagnetism that we need is introduced now: There is a maximum speed at which objects can travel. This is coincidentally equal to the speed of light in a vacuum, $c = 3 \times 10^8$ m s^{-1}. Furthermore, a measurement of the speed of a particular light beam yields the same answer regardless of the speed of the light source or the speed at which the measuring instrument is moving.

This rather bizarre experimental result is in contrast to what occurs in Galilean relativity. If two cars pass a pedestrian standing on a curb, one at 20 m s^{-1} and the other at 50 m s^{-1}, the faster car appears to be moving at 30 m s^{-1} relative to the slower car. However, if a light beam moving at 3×10^8 m s^{-1} passes an interstellar spaceship moving at 2×10^8 m s^{-1}, then the light beam appears to occupants of the spaceship to be moving at 3×10^8 m s^{-1}, not 1×10^8 m s^{-1}. Furthermore, if the spaceship beams a light signal forward to its (stationary) destination planet, then the resulting beam appears to be moving at 3×10^8 m s^{-1} to instruments at the destination, not 5×10^8 m s^{-1}.

The fact that we are talking about light beams is only for convenience. Any other means of sending a signal at the maximum allowed speed would result in the same behavior. We therefore cannot seek the answer to this apparent paradox in the special properties of light. Instead we have to look to the basic nature of space and time.

Galilean Spacetime Thinking

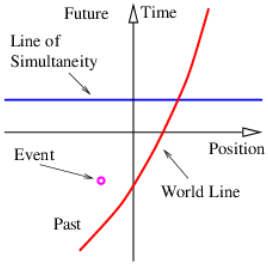


Figure 1.1.1:: Spacetime diagram showing an **event** , a **world line** , and a **line of simultaneity** .

In order to gain an understanding of both Galilean and Einsteinian relativity it is important to begin thinking of space and time as being different dimensions of a four-dimensional space called spacetime. Actually, since we can't visualize four dimensions very well, it is easiest to start with only one space dimension and the time dimension. Figure 1.1.1: shows a graph with time plotted on the vertical axis and the one space dimension plotted on the horizontal axis. An **event** is something that occurs at a particular time and a particular point in space. ("Julius X. wrecks his car in Lemitar, NM on 21 June at 6:17 PM.") A **world line** is a plot of the position of some object as a function of time on a spacetime diagram, although it is conventional to put time on the vertical axis. Thus, a world line is really a line in spacetime, while an event is a point in spacetime. A horizontal line parallel to the position axis is a **line of simultaneity** in Galilean relativity — i. e., all events on this line occur simultaneously.

In a spacetime diagram the slope of a world line has a special meaning. Notice that a vertical world line means that the object it represents does not move — the velocity is zero. If the object moves to the right, then the world line tilts to the right, and the faster it moves, the more the world line tilts. Quantitatively, we say that

$$----------1----------velocity = slope of world line \tag{1.1.84}$$

in Galilean relativity. Notice that this works for negative slopes and velocities as well as positive ones. If the object changes its velocity with time, then the world line is curved, and the instantaneous velocity at any time is the inverse of the slope of the tangent to the world line at that time.

The hardest thing to realize about spacetime diagrams is that they represent the past, present, and future all in one diagram. Thus, spacetime diagrams don't change with time — the evolution of physical systems is represented by looking at successive horizontal slices in the diagram at successive times. **Spacetime diagrams represent evolution, but they don't evolve themselves.**

The principle of relativity states that the laws of physics are the same in all inertial reference frames. An inertial reference frame is one that is not accelerated. Reference frames attached to a car at rest and to a car moving steadily down the freeway at 30 m s^{-1} are both inertial. A reference frame attached to a car accelerating away from a stop light is not inertial.

The principle of relativity is an educated guess or hypothesis based on extensive experience. If the principle of relativity weren't true, we would have to do all our calculations in some preferred reference frame. This would be very annoying. However, the more fundamental problem is that we have no idea what the velocity of this preferred frame might be. Does it move with the earth? That would be very earth-centric. How about the velocity of the center of our galaxy or the mean velocity of all the galaxies? Rather than face the issue of a preferred reference frame, physicists have chosen to stick with the principle of relativity.

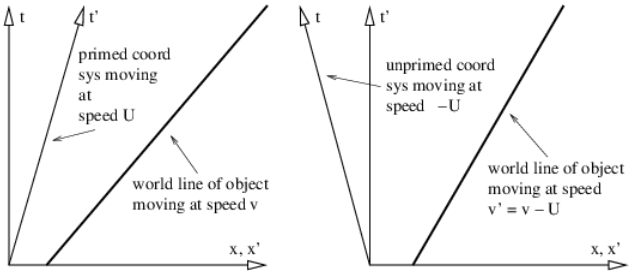


Figure 1.1.2:: The left panel shows the world line in the unprimed reference frame, while the right panel shows it in the primed frame, which moves to the right at speed U relative to the unprimed frame. (The "prime" is just a label that allows us to distinguish the axes corresponding to the two reference frames.)

If an object is moving to the left at velocity v relative to a particular reference frame, it appears to be moving at a velocity $v' = v - U$ relative to another reference frame which itself is moving at velocity U. This is the Galilean velocity transformation law, and it is based on everyday experience. If you are traveling 30 m s^{-1} down the freeway and another car passes you doing 40 m s^{-1}, then the other car moves past you at 10 m s^{-1} relative to your car.

Figure 1.1.2: shows how the world line of an object is represented differently in the unprimed (x, t) and primed (x', t') reference frames. The difference between the velocity of the object and the velocity of the primed frame (i. e., the difference in the inverses of the slopes of the corresponding world lines) is the same in both reference frames in this Galilean case. This illustrates the difference between a physical law independent of reference frame (the difference between velocities in Galilean relativity) and the different motion of the object in the two different reference frames.

Spacetime Thinking in Special Relativity

In special relativity we find that space and time "mix" in a way that they don't in Galilean relativity. This suggests that space and time are different aspects of the same "thing", which we call **spacetime** .

If time and position are simply different dimensions of the same abstract space, then they should have the same units. The easiest way to arrange this is to multiply time by the maximum speed, c, resulting in the kind of spacetime diagram shown in Figure 1.1.3:. Notice that world lines of light have slope ±1 when the time axis is scaled this way. Furthermore, the relationship between speed and the slope of a world line must be revised to read

$$---c--v = slope(worldline). \tag{1.1.85}$$

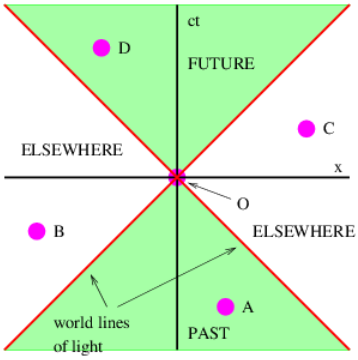


Figure 1.1.3:: Scaled spacetime diagram showing world lines of light passing left and right through the origin.

Notice that it is physically possible for an object to have a world line which connects event O at the origin and the events A and D in Figure 1.1.3:, since the slope of the resulting world line would exceed unity, and thus represent a velocity less than the speed of light. Events which can be connected by a world line are called **timelike** relative to each other. On the other hand, event O cannot be connected to events B and C by a world line, since this would imply a velocity greater than the speed of light. Events which cannot be connected by a world line are called **spacelike** relative to each other. Notice the terminology in Figure 1.1.3:: Event A is in the **past** of event O, while event D is in the **future** . Events B and C are **elsewhere** relative to event O.

Postulates of Special Relativity

As we learned previously, the **principle of relativity** states that the laws of physics are the same in all inertial reference frames. The principle of relativity applies to Einsteinian relativity just as it applies to Galilean relativity.

Notice that the constancy of the speed of light in all reference frames is consistent with the principle of relativity. However, as noted above, it is inconsistent with our notions as to how velocities add, or alternatively, how we think the world should look from reference frames moving at different speeds. We have called the classical way of understanding the view from different reference frames Galilean relativity. The new way that reconciles the behavior of objects moving at very high speeds is called **Einsteinian relativity** . Einstein's great contribution was to discover the laws that tell us how the world looks from reference frames moving at high speeds relative to each other. These laws constitute a **geometry of spacetime** , and from them all of special relativity can be derived.

All of the observed facts about spacetime can be derived from two postulates:

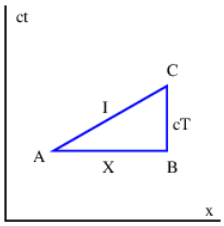


Figure 1.1.4:: Triangle for Pythagorean theorem in spacetime.

- Whether two events are simultaneous depends on the reference frame from which they are viewed.
- Spacetime obeys a modified Pythagorean theorem, which gives the distance, I, in spacetime or **spacetime interval** as
- $$2222I = X - cT, \tag{1.1.86}$$

  where X, T, and I are defined in Figure 1.1.4:.

Let us discuss these postulates in turn.

Simultaneity

The classical way of thinking about **simultaneity** is so ingrained in our everyday habits that we have a great deal of difficulty adjusting to what special relativity has to say about this subject. Indeed, understanding how relativity changes this concept is the single most difficult part of the theory — once you understand this, you are well on your way to mastering relativity!

Before tackling simultaneity, let us first think about **collocation** . Two events (such as A and E in Figure 1.1.5:) are collocated if they have the same x value. However, collocation is a concept that depends on the reference frame. For instance, George is driving from Boston to Washington, with the line passing through events A and D being his world line. Just as he passes New York he sneezes (event A in Figure 1.1.5:). As he drives by Baltimore, he sneezes again (event D). In the reference frame of the earth, these two sneezes are not collocated, since they are separated by many kilometers. However, in the reference frame of George's car, they occur in the same place — assuming that George hasn't left the driver's seat!

Notice that any two events separated by a timelike interval are collocated in **some** reference frame. The speed of the reference frame is given by equation 4.2}, where the slope is simply the slope of the world line connecting the two events.

In Galilean relativity, if two events are simultaneous, we consider them to be simultaneous in all reference frames. For instance, if two clocks, one in New York and one in Los Angeles, strike the hour at the same time in the earth reference frame, then in Galilean relativity these events also appear to be simultaneous to instruments in the space shuttle as it flies over the United States. However, if the space shuttle is moving from west to east, i. e., from Los Angeles toward New York, careful measurements will show that the clock in New York strikes the hour **before** the clock in Los Angeles! Thus, the Galilean point of view is not accurate.

Just as collocation depends on one's reference frame, this result shows that simultaneity also depends on the reference frame. Figure 1.1.5: shows how this works. In Figure 1.1.5: events A and B are simultaneous in the rest or unprimed reference frame. However, in the primed reference frame, events A and C are simultaneous, and event B occurs at an earlier time. If A and B correspond to the clocks striking in Los Angeles and New York respectively, then it is clear that B must occur at an earlier time in the primed frame if indeed A and C are simultaneous in that frame.

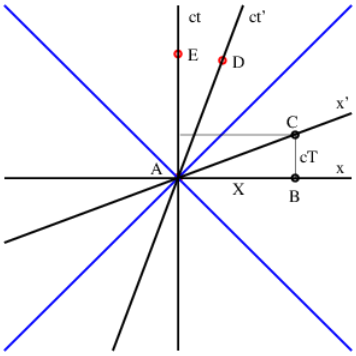


Figure 1.1.5:: Sketch of coordinate axes for a moving reference frame, x', ct'. The meanings of the events A-E are discussed in the text. The lines tilted at $\pm 45^\circ$ are the world lines of light passing through the origin.

The tilted line passing through events A and C in Figure 1.1.5: is called the **line of simultaneity** for the primed reference frame. Its slope is related to the speed, U, of the reference frame by

$$slope = U\ /\ c(lineofsimultaneity). \tag{1.1.87}$$

Notice that this is the inverse of the slope of the world line attached to the primed reference frame. There is thus a symmetry between the world line and the line of simultaneity of a moving reference frame — as the reference frame moves faster to the right, these two lines close like the blades of a pair of scissors on the 45° line.

In Galilean relativity it is fairly obvious what we mean by two events being simultaneous — it all boils down to coordinating portable clocks which are sitting next to each other, and then moving them to the desired locations. Two events separated in space are simultaneous if they occur at the same time on clocks located near each event, assuming that the clocks have been coordinated in the above manner.

In Einsteinian relativity this doesn't work, because the very act of moving the clocks changes the rate at which the clocks run. Thus, it is more difficult to determine whether two distant events are simultaneous.

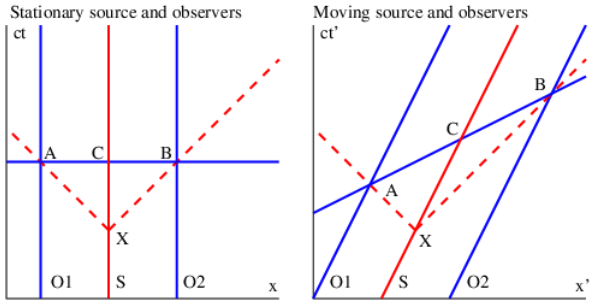


Figure 1.1.6:: World lines of two observers (O1, O2) and a pulsed light source (S) equidistant between them. In the left frame the observers and the source are all stationary. In the right frame they are all moving to the right at half the speed of light. The dashed lines show pulses of light emitted simultaneously to the left and the right.

An alternate way of experimentally determining simultaneity is shown in Figure 1.1.6:. Since we know from observation that light travels at the same speed in all reference frames, the pulses of light emitted by the light sources in Figure 1.1.6: will reach the two equidistant observers simultaneously in both cases. The line passing through these two events, A and B, defines a line of simultaneity for both stationary and moving observers. For the stationary observers this line is horizontal, as in Galilean relativity. For the moving observers the light has to travel farther in the rest frame to reach the observer receding from the light source, and it therefore takes longer in this frame. Thus, event B in the right panel of Figure 1.1.6: occurs later than event A in the stationary reference frame and the line of simultaneity is tilted. We see that the postulate that light moves at the same speed in all reference frames leads inevitably to the dependence of simultaneity on reference frame.

Spacetime Pythagorean Theorem

The Pythagorean theorem of spacetime differs from the usual Pythagorean theorem in two ways. First, the vertical side of the triangle is multiplied by c. This is a trivial scale factor that gives time the same units as space. Second, the right side of equation 4.3} has a minus sign rather than a plus sign. This highlights a fundamental difference between spacetime and the ordinary xyz

space in which we live. Spacetime is said to have a **non-Euclidean** (but not curved) geometry — in other words, the normal rules of geometry that we learn in high school don't always work for spacetime!

The main consequence of the minus sign in equation 4.3} is that I^2 can be negative and therefore I can be imaginary. Furthermore, in the special case where $X = \pm cT$, we actually have $I = 0$ even though $X, T \neq 0$ — i. e., the "distance" between two well-separated events can be zero. Clearly, spacetime has some weird properties!

The quantity I is usually called an **interval** in spacetime. Generally speaking, if I^2 is positive, the interval is called **spacelike** , while for a negative I^2, the interval is called **timelike** .

A concept related to the spacetime interval is the **proper time** τ. The proper time between the two events A and C in Figure 1.1.4: is defined by the equation

$$\tau 2 = T2 - X2 / c2. \tag{1.1.88}$$

Notice that I and τ are related by

$$\tau 2 = -I2 / c2, \tag{1.1.89}$$

so the spacetime interval and the proper time are not independent concepts. However, I has the dimensions of length and is real when the events defining the interval are spacelike relative to each other, whereas τ has the dimensions of time and is real when the events are timelike relative to each other. Both equation 4.3} and equation (4.5) express the spacetime Pythagorean theorem.

If two events defining the end points of an interval have the same t value, then the interval is the ordinary space distance between the two events. On the other hand, if they have the same x value, then the proper time is just the time interval between the events. If the interval between two events is spacelike, but the events are not simultaneous in the initial reference frame, they can always be made simultaneous by choosing a reference frame in which the events lie on the same line of simultaneity. Thus, the meaning of the interval in that case is just the distance between the events in the new reference frame. Similarly, for events separated by a timelike interval, the proper time is just the time between two events in a reference frame in which the two events are collocated.

Dilation

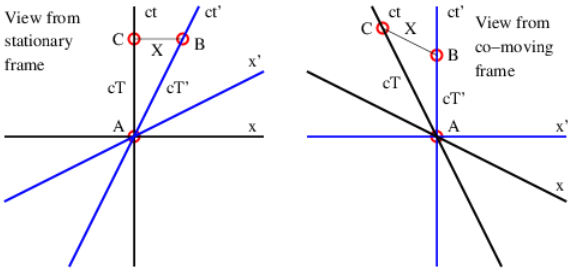


Figure 1.1.7:: Two views of the relationship between three events, A, B, and C. The left panel shows the view from the unprimed reference frame, in which A and C are collocated, while the right panel shows the view from the primed frame, in which A and B are collocated.

Stationary and moving clocks run at different rates in relativity. This is illustrated in Figure 1.1.7:. The triangle ABC in the left panel of Figure 1.1.7: can be used to illustrate this point. Suppose that the line passing through the events A and C in this figure is the world line of a stationary observer. At zero time another observer moving with velocity V passes the stationary observer. The moving observer's world line passes through events A and B.

We assume that events B and C are simultaneous in the rest frame, so ABC is a right triangle. Application of the spacetime Pythagorean theorem thus yields

$$2'2222cT = cT - X. \tag{1.1.90}$$

Since the second observer is moving at velocity V, the slope of his world line is

$$ccTV - = X - -, \tag{1.1.91}$$

where the right side of the above equation is the slope calculated as the rise of the world line cT over the run X between events A and B. Eliminating X between the above two equations results in a relationship between T and T':

$$T' = T(1 - V2 / c2)1 / 2 \equiv T / \gamma, \tag{1.1.92}$$

where

$$\gamma = - - - - - - - 1 - - - - - - - - -.(1 - V2 \ / \ c2)1 \ / \ 2 \tag{1.1.93}$$

The quantity γ occurs so often in relativistic calculations that we give it this special symbol. Note by its definition that $\gamma \geq 1$.

Equation (4.9) tells us that the time elapsed for the moving observer is less than that for the stationary observer, which means that the clock of the moving observer runs more slowly. This is called the **time dilation** effect.

Let us view this situation from the reference frame of the moving observer. In this frame the moving observer becomes stationary and the stationary observer moves in the opposite direction, as illustrated in the right panel of Figure 1.1.7:. By symmetric arguments, one infers that the clock of the initially stationary observer who is now moving to the left runs more slowly in this reference frame than the clock of the initially moving observer. One might conclude that this contradicts the previous results. However, examination of the right panel of Figure 1.1.7: shows that this is not so. The interval cT is still greater than the interval cT ′, because such intervals are relativistically invariant quantities. However, events B and C are no longer simultaneous, so one cannot use these results to infer anything about the rate at which the two clocks run in this frame. Thus, the relative nature of the concept of simultaneity saves us from an incipient paradox, and we see that the relative rates at which clocks run depends on the reference frame in which these rates are observed.

Lorentz Contraction

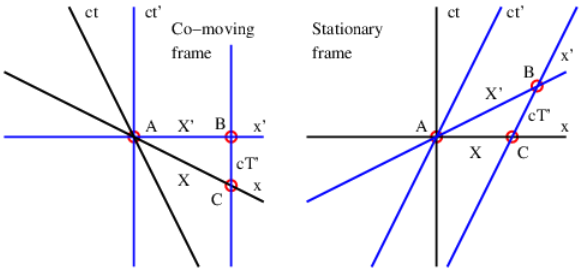


Figure 1.1.8:: Definition sketch for understanding the Lorentz contraction. The parallel lines represent the world lines of the front and the rear of a moving object. The left panel shows a reference frame moving with the object, while the right panel shows a stationary reference frame.

A similar argument can be made to show how the postulates of relativity result in the **Lorentz contraction** . Figure 1.1.8: compares the length X ′ of a moving object measured in its own reference frame (left panel) with its length X as measured in a stationary reference frame (right panel). The length of a moving object is measured by **simultaneously** measuring the positions of the front and the rear of the object and subtracting these two numbers. Events A and C correspond to these position measurements for the stationary reference frame since they are respectively on the rear and front world lines of the object. Thus, the interval AC, which is equal to X, is the length of the object as measured in the stationary frame.

In the left panel, X is the hypotenuse of a right triangle. Therefore, by the Pythagorean theorem of spacetime, we have

$$X = I = (X'2 - c2T'2)1 \ / \ 2. \tag{1.1.94}$$

Now, the line passing through A and C in the left panel is the line of simultaneity of the stationary reference frame. The slope of this line is -V / c, where V is the speed of the object relative to the stationary reference frame. Geometrically in Figure 1.1.8:, the slope of this line is -cT ′/X ′, so we find by equating these two expressions for the slope that

$$''2T = VX \ / \ c. \tag{1.1.95}$$

Finally, eliminating T ′ between (4.11) and (4.12) results in

$$X = X'(1 - V2 \ / \ c2)1 \ / \ 2 = X'/\gamma. \tag{1.1.96}$$

This says that the length of a moving object as measured in a stationary reference frame (X) is **less** than the actual length of the object as measured in its own reference frame (X ′). This reduction in length is called the Lorentz contraction.

Note that the Lorentz contraction only occurs in the direction of motion. The dimensions of a moving object perpendicular to the motion remain unchanged.

Paradox

An interesting application of time dilation is the so-called **twin paradox** , which turns out not to be a paradox at all. Two twins are initially the same age. One twin travels to a distant star on an interstellar spaceship which moves at speed V, which is close to the

speed of light. Upon reaching the star, the traveling twin immediately turns around and heads home. When reaching home, the traveling twin has aged less than the twin that stayed home. This is easily explained by the time dilation effect, which shows that the proper time elapsed along the world line of the traveling twin is $(1 - V^2/c^2)^{1/2}$ times the proper time elapsed along the world line of the other twin.

The “paradox” part of the twin paradox arises from making the symmetric argument in which one assumes the reference frame of the traveling twin to be stationary. The frame of the earth-bound twin must then travel in the sense opposite that of the erstwhile traveling twin, which means that the earth-bound twin must age less rather than more. This substitution is justified on the basis of the principle of relativity, which states that the laws of physics must be the same in all inertial reference frames.

However, the above argument is fallacious, because the reference frame of the traveling twin is **not** inertial throughout the entire trip, since at various points the spaceship has to accelerate and decelerate. Thus, the principle of relativity cannot be used to assert the equivalence of the traveler’s reference frame to the stationary frame.

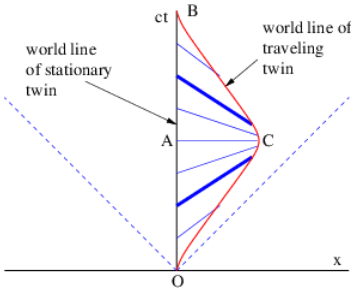


Figure 1.1.9:: Definition sketch for the twin paradox. The vertical line is the world line of the twin that stays at home, while the traveling twin has the curved world line to the right. The slanted lines between the world lines are lines of simultaneity at various times for the traveling twin. The heavy lines of simultaneity bound the period during which the traveling twin is decelerating to a stop and accelerating toward home.

Of particular importance is the period of deceleration and acceleration near the distant star. During this interval, the line of simulaneity of the traveling twin rotates, as illustrated in Figure 1.1.9:, such that the twin staying at home rapidly ages in the reference frame of the traveler. Thus, even though the acceleration of the traveling twin may occupy only a negligible segment of the twin’s world line, the overall effect is not negligible. In fact, the shorter and more intense the period of acceleration, the more rapidly the earth-bound twin ages in the traveling frame during this interval!

Problems

1. Sketch your personal world line on a spacetime diagram for the last 24 hours, labeling by time and location special events such as meals, physics classes, etc. Relate the slope of the world line at various times to how fast you were walking, riding in a car, etc.
2. Spacetime conversions:
    1. What is the distance from New York to Los Angeles in seconds? From here to the moon? From here to the sun?
    2. What is one nanosecond in meters? One second? One day? One year?
3. Three events have the following spacetime coordinates: A is at (x, ct) = (2 m, 1 m); B is at (x, ct) = (-2 m, 0 m); C is at (x, ct) = (0 m, 3 m).
    1. A world line for an object passes through events B and C. How fast and in which direction is the object moving?
    2. A line of simultaneity for a coordinate system passes through events A and B. How fast and in which direction is the coordinate system moving?
    3. What is the invariant interval between events A and B? B and C? A and C?
    4. Can a signal from event B reach event A? Can it reach event C? Explain.

    Hint: Draw a spacetime diagram with all the events plotted before trying to answer the above questions.
4. In the following problem be sure to indicate the slope of all pertinent lines drawn.
    1. In a spacetime diagram, sketch a line of simultaneity for a reference frame moving to the left at $V = c/2$, where c is the speed of light.
    2. Sketch the world line of an object which is initially stationary, but which accelerates to a velocity of $v = c/3$.
5. If the slopes of the world lines of the observers in the right panel of Figure 1.1.6: are both $1/\beta$, find the slope of their line of simultaneity, AB.

6. Suppose that an interstellar spaceship goes a distance X = 100 light years relative to the rest frame in T' = 10 years of its own time.
    1. Draw a spacetime diagram in the rest frame showing X, T', and the time T needed for this journey relative to the rest frame.
    2. Compute T, using your spacetime diagram as an aid.
    3. Compute the speed of the spaceship.
7. If an airline pilot flies 80 hr per month (in the rest frame) at 300 m s^{-1} for 30 years, how much younger will she be than her twin brother (who handles baggage) when she retires? Hint: Use $(1 + \epsilon)^x \approx 1 + x\epsilon$ for small ϵ.
8. A mu particle normally lives about 2×10^{-6} sec before it decays. However, muons created by cosmic rays 20 km up in the atmosphere reach the Earth's surface. How fast must they be going?
9. The Stanford Linear Accelerator accelerates electrons to a speed such that the 3 km long accelerator appears to be 8 cm long to the electron, due to the Lorentz contraction. How much less than the speed of light is the electron traveling? Hint: It is best to first develop an approximation for the relationship between $\gamma = (1 - v^2/c^2)^{-1/2}$ and the **difference** between c and v for a particle moving close to the speed of light.
10. How fast do you have to go to reach the center of our galaxy in your expected lifetime? At this speed, what does this distance appear to be? (We are about 30000 light years from the galactic center.)
11. Two identical spaceships pass each other going in the opposite direction at the same speed.
    1. Sketch a spacetime diagram showing the world lines of the front and rear of each spaceship as well as lines of simultaneity for each spaceship.
    2. Indicate an interval on the diagram corresponding to the rightward-moving spaceship's length in its own reference frame.
    3. Indicate an interval corresponding to the leftward-moving spaceship's length in the reference frame of the rightward-moving spaceship.
    4. Indicate an interval equal to the length of either spaceship in the rest frame.

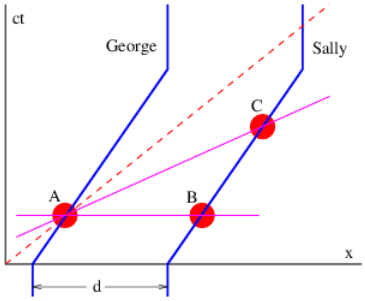


Figure 1.1.10:: Sketch for moving twins. Line AC is the line of simultaneity for a reference frame moving with Sally and George.

---

12. George and Sally are twins initially separated by a distance d and at rest. In the rest frame they are initially the same age. At time t = 0 both George and Sally get in their spaceships and head to the right at velocity U. Both move a distance d to the right and decelerate to a halt. (See Figure 1.1.10:.)
    1. When both are moving, how far away is Sally according to George?
    2. How much older or younger is Sally relative to George while both are moving?
    3. How much older or younger is Sally relative to George after both stop?

    Hint: Draw the triangle ABC in a reference frame moving with George and Sally.

Chapter 5 Applications of Special Relativity

In this chapter we continue the study of special relativity. Three important applications of the ideas developed in the previous chapter are made here. First, we show how to describe waves in the context of spacetime. We then see how waves which have no preferred reference frame (such as that of a medium supporting them) are constrained by special relativity to have a dispersion relation of a particular form. This dispersion relation turns out to be that of the relativistic matter waves of quantum mechanics. Second, we investigate the Doppler shift phenomenon, in which the frequency of a wave takes on different values in different coordinate systems. Third, we show how to add velocities in a relativistically consistent manner.

A new mathematical idea is presented in the context of relativistic waves, namely the spacetime vector or four-vector. Writing the laws of physics totally in terms of relativistic scalars and four-vectors insures that they will be valid in all inertial reference frames.

Waves in Spacetime

We now look at the characteristics of waves in spacetime. Recall that a sine wave moving to the right in one space dimension can be represented by

$$A(x,t) = A0sin(kx - \omega t), \tag{1.1.97}$$

where A_0 is the (constant) amplitude of the wave, k is the wavenumber, and ω is the angular frequency, and that the quantity $\phi = kx - \omega t$ is called the **phase** of the wave. For a plane wave in three space dimensions, the wave is represented in a similar way,

$$A(x,t) = Asin(k \cdot x - \omega t), 0 \tag{1.1.98}$$

where x is now the position vector and k is the wave vector. The magnitude of the wave vector, $|k| = k$ is just the wavenumber of the wave and the direction of this vector indicates the direction the wave is moving. The phase of the wave in this case is $\phi = k \cdot x - \omega t$.

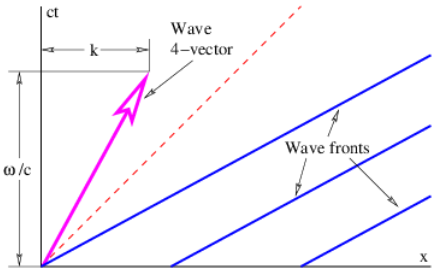


Figure 1.1.1:: Sketch of wave fronts for a wave in spacetime. The large arrow is the associated wave four-vector, which has slope ω / ck. The slope of the wave fronts is the inverse, ck / ω. The phase speed of the wave is greater than c in this example. (Can you tell why?)

In the case of a one-dimensional wave moving to the right $\phi = kx - \omega t$. A wave front has constant phase ϕ, so solving this equation for t and multiplying by c, the speed of light in a vacuum, gives us an equation for the world line of a wave front:

$$ckx - c\phi - cx - c\phi - ct = \omega - \omega = up - \omega(wavefront). \tag{1.1.99}$$

The slope of the world line in a spacetime diagram is the coefficient of x, or c / u_p, where $u_p = \omega / k$ is the phase speed. The world lines of the wave fronts of a wave are illustrated in Figure 1.1.1:.

Tutorial – Four-Vectors

Also shown in Figure 1.1.1: is a spacetime vector or **four-vector** which represents the frequency and wavenumber of the wave, which we refer to as the **wave four-vector** . It is called a four-vector because it has 3 spacelike components and one timelike component when there are 3 space dimensions. In the case shown, there is only a single space dimension. The spacelike component of the wave four-vector is just k (or k when there are 3 space dimensions), while the timelike component is ω / c. The c is in the denominator to give the timelike component the same dimensions as the spacelike component. From Figure 1.1.1: it is clear that the slope of the line representing the four-vector is ω / ck, which is just the inverse of the slope of the wave fronts.

Let us define some terminology. We indicate a four-vector by underlining and write the components in the following way: $k = (k, \omega / c)$, where k is the wave four-vector, k is its spacelike component, and ω / c is its timelike component. For three space dimensions, where we have a wave vector rather than just a wavenumber, we write $k = (k, \omega / c)$.

Another example of a four-vector is simply the position vector in spacetime, $x = (x, ct)$, or $x = (x, ct)$ in three space dimensions. The c **multiplies** the timelike component in this case, because that is what is needed to give it the same dimensions as the spacelike component.

In three dimensions we define a vector as a quantity with magnitude and direction. Extending this to spacetime, a four-vector is a quantity with magnitude and direction **in spacetime** . Implicit in this definition is the notion that the vector's magnitude is a quantity independent of coordinate system or reference frame. We have seen that the invariant interval in spacetime from the origin to the point (x, ct) is $I = (x^2 - c^2 t^2)^{1/2}$, so it makes sense to identify this as the magnitude of the position vector. This leads to a way of defining a dot product of four-vectors. Given two four-vectors $A = (A, A_t)$ and $B = (B, B_t)$, the dot product is

$$A \cdot B = A \cdot B - AB(dotproductinspacetime). ----tt \tag{1.1.100}$$

This is consistent with the definition of invariant interval if we set $A = B = x$, since then $x \cdot x = x^2 - c^2 t^2 = I^2$.

In the odd geometry of spacetime it is not obvious what “perpendicular” means. We therefore **define** two four-vectors A and B to be perpendicular if their dot product is zero, $A \cdot B = 0$, in analogy with ordinary vectors.

The dot product of two four-vectors is a **scalar** result, i. e., its value is independent of coordinate system. This can be used to our advantage on occasion. For instance, consider the dot product of a four-vector A which resolves into (A_x, A_t) in the unprimed frame. Let us further suppose that the spacelike component is zero in some primed frame, so that the components in this frame are $(0, A_t')$. The fact that the dot product is independent of coordinate system means that

$$22'2A - \cdot A - = Ax - At = -At. \tag{1.1.101}$$

This constitutes an extension of the spacetime Pythagorean theorem to four-vectors other than the position four-vector. Thus, for instance, the wavenumber for some wave may be zero in the primed frame, which means that the wavenumber and frequency in the unprimed frame are related to the frequency in the primed frame by

$$k2 - \omega 2 / c2 = -\omega' 2 / c2. \tag{1.1.102}$$

Principle of Relativity Applied

Returning to the phase of a wave, we immediately see that

$$\phi = k \cdot x - \omega t = k \cdot x - (\omega / c)(ct) = k \cdot x. \tag{1.1.103}$$

Thus, a compact way to rewrite equation 5.2} is

$$A(x) = A0sin(k - \cdot x). \tag{1.1.104}$$

Since x is known to be a four-vector and since the phase of a wave is known to be a scalar independent of reference frame, it follows that k is also a four-vector rather than just a set of numbers. Thus, the square of the length of the wave four-vector must also be a scalar independent of reference frame:

$$22k \cdot k - = k \cdot k - \omega / c = const. \tag{1.1.105}$$

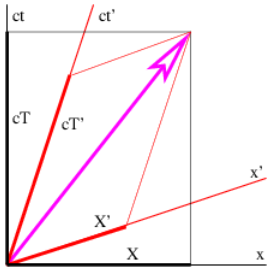


Figure 1.1.2:: Resolution of a four-vector into components in two different reference frames.

Let us review precisely what this means. As Figure 1.1.2: shows, we can resolve a position four-vector into components in two different reference frames, $x = (X, cT) = (X', cT')$. However, even though $X \neq X'$ and $T \neq T'$, the vector lengths computed from these two sets of components are necessarily the same: $x \cdot x = X^2 - c^2 T^2 = X'^2 - c^2 T'^2$.

Applying this to the wave four-vector, we infer that

$$222'2'22k - \omega / c = k - \omega / c = const., \tag{1.1.106}$$

where the unprimed and primed values of k and ω refer to the components of the wave four-vector in two different reference frames.

Up to now, this argument applies to **any** wave. However, waves can be divided into two categories, those for which a “special” reference frame exists, and those for which there is no such special frame. As an example of the former, sound waves look simplest in the reference frame in which the gas carrying the sound is stationary. The same is true of light propagating through a material medium with an index of refraction not equal to unity. In both cases the speed of the wave is the same in all directions **only** in the frame in which the material medium is stationary.

Suppose we have a machine that produces a wave with wavenumber k and frequency ω in its own rest frame. If we observe the wave from a moving reference frame, the wavenumber and frequency will be different, say, k' and ω'. However, these quantities will be related by equation 5.10}.

Up to this point the argument applies to **any** wave whether a special reference frame exists or not; the observed changes in wavenumber and frequency have nothing to do with the wave itself, but are just consequences of how we have chosen to observe it. However, if there is no special reference frame for the type of wave under consideration, then the same result can be obtained by keeping the observer stationary and moving the wave-producing machine in the opposite direction. By moving it at various speeds, any desired value of k' can be obtained **in the initial reference frame** (as opposed to some other frame), and the resulting value of ω' can be computed using equation 5.10}.

This is actually an amazing result. We have shown on the basis of the principle of relativity that any wave type for which no special reference frame exists can be made to take on a full range of frequencies and wavenumbers in any given reference frame, and furthermore that these frequencies and wavenumbers obey

$$2222\omega = kc + \mu. \tag{1.1.107}$$

Equation (5.11) comes from solving equation 5.10} for ω^2 and the constant μ^2 equals the constant in equation (5.10) times $-c^2$. Equation (5.11) relates frequency to wavenumber and therefore is the **dispersion relation** for such waves. We call waves which have no special reference frame and therefore necessarily obey equation (5.11) **relativistic waves** . The only difference in the dispersion relations between different types of relativistic waves is the value of the constant μ. The meaning of this constant will become clear later.

Characteristics of Relativistic Waves

Light in a vacuum is an example of a wave for which no special reference frame exists. For light, $\mu = 0$, and we have (taking the positive root) $\omega = ck$. This simply states what we know already, namely that the phase speed of light in a vacuum is c.

If $\mu \neq 0$, waves of this type are dispersive. The phase speed is

$$\omega up = -- = (c2 + \mu 2 / k2)1 / 2.k \tag{1.1.108}$$

This phase speed always exceeds c, which at first seems like an unphysical conclusion. However, the group velocity of the wave is

$$222ug = d - \omega = - - - - - -c - k - - - - - - - = kc - - = c-, \tag{1.1.109}$$
$$dk(k2c2 + \mu 2)1 / 2\omega up$$

which is always less than c. Since wave packets and hence signals propagate at the group velocity, waves of this type are physically reasonable even though the phase speed exceeds the speed of light.

Another interesting property of such waves is that the wave four-vector is parallel to the world line of a wave packet in spacetime. This is easily shown by the following argument. As Figure 1.1.1: shows, the spacelike component of a wave four-vector is k, while the timelike component is ω / c. The slope of the four-vector on a spacetime diagram is therefore ω / kc. However, the slope of the world line of a wave packet moving with group velocity u_g is $c / u_g = \omega / (kc)$, which is the same as the slope of the k four-vector.

Note that when $k = 0$ we have $\omega = \mu$. In this case the group velocity of the wave is zero. For this reason we call μ the **rest frequency** of the wave.

Doppler Shift

You have probably heard how the pitch of a train horn changes as it passes you. When the train is approaching, the pitch or frequency is higher than when it is moving away from you. This is called the Doppler shift. A similar, but distinct shift occurs if you are moving past a source of sound. If a stationary whistle is blowing, the pitch perceived from a moving car is higher while moving toward the source than when moving away. The first case thus has a **moving source** , while the second case has a **moving observer** .

In this section we compute the Doppler shift as it applies to light moving through a vacuum. Figure 1.1.3: shows the geometry for computing the time between wave fronts of light for a stationary and a moving reference frame. The time in the stationary frame is just T. Since the world lines of the wave fronts have a slope of unity, the sides of the shaded triangle have the same value, X. If the observer is moving at speed U, the slope of the observer's world line is c / U, which means that $c / U = (cT + X) / X$. Solving this for X yields $X = UT / (1 - U / c)$, which can then be used to compute $T' = T + X / c = T / (1 - U / c)$. This formula as it stands leads to the classical Doppler shift for a moving observer. However, with relativistic velocities, one additional factor needs to be

taken into account: The observer experiences time dilation since he or she is moving. The actual time measured by the observer between wave fronts is actually

$$()1 / 2'2221 / 2(1 - -U - 2 / c2)1 / 2 - 1 + -U - /c\tau \\ = (T - X / c) = T1 - U / c = T1 - U / c, \tag{1.1.110}$$

where the last step uses 1 - U^2/c^2 = (1 - U / c)(1 + U / c). From this we infer the relativistic Doppler shift formula for light in a vacuum:

$$(1 - U / c)1 / 2\omega' = \omega - - - - - - - - -, 1 + U / c \tag{1.1.111}$$

where the frequency measured by the moving observer is ω' = $2\pi / \tau$ and the frequency observed in the stationary frame is ω = $2\pi / T$.

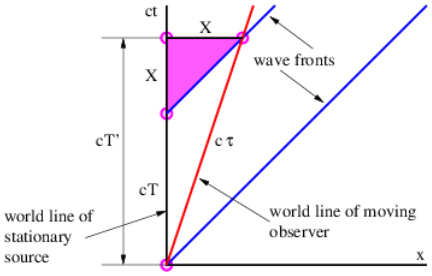


Figure 1.1.3:: Definition sketch for computing the Doppler shift for light.

We could go on to determine the Doppler shift resulting from a moving source. However, by the principle of relativity, the laws of physics should be the same in the reference frame in which the observer is stationary and the source is moving. Furthermore, the speed of light is still c in that frame. Therefore, the problem of a stationary observer and a moving source is conceptually the same as the problem of a moving observer and a stationary source when the wave is moving at speed c. This is unlike the case for, say, sound waves, where the stationary observer and the stationary source yield different formulas for the Doppler shift.

Addition of Velocities

Figure 1.1.4: shows the world line of a moving object from the point of view of two different reference frames, with the primed frame (left panel) moving to the right at speed U relative to the unprimed frame (right panel). The goal is to calculate the velocity of the object relative to the unprimed frame, v, assuming its velocity in the primed frame, v' is known. The classical result is simply

$$v = U + v'(classicalresult). \tag{1.1.112}$$

However, this is inconsistent with the speed of light being constant in all reference frames, since if we substitute c for v', this formula predicts that the speed of light in the unprimed frame is U + c.

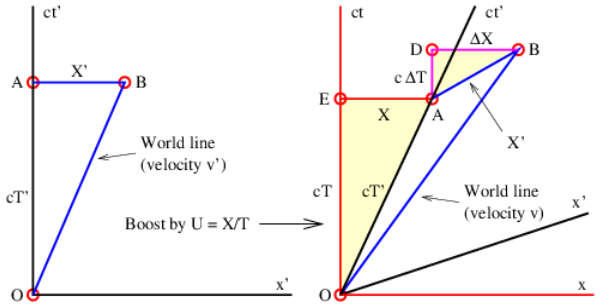


Figure 1.1.4:: Definition sketch for relativistic velocity addition. The two panels show the world line of a moving object relative to two different reference frames moving at velocity U with respect to each other. The velocity of the world line in the left panel is v' while its velocity in the right panel is v.

We can use the geometry of Figure 1.1.4: to come up with the correct relativistic formula. From the right panel of this figure we infer that

$$v - - = X - - + -\Delta X - - - = X - /(cT) - + - \Delta X - - / (cT) \\ .ccT + c\Delta T1 + \Delta T / T \tag{1.1.113}$$

This follows from the fact that the slope of the world line of the object in this frame is c / v. The slope is calculated as the ratio of the rise, $c\,(T + \Delta T\,)$, to the run, $X + \Delta X$.

From the left panel of Figure 1.1.4: we similarly see that

$$v'X'-- = --'.ccT \tag{1.1.114}$$

However, we can apply our calculations of Lorentz contraction and time dilation from the previous chapter to triangles ABD and OAE in the right panel. The slope of AB is U / c because AB is horizontal in the left panel, so $X' = \Delta X\,(1 - U^{2}/c^{2})^{1/2}$. Similarly, the slope of OA is c / U since OA is vertical in the left panel, and $T' = T\,(1 - U^{2}/c^{2})^{1/2}$. Substituting these formulas into the equation for v'/c yields

$$v'\Delta X--= ----.ccT \tag{1.1.115}$$

Again using what we know about the triangles ABD and OAE, we see that

$$U-= c\Delta T--= X--.c\Delta XcT \tag{1.1.116}$$

Finally, we calculate $\Delta T / T$ by noticing that

$$\begin{aligned}\Delta T\Delta Tc\Delta X(c\Delta T)(\Delta X)Uv'---- &= ----------\\ = ----------= ----.TTc\Delta X\Delta XcTcc\end{aligned} \tag{1.1.117}$$

Substituting equations (5.19), (5.20), and (5.21) into equation 5.17} and simplifying yields the relativistic velocity addition formula:

$$'v = --U -+- v----(specialrelativity).1 + Uv'/c2 \tag{1.1.118}$$

Notice how this equation behaves in various limits. If $|Uv'| \ll c^{2}$, the denominator of equation (5.22) is nearly unity, and the special relativistic formula reduces to the classical case. On the other hand, if $v' = c$, then equation (5.22) reduces to $v = c$. In other words, if the object in question is moving at the speed of light in one reference frame, it is moving at the speed of light in all reference frames, i. e., for all possible values of U. Thus, we have found a velocity addition formula that 1) reduces to the classical formula for low velocities and 2) gives the observed results for very high velocities as well.

Equation (5.22) is valid even if v' is negative, i. e., if the object is moving to the right less rapidly than the primed reference frame, or even if it is moving to the left in the unprimed frame.

Problems

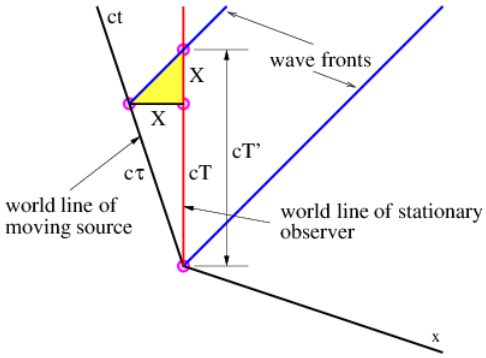


Figure 1.1.5:: Doppler shift for a moving light source.

1. Sketch the wave fronts and the k four-vector in a spacetime diagram for the case where $\omega / k = 2c$. Label your axes and space the wave fronts correctly for the case $k = 4\pi$ m^{-1}.
2. If the four-vector $k = (0, 1\text{ nm}^{-1})$ in the rest frame, find the space and time components of k in a frame moving to the **left** at speed $c/2$.
3. Let's examine the four-vector $u = (u_{g}, c)/\,(1 - \beta^{2})^{1/2}$ where $\beta = u_{g}/c$, u_{g} being the velocity of some object. The four-vector u is called the four-velocity.
    1. Show that u is parallel to the world line of the object.
    2. Show that $u \cdot u = -c^{2}$.
    3. If u_{g} is the group velocity of a relativistic wave packet, show that $k = (\mu / c^{2})u$, where k is the central wave four-vector of the wave packet.

4. Find the Doppler shift for a moving source of light from Figure 1.1.5:, roughly following the procedure used in the text to find the shift for a moving observer. (Assume that the source moves to the left at speed U.) Is the result the same as for the moving observer, as demanded by the principle of relativity?
5. Find the Doppler shift for a stationary source of **sound** ($c \approx 380$ m s^{-1} in a stationary atmosphere) as measured by a moving observer. Follow the procedure as for light in section 5.5 except note that the proper time τ is virtually the same as the time T' in Figure 1.1.3: since speeds are much less than the speed of light.
6. Suppose you shine a laser with frequency ω and wavenumber k on a mirror moving toward you at speed v, as seen in Figure 1.1.6:. What are the frequency ω' and wavenumber k' of the reflected beam? Hint: Find the frequency of the incident beam in the reference frame of the mirror. The frequency of the reflected beam will be the same as that of the incident beam in this frame. Then transform back to the reference frame of the laser.

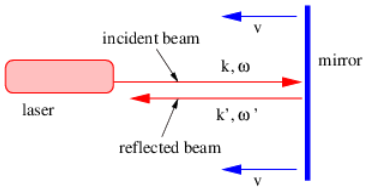


Figure 1.1.6:: Laser beam reflecting off of a moving mirror.

---

7. Suppose the moving twin in the twin paradox has a powerful telescope so that she can watch her twin brother back on earth during the entire trip. Describe how the earthbound twin appears to age to the travelling twin compared to her own rate of aging. Use a spacetime diagram to illustrate your argument and consider separately the outbound and return legs. Remember that light travels at the speed of light! Hint: Does the concept of Doppler shift help here?
8. Find the velocity of an object with respect to the rest frame if it is moving at a velocity of $0.1c$ with respect to another frame which itself is moving in the same direction at $0.1c$ relative to the rest frame using
    1. the Galilean formula and
    2. the formula of special relativity.

    Determine the fractional error made in using the Galilean formula.
9. Each stage of a high performance 3 stage rocket can accelerate to a speed of $0.9c$ from rest. If the rocket starts from rest, how fast does the final stage eventually go?
10. An interstellar spaceship is going from Earth to Sirius with speed $U = 0.8c$ relative to the rest frame. It passes a spaceship which is going from Sirius to Earth at a speed of $0.95c$ in the reference frame of the first spaceship. What is the velocity (direction and speed) of the second spaceship in the rest frame?

Chapter 6 Acceleration and General Relativity

General relativity is Einstein's extension of special relativity to include gravity. An important aspect of general relativity is that spacetime is no longer necessarily flat, but in fact may be curved under the influence of mass. Understanding curved spacetime is an advanced topic which is not easily accessible at the level of this text. However, it turns out that some insight into general relativistic phenomena may be obtained by investigating the effects of acceleration in the flat (but non-Euclidean) space of special relativity.

The central assumption of general relativity is the equivalence principle, which states that gravity is a force which arises from being in an accelerated reference frame. To understand this we must first investigate the concept of acceleration. We then see how this leads to phenomena such as the gravitational red shift, event horizons, and black holes. We also introduce in a preliminary way the notions of force and mass.

Acceleration

Imagine that you are in a powerful luxury car stopped at a stoplight. As you sit there, gravity pushes you into the comfortable leather seat. The light turns green and you "floor it". The car accelerates and an additional force pushes you into the seat back. You round a curve, and yet another force pushes you toward the outside of the curve. (But the well designed seat and seat belt keep you from feeling discomfort!)

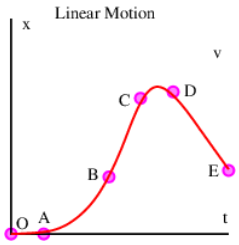


Figure 1.1.1:: Example of linear motion.

Let us examine the idea of **acceleration** more closely. Considering first acceleration in one dimension, Figure 1.1.1: shows the position of an object as a function of time, x (t). The velocity is simply the time rate of change of the position:

$$dx(t)v(t) = dt. \tag{1.1.119}$$

The acceleration is the time rate of change of velocity:

$$dv(t)d2x - (t)a(t) = dt = dt2. \tag{1.1.120}$$

In Figure 1.1.1:, only the segment OA has zero velocity. Velocity is increasing in AB, and the acceleration is positive there. Velocity is constant in BC, which means that the acceleration is zero. Velocity is decreasing in CD, and the acceleration is negative. Finally, in DE, the velocity is negative and the acceleration is zero.

In two or three dimensions, position x, velocity v, and acceleration a are all vectors, so that the velocity is

$$dx(t)v(t) = - - - - - - dt \tag{1.1.121}$$

while the acceleration is

$$dv(t)a(t) = - - - - - .dt \tag{1.1.122}$$

Thus, over some short time interval Δt, the changes in x and v can be written

$$\Delta x = v\Delta t\Delta v = a\Delta t. \tag{1.1.123}$$

These are vector equations, so the subtractions implied by the “delta” operations must be done vectorially. An example where the vector nature of these quantities is important is motion in a circle at constant speed, which is discussed in the next section.

Circular Motion

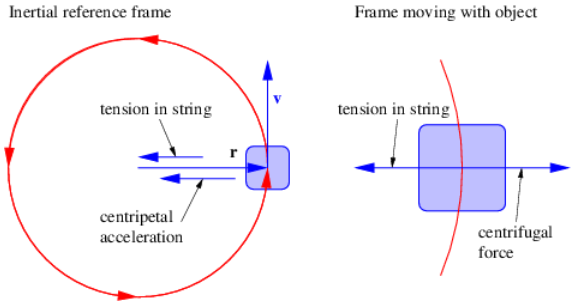


Figure 1.1.2:: Two different views of circular motion of an object. The left panel shows the view from the inertial reference frame at rest with the center of the circle. The tension in the string is the only force and it causes an acceleration toward the center of the circle. The right panel shows the view from an accelerated frame in which the object is at rest. In this frame the tension in the string balances the centrifugal force, which is the inertial force arising from being in an accelerated reference frame, leaving zero net force.

Imagine an object constrained by an attached string to move in a circle at constant speed, as shown in the left panel of Figure 1.1.2:. We now demonstrate that the acceleration of the object is toward the center of the circle. The acceleration in this special case is called the **centripetal acceleration** .

Figure 1.1.3: shows the position of the object at two times spaced by the time interval Δt. The position vector of the object relative to the center of the circle rotates through an angle $\Delta\theta$ during this interval, so the angular rate of revolution of the object about the center is $\omega = \Delta\theta/\ \Delta t$. The magnitude of the velocity of the object is v, so the object moves a distance $v\ \Delta t$ during the time interval. To the extent that this distance is small compared to the radius r of the circle, the angle $\Delta\theta = v\ \Delta t\ /\ r$. Solving for v and using $\omega = \Delta\theta\ /\ \Delta t$, we see that

$$v = \omega r (circular motion). \tag{1.1.124}$$

The direction of the velocity vector changes over this interval, even though the magnitude v stays the same. Figure 1.1.3: shows that this change in direction implies an acceleration a which is directed toward the center of the circle, as noted above. The magnitude of the vectoral change in velocity in the time interval Δt is $a\ \Delta t$. Since the angle between the initial and final velocities is the same as the angle $\Delta\theta$ between the initial and final radius vectors, we see from the geometry of the triangle in Figure 1.1.3: that $a\ \Delta t\ /\ v = \Delta\theta$. Solving for a results in

$$a = \omega v (circular motion). \tag{1.1.125}$$

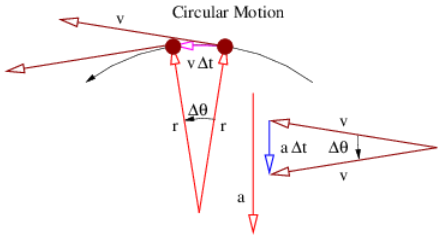


Figure 1.1.3:: Definition sketch for computing centripetal acceleration.

Combining equations (6.6) and (6.7) yields the equation for centripetal acceleration:

$$a = \omega 2r = v2\ /\ r (centripetal acceleration). \tag{1.1.126}$$

The second form is obtained by eliminating ω from the first form using equation 6.6}.

Acceleration, Force, and Mass

We have a good intuitive feel for the concepts of force and mass because they are very much a part of our everyday experience. We think of force as how hard we push on something. Mass is the resistance of an object to acceleration if it is otherwise free to move. Thus, pushing on a bicycle on a smooth, level road causes it to accelerate more readily than pushing on a car. We say that the car has greater mass. We can summarize this relationship with Newton's second law

$$F = ma \tag{1.1.127}$$

where F is the total force on an object, m is its mass, and a is the acceleration resulting from the force.

Three provisos apply to equation 6.9}. First, it only makes sense in unmodified form when the velocity of the object is much less than the speed of light. For relativistic velocities it is best to write this equation in a slightly different form which we introduce later. Second, the force must be the total force, including all frictional and other incidental forces which might otherwise be neglected by an uncritical observer. Third, it only works in a reference frame which itself is unaccelerated, i.e., an **inertial reference frame** . We deal below with accelerated reference frames.

Acceleration in Special Relativity

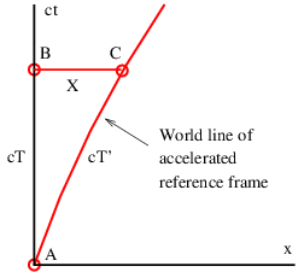


Figure 1.1.4:: World line of the origin of an accelerated reference frame.

As noted above, acceleration is just the time rate of change of velocity. We use the above results to determine how acceleration transforms from one reference frame to another. Figure 1.1.4: shows the world line of an **accelerated reference frame** , with a time-varying velocity U (t) relative to the unprimed inertial rest frame. Defining ΔU = U (T) - U (0) as the change in the velocity of the accelerated frame (relative to the unprimed frame) between events A and C, we can relate this to the change of velocity, ΔU ′, of the accelerated frame **relative to an inertial frame moving with the initial velocity** , U (0). Applying the equation for the relativistic addition of velocities, we find

$$U(0) + \Delta U' U(T) = U(0) + \Delta U = ---------- \tag{1.1.128}$$

$$'--2.1+U(0)\Delta U\,/\,c$$

We now note that the mean acceleration of the reference frame between events A and C in the rest frame is just $a = \Delta U\,/\,T$, whereas the mean acceleration in the primed frame between the same two events is $a' = \Delta U'/T'$. From equation 6.10} we find that

$$\Delta U'[1-U(0)2\,/\,c2]\Delta U = \frac{}{'--2-,1+U(0)\Delta U\,/\,c} \tag{1.1.129}$$

and the acceleration of the primed reference frame as it appears in the unprimed frame is

$$-U(0)2\,/\,c2]a = \frac{\Delta U\Delta U'[1}{+U(0)\Delta U'/c2]} \tag{1.1.130}$$

Since we are interested in the instantaneous rather than the average acceleration, we let T become small. This has three consequences. First, ΔU and $\Delta U'$ become small, which means that the term $U(0)\Delta U'/c^2$ in the denominator of equation 6.12} can be ignored compared to 1. This means that

$$'22a \approx \Delta U--[1---U-(0)-/c-],T \tag{1.1.131}$$

with the approximation becoming exact as $T \to 0$. Second, the "triangle" with the curved side in Figure 1.1.4: becomes a true triangle, with the result that $T' = T\ [1 - U\ (0)^2/c^{\ 2}]^{1/2}$. The acceleration of the primed frame with respect to an inertial frame moving at speed U (0) can therefore be written

$$'\Delta U'\Delta U'a = ---'- = ----------2--21\,/\,2.TT[1-U(0)\,/\,c] \tag{1.1.132}$$

Third, we can replace U (0) with U, since the velocity of the accelerated frame doesn't change very much over a short time interval.

Dividing equation 6.13} by equation (6.14) results in a relationship between the two accelerations:

$$a = a'(1-U2\,/\,c2)3\,/\,2, \tag{1.1.133}$$

which shows that the acceleration of a rapidly moving object, a, as observed from the rest frame, is less than its acceleration relative to an inertial reference frame in which the object is nearly stationary, a', by the factor $(1 - U^{\ 2}/c^{\ 2})^{3/2}$. We call this latter acceleration the **intrinsic acceleration** . This difference in observed acceleration between the two inertial reference frames is purely the result of the geometry of spacetime, but it has interesting consequences.

Identifying a with $dU\,/\,dt$, we can integrate the acceleration equation assuming that the intrinsic acceleration a' is constant and that the velocity $U = 0$ at time $t = 0$. We get the following result (verify this by differentiating with respect to time):

$$a't = -------U--------,(1-U2\,/\,c2)1\,/\,2 \tag{1.1.134}$$

which may be solved for $U\,/\,c$:

$$'U-=-----a-t\,/\,c-----.c[1+(a't\,/\,c)2]1\,/\,2 \tag{1.1.135}$$

This is plotted in Figure 1.1.5:. Classically, the velocity U would reach the speed of light when $a't\,/\,c = 1$. However, as Figure 1.1.5: shows, the rate at which the velocity increases with time slows as the object moves faster, such that U approaches c asymptotically, but never reaches it.

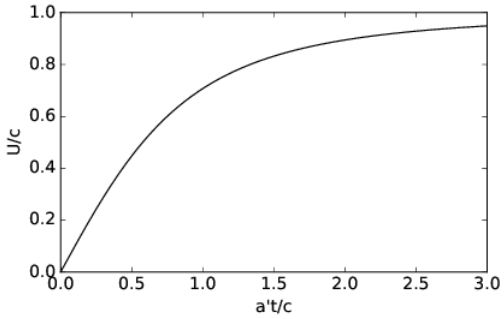


Figure 1.1.5:: Velocity divided by the speed of light as a function of the product of the time and the (constant) acceleration divided by the speed of light.

The results for this section are valid only for acceleration components in the direction of motion. The components perpendicular to this direction behave differently and are treated in more advanced texts.

Accelerated Reference Frames

Referring back to the forces being felt by the occupant of a car, it is clear that the forces associated with accelerations are directed opposite the accelerations and proportional to their magnitudes. For instance, when accelerating away from a stoplight, the acceleration is forward and the perceived force is backward. When turning a corner, the acceleration is toward the corner while the perceived force is away from the corner. Such forces are called **inertial forces** .

The origin of these forces can be understood by determining how acceleration changes when one observes it from a reference frame which is itself accelerated. Suppose that the primed reference frame is accelerating to the right with acceleration A relative to the unprimed frame. The position x' in the primed frame can be related to the position x in the unprimed frame by

$$x = x' + X, \tag{1.1.136}$$

where X is the position of the origin of the primed frame in the unprimed frame. Taking the second time derivative, we see that

$$'a = a + A, \tag{1.1.137}$$

where $a = d^2x/dt^2$ is the acceleration in the unprimed frame and $a' = d^2x'/dt^2$ is the acceleration according to an observer in the primed frame.

We now substitute this into equation 6.9} and move the term involving A to the left side:

$$F - mA = ma'. \tag{1.1.138}$$

This shows that Newton's second law represented by equation 6.9} is not valid in an accelerated reference frame, because the total force F and the acceleration a' in this frame don't balance as they do in the unaccelerated frame; the additional term -mA messes up this balance.

We can fix this problem by considering -mA to be a type of force, in which case we can include it as a part of the total force F. This is the inertial force which we mentioned above. Thus, to summarize, we can make Newton's second law work when objects are observed from accelerated reference frames if we include as part of the total force an inertial force which is equal to -mA, A being the acceleration of the reference frame of the observer and m the mass of the object being observed.

The right panel of Figure 1.1.2: shows the inertial force observed in the reference frame of an object moving in circular motion at constant speed. In the case of circular motion the inertial force is called the **centrifugal force** . It points away from the center of the circle and just balances the tension in the string. This makes the total force on the object zero in its own reference frame, which is necessary since the object cannot move (or accelerate) in this frame.

General relativity says that gravity is nothing more than an inertial force. This was called the **equivalence principle** by Einstein. Since the gravitational force on the Earth points downward, it follows that we must be constantly accelerating upward as we stand on the surface of the Earth! The obvious problem with this interpretation of gravity is that we don't appear to be moving away from the center of the Earth, which would seem to be a natural consequence of such an acceleration. However, relativity has a surprise in store for us here.

It follows from the above considerations that something can be learned about general relativity by examining the properties of accelerated reference frames. In particular, we can gain insight into the above apparent paradox. Equation (6.17) shows that the velocity of an object undergoing constant intrinsic acceleration a (note that we have dropped the "prime" from a to simplify the notation) is

$$dxatv = - - - = - - - - - - - - - - - - - 21/2, dt[1 + (at/c)] \tag{1.1.139}$$

where t is the time and c is the speed of light. A function x (t) which satisfies equation 6.21} is

$$221/2x(t) = (c/a)[1 + (at/c)]. \tag{1.1.140}$$

(Verify this by differentiating it.) The interval OB in Figure 1.1.6: is of length x (0) = c^2/A.

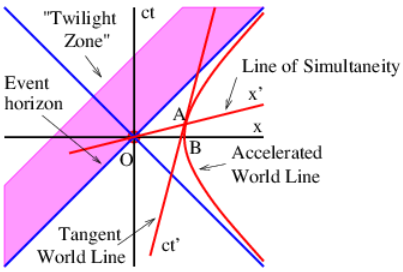


Figure 1.1.6:: Spacetime diagram showing the world line of the origin of a reference frame undergoing constant acceleration.

The slanted line OA is a line of simultaneity associated with the unaccelerated world line tangent to the accelerated world line at point A. This line of simultaneity goes through the origin, as is shown in Figure 1.1.6:. To demonstrate this, multiply equations (6.21) and (6.22) together and solve for v / c :

$$v / c = ct / x. \tag{1.1.141}$$

From Figure 1.1.6: we see that ct / x is the slope of the line OA, where (x, ct) are the coordinates of event A. Equation (6.23) shows that this line is indeed the desired line of simultaneity, since its slope is the inverse of the slope of the world line, c / v. Since there is nothing special about the event A, we infer that **all** lines of simultaneity associated with the accelerated world line pass through the origin.

We now inquire about the length of the invariant interval OA in Figure 1.1.6:. Recalling that $I^2 = x^2 - c^2 t^2$ and using equation 6.22}, we find that the length of OA is

$$2221 / 2421 / 22I = (x - ct) = (c / a) = c / a, \tag{1.1.142}$$

which is the same as the length of the interval OB. By extension, **all** events on the accelerated world line are the same invariant interval from the origin. Recalling that the interval along a line of simultaneity is the distance in the associated reference frame, we reach the astonishing conclusion that even though the object associated with the curved world line in Figure 1.1.6: is accelerating away from the origin, it always remains the same distance (in its own frame) from the origin.

The analogy between this problem and the apparent paradox in which one remains a fixed distance from the center of the earth while accelerating away from it is not perfect. In particular, the earth case depends on the existence of the earth's mass.

Gravitational Red Shift

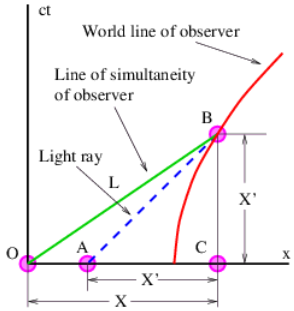


Figure 1.1.7:: Spacetime diagram for explaining the gravitational red shift. Why is the interval AC equal to the interval BC? L is the length of the invariant interval OB.

Light emitted at a lower level in a gravitational field has its frequency reduced as it travels to a higher level. This phenomenon is called the **gravitational red shift** . Figure 1.1.26: shows why this happens. Since experiencing a gravitational force is equivalent to being in an accelerated reference frame, we can use the tools of special relativity to view the process of light emission and absorption from the point of view of the unaccelerated or inertial frame. In this reference frame the observer of the light is accelerating to the right, as indicated by the curved world line in Figure 1.1.26:, which is equivalent to a gravitational force to the left. The light is emitted at point A with frequency ω by a source which is stationary at this instant. At this instant the observer is also stationary in this frame. However, by the time the light gets to the observer, he or she has a velocity to the right which means that the observer measures a Doppler shifted frequency ω' for the light. Since the observer is moving away from the source, $\omega' < \omega$, as indicated above.

The relativistic Doppler shift is given by

$$'()1 / 2\omega - -1 - - - U - /c\omega = 1 + U / c. \tag{1.1.143}$$

so we need to compute U / c. The line of simultaneity for the observer at point B goes through the origin, and is thus given by line segment OB in Figure 1.1.26:. The slope of this line is U / c, where U is the velocity of the observer at point B. From the figure we see that this slope is also given by the ratio X' / X. Equating these, eliminating X in favor of $L = (X^2 - X'^2)^{1/2} = c^2/g$, which is the actual invariant distance of the observer from the origin, and substituting into equation 6.25} results in our gravitational red shift formula:

$$\begin{aligned} &'(')1 / 2(2'21 / 2')1 / 2\omega - - = X - - - - X - - \\ &= (L - - + -X - - -) - - - - - X - -.\omega X + X'(L2 + X'2)1 / 2 + X' \end{aligned} \tag{1.1.144}$$

If $X' = 0$, then there is no redshift, because the source is collocated with the observer. On the other hand, if the source is located at the origin, so $X' = X$, the Doppler shifted frequency is zero. In addition, the light never gets to the observer, since the world line is asymptotic to the light world line passing through the origin. If the source is at a higher level in the gravitational field than the observer, so that $X' < 0$, then the frequency is shifted to a higher value, i. e., it becomes a "blue shift".

Equation (6.26) works for more complex geometries than that associated with an accelerated reference frame, e.g., for the gravitational field g associated with a star, as long as $|X'| \ll c^2/g$. In this case L is no longer the distance to the center of the star but remains equal to c^2/g.

Event Horizons

The 45° diagonal line passing through the origin in Figure 1.1.6: is called the **event horizon** for the accelerated observer in this figure. Notice that light from the "twilight zone" above and to the left of the event horizon cannot reach the accelerated observer. However, the reverse is not true — a light signal emitted to the left by the observer can cross the event horizon into the "twilight zone". The event horizon thus has a peculiar one-way character — it passes signals from right to left, but not from left to right.

Problems

1. An object moves as described in Figure 1.1.8:, which shows its position x as a function of time t.
    1. Is the velocity positive, negative, or zero at each of the points A, B, C, D, E, and F?
    2. Is the acceleration positive, negative, or zero at each of the points A, B, C, D, E, and F?

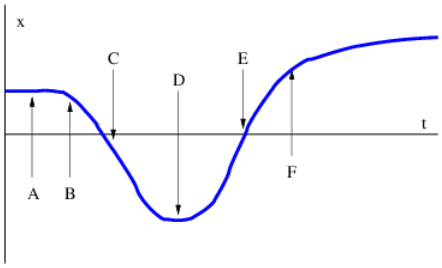


Figure 1.1.8:: Position of an object as a function of time.

---

2. An object is moving counterclockwise at constant speed around the circle shown in Figure 1.1.9: due to the fact that it is attached by a string to the center of the circle at point O.
    1. Sketch the object's velocity vectors at points A, B, and C.
    2. Sketch the object's acceleration vectors at points A, B, and C.
    3. If the string breaks at point A, sketch the subsequent trajectory followed by the object.

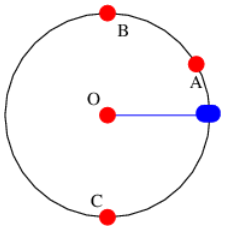


Figure 1.1.9:: Object in circular motion.

3. How fast are you going after accelerating from rest with intrinsic acceleration a = 10 m s^{-2} after the given times measured in the rest frame:
    1. 10 y?
    2. 100000 y?

    Express your answer as the speed of light minus your actual speed. Hint: You may have a numerical problem on the second part, which you should try to resolve using the approximation $(1 + \epsilon)^x \approx 1 + x\epsilon$, which is valid for $|\epsilon| \ll 1$.
4. An object's world line is defined by $x(t) = (d^2 + c^2 t^2)^{1/2}$ where d is a constant and c is the speed of light.
    1. Find the object's velocity as a function of time.
    2. Using the above result, find the slope of the tangent to the world line as a function of time.
    3. Find where the line of simultaneity corresponding to each tangent world line crosses the x axis.
5. A car accelerates in the positive x direction at 3 m s^{-2}.
    1. What is the net force on a 100 kg man in the car as viewed from an inertial reference frame?
    2. What is the inertial force experienced by this man in the reference frame of the car?
    3. What is the net force experienced by the man in the car's (accelerated) reference frame?
6. A person is sitting in a comfortable chair in her home in Bogotá, Colombia, which is essentially on the equator.
    1. What would the rotational period of the earth have to be to make this person weightless?
    2. What is her acceleration according to the equivalence principle in the earth frame in this situation?
7. At time t = 0 a Zork (a creature from the planet Zorkheim) accelerating to the right at $a = 10^3$ m s^{-2} in a spaceship accidently drops its stopwatch from the spaceship just when its velocity is zero.
    1. Describe qualitatively how the hands of the watch appear to move to the Zork as it observes the watch through a powerful telescope.
    2. After a very long time what does the watch read?

    Hint: Draw a spacetime diagram with the world lines of the spaceship and the watch. Then send light rays from the watch to the spaceship.
8. Using a spacetime diagram, show why signals from events on the hidden side of the event horizon from an accelerating spaceship cannot reach the spaceship.
9. Approximate equation 6.26} to first order in X' for the case in which $X' \ll L$.
10. }
11. Imagine two identical clocks, one on top of the volcano Chimborazo in Ecuador (6300 m above sea level), the other in the Ecuadorian city of Guayaquil (at sea level).
    1. From the perspective of Chimborazo, does the clock in Guayaquil appear to be running faster or slower than the Chimborazo clock? Explain.
    2. Compute the fractional frequency difference $(\omega - \omega')/\omega$ in this case, where ω is the freqency of the Guayaquil clock as observed in Guayaquil (and the frequency of the Chimborazo clock on Chimborazo) and ω' is the frequency of the Guayaquil clock as observed from Chimborazo. You may wish to use the results of the previous problem.

Chapter 7 Matter Waves

We begin our study of quantum mechanics by discussing the diffraction undergone by X-rays and electrons when they interact with a crystal. X-rays are a form of electromagnetic radiation with wavelengths comparable to the distances between atoms. Scattering from atoms in a regular crystalline structure results in an interference pattern which is in many ways similar to the pattern from a diffraction grating. We first develop Bragg's law for diffraction of X-rays from a crystal. Two practical techniques for doing X-ray diffraction are then described.

It turns out that electrons have wave-like properties and also undergo Bragg diffraction by crystals. Bragg diffraction thus provides a crucial bridge between the worlds of waves and particles. With this bridge we introduce the classical ideas of momentum and energy by relating them to the wave vector and frequency of a wave. The properties of waves also give rise to the Heisenberg uncertainty principle.

Table 7.1 shows a table of the Nobel prizes associated with the ideas presented in this chapter. This gives us a feel for the chronology of these discoveries and indicates how important they were to the development of physics in the early 20th century.

Year	Recipient	Contribution
1901	W. K. Röntgen	Discovery of X-rays
1906	J. J. Thomson	Discovery of electron
1914	M. von Laue	X-ray diffraction in crystals
1915	W. and L. Bragg	X-ray analysis of crystal structure
1918	M. Planck	Energy quantization
1921	A. Einstein	Photoelectric effect
1922	N. Bohr	Structure of atoms
1929	L.-V. de Broglie	Wave nature of electrons
1932	W. Heisenberg	Quantum mechanics
1933	Schrödinger and Dirac	Atomic theory
1937	Davisson and Thomson	Electron diffraction in crystals

Table 7.1: Selected Nobel prize winners, year of award, and contribution.

Bragg's Law

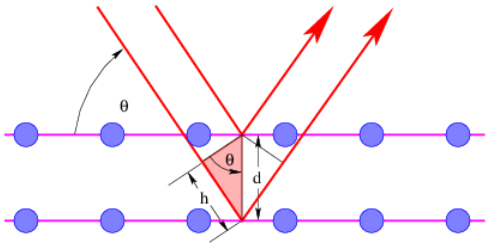


Figure 1.1.1:: Schematic diagram for determining Bragg's law.

Figure 1.1.1: schematically illustrates interference between waves scattering from two adjacent rows of atoms in a crystal. The net effect of scattering from a single row is equivalent to partial reflection from a mirror imagined to be aligned with the row. Thus, the angle of "reflection" equals the angle of incidence for each row. Interference then occurs between the beams reflecting off different rows of atoms in the crystal.

For the two adjacent rows shown in Figure 1.1.1:, the path difference between beams is $2h = 2d \sin\theta$. For constructive interference this must be an integer number of wavelengths, $m\lambda$, where the integer m is called the **order of interference** . The result is Bragg's law of diffraction:

$$m\lambda = 2dsin\theta, m = 1, 2, 3...(Bragg's law). \tag{1.1.145}$$

If only two rows are involved, the transition from constructive to destructive interference as θ changes is gradual. However, if interference from many rows occurs, then the constructive interference peaks become very sharp with mostly destructive interference in between. This sharpening of the peaks as the number of rows increases is very similar to the sharpening of the diffraction peaks from a diffraction grating as the number of slits increases.

X-Ray Diffraction Techniques

Two types of targets are used in Bragg diffraction experiments: single crystals and powder targets.

Single Crystal

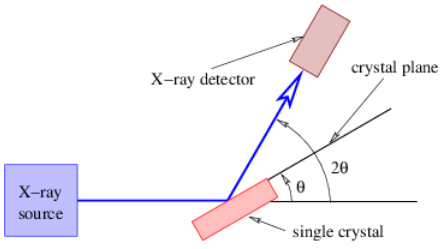


Figure 1.1.2:: Setup for single crystal Bragg diffraction.

In a single crystal setup, an X-ray detector is mounted as shown in Figure 1.1.2:. A mechanical device keeps the detector oriented so that the angle of incidence equals the angle of reflection for the desired crystal plane. Peaks in the X-ray detection rate are sought as the angle θ is varied.

The advantage of this type of apparatus is that diffraction peaks from only the selected crystal plane are observed.

Powder Target

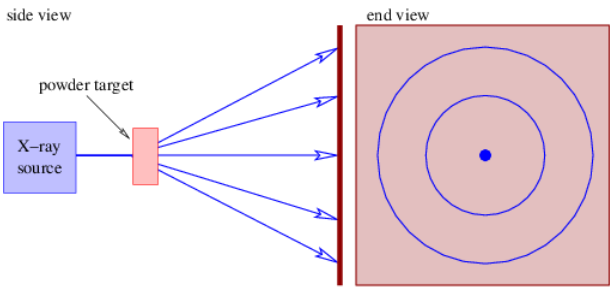


Figure 1.1.3:: Setup for powder target Bragg diffraction.

The powder in a powder target is really a conglomeration of many tiny crystals randomly oriented. Thus, for each possible Bragg diffraction angle there are crystals oriented correctly for Bragg diffraction to take place. The detector is usually a photographic plate or an equivalent electronic device as illustrated in Figure 1.1.3:. For each Bragg diffraction angle one sees a ring on the plate concentric with the axis of the incident X-ray beam.

The advantage of this type of system is that no knowledge is needed of the crystal plane orientations. Furthermore, a single large crystal is not required. However, all possible Bragg scattering angles are seen at once, which can lead to confusion in the interpretation of the results.

Meaning of Quantum Wave Function

Bragg diffraction illustrates the most difficult thing to understand about quantum mechanics, namely that particles can have wave-like properties and waves can have particle-like properties.

The variation of X-ray intensity with angle seen in a Bragg diffraction apparatus is very difficult to explain in any terms other than wave interference. Yet, X-rays are typically detected by a device such as a Geiger counter which produces a pulse of electricity for each X-ray particle, or **photon** , which hits it. Thus, X-rays sometimes act like particles and sometimes like waves.

Light isn't alone in having both particle and wave properties. Davisson and Germer and later G. P. Thomson (son of J. J. Thomson, the discoverer of the electron) showed that electrons also can act like waves. They did this by demonstrating that electrons undergo Bragg diffraction in crystals, much as X-rays do.

Most physicists (including Albert Einstein) have found quantum mechanics to be extremely bizarre, so if you feel the same way, you are in good company! However, there is a useful interpretation of quantum mechanics which at least allows us to get on with using it to solve problems, even though it may not satisfy our intuitive reservations about the theory.

The displacement of the matter wave associated with a particle is usually called the wave function, ψ. It is not at all clear what ψ is a displacement **of** , but its use is straightforward. The absolute square of the wave function, $|\psi\ (x, t\ )|^2$, is proportional to the **probability** of finding the associated particle at position x and time t. The **absolute** square is taken because under many circumstances the wave function is actually **complex** , i. e., it has both real and imaginary parts. The reasons for this will be discussed later.

Due to the interpretation of the wave function, quantum mechanics is a probabilistic theory. It does not tell us with certainty what happens to a particular particle. Instead, it tells us the probability for detecting the particle in any given location. If many experiments are done, with one particle per experiment, the numbers of experiments with particles being detected in the various possible locations are in proportion to the quantum mechanical probabilities.

7.4 Sense and Nonsense in Quantum Mechanics

The essential mystery of quantum mechanics becomes clearer when discussed in the conceptually simpler context of two slit interference. If light and electrons can have both particle and wave properties, then one might ask through which of the two slits the particle passed. However, in physics a question simply doesn't make sense if it cannot be answered by experiment.

One can indeed perform an experiment to determine which slit the X-ray photon or electron passes through in the two slit experiment. However, by the very act of making this measurement, the form of the associated wave is altered. In particular, since the absolute square of the wave displacement represents the probability of finding the particle, once the particle has definitely been found passing through one or the other of the slits, the wave function collapses into a very small wave packet located at the observed position of the particle. Thus, the wave displacement becomes zero at the slit it didn't pass through. However, the

interference pattern results from the superposition of waves emanating from two slits. If no wave comes from one of the slits (because the wave displacement is zero there), then there can be no interference pattern!

We can now make the inverse argument. If there is an interference pattern, then we know that the wave displacement is non-zero at both slits. From the probability interpretation of the wave displacement, we conclude that we cannot say, even in principle, through which slit the particle passed. It is not just that we don't know the answer to this question; there is simply no experiment which can give us an answer without destroying the interference pattern. In other words, the question "Which slit did the particle pass through?" is a nonsensical question in the case where an interference pattern is actually produced.

The American physicist Richard Feynman noticed that the above behavior can be interpreted as violating the normal laws of probability. These laws say that the probability of an event is the sum of the probabilities of alternate independent ways for that event to occur. For instance, the probability for a particle to reach point A on the detection screen of a two slit setup is just the probability P_1 for the particle to reach point A after going through slit 1, plus the probability P_2 for the particle to reach point A after going through slit 2. Thus, if $P_1 = P_2 = 0.1$, then the probability for the particle to reach point A irrespective of which slit it went through should be $P_{\text{total}} = P_1 + P_2 = 0.2$. However, if point A happens to be a point of destructive interference, then we know that $P_{\text{total}} = 0$.

Feynman proposed that the above rule stating that alternate independent probabilities add, is simply incorrect. In its place Feynman asserted that **probability amplitudes** add instead, where the probability amplitude in this case is just the wave function associated with the particle. The total probability for a process is obtained by adding the alternate probability amplitudes together and taking the absolute square of the sum.

Feynman's view is a particularly compact expression of the so-called **Copenhagen interpretation** of quantum mechanics which evolved from the ideas of Niels Bohr, Werner Heisenberg, Max Born, and others in the 1920s. It dispenses with the wave-particle duality and other philosophical baggage by saying "Particles are real in that we can observe them, but the only theory we have is about probability amplitudes for particles." This interpretation of quantum mechanics may be weird, but it appears to be self-consistent and in agreement with experiment.

Mass, Momentum, and Energy

In this section we relate the classical ideas of mass, momentum and energy to what we have done so far. Historically, these connections were first made by Max Planck and Louis de Broglie with help from Albert Einstein. Bragg diffraction of electrons is invoked as an experimental test of the Planck and de Broglie relations.

Technically, we don't need the ideas of mass, momentum, and energy to do physics – the notions of wavenumber, frequency, and group velocity are sufficient to describe and explain all observed phenomena. However, mass, momentum, and energy are so firmly embedded in physics that one couldn't talk to other physicists without an understanding of these quantities!

Planck, Einstein, and de Broglie

Max Planck was the first to develop a theory explaining the energy density of electromagnetic radiation in a box at a fixed temperature. Albert Einstein extended Planck's ideas by postulating that the energy of electromagnetic radiation is quantized into chunks called **photons** . The energy E of a photon is related to the frequency of the electromagnetic radiation by the equation

$$E = hf = \bar{}h\omega(Planck - Einsteinrelation), \tag{1.1.146}$$

where f is the rotational frequency of the associated electromagnetic wave and ω is its angular frequency. The constant $h = 6.63 \times 10^{-34}$ kg m^2 s^{-1} is called **Planck's constant** . The related constant $\hbar = h/2\pi = 1.06 \times 10^{-34}$ kg m^2 s^{-1} is also referred to as Planck's constant, but to avoid confusion with the original constant, we will generally refer to it as "h bar".

Notice that a new physical dimension has appeared, namely mass, with the unit **kilogram** , abbreviated "kg". The physical meaning of mass is much like our intuitive understanding of the concept, i. e., as a measure of the resistance of an object to its velocity being changed. The precise scientific meaning will emerge shortly.

Einstein showed that Planck's idea could be used to explain the emission of electrons which occurs when light impinges on the surface of a metal. This emission, which is called the **photoelectric effect** , can only occur when electrons are supplied with a certain minimum energy E_B required to break them loose from the metal. Experiment shows that this emission occurs only when the frequency of the light exceeds a certain minimum value. This value turns out to equal $\omega_{\text{min}} = E_B/\hbar$, which suggests that electrons gain energy by absorbing a single photon. If the photon energy, $\hbar\omega$, exceeds E_B, then electrons are emitted, otherwise

they are not. It is much more difficult to explain the photoelectric effect from the classical theory of light. The value of E_B, called the binding energy or work function, is different for different metals.

Louis de Broglie proposed that Planck's energy-frequency relationship be extended to all kinds of particles. In addition he hypothesized that the **momentum** Π of the particle and the wave vector k of the corresponding wave were similarly related:

$$\Pi = \bar{}hk(deBroglierelation). \tag{1.1.147}$$

Note that this can also be written in scalar form in terms of the wavelength as $\Pi = h / \lambda$. (We use Π rather than the more common p for momentum, because as we shall see, there are two different kinds of momentum, one related to the wavenumber, the other related to the velocity of a particle. In many cases they are equal, but there are certain important situations in which they are not.)

De Broglie's hypothesis was inspired by the fact that wave frequency and wavenumber are components of the same four-vector according to the theory of relativity, and are therefore closely related to each other. Thus, if the energy of a particle is related to the frequency of the corresponding wave, then there ought to be some similar quantity which is correspondingly related to the wavenumber. It turns out that the momentum is the appropriate quantity. The physical meaning of momentum will become clear as we proceed.

We will also find that the rest frequency, μ, of a particle is related to its mass, m :

$$2Erest \equiv mc = \bar{}h\mu. \tag{1.1.148}$$

The quantity E_{rest} is called the **rest energy** of the particle.

From our perspective, energy, momentum, and rest energy are just scaled versions of frequency, wave vector, and rest frequency, with a scaling factor $\hbar$. We can therefore define a **four-momentum** as a scaled version of the wave four-vector:

$$\Pi - = \bar{}hk. \tag{1.1.149}$$

The spacelike component of Π is just Π, while the timelike part is E / c.

Planck, Einstein, and de Broglie had extensive backgrounds in classical mechanics, in which the concepts of energy, momentum, and mass have precise meaning. In this text we do not presuppose such a background. Perhaps the best strategy at this point is to think of these quantities as scaled versions of frequency, wavenumber, and rest frequency, where the scale factor is $\hbar$. The significance of these quantities to classical mechanics will emerge bit by bit.

Particle Quantities

Let us now recapitulate what we know about relativistic waves, and how this knowledge translates into knowledge about the mass, energy, and momentum of particles. In the following equations, the form on the left is expressed in wave terms, i. e., in terms of frequency, wavenumber, and rest frequency. The form on the right is the identical equation expressed in terms of energy, momentum, and mass. Since the latter variables are just scaled forms of the former, the two forms of each equation are equivalent.

We begin with the dispersion relation for relativistic waves:

$$222222224\omega = kc + \mu E = \Pi c + mc. \tag{1.1.150}$$

Calculation of the group velocity, $u_g = d\omega / dk$, from the dispersion relation yields

$$22ug = c - k - ug = c - \Pi -.\omega E \tag{1.1.151}$$

These two sets of equations represent what we know about relativistic waves, and what this knowledge tells us about the relationships between the mass, energy, and momentum of relativistic particles. When in doubt, refer back to these equations, as they work in all cases, including for particles with zero mass!

It is useful to turn equations (7.6) and (7.7) around so as to express the frequency as a function of rest frequency and group velocity,

$$\begin{gathered} - - - - - - \mu - - - - - - - - - - - - - - - mc2 - - - - - - \omega \\ = (1 - u2 / c2)1 / 2E = (1 - u2 / c2)1 / 2, gg \end{gathered} \tag{1.1.152}$$

and the wavenumber as a similar function of these quantities:

$$\begin{gathered} - - - \mu ug - / c2 - - - - - - - - - - mug - - - - - - k \\ = (1 - u2 / c2)1 / 2\Pi = (1 - u2 / c2)1 / 2.gg \end{gathered} \tag{1.1.153}$$

Note that equations (7.8) and (7.9) work only for particles with non-zero mass! For zero mass particles both the numerators and denominators of equations (7.8) and (7.9) are zero, making these equations undefined, and you need to use equations (7.6) and (7.7) with m = 0 and μ = 0 instead.

The quantity ω - μ indicates how much the frequency exceeds the rest frequency. Notice that if $\omega = \mu$, then from equation 7.6} k = 0. Thus, positive values of $\omega_k \equiv \omega - \mu$ indicate $|k| > 0$, which means that the particle is moving according to equation (7.7). Let us call ω_k the **kinetic frequency** :

$$[][] - - - - - - - 1 - - - - - - - - - - - - - - - 1 - - - - - - - - 2\omega k \quad (1.1.154)$$
$$= (1 - u2 / c2)1 / 2 - 1\mu K = (1 - u2 / c2)1 / 2 - 1mc.gg$$

We call K the **kinetic energy** for similar reasons. Again, equation 7.10} only works for particles with non-zero mass. For zero mass particles the kinetic energy equals the total energy.

Note that the results of this section are valid only for free particles, i. e., particles to which no force is applied. Force in classical and quantum mechanics is treated in the next chapter.

Non-Relativistic Limits

When the mass is non-zero and the group velocity is much less than the speed of light, it is useful to compute approximate forms of the above equations valid in this limit. Using the approximation $(1 + \epsilon)^x \approx 1 + x\epsilon$, we find that the dispersion relation becomes

$$222\omega = \mu + k - c - -E = mc2 + \Pi - --, 2\mu 2m \quad (1.1.155)$$

and the group velocity equation takes the approximate form

$$c2k\Pi ug = - - - - - ug = --.\mu m \quad (1.1.156)$$

The non-relativistic limits for equations (7.8) and (7.9) become

$$\mu u2gmu2g\omega = \mu + - - 2 - E = mc2 + - - - - - 2c2 \quad (1.1.157)$$

and

$$2k = \mu ug / c\Pi = mug, \quad (1.1.158)$$

while the approximate kinetic energy equation is

$$\mu u2mu2\omega k = - - 2gK = - - - g.2c2 \quad (1.1.159)$$

Just a reminder — the equations in this section are not valid for massless particles!

Experimental Test

How can we test the above predictions against experiment? The key point is to be able to relate the wave aspects to the particle aspects of a quantum mechanical wave-particle. Equation (7.9), or equation 7.14} in the non-relativistic case, relates a particle's wavenumber k to its velocity u_g. Both of these quantities can be measured in a Bragg's law experiment with electrons. In this experiment electrons are fired at a crystal with known atomic dimensions at a known speed, which we identify with the group velocity u_g. The Bragg angle which yields constructive interference can be used to calculate the wavelength of the corresponding electron wave, and hence the wavenumber and momentum. If the momentum is plotted against group velocity in the non-relativistic case, a straight line should be found, the slope of which is the particle's mass. In the fully relativistic case one needs to plot momentum versus $u_g / (1 - u_g^2 / c^2)^{1/2}$. Again, a straight line indicates agreement with the theory and the slope of the line is the particle's mass. This particular experiment is difficult to do, but the corresponding theories verify in many other experiments.

Heisenberg Uncertainty Principle

Classically, we consider the location of a particle to be a knowable piece of information. In quantum mechanics the position of a particle is well known if the wave packet representing it is small in size. However, quantum mechanics imposes a price on accurately knowing the position of a particle in terms of the future predictability of its position. This is because a small wave packet, which corresponds to accurate knowledge of the corresponding particle's position, implies the superposition of plane waves corresponding to a broad distribution of wavenumbers. This translates into a large uncertainty in the wavenumber, and hence the momentum of the particle. In contrast, a broad wave packet corresponds to a narrower distribution of wavenumbers, and correspondingly less uncertainty in the momentum.

Referring back to chapters 1 and 2, recall that both the longitudinal (along the direction of motion) and transverse (normal to the direction of motion) dimensions of a wave packet, Δx_{L} and Δx_{T}, can be related to the spread of longitudinal and transverse wavenumbers, Δk_{L} and Δk_{T}:

$$\Delta kL\Delta xL \approx 1, \tag{1.1.160}$$

We have omitted numerical constants which are of order unity in these approximate relations so as to show their essential similarity.

The above equations can be interpreted in the following way. Since the absolute square of the wave function represents the probability of finding a particle, Δx_{L} and Δx_{T} represent the uncertainty in the particle's position. Similarly, Δk_{L} and Δk_{T} represent the uncertainty in the particle's longitudinal and transverse wave vector components. This latter uncertainty leads to uncertainty in the particle's future motion — larger or smaller longitudinal k results respectively in larger or smaller particle speed, while uncertainty in the transverse wavenumber results in uncertainty in the particle's direction of motion. Thus uncertainties in any component of k result in uncertainties in the corresponding component of the particle's velocity, and hence in its future position.

The equations (7.16) and (7.17) show that uncertainty in the present and future positions of a particle are complimentary. If the present position is accurately known due to the small size of the associated wave packet, then the future position is not very predictable, because the wave packet disperses rapidly. On the other hand, a broad-scale initial wave packet means that the present position is poorly known, but the uncertainty in position, poor as it is, doesn't rapidly increase with time, since the wave packet has a small uncertainty in wave vector and thus disperses slowly. This is a statement of the **Heisenberg uncertainty principle** .

The uncertainty principle also applies between frequency and time:

$$\Delta\omega\Delta t \approx 1. \tag{1.1.161}$$

This shows up in the beat frequency equation $1/T_{\text{beat}} = \Delta f = \Delta\omega/2\pi$. The beat period T_{beat} may be thought of as the size of a "wave packet in time". The beat frequency equation may be rewritten as $\Delta\omega T_{\text{beat}} = 2\pi$, which is the same as equation (7.18) if the factor of 2π is ignored and T_{beat} is identified with Δt.

The above forms of the uncertainty principle are not relativistically invariant. A useful invariant form may be obtained by transforming to the coordinate system in which a particle is stationary. In this reference frame the time t becomes the proper time τ associated with the particle. Furthermore, the frequency ω becomes the rest frequency μ. The uncertainty principle thus becomes

$$\Delta\mu\Delta\tau \approx 1 \tag{1.1.162}$$

in this reference frame. However, since $\Delta\mu$ and $\Delta\tau$ are relativistic invariants, this expression of the uncertainty principle is valid in **any** reference frame.

It is more common to express the uncertainty principle in terms of the mass, momentum, and energy by multiplying equations (7.16) - (7.19) by $\hbar$. Lumping the momentum equations, we find

$$\Delta\Pi\Delta x \approx \bar{}h, \tag{1.1.163}$$

and

$$\Delta(mc2)\Delta\tau \approx \bar{}h. \tag{1.1.164}$$

Classical mechanics is the realm of quantum mechanics in which the dimensions of the system of interest are much larger than the wavelengths of the waves corresponding to the particles constituting the system. In this case the uncertainties induced by the uncertainty principle are unimportant. This limit is analogous to the geometrical optics limit for light. Thus, we can say that classical mechanics is the geometrical optics limit of quantum mechanics.

Problems

1. An electron with wavelength $\lambda = 1.2 \times 10^{-10}$ m undergoes Bragg diffraction from a single crystal with atomic plane spacing of $d = 2 \times 10^{-10}$ m.
    1. Calculate the Bragg angles (all of them!) for which constructive interference occurs.
    2. Calculate the speed of the electron.
2. Suppose that electrons impinge on two slits in a plate, resulting in a two slit diffraction pattern on a screen on the other side of the plate. The probability for an electron to pass through either one of the slits and reach point A on the screen is P, assuming that the other slit is blocked.

1. If there are two slits open and A is a point of constructive interference, what is the probability of an electron reaching A? Hint: Remember that amplitudes, not probabilities add.
2. If there are two slits open and A is a point of destructive interference, what is the probability of an electron reaching A?
3. If there are two slits open, what is the probability for an electron to reach point A according to the conventional rule that probabilities add? (This is the result one would expect if, for instance, the particles were machine gun bullets and the slits were, say, 5 cm apart.)
4. If the slit separation is very much greater than the electron wavelength, how does this affect the spacing of regions of constructive and destructive interference? Explain how the results of parts (a) and (b) become approximately consistent with those of part (c) in this case.

3. Compute the (angular) rest frequency of an electron and a neutron. (Look up their masses.)
4. How does the dispersion relation for relativistic waves simplify if the rest frequency (and hence the particle mass) is zero? What is the group velocity in this case?
5. X-rays are photons with frequencies about 2000 times the frequencies of ordinary light photons. From this information and what you know about light, infer the approximate velocity of electrons which have Bragg diffraction properties similar to X-rays. Are the electrons relativistic or non-relativistic?
6. Electrons with velocity $v = 0.6c$ are diffracted with a 0.2 radian half-angle of diffraction when they hit an object. What is the approximate size of the object? Hint: Diffraction of a wave by an object of a certain size is quite similar to diffraction by a hole in a screen of the same size.
7. Work out an approximate formula for the kinetic energy of a particle as a function of mass m and velocity u_g which is valid when $u_g^2 \ll c^2$. Hint: Use the approximation $(1+\epsilon)^x \approx 1 + x\epsilon$, which is valid for $|\epsilon| \ll 1$. As u_g/c becomes larger, how does this approximate formula deviate from the exact formula?
8. Work out an approximate formula for the momentum of a particle as a function of m and u_g in the case where $u_g^2 \ll c^2$. You may wish to use the approximation mentioned in the previous problem.
9. If a photon is localized to within a distance Δx, what is the uncertainty in the photon energy?
10. If an electron is localized to within a distance Δx, what is the uncertainty in the electron kinetic energy? Hint: As long as $\Delta\Pi \ll \Pi$, $\Delta\Pi^2 \approx 2\Pi\Delta\Pi$. To see why, compute $d\,\Pi^2/d\,\Pi$.
11. A grocer dumps some pinto beans onto a scale, estimates their mass as 2 kg, and then dumps them off after 5 s. What is the quantum mechanical uncertainty in this measurement? Assume this occurs in Quantum World where the speed of light is 10 m s^{-1} (speed of a fast buggy) and Planck's constant $\hbar = 1$ kg m^2 s^{-1}.
12. Mary's physics text (mass 0.3 kg) has to be kept on a leash (length 0.5 m) to prevent it from wandering away from her in Quantum World ($\hbar = 1$ kg m^2 s^{-1}).
    1. If the leash suddenly breaks, what is the maximum speed at which the book is likely to move away from its initial location?
    2. In order to reduce this speed, should Mary make the new leash shorter or longer than the old one? Explain.
13. A proton (mass $M = 1.7 \times 10^{-27}$ kg) is confined to an atomic nucleus of diameter $D = 2 \times 10^{-15}$ m.
    1. What is the uncertainty in the proton's momentum?
    2. Roughly what kinetic energy might you expect the proton to have?

    Planck's constant is $\hbar = 1.06 \times 10^{-34}$ kg m^2 s^{-1}. You may use the non-relativistic equation for the energy.

    }

Chapter 8 Geometrical Optics and Newton's Laws

The question that motivates us to study physics is "What makes things go?" The answers we conceive to this question constitute the subject of **dynamics** . This is in contrast to the question we have primarily addressed so far, namely "How do things go?" The latter question is about kinematics. Extensive preparation in the kinematics of waves and particles in relativistic spacetime is needed to intelligently address dynamics. This preparation is now complete.

In this chapter we outline three different dynamical principles based respectively on pre-Newtonian, Newtonian, and quantum mechanical thinking. We first discuss the Newtonian mechanics of conservative forces in one dimension. Certain ancillary concepts in mechanics such as work and power are introduced at this stage. We then show that Newtonian and quantum mechanics are consistent with each other in the realm in which they overlap, i. e., in the geometrical optics limit of quantum mechanics. For simplicity, this relationship is first developed in one dimension in the non-relativistic limit. Higher dimensions require the introduction of partial derivatives, and the relativistic case will be considered later.

Fundamental Principles of Dynamics

Roughly speaking, there have been three eras of physics, characterized by three different answers to the question of what makes things go.

Pre-Newtonian Dynamics

Aristotle expounded a view of dynamics which agrees closely with our everyday experience of the world. Objects only move when a force is exerted upon them. As soon as the force goes away, the object stops moving. The act of pushing a box across the floor illustrates this principle — the box certainly doesn't move by itself!

Newtonian Dynamics

In contrast to earthly behavior, the motions of celestial objects seem effortless. No obvious forces act to keep the planets in motion around the sun. In fact, it appears that celestial objects simply coast along at constant velocity unless something acts on them. The Newtonian view of dynamics — objects change their **velocity** rather than their position when a force is exerted on them — is expressed by **Newton's second law** :

$$F = ma(Newton's second law), \tag{1.1.165}$$

where F is the force exerted on a body, m is its mass, and a is its acceleration. **Newton's first law** , which states that an object remains at rest or in uniform motion unless a force acts on it, is actually a special case of Newton's second law which applies when F = 0.

It is no wonder that the first successes of Newtonian mechanics were in the celestial realm, namely in the predictions of planetary orbits. It took Newton's genius to realize that the same principles which guided the planets also applied to the earthly realm as well. In the Newtonian view, the tendency of objects to stop when we stop pushing on them is simply a consequence of frictional forces opposing the motion. Friction, which is so important on the earth, is negligible for planetary motions, which is why Newtonian dynamics is more obviously valid for celestial bodies.

Note that the principle of relativity is closely related to Newtonian physics and is incompatible with pre-Newtonian views. After all, two reference frames moving relative to each other cannot be equivalent in the pre-Newtonian view, because objects with nothing pushing on them can only come to rest in one of the two reference frames. Newton's second law obeys the principle of relativity because the acceleration of an object is the same when viewed from two different reference frames moving at a constant velocity with respect to each other.

Einstein's relativity is often viewed as a repudiation of Newton, but this is far from the truth — Newtonian physics makes the theory of relativity possible through its invention of the principle of relativity. Compared with the differences between pre-Newtonian and Newtonian dynamics, the changes needed to go from Newtonian to Einsteinian physics constitute minor tinkering.

Quantum Dynamics

In quantum mechanics, particles are represented by matter waves, with the absolute square of the wave displacement yielding the probability of finding the particle. The behavior of particles thus follows from the reflection, refraction, diffraction, and interference of the associated waves. The connection with Newtonian dynamics comes from tracing the trajectories of matter wave packets. Changes in the speed and direction of motion of these packets correspond to the accelerations of classical mechanics. When wavelengths are small compared to the natural length scale of the problem at hand, the wave packets can be made small, thus pinpointing the position of the associated particle, without generating excessive uncertainty in the particle's momentum. This is the geometrical optics limit of quantum mechanics.

Potential Energy

We now address Newtonian mechanics in the case where the force on a particle is **conservative** . A conservative force is one that can be derived from a so-called **potential energy** U. We assume that the potential energy of the particle depends only on its position. The force is obtained from the potential energy by the equation

$$F = -dU-.dx \tag{1.1.166}$$

Using this equation we write Newton's second law as

$$dU - - - - = ma.dx \tag{1.1.167}$$

We then notice that the acceleration can be written in terms of the x derivative along the particle's trajectory of $v^2/2$:

$$2a = dv- = dv - dx- = dv - v = 1 - dv-.dtdxdtdx2dx \tag{1.1.168}$$

The last step in the above derivation can be verified by applying the product rule: $dv^2/dt = d(vv)/dt = v(dv/dt) + (dv/dt)v = 2v(dv/dt)$. Putting equations (8.3) and (8.4) together, we find that $d(mv^2/2 + U)/dt = 0$, which implies that $mv^2/2 + U$ is constant. We call this constant the **total energy** E and the quantity $K = mv^2/2$ the **kinetic energy** . We thus have the principle of conservation of energy for conservative forces:

$$E = K + U = constant. \tag{1.1.169}$$

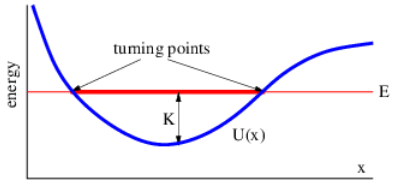


Figure 1.1.1:: Example of spatially variable potential energy $U(x)$ for a particle with fixed total energy E. The kinetic energy $K = E - U$ is zero where the E and U lines cross. These points are called turning points. The thick part of the horizontal line indicates the allowed range for the particle with the given energy.

Recall that in quantum mechanics the momentum is related to the group velocity u_g by

$$\Pi = mug(momentum) \tag{1.1.170}$$

in the nonrelativistic case. Equating the group velocity to v and eliminating it in the kinetic energy results in an alternate expression for this quantity:

$$2K \equiv 1mu2g = -\Pi - -(kineticenergy).22m \tag{1.1.171}$$

Since the total energy E is constant or **conserved** , increases in the potential energy coincide with decreases in the kinetic energy and vice versa, as is illustrated in Figure 1.1.1:. In classical mechanics the kinetic energy cannot be negative, since it is the product of half the mass and the square of the velocity, both of which are positive. Thus, a particle with total energy E and potential energy U is forbidden to venture into regions in which the kinetic energy $K = E - U$ is less than zero.

The points at which the kinetic energy is zero are called **turning points** . This is because a particle decreases in speed as it approaches a turning point, stops there for an instant, and reverses direction. Note also that a particle with a given total energy always has the same speed at some point x regardless of whether it approaches this point from the left or the right:

$$1/2speed = |ug| = |\pm[2(E-U)/m]|. \tag{1.1.172}$$

Gravity as a Conservative Force

An example of a conservative force is gravity. An object of mass m near the surface of the earth has the gravitational potential energy

$$U = mgz(gravitynearearth'ssurface) \tag{1.1.173}$$

where z is the height of the object above some reference point such as the earth's surface and $g = 9.8$ m s^{-2} is the local value of the **gravitational field** near the surface. Notice that the gravitational potential energy increases upward. The speed of the object in this case is $|u_g| = [2(E - mgz)/m]^{1/2}$. If $|u_g|$ is known to equal the constant value u_0 at elevation $z = 0$, then equations (8.8) and (8.9) tell us that $u_0 = (2E/m)^{1/2}$ and $|u_g| = (u_0^2 - 2gz)^{1/2}$.

There are certain types of questions which energy conservation cannot directly answer. For instance, if an object is released at elevation h with zero velocity at $t = 0$, at what time will it reach $z = 0$ under the influence of gravity? In such cases it is often easiest to return to Newton's second law. Since the force on the object is $F = -dU/dz = -mg$ in this case, we find that the acceleration is $a = F/m = -mg/m = -g$. However, $a = du/dt = d^2z/dt^2$, so

$$u = -gt + C1z = -gt2/2 + C1t + C2(constantgravity), \tag{1.1.174}$$

where C_1 and C_2 are constants to be determined by the initial conditions. These results can be verified by differentiating to see if the original acceleration is recovered. Since $u = 0$ and $z = h$ at $t = 0$, we have $C_1 = 0$ and $C_2 = h$. With these results it is easy to show that the object reaches $z = 0$ when $t = (2h/g)^{1/2}$.

Power

When a force is exerted on an object, energy is transferred to the object. The amount of energy transferred is called the **work** done on the object. However, energy is only transferred if the object moves. The work W done is

$$W = F\Delta x \tag{1.1.175}$$

where the distance moved by the object is Δx and the force exerted on it is F. Notice that work can either be positive or negative. The work is positive if the object being acted upon moves in the same direction as the force, with negative work occurring if the object moves opposite to the force.

Equation (8.11) assumes that the force remains constant over the full displacement Δx. If it is not, then it is necessary to break up the displacement into a number of smaller displacements, over each of which the force can be assumed to be constant. The total work is then the sum of the works associated with each small displacement.

If more than one force acts on an object, the works due to the different forces each add or subtract energy, depending on whether they are positive or negative. The total work is the sum of these individual works.

There are two special cases in which the work done on an object is related to other quantities. If F is the total force acting on the object, then $W = F\ \Delta x = ma\ \Delta x$ by Newton's second law. However, $a = dv\ /\ dt$ where v is the velocity of the object, and $\Delta x = (\Delta x/\ \Delta t\ )\Delta t \approx v\ \Delta t$, where Δt is the time required by the object to move through distance Δx. The approximation becomes exact when Δx and Δt become very small. Putting all of this together results in

$$dvd(mv2)Wtotal = m - - - v\Delta t = - - - - - - - - \Delta t = \Delta K(totalwork), dtdt2 \tag{1.1.176}$$

where K is the kinetic energy of the object. Thus, when F is the only force, $W = W_{\text{total}}$ is the total work on the object, and this equals the change in kinetic energy of the object. This is called the **work-energy theorem** , and it demonstrates that work really is a transfer of energy to an object.

The other special case occurs when the force is conservative, but is not necessarily the total force acting on the object. In this case

$$Wcons = -dU - \Delta x = -\Delta U(conservativeforce), dx \tag{1.1.177}$$

where ΔU is the change in the potential energy of the object associated with the force of interest.

The **power** associated with a force is simply the amount of work done by the force divided by the time interval Δt over which it is done. It is therefore the energy per unit time transferred to the object by the force of interest. From equation 8.11} we see that the power is

$$F\Delta xP = - - - - - - = Fv(power), \Delta t \tag{1.1.178}$$

where v is the velocity at which the object is moving. The **total power** is just the sum of the powers associated with each force. It equals the time rate of change of kinetic energy of the object:

$$Ptotal = Wtotal = dK - -(totalpower).\Delta tdt \tag{1.1.179}$$

Mechanics and Geometrical Optics

Louis de Broglie[1] made an analogy between matter waves and light waves, pointing out that wave packets of light change their velocity as the result of spatial variations in the index of refraction of the medium in which they are travelling. This behavior comes about because the dispersion relation for light traveling through a medium with index of refraction n is $\omega = kc\ /\ n$, so that the group velocity, $u_{\text{g}} = d\omega\ /\ dk = c\ /\ n$. Thus, when n increases, u_{g} decreases, and vice versa.[2]

In this section we pursue de Broglie's analogy to see if we can come up with a theory of matter waves which gives the same results as classical mechanics in the geometrical optics limit of these waves. The dispersion relation for free matter waves is $\omega = (k^{\,2}c^{\,2} + \mu^{\,2})^{1/2}$. In the non-relativistic limit $k^{\,2}c^{\,2} \ll \mu^{\,2}$. As done previously, we use $(1 + \epsilon\ )^{\text{n}} \approx 1 + n\epsilon$ for small ϵ. In the non-relativistic limit, the dispersion relation for free waves thus becomes

$$\omega = \mu(1 + k2c2\ /\ \mu2)1\ /\ 2 \approx \mu + k2c2\ /\ (2\mu). \tag{1.1.180}$$

The above equation can be transformed into the total energy equation for a free, non-relativistic particle, $E = mc^{\,2} + K$, where $mc^{\,2}$ is the rest energy and K is the kinetic energy, by multiplying by $\hbar$. We convert the free particle energy equation into the equation for a particle subject to a conservative force by adding the potential energy U the right side. The analogous change to equation (8.16) is to add $S = U/\ \hbar$ to the right side, resulting in a modified dispersion relation:

$$\omega = S(x) + k2c2 \,/\, (2\mu). \tag{1.1.181}$$

(Since the rest energy is just a constant, we have absorbed it into S.) This gives us the dispersion relation for one-dimensional matter waves subject to a spatially varying potential energy. The quantity S, which we see is just a scaled potential energy, plays a role for matter waves which is analogous to the role played by a spatially variable index of refraction for light waves.

Let us now imagine that all parts of the wave governed by this dispersion relation oscillate in phase. The only way this can happen is if ω is constant, i. e., it takes on the same value in all parts of the wave.

If ω is constant, the only way S can vary with x in equation 8.17} is if the wavenumber varies in a compensating way. Thus, constant frequency and spatially varying S together imply that $k = k\,(x\,)$. Solving equation (8.17) for k yields

$$[]1 \,/\, 2k(x) = \pm 2\mu[\omega - - - S - (x-)].c2 \tag{1.1.182}$$

Since ω is constant, the wavenumber becomes smaller and the wavelength larger as the wave moves into a region of increased S.

In the geometrical optics limit, we assume that S doesn't change much over one wavelength so that the wave remains reasonably sinusoidal in shape with approximately constant wavenumber over a few wavelengths. However, over distances of many wavelengths the wavenumber and amplitude of the wave are allowed to vary considerably.

The group velocity calculated from the dispersion relation given by equation 8.17} is

$$2(2)1 \,/\, 2ug = d\omega - = kc - - = \pm 2c - (\omega - - - S-)dk\mu\mu \tag{1.1.183}$$

where k is eliminated in the last step with the help of equation 8.18}. The resulting equation tells us how the group velocity varies as a matter wave traverses a region of slowly varying S. Thus, as S increases, $u_{\,g}$ decreases and vice versa.

We can now calculate the acceleration of a wave packet resulting from the spatial variation in S. We assume that $x\,(t\,)$ represents the position of the wave packet, so that $u_{\,g} = dx \,/\, dt$. Using the chain rule $du_{\,g}/dt = (du_{\,g}/dx\,)(dx \,/\, dt\,) = (du_{\,g}/dx\,)u_{\,g}$, we find

$$22a = dug - = dugu = dug \,/\, 2 - = -c - dS - = -^{-}h - dS-.dtdxgdx\mu dxmdx \tag{1.1.184}$$

The group velocity is eliminated in favor of S by squaring equation 8.19} and substituting the result into equation (8.20).

Recalling that $U = \hbar\, S$, equation 8.20} becomes

$$-1dU - F - a = -mdx = m, \tag{1.1.185}$$

which is just Newton's second law! Thus, the geometrical optics approach to particle motion is completely equivalent to the classical mechanics of a particle moving under the influence of a conservative force, at least in one dimension. We therefore have two ways of solving for the motion of a particle subject to a potential energy $U\,(x\,)$. We can apply the principles of classical mechanics to get the force and the acceleration of the particle, from which we can derive the motion. Alternatively, we can apply the principles of geometrical optics to compute the spatially variable velocity of the wave packet using equation 8.19}. The results are completely equivalent, though the methods are conceptually very different.

Tutorial – Partial Derivatives

In order to understand the generalization of Newtonian mechanics to two and three dimensions, we first need to understand a new type of derivative called the **partial derivative** . The partial derivative is used in functions of more than one variable. It is just like an ordinary derivative, except that when taking the derivative of the function with respect to one of the variables, the other variables are held constant. As an example, let us consider the function

$$4224f(x, y) = Ax + Bxy + Cy \tag{1.1.186}$$

where A, B, and C are constants. The partial derivative of f with respect to x is

$$\partial f - -- = 4Ax3 + 2Bxy2\partial x \tag{1.1.187}$$

and the partial derivative with respect to y is

$$\partial f23 - -- = 2Bxy + 4Cy.\partial y \tag{1.1.188}$$

That's it! Note that a special symbol "∂ " is used in place of the normal "d " for the partial derivative. This is sometimes called a "curly d".

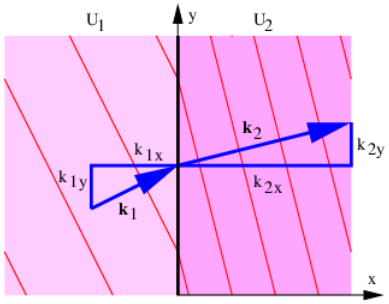


Figure 1.1.2:: Refraction of a matter wave by a discontinuity in potential energy. The component of the wave vector parallel to the discontinuity, k_y, doesn't change, so $k_{1y} = k_{2y}$.

When a matter wave moves through a region of variable potential energy in one dimension, only the wavenumber changes. In two or three dimensions, the wave vector can change in both direction and magnitude. This complicates the calculation of particle movement. However, we already have an example of how to handle this situation, namely, the refraction of light. In that case Snell's law tells us how the direction of the wave vector changes, while the dispersion relation combined with the constancy of the frequency gives us information about the change in the magnitude of the wave vector. For matter waves a similar procedure works, though the details are different, because we seek the consequences of a change in potential energy rather than a change in the index of refraction.

Figure 1.1.2: illustrates the refraction of matter waves at a discontinuity in the potential energy. Let us suppose that the discontinuity occurs at $x = 0$. If the matter wave to the left of the discontinuity is $\psi_1 = \sin(k_{1x}x + k_{1y}y - \omega_1 t)$ and to the right is $\psi_2 = \sin(k_{2x}x + k_{2y}y - \omega_2 t)$, then the wavefronts of the waves will match across the discontinuity for all time only if $\omega_1 = \omega_2 \equiv \omega$ and $k_{1y} = k_{2y} \equiv k_y$. We are already familiar with the first condition from the one-dimensional problem, so the only new ingredient is the constancy of the y component of the wave vector.

In two dimensions the momentum is a vector: $\Pi = m\,u$ when $|u| \ll c$, where u is the particle velocity. Furthermore, the kinetic energy is $K = m\,|u|^2/2 = |\Pi|^2/(2m) = (\Pi_x^2 + \Pi_y^2)/(2m)$. The relationship between kinetic, potential, and total energy is unchanged from the one-dimensional case, so we have

$$22E = U + (\Pi x + \Pi y)\,/\,(2m) = constant. \tag{1.1.189}$$

The de Broglie relationship tells us that $\Pi = \hbar\,k$, so the constancy of k_y across the discontinuity in U tells us that

$$\Pi y = constant \tag{1.1.190}$$

there.

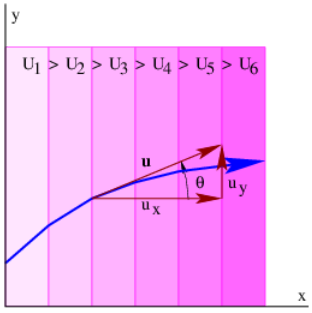


Figure 1.1.3:: Trajectory of a wave packet through a variable potential energy, $U(x)$, which decreases to the right.

Let us now approximate a continuously variable $U(x)$ by a series of steps of constant U oriented normal to the x axis. The above analysis can be applied at the jumps or discontinuities in U between steps, as illustrated in Figure 1.1.3:, with the result that equations (8.25) and (8.26) are valid across all discontinuities. If we now let the step width go to zero, these equations then become valid for U continuously variable in x.

An example from classical mechanics of a problem of this type is a ball rolling down an inclined ramp with an initial velocity component across the ramp, as illustrated in Figure 1.1.4:. The potential energy decreases in the down ramp direction, resulting in a force down the ramp. This accelerates the ball in that direction, but leaves the component of momentum across the ramp unchanged.

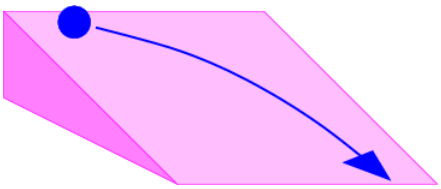


Figure 1.1.4:: Classical mechanics example of the problem illustrated in Figure 1.1.3:.

Using the procedure which we invoked before, we find the force components associated with U (x) in the x and y directions to be F_x = -dU / dx and F_y = 0. This generalizes to

$$()\partial U - -\partial U - -\partial U - -F = -\partial x, \partial y, \partial z(3 - Dconservative force) \tag{1.1.191}$$

in the three-dimensional case where the orientation of constant U surfaces is arbitrary. It is also valid when U (x, y, z) is not limited to a simple ramp form, but takes on a completely arbitrary structure.

The definitions of work and power are slightly different in two and three dimensions. In particular, work is defined as

$$W = F \cdot \Delta x \tag{1.1.192}$$

where Δx is now a vector displacement. The vector character of this expression yields an additional possibility over the one dimensional case, where the work is either positive or negative depending on the direction of Δx relative to F. If the force and the displacement of the object on which the force is acting are perpendicular to each other, the work done by the force is actually zero, even though the force and the displacement both have non-zero magnitudes. The power exhibits a similar change:

$$P = F \cdot u. \tag{1.1.193}$$

Thus the power is zero if an object's velocity is normal to the force being exerted on it.

As in the one-dimensional case, the total work done on a particle equals the change in the particle's kinetic energy. In addition, the work done by a conservative force equals minus the change in the associated potential energy.

Energy conservation by itself is somewhat less useful for solving problems in two and three dimensions than it is in one dimension. This is because knowing the kinetic energy at some point tells us only the magnitude of the velocity, not its direction. If conservation of energy fails to give us the information we need, then we must revert to Newton's second law, as we did in the one-dimensional case. For instance, if an object of mass m has initial velocity u_0 = (u_0, 0) at location (x, z) = (0, h) and has the gravitational potential energy U = mgz, then the force on the object is F = (0, -mg). The acceleration is therefore a = F /m = (0, -g). Since a = d u /dt = d^2x /dt^2 where x = (x, z) is the object's position, we find that

$$2u = (C1, -gt + C2)x = (C1t + C3, -gt / 2 + C2t + C4), \tag{1.1.194}$$

where C_1, C_2, C_3, and C_4 are constants to be evaluated so that the solution reduces to the initial conditions at t = 0. The specified initial conditions tell us that C_1 = u_0, C_2 = 0, C_3 = 0, and C_4 = h in this case. From these results we can infer the position and velocity of the object at any time.

Kinetic and Total Momentum

If you have previously taken a physics course then you have probably noticed that a rather odd symbol is used for momentum, namely Π, rather than the more commonly employed p. The reason for this peculiar usage is that there are actually two kinds of momentum, **kinetic momentum** and **total momentum** , just as there are two kinds of energy, kinetic and total.[3] The symbol Π represents total momentum while p represents kinetic momentum. Normally we don't need to distinguish between the two quantities, as they are generally equal to each other. However, we will find later in the course that it is crucial to make this distinction in the case of charged particles in a magnetic field. As a general rule, the total momentum is related to a particle's wave vector via the de Broglie relation, Π = $\hbar$ k, while the kinetic momentum is related to a particle's velocity, p = m u / (1 - u^2/c^2)$^{1/2}$.

Problems

1. Suppose the dispersion relation for a matter wave under certain conditions is ω = μ + (k - a)^2c^2/ (2μ) where k is the wavenumber of the wave, μ = mc^2/ $\hbar$, m is the associated particle's mass, a is a constant, c is the speed of light, and $\hbar$ is Planck's constant divided by 2π.
    1. Use this disperson relation and the Planck and de Broglie relations to determine the relationship between energy E, momentum Π, and mass m.

2. Compute the group velocity of the wave and use this to determine how the group velocity depends on mass and momentum in this case.

2. A matter wave function associated with a particle of definite (constant) total energy E takes the form shown in Figure 1.1.5:. Make a sketch showing how the kinetic, potential, and total energies of the particle vary with x.

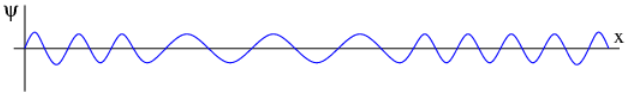


Figure 1.1.5:: A wave function in which the wavelength varies with position.

---

3. Compute $\partial / \partial x$ and $\partial / \partial y$ of the following functions. Other symbols are constants.
   1. $f(x, y) = ax^2 + by^3$
   2. $f(x, y) = ax^2 y^2$
   3. $f(x, y) = (x + a) / (y + b)$
4. Given a potential energy for a particle of mass M of the form $U(x) = Ax^3 - Bx$ where A and B are positive constants:
   1. Find the force on the particle.
   2. Find the values of x where the force is zero.
   3. Sketch $U(x)$ versus x and graphically compare the slope of $U(x)$ to the force computed above. Do the two qualitatively match?
   4. If the total energy of the particle is zero, where are its turning points?
   5. What is the particle's speed as a function of position assuming that the total energy E is known?
5. Given a potential energy function $U(x, y) = A(x^2 + y^2)$ where A is a positive constant:
   1. Sketch lines of constant U in the x-y plane.
   2. Compute the components of force as a function of x and y and draw sample force vectors in the x-y plane on the same plot used above. Do the force vectors point "uphill" or "downhill"?
6. Do the same as in the previous question for the potential energy function $U(x, y) = Axy$.
7. Suppose that the components of the force vector in the x-y plane are $F = (2Axy^3, 3Ax^2 y^2)$ where A is a constant. See if you can find a potential energy function $U(x, y)$ which gives rise to this force.
8. You are standing on top of a cliff of height H with a rock of mass M.
   1. If you throw the rock horizontally outward at speed u_0, what will its speed be when it hits the ground below?
   2. If you throw the rock upward at 45° to the horizontal at speed u_0, what will its speed be when it hits the ground?

   Hint: Can you use conservation of energy to solve this problem? Ignore air friction.
9. A car of mass 1200 kg initially moving 30 m s^{-1} brakes to a stop.
   1. What is the net work done on the car due to all the forces acting on it during the indicated period?
   2. Describe the motion of the car relative to an inertial reference frame initially moving with the car.
   3. In the above reference frame, what is the net work done on the car during the indicated period?

   Is work a relativistically invariant quantity?

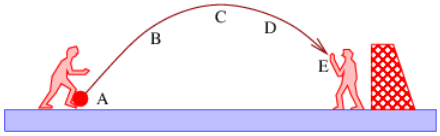


Figure 1.1.6:: The trajectory of a soccer ball.

---

10. A soccer player kicks a soccer ball, which is caught by the goal keeper as shown in Figure 1.1.6:. At various points forces exerted by gravity, air friction, the foot of the offensive player, and the hands of the goal keeper act on the ball.
    1. List the forces acting on the soccer ball at each of the points A, B, C, D, and E.

2. State whether the instantaneous power being applied to the soccer ball due to each of the forces listed above is positive, negative, or zero at each of the labeled points.

11. A cannon located at $(x, z) = (0, 0)$ shoots a cannon ball upward at an angle of θ from the horizontal at initial speed u_0. Hint: In order to solve this problem you must first obtain the x and z components of acceleration from Newton's second law. Second, you must find the velocity components as a function of time from the components of acceleration. Third, you must find x and z as a function of time from the the components of velocity. Only then should you attempt to answer the questions below.
    1. How long does it take the cannon ball to reach its peak altitude?
    2. How high does the cannon ball go?
    3. At what value of x does the cannon ball hit the ground ($z = 0$)?
    4. Determine what value of θ yields the maximum range.

Chapter 9 Symmetry and Bound States

When quantum mechanics was first invented, the dynamical principles used were the same as those underlying classical mechanics. The initial development of the field thus proceeded largely by imposing quantum laws on classical variables such as position, momentum, and energy. However, as quantum mechanics advanced, it became clear that there were many situations in which no classical analogs existed for new types of quantum mechanical systems, especially those which arose in the study of elementary particles. To understand these systems it was necessary to seek guidance from novel sources. One of the most important of these sources was the idea of symmetry, and in particular the relationship between symmetry and conserved variables. This type of relationship was first developed in the early 20th century by the German mathematician Emmy Nöther in the context of classical mechanics. However, her idea is easier to express and use in quantum mechanics than it is in classical mechanics. Emmy Nöther showed that there is a relationship between the symmetries of a system and conserved dynamical variables. This idea is naturally called **N öther's theorem** .

In classical mechanics a particle is bound or confined to a particular region if its total energy exceeds the potential energy only in this region. In quantum mechanics, matter waves can also be confined for the same reason, though the confinement is often less perfect than in the classical case. In this chapter we examine the consequences of wave confinement. We first look at the so-called "particle in a box" in one spatial dimension. We find that confined particles can take on only discrete energy values. When confinement isn't perfect we see how a quantum mechanical particle can leak through a potential energy barrier which is classically impenetrable. Movement of a particle on a circular ring leads us to another form of confinement and the introduction of angular momentum. This brings us finally to a discussion of the intrinsic or spin angular momentum of elementary particles.

Tutorial — Complex Waves

Until now we have represented quantum mechanical plane waves by sine and cosine functions, just as with other types of waves. However, plane matter waves cannot be truly represented by sines and cosines. We need instead mathematical functions in which the wave displacement is complex rather than real. This requires the introduction of a bit of new mathematics, which we tackle first. Using our new mathematical tool, we are then able to explore two crucially important ideas in quantum mechanics; (1) the relationship between symmetry and conservation laws, and (2) the dynamics of spatially confined waves.

A **complex number** z is the sum of a real number and an imaginary number. An imaginary number is just a real number multiplied by $i \equiv (-1)^{1/2}$. Thus, we can write $z = a + ib$ for any complex z, where a and b are real. The quantities a and b are the **real** and **imaginary** parts of z, sometimes written Re(z) and Im(z).

Quantum mechanics requires wave functions to be complex, i. e., to possess real and imaginary parts. Plane waves in quantum mechanics actually take the form $\psi = \exp[i\,(kx - \omega t\,)]$ rather than, say, $\cos(kx - \omega t\,)$. The reason for this is the need to distinguish between waves with positive and negative frequencies. If we replace k and ω with $-k$ and $-\omega$ in the cosine form, we get $\cos(-kx + \omega t\,) = \cos[-(kx - \omega t\,)] = \cos(kx - \omega t\,)$. In other words, changing the sign of k and ω results in no change in a wave expressed as a cosine function. The two quantum mechanical states, one with wavenumber and frequency k and ω and the other with $-k$ and $-\omega$, yield indistinguishable wave functions and therefore would represent physically indistinguishable states. The cosine form is thus insufficiently flexible to represent quantum mechanical waves. On the other hand, if we replace k and ω with their negatives in the complex exponential form of a plane wave we get $\psi = \exp[-i\,(kx - \omega t\,)]$, which is different from $\exp[i\,(kx - \omega t\,)]$. These two wave functions are distinguishable and thus correspond to distinct physical states.

It is not immediately obvious that a complex exponential function provides the oscillatory behavior needed to represent a plane wave. However, the complex exponential can be expressed in terms of sines and cosines using Euler's equation:

$$\exp(i\phi) = \cos(\phi) + i\sin(\phi) \quad (Euler's\ equation) \qquad (1.1.105)$$

$$exp(i\varphi) = cos(\varphi) + isin(\varphi) (Euler\ equation). \tag{1.1.195}$$

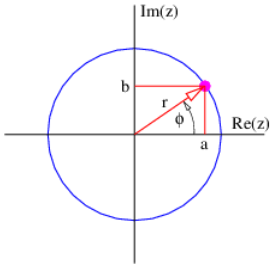


Figure 1.1.1:: Graphical representation of a complex number z as a point in the complex plane. The horizontal and vertical Cartesian components give the real and imaginary parts of z respectively.

If we define $r = (a^2 + b^2)^{1/2}$ and $\phi = \tan^{-1}(b / a)$, then an alternate way of expressing a complex number is $z = r\exp(i\phi)$, which by Euler's equation equals $r\cos(\phi) + ir\sin(\phi)$. Comparison shows that $a = r\cos(\phi)$ and $b = r\sin(\phi)$. Thus, a complex number can be thought of as a point in the a-b plane with Cartesian coordinates a and b and polar coordinates r and ϕ. The a-b plane is called the **complex plane** .

We now see how the complex wave function represents an oscillation. If $\psi = \exp[i(kx - \omega t)]$, the complex function $\psi(x,t)$ moves round and round the unit circle in the complex plane as x and t change, as illustrated in Figure 1.1.1:. This contrasts with the back and forth oscillation along the horizontal axis of the complex plane represented by $\cos(kx - \omega t)$.

We will not present a formal proof of Euler's equation — you will eventually see it in your calculus course. However, it may be helpful to note that the ϕ derivatives of $\exp(i\phi)$ and $\cos(\phi) + i\sin(\phi)$ have the same behavior:

$$d---exp(i\phi) = iexp(i\phi); d\phi \tag{1.1.196}$$

$$dd\phi - [cos(\phi) + isin(\phi)] = -sin(\phi) + icos(\phi) = i[cos(\phi) + isin(\phi)]. \tag{1.1.197}$$

(In the second of these equations we have replaced the minus sign in front of the sine function by i^2 and then extracted a common factor of i.) The ϕ derivative of both of these functions thus yields the function back again times i. This is a strong hint that $\exp(i\phi)$ and $\cos(\phi) + i\sin(\phi)$ are different ways of representing the same function.

We indicate the **complex conjugate** of a complex number z by a superscripted asterisk, i. e., z^*. It is obtained by replacing i by $-i$. Thus, $(a + ib)^* = a - ib$. The **absolute square** of a complex number is the number times its complex conjugate:

$$|z|2 = |a + ib|2 \equiv (a + ib)(a - ib) = a2 + b2 = r2. \tag{1.1.198}$$

Notice that the absolute square of a complex exponential function is one:

$$2|exp(i\phi)| = exp(i\phi)exp(-i\phi) = exp(i\phi - i\phi) = exp(0) = 1. \tag{1.1.199}$$

In quantum mechanics the absolute square of the wave function at any point expresses the relative probability of finding the associated particle at that point. Thus, the probability of finding a particle represented by a plane wave is uniform in space. Contrast this with the relative probability associated with a sine wave: $|\sin(kx - \omega t)|^2 = \sin^2(kx - \omega t)$. This varies from zero to one, depending on the phase of the wave. The "waviness" in a complex exponential plane wave resides in the phase rather than in the magnitude of the wave function.

One more piece of mathematics is needed. The complex conjugate of Euler's equation is

$$exp(-i\phi) = cos(\phi) - isin(\phi). \tag{1.1.200}$$

Taking the sum and the difference of this with the original Euler's equation results in the expression of the sine and cosine in terms of complex exponentials:

$$\begin{aligned} cos(\phi) &= exp(i\phi) + -exp(--i\phi) - sin(\phi) \\ &= exp - (i\phi) - --exp - (--i\phi).22i \end{aligned} \tag{1.1.201}$$

We aren't used to having complex numbers show up in physical theories and it is hard to imagine how we would measure such a number. However, everything observable comes from taking the absolute square of a wave function, so we deal only with real numbers in experiments.

Symmetry and Quantum Mechanics

The idea of symmetry plays a huge role in physics. We have already used symmetry arguments in the theory of relativity — applying the principle of relativity to obtain the dispersion relation for relativistic matter waves is just such an argument. In this section we begin to explore how symmetry can be used to increase our understanding of quantum mechanics.

Particle

For our first example we take the case of a free particle in quantum mechanics, i. e., a particle subject to no force. The wave function for a free particle of definite momentum Π and energy E is given by

$$\psi = exp[i(kx - \omega t)] = exp[i(\Pi x - Et) / \bar{}h \tag{1.1.202}$$

. \label{9.8}\]

For this wave function $|\psi|^2$ = 1 everywhere, so the probability of finding the particle anywhere in space and time is uniform. This contrasts with the probability distribution which arises if we assume a free particle to have the wave function ψ = cos[(Πx - Et)/ $\hbar$]. In this case $|\psi|^2$ = cos ^{2}[(Πx - Et)/ $\hbar$], which varies with position and time, and is inconsistent with a uniform probability distribution.

Symmetry and Definiteness

Quantum mechanics is a probabilistic theory, in the sense that the predictions it makes tell us, for instance, the probability of finding a particle somewhere in space. If we know nothing about a particle's previous history, and if there are no physical constraints that would make it more likely for a particle to be at one point along the x axis than any another, then the probability distribution must be P (x) = $constant$.

This is an example of a **symmetry** argument. Expressed more formally, it states that if the above conditions apply, then the probability distribution ought to be subject to the condition P (x + D) = P (x) for any constant value of D. The only possible P (x) in this case is P = $constant$. In the language of physics, if there is nothing that gives the particle a higher probability of being at one point rather than another, then the probability is independent of position and the system is **invariant under displacement** in the x direction.

The above argument doesn't suffice for quantum mechanics, since as we have learned, the fundamental quantity describing a particle is not the probability distribution, but the wave function ψ (x). Thus, the wave function rather than the probability distribution ought to be the quantity which is invariant under displacement, i. e., ψ (x + D) = ψ (x).

This condition turns out to be too restrictive, because it implies that ψ (x) = $constant$, whereas we know that a one-dimensional plane wave, which describes a particle with a uniform probability of being found anywhere along the x axis, has the form ψ (x) = exp(ikx). (For simplicity we temporarily ignore the time dependence.) If we make the substitution $x \rightarrow x + D$ in a plane wave, we get exp[ik (x + D)] = exp(ikx) exp(ikD). The wave function is thus technically not invariant under displacement, in that the displaced wave function is multiplied by the factor exp(ikD). However, the probability distribution of the displaced wave function still equals one everywhere, so there is no change in what we observe. Thus, in determining invariance under displacement, we are allowed to ignore changes in the wave function which consist only of multiplying it by a complex constant with an absolute value of one. Such a multiplicative constant is called a **phase factor** .

It is easy to convince oneself by trial and error or by more sophisticated means that the only form of wave function ψ (x) which satisfies the condition ψ (x + D) = ψ (x) × (phase factor) is ψ (x) = A exp(ikx) where A is a (possibly complex) constant. This is just in the form of a complex exponential plane wave with wavenumber k. Thus, not only is the complex exponential wave function invariant under displacements in the manner defined above, it is the **only** wave function which is invariant to displacements. Furthermore, the phase factor which appears for a displacement D of such a plane wave takes the form exp(iC) = exp(ikD), where k is the wavenumber of the plane wave.

As an experiment, let us see if a wave packet is invariant under displacement. Let's define a wave packet consisting of two plane waves:

$$\psi(x) = exp(ik1x) + exp(ik2x). \tag{1.1.203}$$

Making the substitution $x \rightarrow x + D$ in this case results in

$$\begin{aligned} &\psi(x+D) = exp[ik1(x+D)] + exp[ik2(x+D)] \\ &= exp(ik1x)exp(ik1D) + exp(ik2x)exp(ik2D)/ \\ &= [exp(ik1x) + exp(ik2x)] \times (phase\,factor) \end{aligned} \tag{1.1.204}$$

$[\exp(ikx) + \exp(ikx)] \times (\text{phase factor})$.

The impossibility of writing $\psi(x + D) = \psi(x) \times$ (phase factor) lends plausibility to the assertion that a single complex exponential is the only possible form of the wave function that is invariant under displacement.

Notice that the wave packet does not have definite wavenumber, and hence, momentum. In particular, the wave packet is a sum of complex exponentials with wavenumbers k_1 and k_2, which means that the associated particle can have either momentum $\Pi_1 = \hbar k_1$ or $\Pi_2 = \hbar k_2$. This makes sense from the point of view of the uncertainty principle – for a single plane wave the uncertainty in position is complete and the uncertainty in momentum is zero. For a wave packet the uncertainty in position is reduced and the uncertainty in the momentum is non-zero. As we have seen, this idea can be carried further: A definite value of momentum must be associated with a completely indefinite probability distribution in position, i. e., with $P = constant$. This corresponds to a wave function which has the form of a complex exponential plane wave. However, such a plane wave is invariant under displacement D, except for the multiplicative phase factor $\exp(ikD)$, which has no physical consequences since it disappears when the probability distribution is obtained. Thus, we see that invariance under displacement of the wave function and a definite value of the momentum are linked, in that each implies the other:

$$invarianceunderdisplacement \Longleftrightarrow definitemomentum \tag{1.1.205}$$

The idea of potential energy was introduced in the previous chapter. In particular, we found that if the total energy is constant, then the momentum cannot be constant in the presence of spatially varying potential energy. This means that the wavenumber, and hence the wavelength of the oscillations in the wave function also vary with position. The spatial inhomogeneity of the potential energy gives rise to spatial inhomogeneity in the wave function, and hence an indefinite momentum.

The above argument can be extended to other variables besides momentum. In particular since the time dependence of a complex exponential plane wave is $\exp(-i\omega t) = \exp(-iEt/\hbar)$, where E is the total energy, we have by analogy with the above argument that

$$invarianceundertimeshift \Longleftrightarrow definiteenergy. \tag{1.1.206}$$

Thus, invariance of the wave function under a displacement in time implies a definite value of the energy of the associated particle.

In the previous chapter we assumed that the frequency (and hence the energy) was definite and constant for a particle passing through a region of variable potential energy. We now see that this assumption is justified only if the potential energy doesn't change with time. This is because a time-varying potential energy eliminates the possibility of invariance under time shift.

Invariance

We have seen a few examples of invariance in quantum mechanics. It is now time to define this concept more precisely. A quantum mechanical wave function is said to be invariant under some transformation if the transformed wave function is observationally indistinguishable from the original.

In the above examples, the transformation is accomplished by replacing x by $x + D$ in the case of displacement in space and similarly by replacing t by $t + T$ for displacement in time. However, the idea of a transformation is much more general; other examples will be discussed as they arise.

The idea of "observationally indistinguishable" can be tricky. For example, if some transformation results in a new wave function which is the old wave function times a constant phase factor, then the new wave function is observationally indistinguishable from the old one. This is because physical measurements capture phase differences between different parts of wave functions (think of how interferometers work), but not absolute phases. The constant phase factor disappears in this difference calculation. However, if the multiplicative phase factor created by some transformation is a function of position, then the phase difference between different parts of a wave function changes as a result of the transformation. The wave function is not invariant under this transformation.

Compatible Variables

We already know that definite values of certain pairs of variables cannot be obtained simultaneously in quantum mechanics. For instance, the indefiniteness of position and momentum are related by the uncertainty principle — a definite value of position implies an indefinite value of the momentum and vice versa. If definite values of two variables can be simultaneously obtained, then we call these variables **compatible** . If not, the variables are **incompatible** .

If the wave function of a particle is invariant under the displacements associated with both variables, then the variables are compatible. For instance, the complex exponential plane wave associated with a free particle is invariant under displacements in both space and time. Since momentum is associated with space displacements and energy with time displacements, the momentum and energy are compatible variables for a free particle.

Compatibility and Conservation

Variables which are compatible with the energy have a special status. The wave function which corresponds to a definite value of such a variable is invariant to displacements in time. In other words, the wave function doesn't change under this displacement except for a trivial phase factor. Thus, if the wave function is also invariant to some other transformation at a particular time, it is invariant to that transformation for all time. The variable associated with that transformation therefore retains its definite value for all time — i. e., it is **conserved** .

For example, the plane wave implies a definite value of energy, and is thus invariant under time displacements. At time t = 0, it is also invariant under x displacements, which corresponds to the fact that it represents a particle with a known value of momentum. However, since momentum and energy are compatible for a free particle, the wave function will represent the **same** value of momentum at all other times. In other words, if the momentum is definite at t = 0, it will be definite at all later times, and furthermore will have the same value. This is how the conservation of momentum (and by extension, the conservation of any other variable compatible with energy) is expressed in quantum mechanics.

Symmetries and Variables

In modern quantum physics, the discovery of new symmetries leads to new dynamical variables. In the problems we show how that comes about for the symmetries of **parity** ($x \rightarrow -x$), **time reversal** $t \rightarrow -t$), and **charge conjugation** (the interchange of particles with antiparticles). One of the key examples of this was the development of the quark theory of matter, which came from the observation that the interchange of certain groups of elementary particles left the universe approximately unchanged, meaning that the universe was (approximately) symmetric under these interchanges.

Confined Matter Waves

Confinement of a wave to a limited spatial region results in rather peculiar behavior — the wave can only fit comfortably into the confined region if the wave frequency, and hence the associated particle energy, takes on a limited set of possible values. This is the origin of the famous quantization of energy, from which the "quantum" in quantum mechanics comes. We will explore two types of confinement, position confinement due to a potential energy well, and rotational confinement due to the fact that rotation of an object through 2π radians returns the object to its original orientation.

Particle in a Box

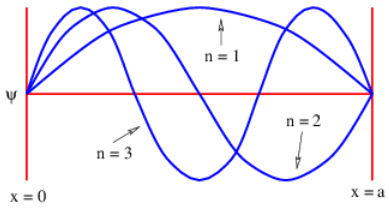


Figure 1.1.2:: First three modes for wave function of a particle in a box.

We now imagine how a particle confined to a region $0 \leq x \leq a$ on the x axis must behave. As with the displacement of a guitar string, the wave function must be zero at x = 0 and a, i. e., at the ends of the guitar string. A single complex exponential plane wave cannot satisfy this condition, since $|\exp[i\ (kx\ -\ \omega t\ )]|^2$ = 1 everywhere. However, a superposition (with a minus sign) of leftward and rightward traveling waves creates a **standing wave** , in which the the wave function separates into a function of space alone times a function of time alone.

$$\psi = exp[i(kx - \omega t)] - exp[i(-kx - \omega t)] = 2iexp(-i\omega t)sin(kx). \tag{1.1.207}$$

Notice that the time dependence is still a complex exponential, which means that $|\psi\ |^2$ is independent of time. This insures that the probability of finding the particle **somewhere** in the box remains constant with time. It also means that the wave packet corresponds to a definite energy $E = \hbar\ \omega$.

Because we took a difference rather than a sum of plane waves, the condition ψ = 0 is already satisfied at x = 0. To satisfy it at x = a, we must have $ka = n\pi$, where n = 1, 2, 3, ... Thus, the absolute value of the wavenumber must take on the discrete values

$$n\pi - kn = a, n = 1, 2, 3, ... \tag{1.1.208}$$

(The wavenumbers of the two plane waves equal plus or minus this absolute value respectively.) This implies that the absolute value of the particle momentum is $\Pi_n = \hbar\ k_n = n\pi\ \hbar\ /a$, which in turn means that the energy of the particle must be

$$En = (\Pi 2nc2 + m2c4)1\ /\ 2 = (n2\pi 2^{-}h2c2\ /\ a2 + m2c4)1\ /\ 2, \tag{1.1.209}$$

where m is the particle mass. In the non-relativistic limit this becomes

$$2222En = \Pi - n - = n - \pi - \bar{}h - -(non - relativistic)2m2ma2 \tag{1.1.210}$$

where we have dropped the rest energy mc^2 since it is a constant offset. In the ultra-relativistic case where we can ignore the particle mass, we find

$$E = |\Pi|c = n - \pi\bar{}hc(zeromass).nna \tag{1.1.211}$$

The shapes of the wave functions for the first three values of n for the particle in the box are illustrated in Figure 1.1.2:.

In both limits the energy takes on only a certain set of possible values. This is called **energy quantization** and the integer n is called the **energy quantum number** . In the non-relativistic limit the energy is proportional to n^2, while in the ultra-relativistic case the energy is proportional to n.

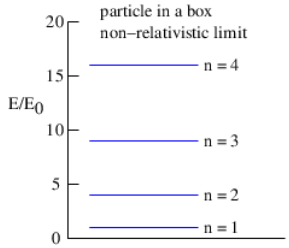


Figure 1.1.3:: Allowed energy levels for the non-relativistic particle in a box. The constant $E_0 = \pi^2\hbar^2/(2ma^2)$. See text for the meanings of symbols.

We can graphically represent the allowed energy levels for the particle in a box by an **energy level diagram** . Such a diagram is shown in Figure 1.1.3: for the non-relativistic case.

One aspect of this problem deserves a closer look. Equation (9.13) shows that the wave function for this problem is a superposition of two plane waves corresponding to momenta $\Pi_1 = +\hbar\ k$ and $\Pi_2 = -\hbar\ k$ and is therefore a kind of wave packet. Thus, the wave function is not invariant under displacement and does not correspond to a definite value of the momentum — the momentum's absolute value is definite, but its sign is not. Following Feynman's prescription, equation 9.13} tells us that the amplitude for the particle in the box to have momentum $+\hbar\ k$ is exp[i (kx - ωt)], while the amplitude for it to have momentum -$\hbar\ k$ is - exp[i (-kx - ωt)]. The absolute square of the sum of these amplitudes gives us the relative probability of finding the particle at position x :

$$P(x) = |2iexp(-i\omega t)sin(kx)|2 = 4sin2(kx). \tag{1.1.212}$$

Which of the two possible values of the momentum the particle takes on is unknowable, just as it is impossible in principle to know which slit a particle passes through in two slit interference. If an experiment is done to measure the momentum, then the wave function is irreversibly changed, just as the interference pattern in the two slit problem is destroyed if the slit through which the particle passes is unambiguously determined.

Barrier Penetration

Unlike the situation in classical mechanics, quantum mechanics allows the kinetic energy K to be negative. This makes the momentum Π (equal to $(2mK)^{1/2}$ in the nonrelativistic case) imaginary, which in turn gives rise to an imaginary wavenumber.

Let us investigate the nature of a wave with an imaginary wavenumber. Let us assume that $k = i\kappa$ in a complex exponential plane wave, where κ is real:

$$\psi = exp[i(kx - \omega t)] = exp(-\kappa x - i\omega t) = exp(-\kappa x)exp(-i\omega t). \tag{1.1.213}$$

The wave function doesn't oscillate in space when $K = E - U < 0$, but grows or decays exponentially with x, depending on the sign of κ.

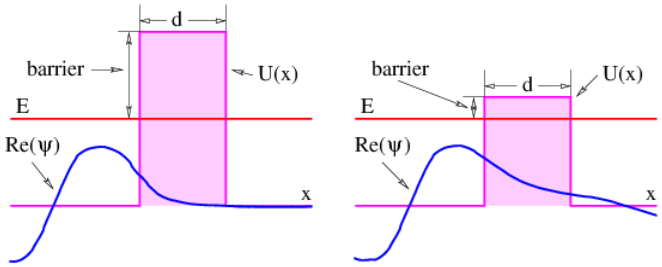


Figure 1.1.4:: Real part of wave function Re[ψ (x)] for barrier penetration. The left panel shows weak penetration occurring for a large potential energy barrier, while the right panel shows stronger penetration which occurs when the barrier is small.

For a particle moving to the right, with positive k in the allowed region, κ turns out to be positive, and the solution decays to the right. Thus, a particle impingent on a potential energy barrier from the left (i. e., while moving to the right) will have its wave amplitude decay in the classically forbidden region, as illustrated in Figure 1.1.4:. If this decay is very rapid, then the result is almost indistinguishable from the classical result — the particle cannot penetrate into the forbidden region to any great extent. However, if the decay is slow, then there is a reasonable chance of finding the particle in the forbidden region. If the forbidden region is finite in extent, then the wave amplitude will be small, but non-zero at its right boundary, implying that the particle has a finite chance of completely passing through the classical forbidden region. This process is called **barrier penetration** .

The probability for a particle to penetrate a barrier is the absolute square of the amplitude after the barrier divided by the square of the amplitude before the barrier. Thus, in the case of the wave function illustrated in equation 9.19}, the probability of penetration is

$$P = |\psi(d)|2 \,/\, |\psi(0)|2 = exp(-2\kappa d) \tag{1.1.214}$$

where d is the thickness of the barrier.

The rate of exponential decay with x in the forbidden region is related to how negative K is in this region. Since

$$\Pi 2^{-}h2k2^{-}h2\kappa 2 - K = U - E = - - - - - - \\ = - - - - - - - = - - - - - -, 2m2m2m \tag{1.1.215}$$

we find that

$$()1 \,/\, 2\kappa = 2mB - -^{-}h2 \tag{1.1.216}$$

where the **potential energy barrier** is $B \equiv$-K = U - E. The smaller B is, the smaller is κ, resulting in less rapid decay of the wave function with x. This corresponds to stronger barrier penetration. (Note that the way B is defined, it is positive in forbidden regions.)

If the energy barrier is very high, then the exponential decay of the wave function is very rapid. In this case the wave function goes nearly to zero at the boundary between the allowed and forbidden regions. This is why we specify the wave function to be zero at the walls for the particle in a box. These walls act in effect as infinitely high potential barriers.

Barrier penetration is important in a number of natural phenomena. Certain types of radioactive decay and the fissioning of heavy nuclei are governed by this process.

Orbital Angular Momentum

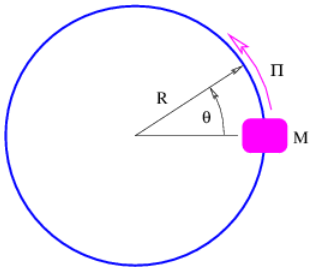


Figure 1.1.5:: Illustration of a bead of mass M sliding (without friction) on a circular loop of wire of radius R with momentum Π.

Another type of bound state motion occurs when a particle is constrained to move in a circle. (Imagine a bead sliding on a circular loop of wire, as illustrated in Figure 1.1.5:.) We can define x in this case as the path length around the wire and relate it to the angle θ : x = $R\theta$. For a plane wave we have

$$\psi = exp[i(kx - \omega t)] = exp[i(kR\theta - \omega t)]. \tag{1.1.217}$$

This plane wave differs from the normal plane wave for motion along a Cartesian axis in that we must have $\psi(\theta) = \psi(\theta + 2\pi)$. This can only happen if the circumference of the loop, $2\pi R$, is an integral number of wavelengths, i. e., if $2\pi R / \lambda = m$ where m is an integer. However, since $2\pi / \lambda = k$, this condition becomes $kR = m$.

Since $\Pi = \hbar\, k$, the above condition can be written $\Pi_m R = m\, \hbar$. The quantity

$$Lm \equiv \Pi mR \tag{1.1.218}$$

is called the **angular momentum** , leading to our final result,

$$Lm = m^{-}h, m = 0, \pm 1, \pm 2, ... \tag{1.1.219}$$

We see that the angular momentum can only take on values which are integer multiples of $\hbar$. This represents the **quantization of angular momentum** , and m in this case is called the **angular momentum quantum number** . Note that this quantum number differs from the energy quantum number for the particle in the box in that zero and negative values are allowed.

The energy of our bead on a loop of wire can be expressed in terms of the angular momentum:

$$\Pi 2L2Em = --m- = ---m - 2-.2M2MR \tag{1.1.220}$$

This means that angular momentum and energy are compatible variables in this case, which further means that angular momentum is a conserved variable. Just as definite values of linear momentum are related to invariance under translations, definite values of angular momentum are related to invariance under rotations. Thus, we have

$$invarianceunderrotation \Longleftrightarrow definiteangularmomentum \tag{1.1.221}$$

for angular momentum.

We need to briefly address the issue of angular momentum in three dimensions. Angular momentum is actually a vector oriented perpendicular to the wire loop in the example we are discussing. The direction of the vector is defined using a variation on the right-hand rule: Curl your fingers in the direction of motion of the bead around the loop (using your right hand!). The orientation of the angular momentum vector is defined by the direction in which your thumb points. This tells you, for instance, that the angular momentum in Figure 1.1.5: points out of the page.

In quantum mechanics it is only possible to measure simultaneously the square of the length of the angular momentum vector and one component of this vector. Two different components of angular momentum cannot be simultaneously measured because of the uncertainty principle. However, the **length** of the angular momentum vector may be measured simultaneously with one component. Thus, in quantum mechanics, the angular momentum is completely specified if the length and one component of the angular momentum vector are known.

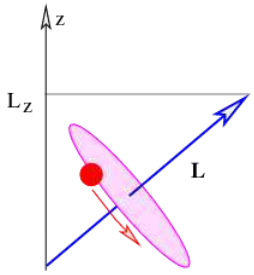


Figure 1.1.6:: Illustration of the angular momentum vector L for a tilted loop and its z component L_z.

Figure 1.1.6: illustrates the angular momentum vector associated with a bead moving on a wire loop which is tilted from the horizontal. One component (taken to be the z component) is shown as well. For reasons we cannot explore here, the square of the length of the angular momentum vector L^2 is quantized with the following values:

$$22Ll = {}^{-}hl(l+1), l = 0, 1, 2, ... \tag{1.1.222}$$

One component (say, the z component) of angular momentum is quantized just like angular momentum in the two-dimensional case, except that l acts as an upper bound on the possible values of $|m|$. In other words, if the square of the length of the angular momentum vector is $\hbar^2 l(l+1)$, then the z component can take on the values

$$Lzm = {}^{-}hm, m = -l, -l+1, ..., l-1, l. \tag{1.1.223}$$

The quantity l is called the **angular momentum quantum number** , while m is called the **orientation** or **magnetic quantum number** , the latter for historical reasons.

Angular Momentum

The type of angular momentum discussed above is associated with the movement of particles in orbits. However, it turns out that even stationary particles can possess angular momentum. This is called **spin** angular momentum. The **spin quantum number** s plays a role analogous to l for spin angular momentum, i. e., the square of the spin angular momentum vector of a particle is

$$22Ls = ¯hs(s + 1). \tag{1.1.224}$$

The **spin orientation quantum number** m_s is similarly related to s :

$$Lzs = ¯hms, ms = −s, −s + 1, ..., s − 1, s. \tag{1.1.225}$$

The spin angular momentum for an elementary particle is absolutely conserved, i. e., it can never change. Thus, the value of s is an intrinsic property of a particle. The major difference between spin and orbital angular momentum is that the spin quantum number can take on more values, i. e., s = 0, 1/ 2, 1, 3/ 2, 2, 5/ 2, ...

Particles with integer spin values s = 0, 1, 2, ... are called **bosons** after the Indian physicist Satyendra Nath Bose. Particles with half-integer spin values s = 1/ 2, 3/ 2, 5/ 2, ... are called **fermions** after the Italian physicist Enrico Fermi. As we shall see later in the course, bosons and fermions play very different roles in the universe.

Problems

1. Suppose that a particle is represented by the wave function ψ = sin(kx - ωt) + sin(-kx - ωt).
    1. Use trigonometry to simplify this wave function.
    2. Compute the x and t dependence of the probability of finding the particle by squaring the wave function.
    3. Explain what this result says about the time dependence of the probability of finding the particle. Does this make sense?
2. Repeat the above problem for a particle represented by the wave function ψ = exp[i (kx - ωt)] + exp[i (-kx - ωt)].
3. Determine if the wavefunction ψ (x) = exp(iCx ^{2}) is invariant under displacement in the sense that the displaced wave function differs from the original wave function by just a phase factor.
4. Just as invariance under the substitution x → x + D is associated with momentum, invariance under the substitution x →-x is associated with a quantum mechanical variable called **parity** , denoted P. However, unlike momentum, which can take on any numerical value, parity can take on only two possible values, ±1. The parity of a wave function ψ (x) is +1 if ψ (-x) = ψ (x), while the parity is -1 if ψ (-x) = -ψ (x). If ψ (x) satisfies neither of these conditions, then it has no definite value of parity.
    1. What is the parity of ψ = sin(kx)? Of ψ = cos(kx)? The quantity k is a constant.
    2. Is ψ (x) = cos(kx) invariant under the substitution x = x + D for all possible values of D ? Does this wave function have a definite value of the momentum?
    3. Show that a wave function with a definite value of the momentum does not have a definite value of parity. Are momentum and parity compatible variables?
5. Realizing that cos(kx - ωt) can be written in terms of complex exponential functions, give a physical interpretation of the meaning of the above cosine wave function. In particular, what are the possible values of the associated particle's momentum and energy?
6. The time reversal operation T makes the substitution t →-t. Similar to parity, time reversal can only take on values ±1. Is symmetry of a wave function under time reversal, i. e., ψ (-t) = ψ (t), consistent with a definite value of the energy? Hint: Any wave function corresponding to a definite value of energy E must have the form ψ = A exp(-iEt/ $\hbar$) where A is not a function of time t. (Why?)
7. The operation C takes the complex conjugate of the wave function, i. e., it makes the substitution i →-i. In modern quantum mechanics this corresponds to interchanging particles and antiparticles, and is called **charge conjugation** . What does the combined operation CPT do to a complex plane wave, i. e., one with definite wave vector and frequency?
8. Make an energy level diagram for the case of a massless particle in a box.
9. Compare |Π| for the ground state of a non-relativistic particle in a box of size a with ΔΠ obtained from the uncertainty principle in this situation. Hint: What should you take for Δx ?
10. Imagine that a billiard table has an infinitely high rim around it. For this problem assume that $\hbar$ = 1 kg m^2 s^{-1}.
    1. If the table is 1. 5 m long and if the mass of a billiard ball is M = 0. 5 kg, what is the billiard ball's lowest or ground state energy? Hint: Even though the billiard table is two dimensional, treat this as a one-dimensional problem. Also, treat the

problem nonrelativistically and ignore the contribution of the rest energy to the total energy.

2. The energy required to lift the ball over a rim of height H against gravity is $U = MgH$ where $g = 9.8$ m s^{-2}. What rim height makes the gravitational potential energy equal to the ground state energy of the billiard ball calculated above?
3. If the rim is actually twice as high as calculated above but is only 0.1 m thick, determine the probability of the ball penetrating the rim.

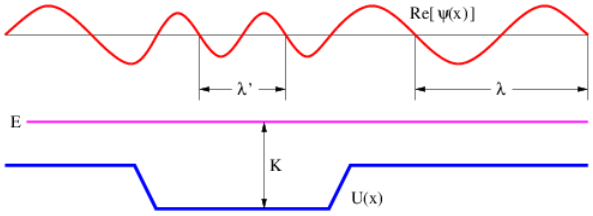


Figure 1.1.7:: Real part of the wave function ψ, corresponding to a fixed total energy E, occurring in a region of spatially variable potential energy U (x). Notice how the wavelength λ changes as the kinetic energy $K = E - U$ changes.

---

11. The real part of the wave function of a particle with positive energy E passing through a region of negative potential energy is shown in Figure 1.1.7:.
    1. If the total energy is definitely E, what is the dependence of this wave function on time?
    2. Is the wave function invariant under displacement in space in this case? Why or why not?
    3. Does this wave function correspond to a definite value of momentum? Why or why not?
    4. Is the momentum compatible with the energy in this case? Why or why not?
12. Assuming again that $\hbar = 1$ kg m^2 s^{-1}, what are the possible speeds of a toy train of mass 3 kg running around a circular track of radius 0.8 m?
13. If a particle of zero mass sliding around a circular loop of radius R can take on angular momenta $L_m = m\hbar$ where m is an integer, what are the possible kinetic energies of the particle? Hint: Remember that $L = \Pi R$.

Chapter 10 Dynamics of Multiple Particles

So far we have considered only the dynamics of a single particle subject to an externally imposed potential energy. The particle has no way of influencing this external agent. In the real world particles interact with each other. In this chapter we learn how this happens.

We first rewrite Newton's second law in terms of momentum. This is useful in the subsequent consideration of Newton's third law, which leads to the principle of the conservation of momentum. Collisions between particles and the behavior of rockets and conveyor belts are then studied as applications of the conservation laws to more than one particle.

Momentum and Newton's Second Law

Up to this point we have stated Newton's second law in its conventional form, $F = ma$. However, in the non-relativistic case $ma = md\,u\,/dt = d\,(m\,u\,)/dt$, so we can also write Newton's second law as

$$F = dp - (Newton's second law)dt \tag{1.1.226}$$

where $p = m\,u$ is the non-relativistic kinetic momentum. This form of Newton's second law is actually closer to Newton's original statement of the law. It also has the advantage that it is correct even in the relativistic case when the relativistic definition of kinetic momentum, $p = m\,u\,/\,(1 - u^{\,2}/c^{\,2})^{1/2}$ (as defined earlier), is substituted.

Newton's Third Law

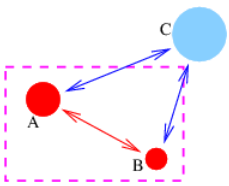


Figure 1.1.1:: Interactions between three particles, A, B, and C. A and B are considered to be part of the system defined by the dashed line.

Newton's third law states that if particle A exerts a force F on particle B, then particle B exerts a force -F on particle A. Newton's third law makes it possible to apply Newton's second law to systems of particles without considering the detailed interactions between particles within the system. For instance, if we (arbitrarily) define the system in Figure 1.1.1: to be the particles A and B inside the dashed lines, then we can divide the forces acting on these particles into internal and external parts,

$$FA = FA - internal + FA - external = dpA -, dt \tag{1.1.227}$$

$$FB = FB - internal + FB - external = dpB -.dt \tag{1.1.228}$$

Adding these equations together results in the net force F_{net} being equal to

$$\begin{aligned} d - FA - internal + FA - external + FB - internal + FB - external \\ = dt(pA + pB). \end{aligned} \tag{1.1.229}$$

However, the internal interactions in this case are A acting on B and B acting on A. These forces are equal in magnitude but opposite in direction, so they cancel out, leaving us with the net force equal to the sum of the external parts, $F_{\text{net}} = F_{\text{A-external}} + F_{\text{B-external}}$. The external forces in Figure 1.1.1: are the force of C on A and the force of C on B. Defining the total kinetic momentum of the system as the sum of the A and B momenta, $p_{\text{tot}} = p_{\text{A}} + p_{\text{B}}$, the above equation becomes

$$dptotFnet = - - - - -, dt \tag{1.1.230}$$

which looks just like Newton's second law for a single particle, except that it now applies to the system of particles (A and B in the present case) as a whole. This argument easily generalizes to any number of particles inside and outside the system. Thus, for instance, even though a soccer ball consists of billions of atoms, we are sure that the forces between atoms within the soccer ball cancel out, and the trajectory of the ball as a whole is determined solely by external forces such as gravity, wind drag, friction with the ground, and the kicks of soccer players.

Remember that for two forces to be a third law pair, they have to be acting on different particles. Furthermore, if one member of the pair is the force of particle A acting on particle B, then the other must be the force of particle B acting on particle A. A counterexample would be gravity and the upward normal force acting on a mass sitting on a table; these forces are equal and opposite in the stationary case, but act on the same object, and therefore **do not** constitute a third law pair. However, the upward normal force of the table on the mass and the downward normal force of the mass on the table **would be** a third law pair.

Conservation of Momentum

If all external forces on a system are zero, then equation 10.5} reduces to

$$ptot = const(isolatedsystem). \tag{1.1.231}$$

A system of particles with no external forces acting on it is called **isolated** . Newton's third law thus tells us that the kinetic momentum of an isolated system doesn't change with time. This law is called the **conservation of momentum** .

Collisions

Let us now consider the situation in which two particles collide with each other. There can be several outcomes to this collision, of which we will study two:

- The two particles collide **elastically** , in essence bouncing off of each other.
- The two particles stick together, resulting in the production of a single particle, or a single particle breaks apart into two particles. These are **inelastic processes** .

In both of the above cases energy and momentum are conserved. We assume that the forces acting between the particles are short range, so that except in the instant of collision, we need not worry about potential energy or potential momentum — all energy is in the form of rest plus kinetic energy except in this short interval, and all momenta are kinetic momenta.

Because of the principle of relativity, we are free to consider collisions in any convenient reference frame. We can then transform the results to any reference frame we please. Generally speaking, the most convenient reference frame to consider is the one in which the total momentum of the two particles is zero. For the sake of simplicity we only consider collisions in one dimension.

Elastic Collisions

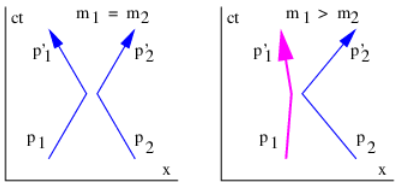


Figure 1.1.2:: One-dimensional elastic collisions of two particles in the center of momentum frame as seen in spacetime diagrams.

Suppose a particle with mass m_1 and initial velocity u_1 in the **center of momentum** frame, i. e., the reference frame in which the total momentum is zero, collides elastically with another particle of mass m_2 with initial velocity u_2. The momenta of the two particles are

$$p = - - - - - m1u1 - - - - - - - p = - - - - - m2u2 - - - - - - - \\ .1(1 - u21 / c2)1 / 22(1 - u22 / c2)1 / 2 \tag{1.1.232}$$

In the center of momentum frame we must have

$$p1 = -p2. \tag{1.1.233}$$

Figure 1.1.2: shows what happens when these two particles collide. The first particle acquires momentum p_1' while the second acquires momentum p_2'. The conservation of momentum tells us that the total momentum after the collision is the same as before the collision, namely zero, so

$$''p1 = -p2. \tag{1.1.234}$$

In the center of momentum frame we know that $|p_1|$ = $|p_2|$ and we know that the two momentum vectors point in opposite directions. Similarly, $|p_1'|$ = $|p_2'|$. However, we as yet don't know how p_1' is related to p_1. Conservation of energy,

$$E1 + E2 = E'1 + E'2, \tag{1.1.235}$$

gives us this information. Notice that if p_1' = -p_1, then $E_1'^2 = p_1'^2c^2 + m_1^2c^4 = p_1^2c^2 + m_1^2c^4 = E_1^2$. Assuming positive energies, we therefore have $E_1' = E_1$. If p_2' = -p_2, then we can similiarly infer that $E_2' = E_2$. If these conditions are satisfied, then so is equation 10.10}. Therefore, a complete solution to the problem is

$$''p1 = -p1 = -p2 = p2 \equiv p \tag{1.1.236}$$

and

$$E = E' = (p2c2 + m2c4)1 / 2E = E' = (p2c2 + m2c4)1 / 2.111222 \tag{1.1.237}$$

In other words, the particles just exchange momenta.

The left panel of Figure 1.1.2: shows what happens in a collision when the masses of the two colliding particles are equal. If m_1 = m_2, then the incoming and outgoing velocities of the two particles are the same, as indicated by the inverse slopes of the world lines. On the other hand, if $m_1 > m_2$, then the velocity of particle 2 is greater than the velocity of particle 1, as is illustrated in the right panel of Figure 1.1.2:.

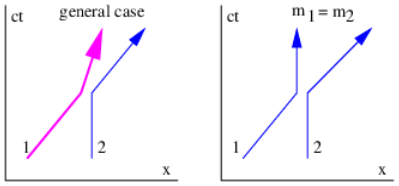


Figure 1.1.3:: Elastic collisions viewed from a reference frame in which one particle is initially stationary.

Suppose we wish to view the results of an elastic collision in a reference frame in which particle 2 is initially stationary. All we have to do is to transform the velocities into a reference frame moving with the initial velocity of particle 2, as illustrated in Figure 1.1.3:. We do this by relativistically adding U = -u_2 to each velocity. (Note that the velocity U of the moving frame is positive since u_2 is negative.) Using the relativistic velocity translation formula, we find that

$$''v1 = - - u1 + -U - - - v' = - - u1 - + - U - - - v \\ ' = - - u2 + -U - - - 1 + u1U / c211 + u'1U / c221 + u'2U / c2 \tag{1.1.238}$$

https://phys.libretexts.org/@go/page/32918

where u_1, u_1', u_2, and u_2' indicate velocities in the original, center of momentum reference frame and v_1, v_1', etc., indicate velocities in the transformed frame.

In the special case where the masses of the two particles are equal to each other, we have $v_1 = 2U/(1 + U^2/c^2)$, $v_1' = 0$, and $v_2' = 2U/(1 + U^2/c^2) = v_1$. Thus, when the masses are equal, the particles simply exchange velocities.

If the velocities are nonrelativistic, then the simpler Galilean transformation law $v = u + U$ can be used in place of the relativistic equations invoked above.

Inelastic Collisions

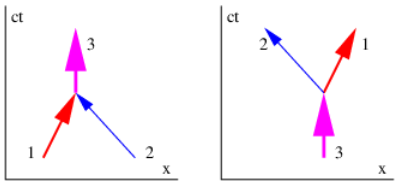


Figure 1.1.4:: Building blocks of inelastic collisions. In the left panel two particles collide to form a third particle. In the right panel a particle breaks up, forming two particles.

An inelastic collision is one in which the particles coming out of the collision are not the same as the particles going into it. Inelastic collisions conserve both total momentum and energy just as elastic collisions do. However, unlike elastic collisions, inelastic collisions generally do not conserve the total **kinetic** energy of the particles, as some rest energy is generally created or destroyed.

Figure 1.1.4: shows the fundamental building blocks of inelastic collisions. We can consider even the most complex inelastic collisions to be made up of composites of only two processes, the creation of one particle from two, and the disintegration of one particle into two.

Let us consider each of these in the center of momentum frame. In both cases the single particle must be stationary in this frame since it carries the total momentum of the system, which has to be zero. By conservation of momentum, if particle 1 in the left panel of Figure 1.1.4: has momentum p, then the momentum of particle 2 is -p. If the two particles have masses m_1 and m_2, then their energies are $E_1 = (p^2c^2 + m_1^2c^4)^{1/2}$ and $E_2 = (p^2c^2 + m_2^2c^4)^{1/2}$. The energy of particle 3 is therefore $E_3 = E_1 + E_2$, and since it is at rest, all of its energy is in the form of "mc^2" or rest energy, and so the mass of this particle is

$$\begin{aligned} m3 = (E1 + E2)/c2(centerofmomentumframe)2221/22221/2 \\ = (p/c + m1) + (p/c + m2) \\ = m1[1 + p2/(m21c2)]1/2 + m2[1 + p2/(m22c2)]1/2. \end{aligned} \tag{1.1.239}$$

The last line in the above equation shows that $m_3 > m_1 + m_2$ because it is in the form $m_1A + m_2B$ where both A and B are greater than one. Thus, rest energy is created in the amount $\Delta E_{rest} = (m_3 - m_1 - m_2)c^2$.

Actually, it is easy to calculate the mass of particle 3 in the above case from any reference frame as long as the momenta and energies of particles 1 and 2 are known in this frame. By conservation of energy and momentum, $E_3 = E_1 + E_2$ and $p_3 = p_1 + p_2$. Furthermore, $E_3^2 = p_3^2c^2 + m_3^2c^4$, so we can solve for m_3:

$$m = [(E + E)2/c4 - (p + p)\cdot(p + p)/c2]1/2(anyframe).3121212 \tag{1.1.240}$$

The right panel of Figure 1.1.4: shows the process of particle decay. This is just the inverse of the particle creation process, and all of the analysis we have done for creation is valid for particle decay except that rest energy is converted to kinetic energy rather than vice versa.

Rockets and Conveyor Belts

Normally when we define a system to which Newton's second law is to be applied, the system is **closed** in the sense that mass cannot enter or exit the system. However, sometimes it is convenient to work with **open** systems for which this is not true. The classic example is the rocket, where exhaust gases leave the system, thus decreasing the mass of the rocket with time.

Open systems can be analyzed if momentum is considered to be a quantity which is accounted for much as money is accounted for in a bank account. The bank account can change in three ways: money can be deposited in the account, it can be withdrawn from the account, and the amount can grow or shrink as a consequence of interest payments or fees. Similarly, the amount of momentum in a system can change as the result of mass entering the system, mass leaving the system, and forces acting on the system. The time rate of change of momentum in a system is therefore

$$()()dpdpdpdt- = F + dt - -dt-, inout \tag{1.1.241}$$

where F is the net force on the system, $(dp / dt)_{in}$ is the momentum per unit time added by mass entering the system, and $(dp / dt)_{out}$ is the amount lost per unit time by mass exiting the system. In the non-relativistic case, $(dp / dt)_{in} = u_{in}(dm / dt)_{in}$ and $(dp / dt)_{out} = u_{out}(dm / dt)_{out}$, where $(dm / dt)_{in}$ is the mass entering the system per unit time with velocity u_{in} and $(dm / dt)_{out}$ is the mass per unit time exiting the system with velocity u_{out}.

For non-relativistic velocities, the momentum of the system can be written as $p = mu$ so that

$$dp- = dm - u + mdu-.dtdtdt \tag{1.1.242}$$

To complete the analysis, we need an accounting of the mass entering and leaving the system:

$$()()dm - -dm - -dm - -dt = dt - dt.inout \tag{1.1.243}$$

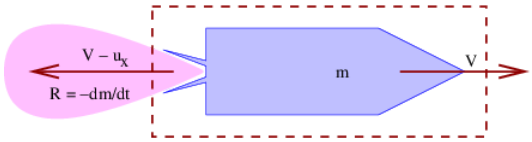


Figure 1.1.5:: Rocket moving with velocity V while expelling gas at a rate R with velocity V - u_x.

Let us see how to apply this to a rocket for which all velocities are non-relativistic. As Figure 1.1.5: indicates, a rocket spews out a stream of exhaust gas. The system is defined by the dashed box and includes the rocket and the part of the exhaust gas inside the box. The reaction to the momentum carried off in this stream of gas is what causes the rocket to accelerate. We note that $(dm / dt)_{in}$ = 0 since no mass is entering the system, and $(dm / dt)_{out} = R$, the rate at which mass is ejected by the rocket in the form of exhaust gas. The rocket is assumed to be moving to the right at speed V and the gas is ejected at a speed u_x relative to the rocket, which means that its actual velocity after ejection is V - u_x. We call u_x the **exhaust velocity** . Notice that V - u_x may be either positive or negative, depending on how big V is.

Equating the mass of the rocket to the system mass, we find that R = -dm / dt. The momentum balance equation 10.16} becomes dp / dt = -(V - u_x)R. The force on the rocket is actually zero, so the force term does not enter the momentum balance equation. This is non-intuitive, because we are used to acceleration being the result of a force. However, nothing, including the ejected gas, is actually **pushing** on the system, so we must indeed conclude that there is no force — all of the change in the system's momentum arises from the ejection of gas with the opposite momentum.[1]

Finally, we see that $dp / dt = (dm / dt)V + m\,(dV / dt)$ = -RV + m (dV / dt). Equating this to the results of the momentum balance calculation gives us -RV + m (dV / dt) = -(V - u_x)R. Solving for the acceleration dV / dt results in

$$dV- = uxR - -(rocketacceleration).dtm \tag{1.1.244}$$

Thus, the acceleration of the rocket depends on the exhaust velocity of the ejected gas, the rate at which the gas is being ejected, and the mass of the rocket.

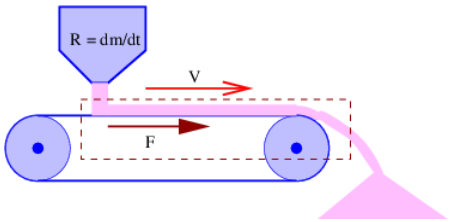


Figure 1.1.6:: Sand is dumped on a conveyor belt and in turn is dumped off the end of the belt.

Figure 1.1.6: illustrates another type of open system problem. A hopper dumps sand on a conveyor belt at a rate of R kilograms per second. The conveyor belt is moving to the right at (non-relativistic) speed V and the sand is dumped off at the end. What force F is needed to keep the conveyor belt moving at a constant speed, assuming that the conveyor belt mechanism itself is frictionless? In this case $(dm / dt)_{in} = (dm / dt)_{out} = R$. Furthermore, since the system outlined by the dashed line is in a steady state, dp / dt = 0.

The key to understanding this problem is that the sand enters the system with zero horizontal velocity, but exits the system with the horizontal velocity of the conveyor belt, V. The momentum balance equation is thus

$$0 = F - VR \tag{1.1.245}$$

and the force is

$$F = VR(forceonconveyorbelt). \tag{1.1.246}$$

This force serves to accelerate the sand up to the velocity of the conveyor belt.

Problems

1. Imagine a block of mass M resting on a plate under the influence of gravity, as shown in Figure 1.1.7:.
   1. Determine the force of the plate on the block, N_b, and the force of the block on the plate, N_p.
   2. State which of the three forces, Mg, N_b, and N_p, form a Newton's third law pair.

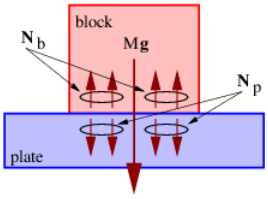


Figure 1.1.7:: Block of mass M subject to gravitational force Mg while resting on a plate. The force of the block on the plate is N_p while the force of the plate on the block is N_b.

---

2. Repeat the previous problem assuming that the block and the plate are in an elevator accelerating upward with acceleration a.
3. Straighten out the misunderstanding of Newton's third law implicit in the question "If the force of the horse on the cart equals the force of the cart on the horse, why does anything ever go anywhere"? Examine in particular the conditions under which the horse-cart system accelerates.
4. A pusher boat (mass M) on the Mississippi is pushing two barges (each mass m) at a steady speed as shown in Figure 1.1.8:. Each barge is subject to a drag force by the water of F_B. Consider only horizontal force components in the following.
   1. What is the total horizontal force of the water on the barge-boat system? Explain.
   2. What is the direction and magnitude of the force of the pusher boat on barge 1? Explain.

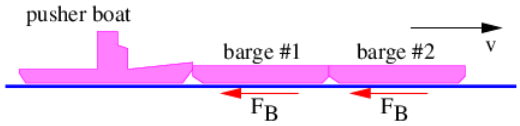


Figure 1.1.8:: Barges being pushed by a pusher boat on the Mississippi. Each barge experiences a drag force F_b.

---

5. A train with an engine of mass M and 2 freight cars, each of mass m, is accelerating to the right with acceleration a on a horizontal track as shown in Figure 1.1.9:. Assume that the two freight cars roll with negligible friction. Consider only horizontal force components below.
   1. Find the direction and magnitude of the force of the rails on the engine and specify the system to which Newton's second law is applied.
   2. Find the direction and magnitude of the force of the engine on the first car and specify the system to which Newton's second law is applied.
   3. Find the direction and magnitude of the force of the first car on the second car and specify the system to which Newton's second law is applied.
   4. Find the direction and magnitude of the force of the second car on the first car and specify the law used to obtain this force.

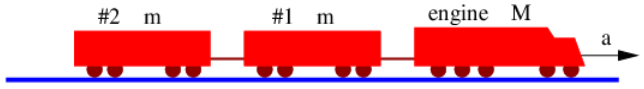


Figure 1.1.9:: An engine and two freight cars accelerating to the right.

---

6. A car and trailer are descending a hill as shown in Figure 1.1.10:. Assume that the trailer rolls without friction and that air friction can be ignored. Consider only forces parallel to the road surface.
    1. Compute the force of the road on the car if the car-trailer system shown in Figure 1.1.10: is moving down the hill at constant speed.
    2. Compute the force of the trailer on the car in the above conditions.
    3. If the driver takes his foot off the brake and lets the car coast frictionlessly, recompute the force of the trailer on the car.

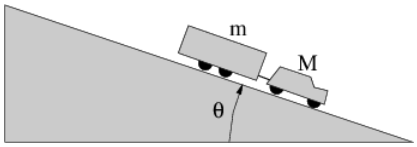


Figure 1.1.10:: A car and a trailer going down a hill.

---

7. Consider a one-dimensional elastic collision between particles of masses m_1 and m_2. If particle 2 is initially stationary, what range of values must m_1/m_2 have for the initial particle to rebound backwards along its initial track after the collision? (Do this problem non-relativistically.)
8. A stationary pion (mass M) decays into a muon (mass $m < M$) and a neutrino (massless).
    1. What is the (fully relativistic) momentum of the muon after the decay?
    2. What is the energy of the neutrino?
9. In an elastic collision viewed in the center of momentum frame, the energy of each particle is conserved individually. Is this true for the same process viewed from a reference frame in which one of the particles is initially stationary?
10. A space probe approaches a planet in the -x direction, curves around it under the influence of the planet's powerful gravity (a conservative force) and recedes from the planet in the +x direction, as seen in Figure 1.1.11:. The planet is moving in the +x direction at speed V, while the space probe is initially moving in the -x direction at speed u_1. What is its speed u_2 in the +x direction after this close approach to the planet? Treat this problem nonrelativistically. Hint: First transform to the center of mass frame in which the planet is essentially stationary. Work out the interaction between the probe and the planet in this frame. Then transform back to the original reference frame. Assume that the mass of the probe is negligible compared to that of the planet.

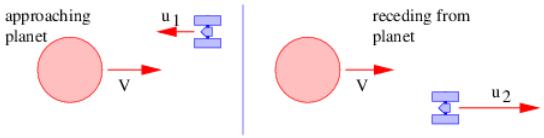


Figure 1.1.11:: A space probe approaches a planet, curves around it, and heads off in the opposite direction.

---

11. Two asteroids, each with mass 10^{10} kg and initial speed 10^5 m s^{-1}, collide head on. The whole mess congeals into one large mass. How much rest mass (rest energy divided by c^2) is created?
12. Two equal objects, both with mass m, collide and stick together. Before the collision, one mass is stationary and the other is moving at speed v. In the following, assume that velocities are fully relativistic.
    1. Compute the total momentum and energy (including rest energy) of the two masses before the collision.
    2. Compute the mass M of the combined system after the collision, taking the conversion of energy into mass into account.
13. Explain qualitatively why a fireman needs to push forward on a firehose to keep it stationary. Hint: The water is flowing faster after it comes out of the nozzle of the hose than before.
14. Solve equation 10.19} for V as a function of m, assuming that $V = 0$ and $m = m_0$ at $t = 0$. Hint: Since $R = -dm/dt$, we have $R/m = -d\ln(m)/dt$.
15. }
16. Bottles are filled with soft drink at a bottling plant as shown in Figure 1.1.12:. The bottles sit on a scale which is used to determine when to shut off the flow of soft drink. If the desired mass of the bottle plus soft drink after filling is M, what weight should the scale read when the bottle is full? The rate at which mass is being added to the bottle is R and its velocity entering the bottle is V.

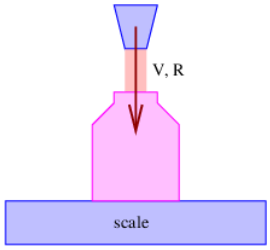


Figure 1.1.12:: A bottle being filled with soft drink at a rate R. The liquid enters the bottle with velocity V.

---

17. An interstellar space probe has frontal area A, initial mass M_0, and initial velocity V_0, which is non-relativistic. The tenuous gas between the stars has mass density ρ. These gas molecules stick to the probe when they hit it. Find the probe's acceleration. Hint: In a frame of reference in which the gas is stationary, does the momentum of the space probe change with time? Does its mass?
18. A light beam with power J hits a plate which is oriented normally to the beam. Compute the force required to hold the plate in place if
    1. the plate completely absorbs the light, and
    2. the plate completely reflects the light.

    Hint: Photons are massless, so the momentum of a photon with energy E is E/c. Thus, the momentum per unit time hitting the plate is J/c.
19. Find the acceleration of a rocket when the exhaust "gas" is actually a laser beam of power J. Assume that the rocket moves at non-relativistic velocities and that the decrease in mass due to the loss of energy in the laser beam is negligible.

Chapter 11 Rotational Dynamics

We have already seen the quantum mechanical treatment of angular momentum and rotational dynamics. In this section we study these subjects in a classical, non-relativistic context. We first define the concepts of torque and angular momentum in order to understand the orbital motion of a single particle. Next we examine two particles in arbitrary motion and learn how kinetic energy and angular momentum are partitioned between orbital and internal components. Two particles fixed to the ends of a light rod constitute a dumbbell, which serves as a prototype for the rotation of rigid bodies. We then see how what we learned for two particles extends to an arbitrary number of particles. Finally, we explore the physics of structures in static equilibrium.

Before we begin, we need to extend our knowledge of vectors to the cross product.

Tutorial — Cross Product

There are two ways to multiply two vectors together, the dot product and the cross product. We have already studied the dot product of two vectors, which results in a scalar or single number.

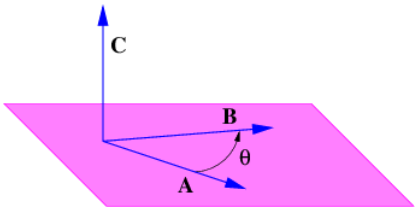


Figure 1.1.1:: Illustration of the cross product of two vectors A and B. The resulting vector C is perpendicular to the plane defined by A and B.

The cross product of two vectors results in a third vector, and is written symbolically as follows:

$$C = A \times B. \tag{1.1.247}$$

As illustrated in Figure 1.1.1:, the cross product of two vectors is perpendicular to the plane defined by these vectors. However, this doesn't tell us whether the resulting vector in Figure 1.1.1: points upward out of the plane or downward. This ambiguity is resolved using the **right-hand rule** :

1. Point the uncurled fingers of your right hand along the direction of the first vector A.

2. Rotate your arm until you can curl your fingers in the direction of the second vector B.
3. Your stretched out thumb now points in the direction of the cross product vector C.

The magnitude of the cross product is given by

$$|C| = |A||B|sin(\theta), \tag{1.1.248}$$

where $|A|$ and $|B|$ are the magnitudes of A and B, and θ is the angle between these two vectors. Note that the magnitude of the cross product is zero when the vectors are parallel or anti-parallel, and maximum when they are perpendicular. This contrasts with the dot product, which is maximum for parallel vectors and zero for perpendicular vectors.

Notice that the cross product does not commute, i. e., the order of the vectors is important. In particular, it is easy to show using the right-hand rule that

$$A \times B = -B \times A. \tag{1.1.249}$$

An alternate way to compute the cross product is most useful when the two vectors are expressed in terms of components, i. e., $A = (A_x, A_y.A_z)$ and $B = (B_x, B_y, B_z)$:

$$Cx = AyBz - AzByCy = AzBx - AxBzCz = AxBy - AyBx. \tag{1.1.250}$$

Notice that once you have the first of these equations, the other two can be obtained by cyclically permuting the indices, i. e., $x \rightarrow y$, $y \rightarrow z$, and $z \rightarrow x$. This is useful as a memory aid.

Torque and Angular Momentum

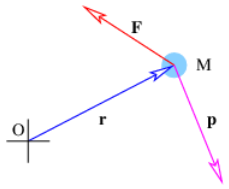


Figure 1.1.2:: A mass M located at r relative to the origin O has momentum p and has a force F applied to it. By the right-hand rule the torque $\tau = r \times F$ points out of the page, while the angular momentum $L = r \times p$ points into the page.

Torque is the action of a force F on a mass M which induces it to revolve about some point, called the **origin** . It is defined

$$\tau = r \times F, \tag{1.1.251}$$

where r is the position of the mass relative to the origin, as illustrated in Figure 1.1.2:.

Notice that the torque is zero in a number of circumstances. If the force points directly toward or away from the origin, the cross product is zero, resulting in zero torque, even though the force is non-zero. Likewise, if $r = 0$, the torque is zero. Thus, a force acting at the origin produces no torque. Both of these limits make sense intuitively, since neither induces the mass to revolve around the origin.

The **angular momentum** of a mass M relative to a point O is

$$L = r \times p, \tag{1.1.252}$$

where p is the ordinary kinetic momentum of the mass.[1] The angular momentum is zero if the motion of the object is directly towards or away from the origin, or if it is located at the origin.

If we take the cross product of the position vector and Newton's second law, we obtain an equation that relates torque and angular momentum:

$$dpddrr \times F = r \times ---= --(r \times p) ----\times p.dtdtdt \tag{1.1.253}$$

The second term on the right side of the above equation is zero because $d\,r\,/dt$ equals the velocity of the mass, which is parallel to its momentum and the cross product of two parallel vectors is zero. This equation can therefore be written

$$dL\tau = -dt(Newton'ssecondlawforrotation). \tag{1.1.254}$$

It is the rotational version of Newton's second law.

For both torque and angular momentum the location of the origin is arbitrary, and is generally chosen for maximum convenience. However, it is necessary to choose the same origin for both the torque and the angular momentum.

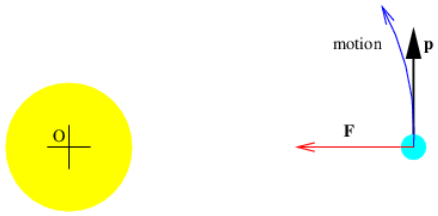


Figure 1.1.3:: A convenient choice of origin for a planet (right-hand sphere) revolving around the sun is simply the center of the sun. In this case the torque of the sun's gravitational force on the planet is zero.

For the case of a **central force** , i. e., one which acts along the line of centers between two objects (such as gravity), there often exists a particularly convenient choice of origin. Imagine a planet revolving around the sun, as illustrated in Figure 1.1.3:. If the origin is placed at the center of the sun (which is assumed not to move under the influence of the planet's gravity), then the torque exerted on the planet by the sun's gravity is zero, which means that the angular momentum of the planet about the center of the sun is constant in time. No other choice of origin would yield this convenient result.

We already know about two fundamental conservation laws — those of energy and linear momentum. We believe that angular momentum is similarly conserved in isolated systems. In other words, particles can exchange angular momentum between themselves, but the vector sum of the angular momentum of all the particles in a system isolated from outside influences must remain constant.

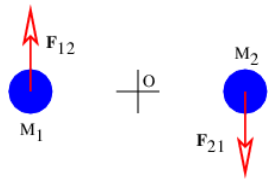


Figure 1.1.4:: Scenario for the non-conservation of angular momentum. F_{12} is the force of mass M_2 on mass M_1 and vice versa.

Conservation of angular momentum is not an automatic consequence of the conservation of linear momentum, even though the governing equation 11.8} for angular momentum is derived from Newton's second law. As an example, Figure 1.1.4: shows a hypothetical situation in which the force F_{21} of M_1 on M_2 is equal in magnitude but opposite in sign to the force F_{12} of M_2 on M_1, i. e., Newton's third law holds, and the sum of the momenta of the two masses is conserved. However, because the forces are non-central, the angular momentum of the masses is not conserved. This scenario is impossible if the forces are central.

Particles

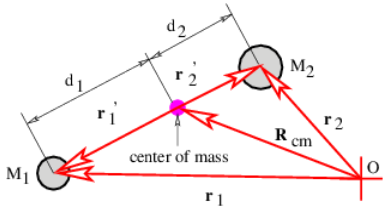


Figure 1.1.5:: Two particles of mass M_1 and M_2 with $M_2 > M_1$.

Suppose we wish to apply Newton's second law to two particles considered together as a single system. As we showed previously, only external forces act on the total momentum, $p_{total} = p_1 + p_2$, of the two particles:

$$F = dptotal.externaldt \tag{1.1.255}$$

Let's write the total non-relativistic momentum of the two particles in a special way:

$$(M1v1 + M2v2)ptotal = M1v1 + M2v2 \tag{1.1.256}$$
$$= Mtotal - - - - M - - - - - - - - - \equiv MtotalVcm, total$$

where $M_{total} = M_1 + M_2$. The quantity V_{cm} is the **velocity of the center of mass** and can be expressed as the time derivative of the **position of the center of mass** , R_{cm},

$$dRcmVcm = - - - - -, dt \tag{1.1.257}$$

where

$$M1r1 + M2r2Rcm = - - - - - - - - - - - - - - - - .Mtotal \tag{1.1.258}$$

We now see how the kinetic energy and the angular momentum of the two particles may be split into two parts, one having to do with the motion of the center of mass of the two particles, the other having to do with the motion of the two particles relative their center of mass. Figure 1.1.5: shows graphically how the vectors $r\,'_1 = r_1 - R_{cm}$ and $r\,'_2 = r_2 - R_{cm}$ are defined. These vectors represent the positions of the two particles relative to the center of mass. Substitution into equation 11.12} shows that $M_1 r\,'_1 + M_2 r\,'_2 = 0$. This leads to the conclusion that $M_1 d_1 = M_2 d_2$ in Figure 1.1.5:. We also define the velocity of each mass relative to the center of mass as $v\,'_1 = d\,r\,'_1/dt$ and $v\,'_2 = d\,r\,'_2/dt$, and we therefore have $M_1 v\,'_1 + M_2 v\,'_2 = 0$.

The total kinetic energy is just the sum of the kinetic energies of the two particles, $K = M_1 v_1^2/2 + M_2 v_2^2/2$, where v_1 and v_2 are the magnitudes of the corresponding velocity vectors. Substitution of $v_1 = V_{cm} + v\,'_1$ etc., into the kinetic energy formula and rearranging yields

$$2\,'2\,'2Ktotal = Ktrans + Kintern = [MtotalVcm/2] + [M1v1/2 + M2v2/2]. \tag{1.1.259}$$

Terms like $V_{cm} \cdot v\,'_1$ cancel out because $M_1 v\,'_1 + M_2 v\,'_2 = 0$.

The first term on the right side of equation 11.13} in square brackets is the kinetic energy the two particles would have if all of the mass were concentrated at the center of mass. The second term is the kinetic energy computed relative to the motion of the center of mass. The first is called the **translational kinetic energy** of the system while the second is called the **internal kinetic energy** .

The angular momentum of the system is just the sum of the angular momenta of the two particles: $L_{total} = M_1 r_1 \times v_1 + M_2 r_2 \times v_2$. By reasoning similar to the case of kinetic energy, we can rewrite this as

$$L = L + L = [MR \times V] + [Mr' \times v' + Mr' \times v'].totalorbspintotalcmcm111222 \tag{1.1.260}$$

The first term in square brackets on the right is called the **orbital angular momentum** while the second term is called the **spin angular momentum** . The former is the angular momentum the system would have if all the mass were concentrated at the center of mass, while the latter is the angular momentum of motion about the center of mass.

Interestingly, the idea of center of mass and the corresponding split of kinetic energy and angular momentum into orbital and spin parts has no useful relativistic generalization. This is due to the factor of $\gamma \equiv (1 - v^2/c^2)^{-1/2}$ in the relativistic definition of momentum, $p = m\,v\,\gamma$, which means that

$$dmr - - - - - / = p(relativisticcase).dt \tag{1.1.261}$$

Uneven Dumbbell

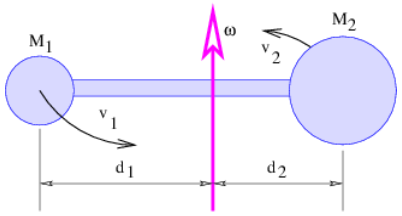


Figure 1.1.6:: Perspective view of the rotating dumbbell attached to an axle labeled ω. The axle attaches to the crossbar at the center of mass. The velocity v_1 is out of the page and v_2 is into the page.

So far we have put no restrictions on the movements of the two particles. An interesting special case occurs when the particles are connected by a lightweight, rigid rod, giving us a dumbbell. In order to further simplify things, we assume that the rod is connected rigidly to a fixed axle at the center of mass of the two particles, as shown in Figure 1.1.6:. The masses constituting the ends of the dumbbell are therefore free to revolve in circles about the axle, but they are prevented from executing any other motion. The key effect of this constraint is that both masses rotate about the axle with the same angular frequency ω.

If the particles are respectively distances d_1 and d_2 from the axle, then their speeds are $v_1 = d_1\omega$ and $v_2 = d_2\omega$. Thus the kinetic energy of the rotating dumbbell is

$$Kintern = 1M1v2 + 1 - M2v2 = 1I\omega2(fixedaxle), 21222 \tag{1.1.262}$$

where $I = M_1 d_1^2 + M_2 d_2^2$ is called the **moment of inertia** . Similarly, the magnitude of the spin angular momentum, which is a vector parallel to the axle, is

$$I_{spin} = M1d_1^2 + M2d_2^2 = I_{cm}(fixedaxle) \tag{1.1.263}$$

$$Lspin = M1d1v1 + M2d2v2 = I\omega(fixedaxle). \tag{1.1.263}$$

Finally, Newton's second law for rotation becomes

$$\tau = dLspin = dI\omega - = Id\omega - (fixedaxle), dtdtdt \tag{1.1.264}$$

where τ is the component of torque along the rotation axis.

Note that the rightmost expression in equation 11.18} assumes that I is constant, which only is true if d_1 and d_2 are constant – i. e., the dumbbell must truly be rigid.

Particles

The generalization from two particles to many particles is quite easy in principle. If a subscripted i indicates the value of a quantity for the **i** th particle, then the center of mass is given by

$$--1---\sum Rcm = MMiritotali \tag{1.1.265}$$

where

$$M = \sum M.totalii \tag{1.1.266}$$

Furthermore, if we define $r'_i = r_i$ - R_{cm}, etc., then the kinetic energy is just

$$K = MV2 / 2 + \sum Mv'2 / 2totaltotalcmiii \tag{1.1.267}$$

and the angular momentum is

$$\sum'' Ltotal = MtotalRcm \times Vcm + Miri \times vi.i \tag{1.1.268}$$

In other words, both the kinetic energy and the angular momentum can be separated into two parts: one part is related to the overall motion of the system and the other is due to motions of system components relative to the center of mass, just as for the case of the dumbbell.

Rigid Bodies

For a rigid body rotating about a fixed axle, the moment of inertia is

$$\sum I = Mid2i, i \tag{1.1.269}$$

where d_i is the perpendicular distance of the **i** th particle from the axle. Equations (11.16)-(11.18) are valid for a rigid body consisting of many particles. Furthermore, the moment of inertia is constant in this case, so it can be taken out of the time derivative:

$$dI\omega d\omega\tau = ----- = I --- = I\alpha(fixedaxle, constantI).dtdt \tag{1.1.270}$$

The quantity $\alpha = d\omega / dt$ is called the **angular acceleration** .

The sum in the equation for the moment of inertia can be converted to an integral for a continuous distribution of mass. We shall not pursue this here, but simply quote the results for a number of solid objects of uniform density:

- For rotation of a sphere of mass M and radius R about an axis piercing its center: $I = 2MR^2/5$.
- For rotation of a cylinder of mass M and radius R about its axis of symmetry: $I = MR^2/2$.
- For rotation of a thin rod of mass M and length L about an axis perpendicular to the rod passing through its center: $I = ML^2/12$.
- For rotation of an annulus of mass M, inner radius R_a, and outer radius R_b about its axis of symmetry: $I = M(R_a^2 + R_b^2)/2$.

Statics

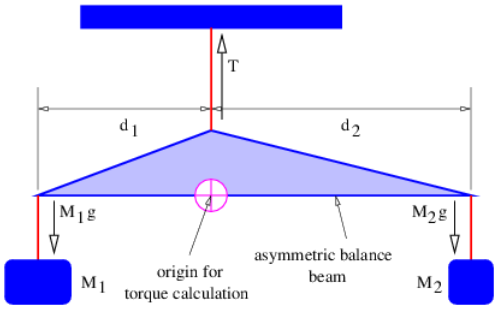


Figure 1.1.7:: Asymmetric mass balance. We assume that the balance beam is massless.

If a rigid body is initially at rest, it will remain at rest if and only if the sum of all the forces and the sum of all the torques acting on the body are zero. As an example, a mass balance with arms of differing length is shown in Figure 1.1.7:. The balance beam is subject to three forces pointing upward or downward, the tension T in the string from which the beam is suspended and the weights $M_1 g$ and $M_2 g$ exerted on the beam by the two suspended masses. The parameter g is the local gravitational field and the balance beam itself is assumed to have negligible mass. Taking upward as positive, the force condition for static equilibrium is

$$T - Mg - Mg = 0(zeronetforce).12 \tag{1.1.271}$$

Defining a counterclockwise torque to be positive, the torque balance computed about the pivot point in Figure 1.1.7: is

$$\tau = Mgd - Mgd = 0(zerotorque), 1122 \tag{1.1.272}$$

where d_1 and d_2 are the lengths of the beam arms.

The first of the above equations shows that the tension in the string must be

$$T = (M1 + M2)g, \tag{1.1.273}$$

while the second shows that

$$Md - - - 1 = -2.M2d1 \tag{1.1.274}$$

Thus, the tension in the string is just equal to the weight of the masses attached to the balance beam, while the ratio of the two masses equals the inverse ratio of the associated beam arm lengths.

Problems

1. Show using the component form of the cross product given by equation 11.4} that $A \times B$ = -$B \times A$.
2. }
3. A mass M is sliding on a frictionless table, but is attached to a string which passes through a hole in the center of the table as shown in Figure 1.1.8:. The string is gradually drawn in so the mass traces out a spiral pattern as shown in Figure 1.1.8:. The initial distance of the mass from the hole in the table is R and its initial tangential velocity is v. After the string is drawn in, the mass is a distance R' from the hole and its tangential velocity is v'.
    1. Given R, v, and R', find v'.
    2. Compute the change in the kinetic energy of the mass in going from radius R to radius R'.
    3. If the above change is non-zero, determine where the extra energy came from.

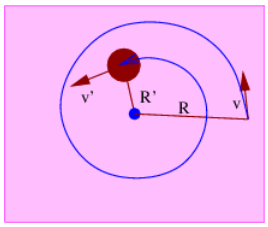


Figure 1.1.8:: Trajectory of a mass on a frictionless table attached to a string which passes through a hole in the table. The string is drawing the mass in.

---

4. A car of mass 1000 kg is heading north on a road at 30 m s^{-1} which passes 2 km east of the center of town.

1. Compute the angular momentum of the car about the center of town when the car is directly east of the town.
2. Compute the angular momentum of the car about the center of town when it is 3 km north of the above point.

5. The apparatus illustrated in Figure 1.1.9: is used to raise a bucket of mass M out of a well.
   1. What force F must be exerted to keep the bucket from falling back into the well?
   2. If the bucket is slowly raised a distance d, what work is done on the bucket by the rope attached to it?
   3. What work is done by the force F on the handle in the above case?

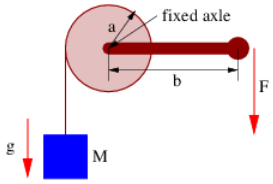


Figure 1.1.9:: A crank on a fixed axle turns a drum, thus winding the rope around the drum and raising the mass.

---

6. Derive equations (11.13) and (11.14).
7. A mass M is held up by the structure shown in Figure 1.1.10:. The support beam has negligible mass. Find the tension T in the diagonal wire. Hint: Compute the net torque on the support beam about point A due to the tension T and the weight of the mass M.

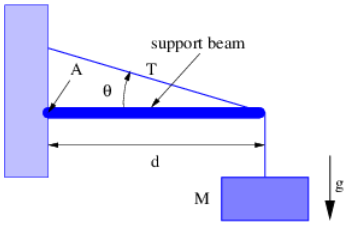


Figure 1.1.10:: A mass is supported by the tension in the diagonal wire. The support beam is free to pivot at point A.

---

8. A system consists of two stars, one of mass M moving with velocity $v_1 = (0, v, 0)$ at position $r_1 = (d, 0, 0)$, the other of mass 2 M with zero velocity at the origin.
   1. Find the center of mass position and velocity of the system of two stars.
   2. Find the spin angular momentum of the system.
   3. Find the internal kinetic energy of the system.

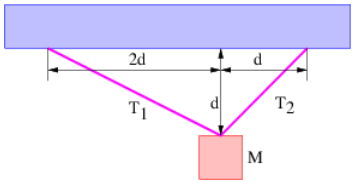


Figure 1.1.11:: A mass is supported by two strings.

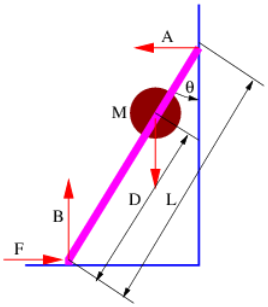


Figure 1.1.12:: A ladder leaning against a wall is held in place the force F acting on the base of the ladder.

---

9. A solid disk is rolling down a ramp tilted an angle θ from the horizontal. Compute the acceleration of the disk down the ramp and compare it with the acceleration of a block sliding down the ramp without friction.
10. A mass M is suspended from the ceiling by two strings as shown in Figure 1.1.11:. Find the tensions in the strings.
11. A man of mass M is a distance D up a ladder of length L which makes an angle θ with respect to the vertical wall as shown in Figure 1.1.12:. Take the mass of the ladder to be negligible. Find the force F needed to keep the ladder from sliding if the wall and floor are frictionless and therefore can only exert normal forces A and B on the ladder.

Chapter 12 Harmonic Oscillator

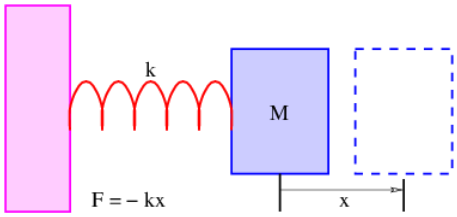


Figure 1.1.1:: Illustration of a mass-spring system.

Figure 1.1.1: illustrates the prototypical harmonic oscillator, the mass-spring system. A mass M is attached to one end of a spring. The other end of the spring is attached to something rigid such as a wall. The spring exerts a restoring force F = -kx on the mass when it is stretched by an amount x, i. e., it acts to return the mass to its initial position. This is called **Hooke's law** and k is called the **spring constant** .

Energy Analysis

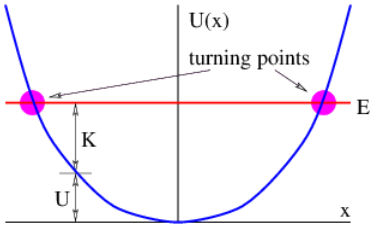


Figure 1.1.2:: Potential, kinetic, and total energy of a harmonic oscillator plotted as a function of spring displacement x.

The potential energy of the mass-spring system is

$$2U(x) = kx / 2 \tag{1.1.275}$$

which may be verified by noting that the Hooke's law force is derived from this potential energy: F = -d (kx ^{2}/ 2)/dx = -kx. This is shown in Figure 1.1.2:. Since a potential energy exists, the total energy $E = K + U$ is conserved, i. e., is constant in time. If the total energy is known, this provides a useful tool for determining how the kinetic energy varies with the position x of the mass M : K (x) = E - U (x). Since the kinetic energy is expressed (non-relativistically) in terms of the velocity u as $K = Mu^2/2$, the velocity at any point on the graph in Figure 1.1.2: is

$$()1 / 22(E - - - U -) - u = \pm M. \tag{1.1.276}$$

Given all this, it is fairly evident how the mass moves. From Hooke's law, the mass is always **accelerating** toward the **equilibrium position** , x = 0. However, at any point the **velocity** can be either to the left or the right. At the points where U (x) = E, the kinetic

energy is zero. This occurs at the turning points

$$(2E)1 / 2xTP = \pm - - -.k \tag{1.1.277}$$

If the mass is moving to the left, it slows down as it approaches the left turning point. It stops when it reaches this point and begins to move to the right. It accelerates until it passes the equilibrium position and then begins to decelerate, stopping at the right turning point, accelerating toward the left, etc. The mass thus oscillates between the left and right turning points. (Note that equations (12.2) and (12.3) are only true for the harmonic oscillator.)

How does the period of the oscillation depend on the total energy of the system? Notice that from equation 12.2} the maximum speed of the mass (i. e., the speed at x = 0) is equal to u_{max} = $(2E / M)^{1/2}$. The average speed must be some fraction of this maximum value. Let us guess here that it is half the maximum speed:

$$()1 / 2uaverage \approx umax - = -E - -(approximate).22M \tag{1.1.278}$$

However, the distance d the mass has to travel for one full oscillation is twice the distance between turning points, or d = $4(2E / k)^{1/2}$. Therefore, the period of oscillation must be approximately

$$\begin{aligned} d(2E)1 / 2(2M)1 / 2(M)1 / 2T = u - - - - - - - - \approx 4 - k - - E - - \\ = 8k - -(approximate).average \end{aligned} \tag{1.1.279}$$

Analysis Using Newton's Laws

The acceleration of the mass at any time is given by Newton's second law:

$$d2xFkxa = - - - 2 = - - - - = - - - - .dtMM \tag{1.1.280}$$

An equation of this type is known as a **differential equation** since it involves a derivative of the dependent variable x. Equations of this type are generally more difficult to solve than algebraic equations, as there are no universal techniques for solving all forms of such equations. In fact, it is fair to say that the solutions of most differential equations were originally obtained by **guessing** !

We already have the basis on which to make an intelligent guess for the solution to equation 12.6} since we know that the mass oscillates back and forth with a period that is independent of the amplitude of the oscillation. A function which might fill the bill is the cosine function. Let us try substituting x = cos(ωt), where ω is a constant, into this equation. The second derivative of x with respect to t is -ω^2 cos(ωt), so performing this substitution results in

$$2 - k - -\omega cos(\omega t) = -Mcos(\omega t). \tag{1.1.281}$$

Notice that the cosine function cancels out, leaving us with -ω^2 = -k / M. The guess thus works if we set

$$()1 / 2\omega = -k - .M \tag{1.1.282}$$

The constant ω is the angular oscillation frequency for the oscillator, from which we infer the period of oscillation to be T = 2π $(M / k)^{1/2}$. This agrees with the earlier approximate result of equation 12.5}, except that the approximation has a numerical factor of 8 rather than $2\pi \approx 6$. Thus, the earlier guess is only off by about 30%!

It is easy to show that x = B cos(ωt) is also a solution of equation 12.6}, where B is any constant and ω = $(k / M)^{1/2}$. This confirms that the oscillation frequency and period are independent of amplitude. Furthermore, the sine function is equally valid as a solution: x = A sin(ωt), where A is another constant. In fact, the most general possible solution is just a combination of these two, i. e.,

$$x = Asin(\omega t) + Bcos(\omega t) = Ccos(\omega t - \phi). \tag{1.1.283}$$

The values of A and B depend on the position and velocity of the mass at time t = 0. The right side of equation 12.9} shows an alternate way of writing the general harmonic oscillator solution that uses a cosine function with a phase factor ϕ.

Forced Oscillator

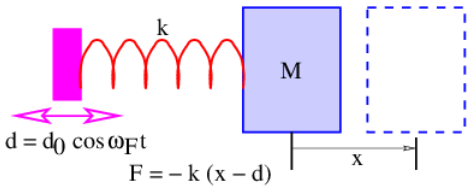


Figure 1.1.3:: Illustration of a forced mass-spring oscillator. The left end of the spring is wiggled back and forth with an angular frequency ω_F and a maximum amplitude d_0.

If we wiggle the left end of the spring by the amount $d = d_0 \cos(\omega_F t)$, as in Figure 1.1.3:, rather than rigidly fixing it as in Figure 1.1.1:, we have a **forced** harmonic oscillator. The constant d_0 is the amplitude of the imposed wiggling motion. The **forcing frequency** ω_F is not necessarily equal to the natural or **resonant** frequency $\omega = (k/M)^{1/2}$ of the mass-spring system. Very different behavior occurs depending on whether ω_F is less than, equal to, or greater than ω.

Given the above wiggling, the force of the spring on the mass becomes $F = -k(x - d) = -k[x - d_0 \cos(\omega_F t)]$ since the length of the spring is the difference between the positions of the left and right ends. Proceeding as for the unforced mass-spring system, we arrive at the differential equation

$$2d - -x + kx - = kd0 - cos(\omega Ft).dt2MM \tag{1.1.284}$$

The solution to this equation turns out to be the sum of a forced part in which x is proportional to $\cos(\omega_F t)$ and a free part which is the same as the solution to the unforced equation (12.9). We are primarily interested in the forced part of the solution, so let us set $x = x_0 \cos(\omega_F t)$ and substitute this into equation (12.10):

$$kx0kd0 - \omega 2Fx0cos(\omega Ft) + - - - - cos(\omega Ft) = - - - - cos(\omega Ft).MM \tag{1.1.285}$$

Again the cosine factor cancels and we are left with an algebraic equation for x_0, the amplitude of the oscillatory motion of the mass.

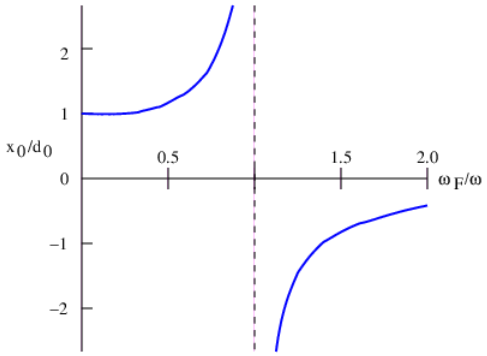


Figure 1.1.4:: Plot of the ratio of response to forcing vs. the ratio of forced to free oscillator frequency for the mass-spring system.

Solving for the ratio of the oscillation amplitude of the mass to the amplitude of the wiggling motion, x_0/d_0, we find

$$x0 - = - - - - - 1 - - - - -, d01 - \omega 2F / \omega 2 \tag{1.1.286}$$

where we have recognized that $k/M = \omega^2$, the square of the frequency of the **free** oscillation. This function is plotted in Figure 1.1.4:.

Notice that if $\omega_F < \omega$, the motion of the mass is in phase with the wiggling motion and the amplitude of the mass oscillation is greater than the amplitude of the wiggling. As the forcing frequency approaches the natural frequency of the oscillator, the response of the mass grows in amplitude. When the forcing is at the resonant frequency, the response is technically infinite, though practical limits on the amplitude of the oscillation will intervene in this case — for instance, the spring cannot stretch or shrink an infinite amount. In many cases friction will act to limit the response of the mass to forcing near the resonant frequency. When the forcing frequency is greater than the natural frequency, the mass actually moves in the opposite direction of the wiggling motion — i. e., the response is out of phase with the forcing. The amplitude of the response decreases as the forcing frequency increases above the resonant frequency.

Forced and free harmonic oscillators form an important part of many physical systems. For instance, any elastic material body such as a bridge or an airplane wing has harmonic oscillatory modes. A common engineering problem is to ensure that such modes are damped by friction or some other physical mechanism when there is a possibility of exitation of these modes by naturally occurring processes. A number of disasters can be traced to a failure to properly account for oscillatory forcing in engineered structures.

Complex Exponential Solutions

Complex exponential functions of the form $x = \exp(\pm i\omega t)$ also constitute solutions to the free harmonic oscillator governed by equation 12.6}. This makes sense, as the complex exponential is the sum of sines and cosines. However, for the frictionless harmonic oscillator, the exponential solutions provide no particular advantage over sines and cosines. Furthermore, oscillator displacements are real, not complex quantities.

The superposition principle solves the problem of complex versus real solutions. For an equation like (12.6) which has real coefficients, if exp($i\omega t$) is a solution, then so is exp(-$i\omega t$), so the superposition of these two solutions is also a solution. Furthermore

$$exp(i\omega t) + exp(-i\omega t) = 2cos(\omega t) = 2Re[exp(i\omega t)]. \tag{1.1.287}$$

This shows a shortcut for getting the physical part of a complex exponential solution to equations like the harmonic oscillator equation; simply take the real part.

Complex exponential solutions come into their own for more complicated equations. For instance, suppose the force on the mass in the mass-spring system takes the form

$$dxF = -kx - b - --.dt \tag{1.1.288}$$

The term containing b represents a frictional damping effect on the harmonic oscillator and the governing differential equation becomes

$$d2xbdxk - - - - + - - - - - - - + - - - x = 0.dt2MdtM \tag{1.1.289}$$

Trying the exponential function exp(σt) in this equation results in

$$\lfloor(2)1 / 2\rfloor\sigma = 1 - \lceil - - b - \pm - b - - - 4k - \rceil = -\beta \pm i(\omega 2 - \beta 2)1 / 22MM2M0 \tag{1.1.290}$$

where we have set

$$()1 / 2 - b - - - k - \beta = 2M\omega 0 = M. \tag{1.1.291}$$

The quantity $\omega \equiv (\omega_0^2 - \beta^2)^{1/2}$ is the actual frequency of oscillation of the damped oscillator, which one can see is less than the oscillation frequency ω_0 that occurs with the damping turned off. The physical solution to the damped oscillator is thus

$$x(t) = Re[exp(\sigma t)] = Re[exp(i\omega t)exp(-\beta t)] = cos(\omega t)exp(-\beta t) \tag{1.1.292}$$

as long as $\omega_0^2 > \beta^2$. Notice that this solution is in the form of an oscillation cos(ωt) multiplied by a decaying exponential exp(-βt). This confirms that the b term decreases the amplitude of the oscillation with time.

Quantum Mechanical Harmonic Oscillator

The quantum mechanical harmonic oscillator shares the characteristic of other quantum mechanical bound state problems in that the total energy can take on only discrete values. Calculation of these values is too difficult for this book, but the problem is sufficiently important to warrant reporting the results here. The energies accessible to a quantum mechanical mass-spring system are given by the formula

$$En = (n + 1 / 2)^{-}h(k / M)1 / 2, n = 0, 1, 2, ... \tag{1.1.293}$$

In other words, the energy **difference** between successive quantum mechanical energy levels in this case is constant and equals the classical resonant frequency for the oscillator, $\omega = (k / M)^{1/2}$, times $\hbar$.

Problems

1. An oscillator (non-harmonic) has the potential energy function $U(x) = Cx^4$, where C is a constant. How does the oscillation frequency depend on energy? Explain your reasoning.
2. Show that C cos(ωt - ϕ) is an alternate way of writing A sin(ωt) + B cos(ωt) by finding the values of A and B in terms of the constants C and ϕ. Hint: Expand cos(ωt -ϕ) by using the trigonometric identity for the cosine of the sum of two angles.
3. If a mass-spring harmonic oscillator has displacement x = 0 and velocity $dx / dt = V$ at time t = 0, determine the values of A and B as well as those of C and ϕ in equation 12.9}.
4. }
5. A mass M is suspended against gravity by a spring of spring constant k. The unstretched length of the spring is x_0 and under the influence of gravity the spring is stretched to a resting length $x_1 > x_0$.
    1. Compute the length of the spring x_1 in the steady, resting case.
    2. Set up the equation of motion for the mass moving under the influence of the two forces, gravity and spring. Solve the equation for the frequency of the oscillation and the position of the spring as a function of time x (t). Does the oscillation frequency change from the case without gravity?
    3. }

6. Determine the two real solutions to the damped harmonic oscillator problem in the case in which $\omega_0^2 < \beta^2$.

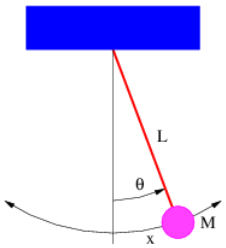


Figure 1.1.5:: The pendulum as a harmonic oscillator.

---

7. Consider the pendulum in Figure 1.1.5:. The mass M moves along an arc with x denoting the distance along the arc from the equilibrium point.
    1. Find the component of the gravitational force tangent to the arc (and thus in the direction of motion of the mass) as a function of the angle θ. Use the small angle approximation on $\sin(\theta)$ to simplify this answer.
    2. Get the force in terms of x rather than θ. (Recall that $\theta = x / L$.)
    3. Use Newton's second law for motion in the x direction (i. e., along the arc followed by the mass) to get the equation of motion for the mass.
    4. }
    5. Solve the equation of motion using the solution to the mass-spring problem as a guide.
    6. }
8. Forced damped oscillator:
    1. Add a damping term to the forced harmonic oscillator equation 12.10} and solve for the forced part of the solution using complex exponential methods. Hint: Change the cosine on the right side of this equation to $\exp(i\omega_F t)$ to convert the equation to complex form and then try the solution $x = x_0 \exp(i\omega_F t)$ where x_0 will depend on ω_F. Also, write the equation in terms of $\beta = b/(2M)$ and $\omega^2 = k/M$.
    2. }
    3. Find the physical part of this solution by taking the real part of $x(t)$. Hint: While taking the real part of x, it may be helpful to recall that the inverse of any complex number can be written $1/(a + ib) = (a - ib)/(a^2 + b^2)$.
    4. Determine how x_0 differs from that in the undamped case when ω_F is near the resonant frequency of the unforced oscillator. In particular, show how the phase and amplitude of the oscillation change as the forcing frequency changes from less than to greater than the resonant frequency.
9. A **massless** particle is confined to a box of length a. (Think of a photon between two mirrors.) Treating the particle classically, compute the period of one round trip from one end of the box to the other and back again. From this compute an angular frequency for the oscillation of this particle in the box. Does this frequency depend on the particle's energy?
10. Compute the ground state energy E_{ground} of a massless particle in a box of length a using quantum mechanics. Compare $E_{\text{ground}}/\hbar$ with the angular frequency computed in the previous problem.

This page titled 1.1: Transverse and Longitudinal Waves is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

1.2: Sine Waves

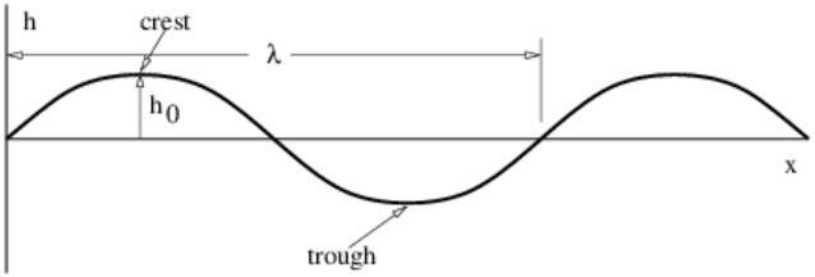


Figure 1.2.1: .2: Definition sketch for a sine wave, showing the wavelength λ and the amplitude h_0 and the phase ϕ at various points.

A particularly simple kind of wave, the sine wave, is illustrated in Figure 1.2.2:. This has the mathematical form

$$h(x) = h_0 \sin(2\pi x/\lambda) \tag{1.2.1}$$

where h is the displacement (which can be either longitudinal or transverse), h_0 is the maximum displacement, also called the amplitude of the wave, and λ is the wavelength. The oscillatory behavior of the wave is assumed to carry on to infinity in both positive and negative x directions. Notice that the wavelength is the distance through which the sine function completes one full cycle. The crest and the trough of a wave are the locations of the maximum and minimum displacements, as seen in Figure 1.2.2:.

So far we have only considered a sine wave as it appears at a particular time. All interesting waves move with time. The movement of a sine wave to the right a distance d may be accounted for by replacing x in the above formula by $x - d$. If this movement occurs in time t, then the wave moves at velocity $c = d / t$. Solving this for d and substituting yields a formula for the displacement of a sine wave as a function of both distance x and time t:

$$h(x,t) = h_0 \sin[2\pi(x - ct)/\lambda \tag{1.2.2}$$

The time for a wave to move one wavelength is called the period of the wave: $T = \lambda/c$. Thus, we can also write

$$h(x,t) = h_0 \sin[2\pi(x/\lambda - t/T)] \tag{1.2.3}$$

Physicists actually like to write the equation for a sine wave in a slightly simpler form. Defining the wavenumber as $k = 2\pi\lambda$ and the angular frequency as $\omega = 2\pi T$, we write

$$h(x,t) = h_0 \sin(kx - \omega t) \tag{1.2.4}$$

We normally think of the frequency of oscillatory motion as the number of cycles completed per second. This is called the rotational frequency, and is given by $f = 1/T$. It is related to the angular frequency by $\omega = 2\pi f$. The rotational frequency is usually easier to measure than the angular frequency, but the angular frequency tends to be used more often in theoretical discussions. As shown above, converting between the two is not difficult. Rotational frequency is measured in units of hertz, abbreviated Hz; $1\ \text{Hz} = 1\ \text{cycle s}^{-1}$. Angular frequency also has the dimensions of inverse time, e. g., radian s^{-1}, but the term “hertz” is generally reserved only for rotational frequency.

The argument of the sine function is by definition an angle. We refer to this angle as the phase of the wave, $\phi = kx - \omega t$. The difference in the phase of a wave at fixed time over a distance of one wavelength is 2π, as is the difference in phase at fixed position over a time interval of one wave period.

Since angles are dimensionless, we normally don’t include this in the units for frequency. However, it sometimes clarifies things to refer to the dimensions of rotational frequency as “rotations per second” or angular frequency as “radians per second”.

As previously noted, we call h_0, the maximum displacement of the wave, the amplitude. Often we are interested in the intensity of a wave, which is proportional to the square of the amplitude. The intensity is often related to the amount of energy being carried by a wave.

The wave speed we have defined above, $c = \lambda/T$, is actually called the phase speed. Since $\lambda = 2\pi k$ and $T = 2\pi/\omega$, we can write the phase speed in terms of the angular frequency and the wavenumber:

$$c = \frac{\omega}{\iota} \quad (\text{ phase speed }) \tag{1.2.5}$$

κ

This page titled 1.2: Sine Waves is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

1.3: Types of Waves

In order to make the above material more concrete, we now examine the characteristics of various types of waves which may be observed in the real world.

Ocean Surface Waves

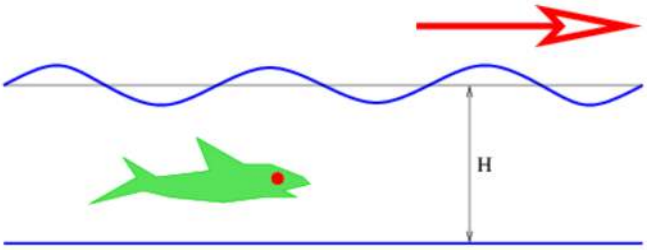


Figure 1.3.1: .3: Wave on an ocean of depth H. The wave is moving to the right and the particles of water at the surface oscillate in elliptical trajectories as the wave crests and troughs pass.

These waves are manifested as undulations of the ocean surface as seen in Figure 1.3.3:. The speed of ocean waves is given by the formula

$$c = \left(\frac{g \tanh(kH)}{k} \right)^{1/2} \tag{1.3.1}$$

where g = 9.8 m s^{-2} is the earth's gravitational force per unit mass, H is the depth of the ocean, and the hyperbolic tangent is defined as[1]

$$\tanh(x) = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)} \tag{1.3.2}$$

The equation for the speed of ocean waves comes from the theory for oscillations of a fluid surface in a gravitional field.

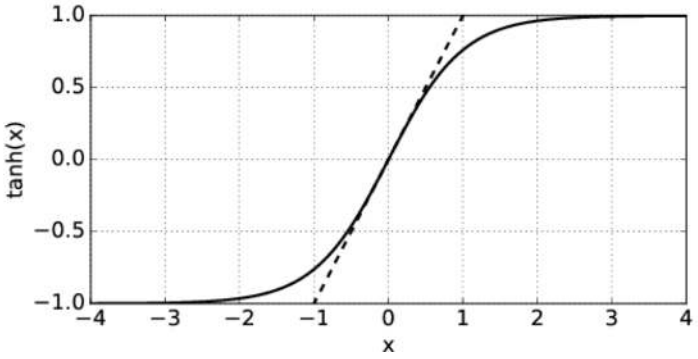


Figure 1.3.1: .4: Plot of the function $\tanh(x)$. The dashed line shows our approximation $\tanh(x) \approx x$ for $|x| < 1$.

As figure 1.4 shows, for $|x| \ll 1$, we can approximate the hyperbolic tangent by $tanh(x) \approx x$, while for $|x| \gg 1$ it is +1 for $x > 0$ and -1 for $x < 0$. This leads to two limits: Since $x = kH$, the shallow water limit, which occurs when $kH \ll 1$, yields a wave speed of

$$c \approx (gH)^{1/2}, \quad \text{(shallow water waves)} \tag{1.3.3}$$

while the deep water limit, which occurs when $kH \gg 1$, yields

$$c \approx (g/k)^{1/2}, \quad \text{(deep water waves)}. \tag{1.3.4}$$

Notice that the speed of shallow water waves depends only on the depth of the water and on *g*. In other words, all shallow water waves move at the same speed. On the other hand, deep water waves of longer wavelength (and hence smaller wavenumber) move more rapidly than those with shorter wavelength. Waves for which the wave speed varies with wavelength are called *dispersive*. Thus, deep water waves are dispersive, while shallow water waves are non-dispersive.

For water waves with wavelengths of a few centimeters or less, surface tension becomes important to the dynamics of the waves. In the deep water case, the wave speed at short wavelengths is given by the formula

$$c = (g/k + Ak)^{1/2} \tag{1.3.5}$$

where the constant A is related to an effect called surface tension. For an air-water interface near room temperature, $A \approx 74 \text{ cm}^3 \text{ s}^{-2}$

Sound Waves

Sound is a longitudinal compression-expansion wave in a fluid. The wave speed for sound in an ideal gas is

$$c = (\gamma R T_{abs})^{1/2} \tag{1.3.6}$$

where γ and R are constants and T_{abs} is the **absolute temperature** . The absolute temperature is measured in Kelvins and is numerically given by

$$T_{abs} = T_C + 273^\circ \tag{1.3.7}$$

where T_{C} is the temperature in Celsius degrees. The angular frequency of sound waves is thus given by

$$\omega = ck = (\gamma R T_{abs})^{1/2} k \tag{1.3.8}$$

The speed of sound in air at normal temperatures is about 340 m s^{-1}.

Light

Light moves in a vacuum at a speed of $c_{vac} = 3 \times 10^8 \text{ m s}^{-1}$. In transparent materials it moves at a speed less than c_{vac} by a factor n which is called the **refractive index** of the material:

$$c = c_{vac}/n \tag{1.3.9}$$

Often the refractive index takes the form

$$n^2 \approx 1 + \frac{A}{1 - (k/k_R)^2} \tag{1.3.10}$$

where k is the wavenumber and k_R and A are positive constants characteristic of the material. The angular frequency of light in a transparent medium is thus

$$\omega = kc = kc_{vac}/n \tag{1.3.11}$$

This page titled 1.3: Types of Waves is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

1.4: Superposition Principle

It is found empirically that as long as the amplitudes of waves in most media are small, two waves in the same physical location don't interact with each other. Thus, for example, two waves moving in the opposite direction simply pass through each other without their shapes or amplitudes being changed. When collocated, the total wave displacement is just the sum of the displacements of the individual waves. This is called the *superposition principle*. At sufficiently large amplitude the superposition principle often breaks down — interacting waves may scatter off of each other, lose amplitude, or change their form.

Interference is a consequence of the superposition principle. When two or more waves are superimposed, the net wave displacement is just the algebraic sum of the displacements of the individual waves. Since these displacements can be positive or negative, the net displacement can either be greater or less than the individual wave displacements. The former case, which occurs when both displacements are of the same sign, is called *constructive interference*, while *destructive interference* occurs when they are of opposite sign.

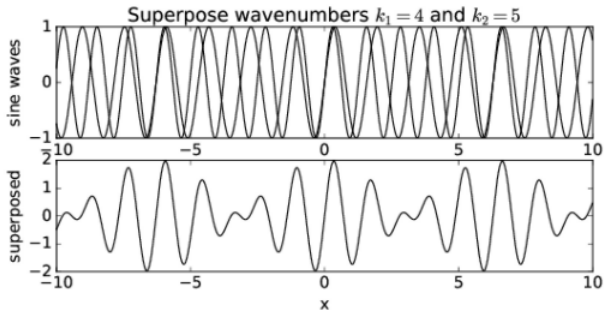


Figure 1.4.1: .5: Superposition (lower panel) of two sine waves (shown individually in the upper panel) with equal amplitudes and wavenumbers $k_1 = 4$ and $k_2 = 5$.

Let us see what happens when we superimpose two sine waves with different wavenumbers. Figure 1.4.5: shows the superposition of two waves with wavenumbers k_1 = 4 and k_2 = 5. Notice that the result is a wave with about the same wavelength as the two initial waves, but which varies in amplitude depending on whether the two sine waves are interfering constructively or destructively. We say that the waves are *in phase* if they are interfering constructively, and they are *out of phase* if they are interfering destructively.

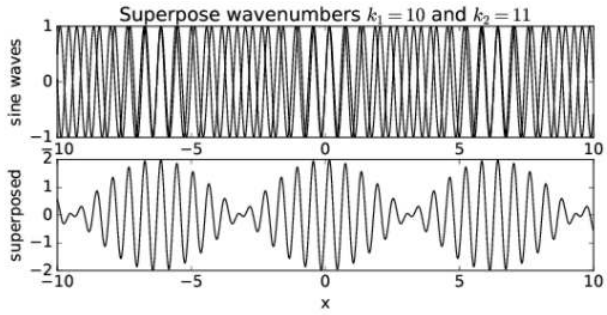


Figure 1.4.1: .6: Superposition of two sine waves with equal amplitudes and wavenumbers k_1 = 10 and k_2 = 11.

What happens when the wavenumbers of the two sine waves are changed? Figure 1.4.6: shows the result when k_1 = 10 and k_2 = 11. Notice that though the wavelength of the resultant wave is decreased, the locations where the amplitude is maximum have the same separation in x as in Figure 1.4.5:.

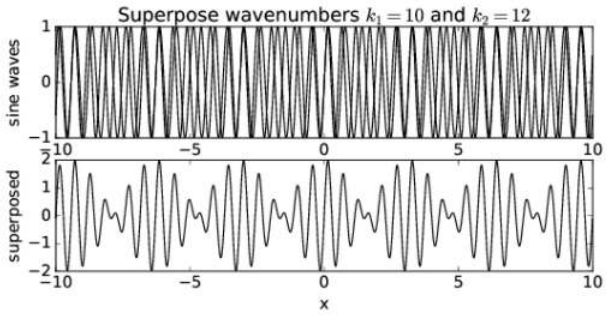


Figure 1.4.1: .7: Superposition of two sine waves with equal amplitudes and wavenumbers k_1 = 10 and k_2 = 12.

If we superimpose waves with k_1 = 10 and k_2 = 12, as is shown in Figure 1.4.7:, we see that the x spacing of the regions of maximum amplitude has decreased by a factor of two. Thus, while the wavenumber of the resultant wave seems to be related to something like the *average* of the wavenumbers of the component waves, the spacing between regions of maximum wave amplitude appears to go inversely with the *difference* of the wavenumbers of the component waves. In other words, if k_1 and k_2 are close together, the amplitude maxima are far apart and vice versa.

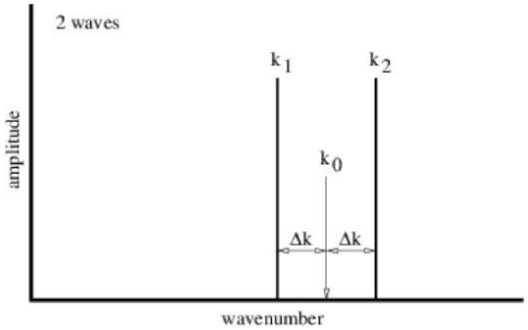


Figure 1.4.1: .8: Representation of the wavenumbers and amplitudes of two superimposed sine waves.

We can symbolically represent the sine waves that make up figures 1.5, 1.6, and 1.7 by a plot such as that shown in Figure 1.4.8:. The amplitudes and wavenumbers of each of the sine waves are indicated by vertical lines in this figure.

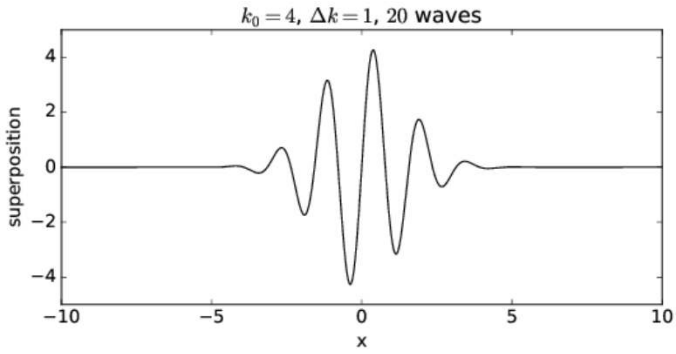


Figure 1.4.1: .9: Superposition of twenty sine waves with k_0 = 4 and Δk = 1.

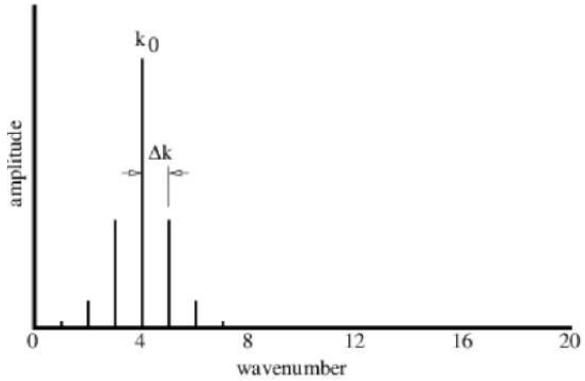


Figure 1.4.1: .10: Representation of the distribution of wavenumbers and amplitudes of 20 superimposed sine waves with maximum at $k_0 = 4$ and half-width $\Delta k = 1$.

The regions of large wave amplitude are called wave packets. Wave packets will play a central role in what is to follow, so it is important that we acquire a good understanding of them. The wave packets produced by only two sine waves are not well separated along the x-axis. However, if we superimpose many waves, we can produce an isolated wave packet. For example, Figure 1.4.9: shows the results of superimposing 20 sine waves with wavenumbers $k = 0.4m, m = 1, 2, \ldots, 20$, where the amplitudes of the waves are largest for wavenumbers near $k = 4$. In particular, we assume that the amplitude of each sine wave is proportional to $\exp\left[-(k-k_0)^2/\Delta k^2\right]$, where $k_0 = 4$ defines the maximum of the distribution of wavenumbers and $\Delta k = 1$ defines the half-width of this distribution. The amplitudes of each of the sine waves making up the wave packet in Figure 1.4.9: are shown schematically in Figure 1.4.10:.

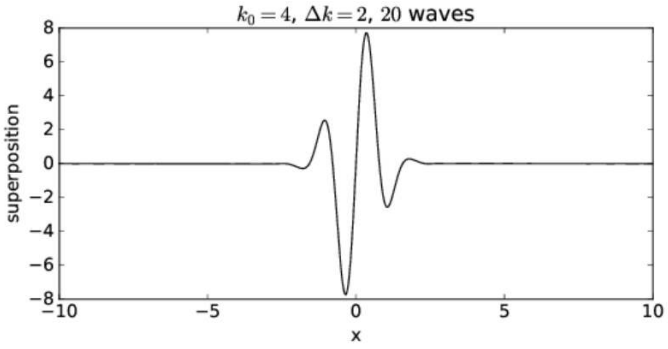


Figure 1.4.1: .11: Superposition of twenty sine waves with $k_0 = 4$ and $\Delta k = 2$.

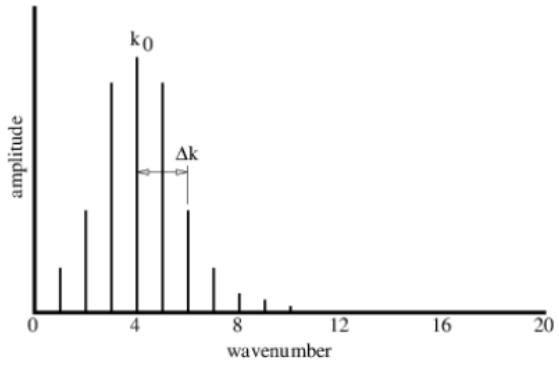


Figure 1.4.1: .12: Representation of the distribution of wavenumbers and amplitudes of 20 superimposed sine waves with maximum at $k_0 = 4$ and half-width $\Delta k = 2$.

The quantity Δk controls the distribution of the sine waves being superimposed — only those waves with a wavenumber k within approximately Δk of the central wavenumber k_0 of the wave packet, i. e., for $3 \leq k \leq 5$ in this case, contribute significantly to the sum. If Δk is changed to 2, so that wavenumbers in the range $2 \leq k \leq 6$ contribute significantly, the wavepacket becomes narrower, as is shown in figures 1.11 and 1.12. Δk is called the wavenumber spread of the wave packet, and it evidently plays a role similar to the difference in wavenumbers in the superposition of two sine waves — the larger the wavenumber spread, the smaller the

physical size of the wave packet. Furthermore, the wavenumber of the oscillations within the wave packet is given approximately by the central wavenumber.

We can better understand how wave packets work by mathematically analyzing the simple case of the superposition of two sine waves. Let us define $k_0 = (k_1 + k_2)/2$ where k_1 and k_2 are the wavenumbers of the component waves. Furthermore let us set $\Delta k = (k_2 - k_1)/2$. The quantities k_0 and Δk are graphically illustrated in Figure 1.4.8:. We can write $k_1 = k_0 - \Delta k$ and $k_2 = k_0 + \Delta k$ and use the trigonometric identity $sin(a+b) = sin(a)cos(b) + cos(a)sin(b)$ to find

$$\begin{aligned}\sin(k_1 x) + \sin(k_2 x) =& \sin[(k_0 - \Delta k)x] + \sin[(k_0 + \Delta k)x] \\ =& \sin(k_0 x)\cos(\Delta k x) - \cos(k_0 x)\sin(\Delta k x) + \\ & \sin(k_0 x)\cos(\Delta k x) + \cos(k_0 x)\sin(\Delta k x) \\ =& 2\sin(k_0 x)\cos(\Delta k x)\end{aligned} \tag{1.4.1}$$

The sine factor on the bottom line of the above equation produces the oscillations within the wave packet, and as speculated earlier, this oscillation has a wavenumber k_0 equal to the average of the wavenumbers of the component waves. The cosine factor modulates this wave with a spacing between regions of maximum amplitude of

$$\Delta x = \pi/\Delta k \tag{1.4.2}$$

Thus, as we observed in the earlier examples, the length of the wave packet Δx is inversely related to the spread of the wavenumbers Δk (which in this case is just the difference between the two wavenumbers) of the component waves. This relationship is central to the uncertainty principle of quantum mechanics.

This page titled 1.4: Superposition Principle is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

1.5: Beats

Suppose two sound waves of different frequency but equal amplitude impinge on your ear at the same time. The displacement perceived by your ear is the superposition of these two waves, with time dependence

$$h(t) = \sin(\omega_1 t) + \sin(\omega_2 t) = 2\sin(\omega_0 t)\cos(\Delta\omega t) \tag{1.5.1}$$

where we have used the above math trick, and where $\omega_0 = (\omega_1 + \omega_2)/2$ and $\Delta\omega = (\omega_2 - \omega_1)/2$. What you actually hear is a tone with angular frequency ω_0 which fades in and out with period

$$T_{\text{beat}} = \pi/|\Delta\omega| = 2\pi/|\omega_2 - \omega_1| = 1/|f_2 - f_1| \tag{1.5.2}$$

The *beat frequency* is simply

$$f_{\text{beat}} = 1/T_{\text{beat}} = |f_2 - f_1| \tag{1.5.3}$$

Note how beats are the time analog of wave packets — the mathematics are the same except that frequency replaces wavenumber and time replaces space.

This page titled 1.5: Beats is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

1.6: Interferometers

An interferometer is a device which splits a beam of light (or other wave) into two sub-beams, shifts the phase of one sub-beam with respect to the other, and then superimposes the sub-beams so that they interfere constructively or destructively, depending on the magnitude of the phase shift between them. In this section we study the Michelson interferometer and interferometric effects in thin films.

Michelson Interferometer

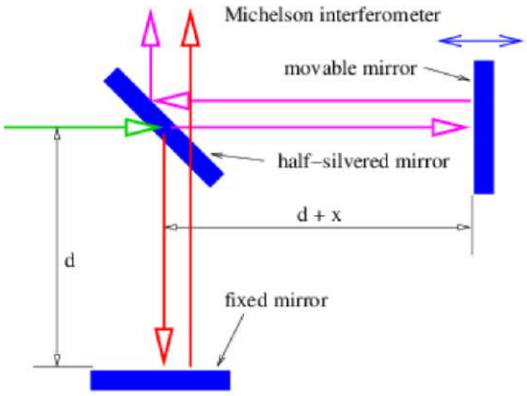


Figure 1.6.1: .13: Sketch of a Michelson interferometer.

The American physicist Albert Michelson invented the optical interferometer illustrated in Figure 1.6.13:. The incoming beam is split into two beams by the half-silvered mirror. Each sub-beam reflects off of another mirror which returns it to the half-silvered mirror, where the two sub-beams recombine as shown. One of the reflecting mirrors is movable by a sensitive micrometer device, allowing the path length of the corresponding sub-beam, and hence the phase relationship between the two sub-beams, to be altered. As Figure 1.6.13: shows, the difference in path length between the two sub-beams is 2x because the horizontal sub-beam traverses the path twice. Thus, constructive interference occurs when this path difference is an integral number of wavelengths, i. e.,

$$2x = m\lambda, \quad m = 0, \pm 1, \pm 2, \ldots \quad \text{(Michelson interferometer)} \tag{1.6.1}$$

where λ is the wavelength of the wave and m is an integer. Note that m is the number of wavelengths that fits evenly into the distance $2x$.

This page titled 1.6: Interferometers is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

1.7: Thin Films

One of the most revealing examples of interference occurs when light interacts with a thin film of transparent material such as a soap bubble. Figure 1.7.14: shows how a plane wave normally incident on the film is partially reflected by the front and rear surfaces. The waves reflected off the front and rear surfaces of the film interfere with each other. The interference can be either constructive or destructive depending on the phase difference between the two reflected waves.

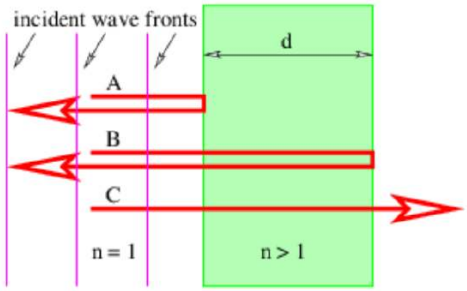


Figure 1.7.1: .14: Plane light wave normally incident on a transparent thin film of thickness d and index of refraction n > 1. Partial reflection occurs at the front surface of the film, resulting in beam A, and at the rear surface, resulting in beam B. Much of the wave passes completely through the film, as with C.

If the wavelength of the incoming wave is λ, one would naively expect constructive interference to occur between the A and B beams if $2d$ were an integral multiple of λ.

Two factors complicate this picture. First, the wavelength inside the film is **not** λ, but λ / n, where n is the index of refraction of the film. Constructive interference would then occur if $2d = m\lambda / n$. Second, it turns out that an additional phase shift of half a wavelength occurs upon reflection when the wave is incident on material with a higher index of refraction than the medium in which the incident beam is immersed. This phase shift doesn't occur when light is reflected from a region with lower index of refraction than felt by the incident beam. Thus beam B doesn't acquire any additional phase shift upon reflection. As a consequence, constructive interference actually occurs when

$$2d = (m + 1 / 2)\lambda / n, m = 0, 1, 2, \ldots \quad (\text{constructive interference}) \tag{1.7.1}$$

while destructive interference results when

$$2d = m\lambda / n, m = 0, 1, 2, \ldots \quad (\text{destructive interference}). \tag{1.7.2}$$

When we look at a soap bubble, we see bands of colors reflected back from a light source. What is the origin of these bands? Light from ordinary sources is generally a mixture of wavelengths ranging from roughly $\lambda = 4.5 \times 10^{-7}$ m (violet light) to $\lambda = 6.5 \times 10^{-7}$ m (red light). In between violet and red we also have blue, green, and yellow light, in that order. Because of the different wavelengths associated with different colors, it is clear that for a mixed light source we will have some colors interfering constructively while others interfere destructively. Those undergoing constructive interference will be visible in reflection, while those undergoing destructive interference will not.

Another factor enters as well. If the light is not normally incident on the film, the difference in the distances traveled between beams reflected off of the front and rear faces of the film will not be just twice the thickness of the film. To understand this case quantitatively, we need the concept of refraction, which will be developed later in the context of geometrical optics. However, it should be clear that different wavelengths will undergo constructive interference for different angles of incidence of the incoming light. Different portions of the thin film will in general be viewed at different angles, and will therefore exhibit different colors under reflection, resulting in the colorful patterns normally seen in soap bubbles.

1.7: Thin Films is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

1.8: Math Review — Derivatives

This section provides a quick review of the idea of the derivative. Often we are interested in the slope of a line tangent to a function $y(x)$ at some value of x. This slope is called the *derivative* and is denoted dy/dx. Since a tangent line to the function can be defined at any point x, the derivative itself is a function of x:

$$g(x) = \frac{dy(x)}{dx} \tag{1.8.1}$$

As Figure 1.8.15: illustrates, the slope of the tangent line at some point on the function may be approximated by the slope of a line connecting two points, A and B, set a finite distance apart on the curve:

$$\frac{dy}{dx} \approx \frac{\Delta y}{\Delta x} \tag{1.8.2}$$

As B is moved closer to A, the approximation becomes better. In the limit when B moves infinitely close to A, it is exact.

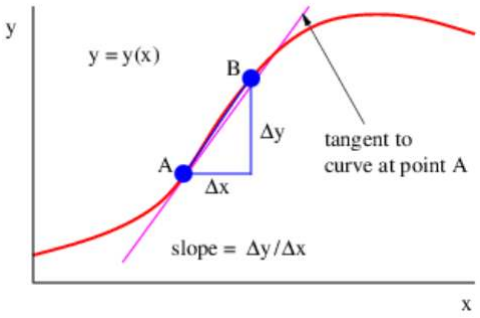


Figure 1.8.1: Estimation of the derivative, which is the slope of the tangent line. When point B approaches point A, the slope of the line AB approaches the slope of the tangent to the curve at point A.

Derivatives of some common functions are now given. In each case a is a constant.

$$\frac{dx^a}{dx} = ax^{a-1} \tag{1.8.3}$$

$$\frac{d}{dx}\exp(ax) = a\exp(ax) \tag{1.8.4}$$

$$\frac{d}{dx}\log(ax) = \frac{1}{x} \tag{1.8.5}$$

$$\frac{d}{dx}\sin(ax) = a\cos(ax) \tag{1.8.6}$$

$$\frac{d}{dx}\cos(ax) = -a\sin(ax) \tag{1.8.7}$$

$$\frac{daf(x)}{dx} = a\frac{df(x)}{dx} \tag{1.8.8}$$

$$\frac{d}{dx}[f(x)+g(x)] = \frac{df(x)}{dx} + \frac{dg(x)}{dx} \tag{1.8.9}$$

$$\frac{d}{dx}f(x)g(x) = \frac{df(x)}{dx}g(x) + f(x)\frac{dg(x)}{dx} \quad \text{(product rule)} \tag{1.8.10}$$

$$\frac{d}{dx}f(y) = \frac{df}{dy}\frac{dy}{dx} \quad \text{(chain rule)} \tag{1.8.11}$$

The product and chain rules are used to compute the derivatives of complex functions. For instance,

$\frac{d}{dx}(\sin(x)\cos(x)) = \frac{d\sin(x)}{dx}\cos(x) + \sin(x)\frac{d\cos(x)}{dx} = \cos^2(x) - \sin^2(x)$

and

$\frac{d}{dx}\log(\sin(x)) = \frac{1}{\sin(x)}\frac{d\sin(x)}{dx} = \frac{\cos(x)}{\sin(x)}.$

1.8: Math Review — Derivatives is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

1.9: Group Velocity

We now ask the following question: How fast do wave packets move? Surprisingly, we often find that wave packets move at a speed very different from the phase speed, $c = \omega/k$ of the wave composing the wave packet.

We shall find that the speed of motion of wave packets, referred to as the **group velocity**, is given by

$$u = \left.\frac{d\omega}{dk}\right|_{k=k_0} \quad \text{(group velocity)}. \tag{1.9.1}$$

The derivative of $\omega(k)$ with respect to k is first computed and then evaluated at $k = k_0$ the central wavenumber of the wave packet of interest.

The relationship between the angular frequency and the wavenumber for a wave, $\omega = \omega(k)$ depends on the type of wave being considered. Whatever this relationship turns out to be in a particular case, it is called the **dispersion relation** for the type of wave in question.

As an example of a group velocity calculation, suppose we want to find the velocity of deep ocean wave packets for a central wavelength of $\lambda_0 = 60$ m. This corresponds to a central wavenumber of $k_0 = 2\pi\lambda_0 \approx 0.1\ \text{m}^{-1}$. The phase speed of deep ocean waves is $c = (g/k)^{12}$. However, since $c \equiv \omega/k$, we find the frequency of deep ocean waves to be $\omega = (gk)^{1/2}$. The group velocity is therefore $u \equiv d\omega/dk = (g/k)^{12}/2 = c/2$. For the specified central wavenumber, we find that $u \approx (9.8\ \text{m s}^{-2}/0.1\ \text{m}^{-1})^{12}/2 \approx 5\ \text{m s}^{-1}$. By contrast, the phase speed of deep ocean waves with this wavelength is $c \approx 10\ \text{m s}^{-1}$.

Dispersive waves are waves in which the phase speed varies with wavenumber. It is easy to show that dispersive waves have unequal phase and group velocities, while these velocities are equal for non-dispersive waves.

Derivation of Group Velocity Formula

We now derive Equation 1.9.1. It is easiest to do this for the simplest wave packets, namely those constructed out of the superposition of just two sine waves. We will proceed by adding two waves with full space and time dependence:

$$h = \sin(k_1 x - \omega_1 t) + \sin(k_2 x - \omega_2 t) \tag{1.9.2}$$

After algebraic and trigonometric manipulations familiar from earlier sections, we find

$$h = 2\sin(k_0 x - \omega_0 t)\cos(\Delta k x - \Delta\omega t) \tag{1.9.3}$$

where as before we have $k_0 = (k_1 + k_2)\,2, \omega_0 = (\omega_1 + \omega_2)\,2, \Delta k = (k_2 - k_1)\,2,$ and $\Delta\omega = (\omega_2 - \omega_1)\,2$.

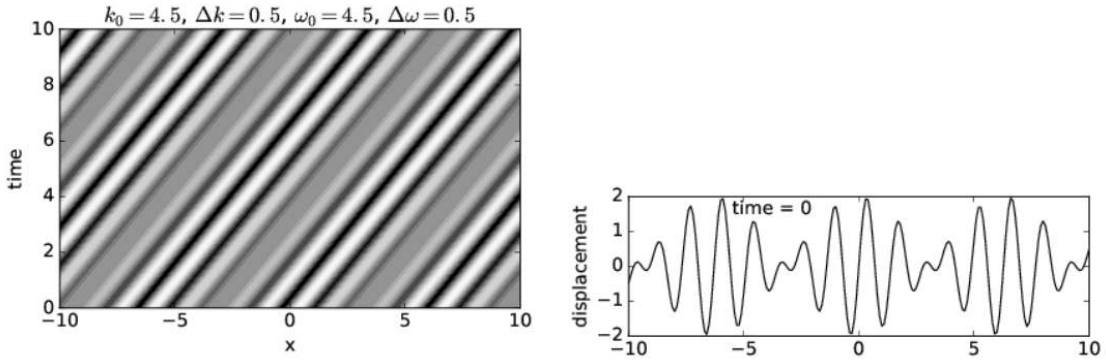


Figure 1.9.1: (left) Net displacement of the sum of two traveling sine waves plotted in the *x* - *t* plane. The white indicates where the displacement is large and positive, while the black indicates where it is large and negative. Non-dispersive case with $u_p = u_g = 1$.. (right) Plot of wave displacement as a function of x at time $t = 0$.

Again think of this as a sine wave of frequency ω_0 and wavenumber k_0 modulated by a cosine function. In this case the modulation pattern moves with a speed so as to keep the argument of the cosine function constant:

$$\Delta k x - \Delta\omega t = \text{ const.} \tag{1.9.4}$$

Differentiating this with respect to *t* while holding Δk and $\Delta\omega$ constant yields

$$u \equiv \frac{dx}{dt} = \frac{\Delta\omega}{\Delta k} \tag{1.9.5}$$

$$\omega \qquad \Delta k$$

In the limit in which the deltas become very small, this reduces to the derivative

$$u = \frac{d\omega}{dk}, \tag{1.9.6}$$

which is the desired result.

Examples

We now illustrate some examples of phase speed and group velocity by showing the displacement resulting from the superposition of two sine waves, as given by Equation 1.9.3, in the *x*-*t* plane. This is an example of a space time diagram, of which we will see many examples later on.

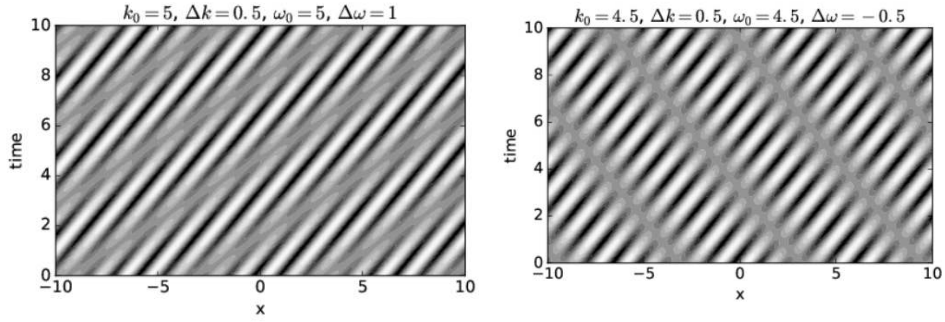


Figure 1.9.2: (A; left) As in the upper panel of Figure 1.9.16: except a dispersive case with u_p = 1, u_g = 2. (B; right) As in the upper panel of Figure 1.9.16: except a dispersive case with phase and group velocities in opposite directions, u_p = 1, u_g = -1.

The upper panel of Figure 1.9.2*A* shows a non-dispersive case in which the phase speed equals the group velocity. The white and black regions indicate respectively strong wave crests and troughs (i. e., regions of large positive and negative displacements), with grays indicating a displacement near zero. Regions with large displacements indicate the location of wave packets. The positions of waves and wave packets at any given time may therefore be determined by drawing a horizontal line across the graph at the desired time and examining the variations in wave displacement along this line. The lower panel of this figure shows the wave displacement as a function of \(*x*\) at time \(*t* = 0\) as an aid to interpretation of the upper panel.

Notice that as time increases, the crests move to the right. This corresponds to the motion of the waves within the wave packets. Note also that the wave packets, i. e., the broad regions of large positive and negative amplitudes, move to the right with increasing time as well.

Since velocity is distance moved Δ*x* divided by elapsed time Δ*t*, the slope of a line in Figure 1.9.2*A*, $\Delta t/\Delta x$, is one over the velocity of whatever that line represents. The slopes of lines representing crests are the same as the slopes of lines representing wave packets in this case, which indicates that the two move at the same velocity. Since the speed of movement of wave crests is the phase speed and the speed of movement of wave packets is the group velocity, the two velocities are equal and the non-dispersive nature of this case is confirmed.

Figure 1.9.2*A* shows a dispersive wave in which the group velocity is twice the phase speed, while Figure 1.9.2*B* shows a case in which the group velocity is actually opposite in sign to the phase speed. See if you can confirm that the phase and group velocities seen in each figure correspond to the values for these quantities calculated from the specified frequencies and wavenumbers.

1.9: Group Velocity is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

1.10: Problems

1. Measure your pulse rate. Compute the ordinary frequency of your heart beat in cycles per second. Compute the angular frequency in radians per second. Compute the period.
2. An important wavelength for radio waves in radio astronomy is 21 cm. (This comes from neutral hydrogen.) Compute the wavenumber of this wave. Compute the ordinary and angular frequencies. (The speed of light is 3×10^8 m s^{-1}.)
3. Sketch the resultant wave obtained from superimposing the waves $A = \sin(2x)$ and $B = \sin(3x)$. By using the trigonometric identity given in equation (1.17), obtain a formula for $A+B$ in terms of $\sin(5x2)$ and $\cos(x/2)$. Does the wave obtained from sketching this formula agree with your earlier sketch?
4. Two sine waves with wavelengths λ_1 and λ_2 are superimposed, making wave packets of length L. If we wish to make L larger, should we make λ_1 and λ_2 closer together or farther apart? Explain your reasoning.
5. By examining Figure 1.10.9: versus Figure 1.10.10: and then Figure 1.10.11: versus Figure 1.10.12:, determine whether equation (1.18) works at least in an approximate sense for isolated wave packets.
6. The frequencies of the chromatic scale in music are given by

$$f_i = f_0 2^{i/12}, \quad i = 0, 1, 2, \ldots, 11 \tag{1.10.1}$$

   where f_0 is a constant equal to the frequency of the lowest note in the scale.

   1. Compute f_1 through f_{11} if f_0 = 440 Hz (the "A" note)
   2. Using the above results, what is the beat frequency between the "A" (i = 0) and "B" (i = 2) notes? (The frequencies are given here in cycles per second rather than radians per second.)
   3. Which pair of the above frequencies f_0 - f_{11} yields the smallest beat frequency? Explain your reasoning.

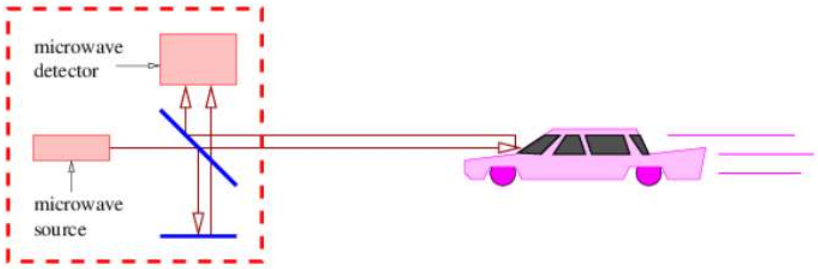


Figure 1.10.19**::** Sketch of a police radar.

7. Large ships in general cannot move faster than the phase speed of surface waves with a wavelength equal to twice the ship's length. This is because most of the propulsive force goes into making big waves under these conditions rather than accelerating the ship.
   1. How fast can a 300 m long ship move in very deep water?
   2. As the ship moves into shallow water, does its maximum speed increase or decrease? Explain.
8. Given the formula for refractive index of light quoted in this section, for what range of k does the phase speed of light in a transparent material take on real values which exceed the speed of light in a vacuum?
9. A police radar works by splitting a beam of microwaves, part of which is reflected back to the radar from your car where it is made to interfere with the other part which travels a fixed path, as shown in Figure 1.10.19:.
   1. If the wavelength of the microwaves is λ, how far do you have to travel in your car for the interference between the two beams to go from constructive to destructive to constructive?
   2. If you are traveling toward the radar at speed v = 30 m s^{-1}, use the above result to determine the number of times per second constructive interference peaks will occur. Assume that λ = 3 cm.

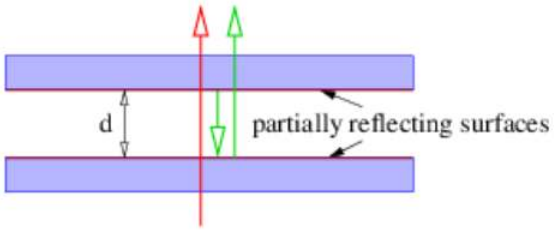


Figure 1.10.20**::** Sketch of a Fabry-Perot interferometer.

10. Suppose you know the wavelength of light passing through a Michelson interferometer with high accuracy. Describe how you could use the interferometer to measure the length of a small piece of material.
11. A Fabry-Perot interferometer (see Figure 1.10.20:) consists of two parallel half-silvered mirrors placed a distance d from each other as shown. The beam passing straight through interferes with the beam which reflects once off of both of the mirrored surfaces as shown. For wavelength λ, what values of d result in constructive interference?
12. A Fabry-Perot interferometer has spacing d = 2 cm between the glass plates, causing the direct and doubly reflected beams to interfere (see Figure 1.10.20:). As air is pumped out of the gap between the plates, the beams go through 23 cycles of constructive-destructive-constructive interference. If the wavelength of the light in the interfering beams is 5×10^{-7} m, determine the index of refraction of the air initially in the interferometer.
13. Measurements on a certain kind of wave reveal that the angular frequency of the wave varies with wavenumber as shown in the following table:

$\omega\,(\mathrm{s}^{-1})$	$k\,(\mathrm{m}^{-1})$
5	1
20	2
45	3
80	4
125	5

(1.10.2)

    1. Compute the phase speed of the wave for k = 3 m^{-1} and for k = 4 m^{-1}.
    2. Estimate the group velocity for k = 3.5 m^{-1} using a finite difference approximation to the derivative.
14. Suppose some type of wave has the (admittedly weird) dispersion relation shown in Figure 1.10.21:.
    1. For what values of k is the phase speed of the wave positive?
    2. For what values of k is the group velocity positive?

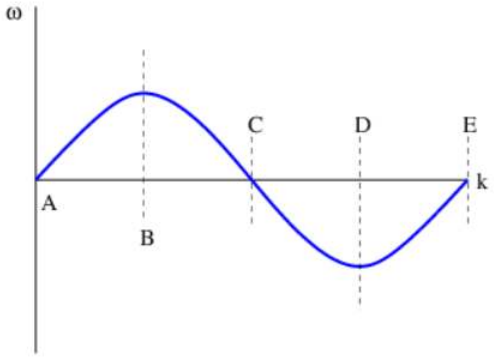


Figure 1.10.21**::** Sketch of a weird dispersion relation.

15. Group velocities of various waves.
    1. Compute the group velocity for shallow water waves. Compare it with the phase speed of shallow water waves. (Hint: You first need to derive a formula for $\omega(k)$ from $c(k)$.)
    2. Repeat the above problem for deep water waves.
    3. Repeat for sound waves. What does this case have in common with shallow water waves?

1.10: Problems is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

CHAPTER OVERVIEW

2: Waves in Two and Three Dimensions

In this chapter we extend the ideas of the previous chapter to the case of waves in more than one dimension. The extension of the sine wave to higher dimensions is the *plane wave*. Wave packets in two and three dimensions arise when plane waves moving in different directions are superimposed.

Diffraction results from the disruption of a wave which is impingent upon an object. Those parts of the wave front hitting the object are scattered, modified, or destroyed. The resulting *diffraction pattern* comes from the subsequent interference of the various pieces of the modified wave. A knowledge of diffraction is necessary to understand the behavior and limitations of optical instruments such as telescopes.

Diffraction and interference in two and three dimensions can be manipulated to produce useful devices such as the *diffraction grating*.

2.1: Math Tutorial — Vectors
2.2: Plane Waves
2.3: Superposition of Plane Waves
2.4: Diffraction Through a Single Slit
2.5: Two Slits
2.6: Diffraction Gratings
2.7: Problems

This page titled 2: Waves in Two and Three Dimensions is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

2.1: Math Tutorial — Vectors

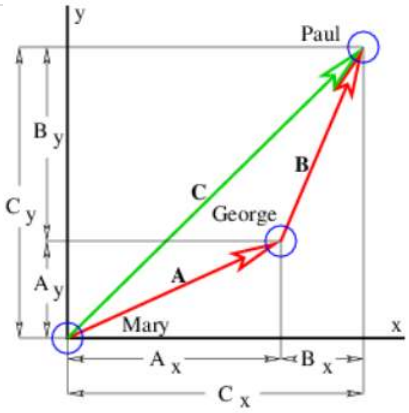


Figure 2.1.2: .1: Displacement vectors in a plane. Vector **A** represents the displacement of George from Mary, while vector **B** represents the displacement of Paul from George. Vector **C** represents the displacement of Paul from Mary and **C** = **A** + **B**. The quantities A_x, A_y, etc., represent the Cartesian components of the vectors.

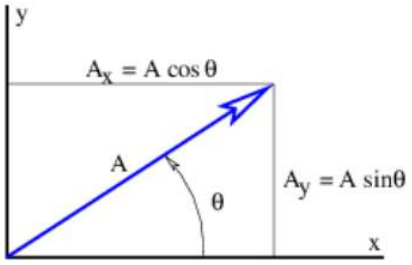


Figure 2.1.2: .2: Definition sketch for the angle θ representing the orientation of a two dimensional vector.

Before we can proceed further we need to explore the idea of a *vector*. A vector is a quantity which expresses both magnitude and direction. Graphically we represent a vector as an arrow. In typeset notation a vector is represented by a boldface character, while in handwriting an arrow is drawn over the character representing the vector.

Figure 2.1.1: shows some examples of *displacement vectors*, i. e., vectors which represent the displacement of one object from another, and introduces the idea of vector addition. The tail of vector **B** is collocated with the head of vector **A**, and the vector which stretches from the tail of **A** to the head of **B** is the sum of **A** and **B**, called **C** in Figure 2.1.1:.

The quantities A_x, A_y etc., represent the *Cartesian components* of the vectors in Figure 2.1.1:. A vector can be represented either by its Cartesian components, which are just the projections of the vector onto the Cartesian coordinate axes, or by its direction and magnitude. The direction of a vector in two dimensions is generally represented by the counterclockwise angle of the vector relative to the *x* axis, as shown in Figure 2.1.2:. Conversion from one form to the other is given by the equations

$$A = \left(A_x^2 + A_y^2\right)^{1/2} \quad \theta = \tan^{-1}\left(A_y/A_x\right) \tag{2.1.1}$$

$$A_x = A\cos(\theta) \quad A_y = A\sin(\theta) \tag{2.1.2}$$

where *A* is the magnitude of the vector. A vector magnitude is sometimes represented by absolute value notation: $A \equiv |\mathbf{A}|$.

Notice that the inverse tangent gives a result which is ambiguous relative to adding or subtracting integer multiples of π. Thus the quadrant in which the angle lies must be resolved by independently examining the signs of A_x and A_y and choosing the appropriate value of θ.

To add two vectors, **A** and **B**, it is easiest to convert them to Cartesian component form. The components of the sum **C** = **A** + **B** are then just the sums of the components:

$$C_x = A_x + B_x \quad C_y = A_y + B_y \tag{2.1.3}$$

Subtraction of vectors is done similarly, e. g., if **A** = **C** - **B**, then

$$A_x = C_x - B_x \quad A_y = C_y - B_y \tag{2.1.4}$$

A *unit vector* is a vector of unit length. One can always construct a unit vector from an ordinary (non-zero) vector by dividing the vector by its length: **n** = **A**/|**A**|. This division operation is carried out by dividing each of the vector components by the number in the denominator. Alternatively, if the vector is expressed in terms of length and direction, the magnitude of the vector is divided by the denominator and the direction is unchanged.

Unit vectors can be used to define a Cartesian coordinate system. Conventionally, **i**, **j**, and **k** indicate the *x*, *y*, and *z* axes of such a system. Note that **i**, **j**, and **k** are mutually perpendicular. Any vector can be represented in terms of unit vectors and its Cartesian components: $\mathbf{A} = A_x\mathbf{i} + A_y\mathbf{j} + A_z\mathbf{k}$. An alternate way to represent a vector is as a list of components: $\mathbf{A} = (A_x,A_y,A_z)$. We tend to use the latter representation since it is somewhat more economical notation.

There are two ways to multiply two vectors, yielding respectively what are known as the *dot product* and the *cross product*. The cross product yields another vector while the dot product yields a number. Here we will discuss only the dot product. The cross product will be presented later when it is needed.

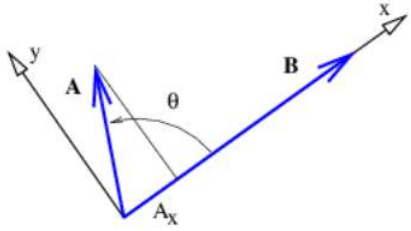


Figure 2.1.3:: Definition sketch for dot product.

Given vectors **A** and **B**, the dot product of the two is defined as

$$\mathbf{A} \cdot \mathbf{B} \equiv |\mathbf{A}||\mathbf{B}| \cos\theta \tag{2.1.5}$$

where θ is the angle between the two vectors. In two dimensions an alternate expression for the dot product exists in terms of the Cartesian components of the vectors:

$$\mathbf{A} \cdot \mathbf{B} = A_x B_x + A_y B_y \tag{2.1.6}$$

It is easy to show that this is equivalent to the cosine form of the dot product when the *x* axis lies along one of the vectors, as in Figure 2.1.3:. Notice in particular that $A_x = |\mathbf{A}| \cos\theta$, while $B_x = |\mathbf{B}|$ and $B_y = 0$. Thus, $\mathbf{A} \cdot \mathbf{B} = |\mathbf{A}| \cos\theta|\mathbf{B}|$ in this case, which is identical to the form given in equation (2.5).

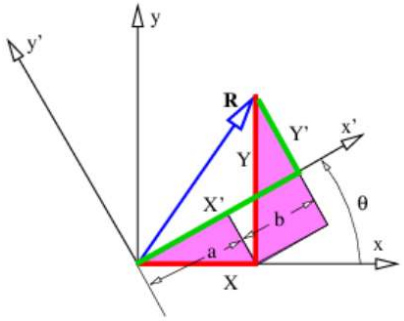


Figure 2.1.2: .4: Definition figure for rotated coordinate system. The vector **R** has components *X* and *Y* in the unprimed coordinate system and components *X*′ and *Y*′ in the primed coordinate system.

All that remains to be proven for Equation 2.1.6 to hold in general is to show that it yields the same answer regardless of how the Cartesian coordinate system is oriented relative to the vectors. To do this, we must show that $A_xB_x + A_yB_y = A_x'B_x' + A_y'B_y'$, where the primes indicate components in a coordinate system rotated from the original coordinate system.

Figure 2.1.4: shows the vector **R** resolved in two coordinate systems rotated with respect to each other. From this figure it is clear that $X' = a + b$. Focusing on the shaded triangles, we see that $a = X\cos\theta$ and $b = Y\sin\theta$. Thus, we find $X' = X\cos\theta + Y\sin\theta$. Similar reasoning shows that $Y' = -X\sin\theta + Y\cos\theta$. Substituting these and using the trigonometric identity $\cos^2\theta + \sin^2\theta = 1$ results in

$$\begin{aligned} A_x'B_x' + A_y'B_y' &= (A_x\cos\theta + A_y\sin\theta)(B_x\cos\theta + B_y\sin\theta) \\ &\quad + (-A_x\sin\theta + A_y\cos\theta)(-B_x\sin\theta + B_y\cos\theta) \\ &= A_{\cdot}B_{\cdot} + A_{\cdot}B_{\cdot} \end{aligned} \tag{2.1.7}$$

thus proving the complete equivalence of the two forms of the dot product as given by equations (2.5) and (2.6). Multiply out the above expression to verify this.

A numerical quantity that doesn't depend on which coordinate system is being used is called a *scalar*. The dot product of two vectors is a scalar. However, the components of a vector, taken individually, are not scalars, since the components change as the coordinate system changes. Since the laws of physics cannot depend on the choice of coordinate system being used, we insist that physical laws be expressed in terms of scalars and vectors, but not in terms of the components of vectors.

In three dimensions the cosine form of the dot product remains the same, while the component form is

$$\mathbf{A} \cdot \mathbf{B} = A_x B_x + A_y B_y + A_z B_z \tag{2.1.8}$$

This page titled 2.1: Math Tutorial — Vectors is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

2.2: Plane Waves

A plane wave in two or three dimensions is like a sine wave in one dimension except that crests and troughs aren't points, but form lines (2-D) or planes (3-D) perpendicular to the direction of wave propagation. Figure 2.2.1 shows a plane sine wave in two dimensions. The large arrow is a vector called the *wave vector*, which defines (1) the direction of wave propagation by its orientation perpendicular to the wave fronts, and (2) the wavenumber by its length.

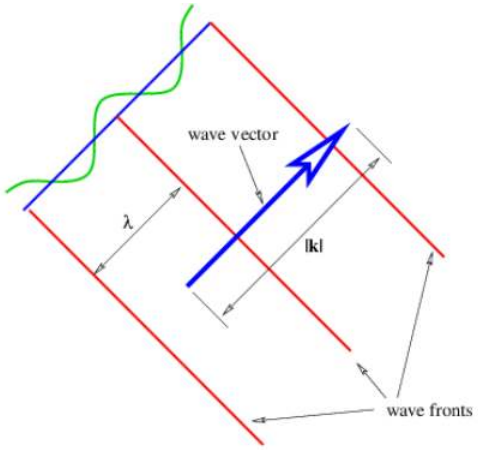


Figure 2.2.1: Definition sketch for a plane sine wave in two dimensions. The wave fronts are constant phase surfaces separated by one wavelength. The wave vector is normal to the wave fronts and its length is the wavenumber.

We can think of a wave front as a line along the crest of the wave. The equation for the displacement associated with a plane sine wave (of unit amplitude) in three dimensions at some instant in time is

$$h(x,y,z) = \sin(\mathbf{k}\cdot\mathbf{x}) = \sin\left(k_x x + k_y y + k_z z\right) \tag{2.2.1}$$

Since wave fronts are lines or surfaces of constant phase, the equation defining a wave front is simply $\mathbf{k}\cdot\mathbf{x} = \text{ const.}$

In the two dimensional case we simply set $k_z = 0$. Therefore, a wave front, or line of constant phase ϕ in two dimensions is defined by the equation

$$\mathbf{k}\cdot\mathbf{x} = k_x x + k_y y = \phi \quad \text{(two dimensions)}. \tag{2.2.2}$$

This can be easily solved for *y* to obtain the slope and intercept of the wave front in two dimensions.

As for one dimensional waves, the time evolution of the wave is obtained by adding a term -ωt to the phase of the wave. In three dimensions the wave displacement as a function of both space and time is given by

$$h(x,y,z,t) = \sin\left(k_x x + k_y y + k_z z - \omega t\right) \tag{2.2.3}$$

The frequency depends in general on all three components of the wave vector. The form of this function, $\omega = \omega\left(k_x, k_y k_z\right)$ which as in the one dimensional case is called the *dispersion relation*, contains information about the physical behavior of the wave.

Some examples of dispersion relations for waves in two dimensions are as follows:

- Light waves in a vacuum in two dimensions obey

$$\omega = c\left(k_x^2 + k_y^2\right)^{1/2} \tag{2.2.4}$$

  where *c* is the speed of light in a vacuum.
- Deep water ocean waves in two dimensions obey

$$\omega = g^{1/2}\left(k_x^2 + k_y^2\right)^{1/4} \quad \text{(ocean waves)} \tag{2.2.5}$$

  where *g* is the strength of the Earth's gravitational field as before.
- Certain kinds of atmospheric waves confined to a vertical *x*-*z* plane called *gravity waves* (not to be confused with the gravitational waves of general relativity) obey

$$\omega = \frac{Nk_x}{\quad} \quad \text{(gravity waves)}, \tag{2.2.6}$$

k_z

where N is a constant with the dimensions of inverse time called the **Brunt-Väisälä frequency**.

Contour plots of these dispersion relations are plotted in the upper panels of Figure 2.2.6:. These plots are to be interpreted like topographic maps, where the lines represent contours of constant elevation. In the case of Figure 2.2.6:, constant values of frequency are represented instead. For simplicity, the actual values of frequency are not labeled on the contour plots, but are represented in the graphs in the lower panels. This is possible because frequency depends only on wave vector magnitude $(k_x^2 + k_y^2)^{1/2}$ for the first two examples, and only on wave vector direction θ for the third.

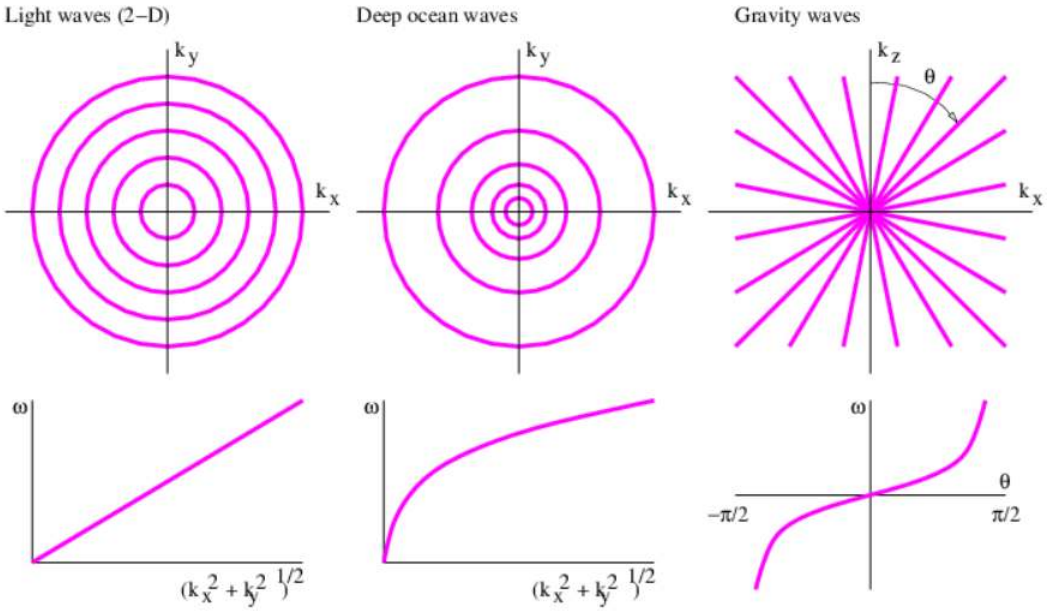


Figure 2.2.2: .6: Contour plots of the dispersion relations for three kinds of waves in two dimensions. In the upper panels the curves show lines or contours along which the frequency ω takes on constant values. Contours are drawn for equally spaced values of ω. For light and ocean waves the frequency depends only on the magnitude of the wave vector, whereas for gravity waves it depends only on the wave vector's direction, as defined by the angle θ in the upper right panel. These dependences for each wave type are illustrated in the lower panels.

This page titled 2.2: Plane Waves is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

2.3: Superposition of Plane Waves

We now study wave packets in two dimensions by asking what the superposition of two plane sine waves looks like. If the two waves have different wavenumbers, but their wave vectors point in the same direction, the results are identical to those presented in the previous chapter, except that the wave packets are indefinitely elongated without change in form in the direction perpendicular to the wave vector. The wave packets produced in this case move in the direction of the wave vectors and thus appear to a stationary observer like a series of passing pulses with broad lateral extent.

Superimposing two plane waves which have the same frequency results in a stationary wave packet through which the individual wave fronts pass. This wave packet is also elongated indefinitely in some direction, but the direction of elongation depends on the dispersion relation for the waves being considered. These wave packets are in the form of steady beams, which guide the individual phase waves in some direction, but don't themselves change with time. By superimposing multiple plane waves, all with the same frequency, one can actually produce a single stationary beam, just as one can produce an isolated pulse by superimposing multiple waves with wave vectors pointing in the same direction.

If the frequency of a wave depends on the magnitude of the wave vector, but not on its direction, the wave's dispersion relation is called isotropic; otherwise it is anisotropic. In the isotropic case, two waves have the same frequency only if the lengths of their wave vectors, and hence their wavelengths, are the same. The first two examples in Figure 2.3.6: satisfy this condition, while the last example is anisotropic.

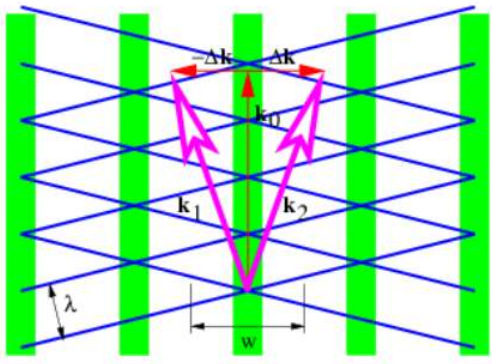


Figure 2.3.2: .7: Wave fronts and wave vectors ($\mathbf{k}_1$ and $\mathbf{k}_2$) of two plane waves with the same wavelength but oriented in different directions. The vertical bands show regions of constructive interference where wave fronts coincide. The vertical regions in between have destructive interference, and hence define the lateral boundaries of the beams produced by the superposition. The quantities $\mathbf{k}_0$ and $\Delta\mathbf{k}$ are also shown.

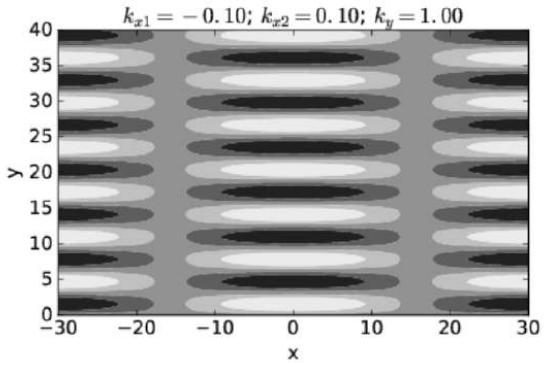


Figure 2.3.2: .8: Example of beams produced by two plane waves with the same wavelength moving in different directions. The wave vectors of the two waves are $\mathbf{k}$ = (±0.1, 1.0). Regions of positive displacement are lighter, while regions of negative displacement are darker.

We now use the language of vectors to investigate the superposition of two plane waves with wave vectors $\mathbf{k}_1$ and $\mathbf{k}_2$:

$$h = \sin(\mathbf{k}_1 \cdot \mathbf{x} - \omega t) + \sin(\mathbf{k}_2 \cdot \mathbf{x} - \omega t) \tag{2.3.1}$$

Applying the trigonometric identity for the sine of the sum of two angles (as we have done previously), equation (2.3.1) can be reduced to

$$h = 2\sin\left(\mathbf{k}_0 \cdot \mathbf{x} - \omega t\right)\cos(\Delta\mathbf{k} \cdot x) \tag{2.3.2}$$

where

$$\mathbf{k}_0 = \left(\mathbf{k}_1 + \mathbf{k}_2\right)/2 \quad \Delta\mathbf{k} = \left(\mathbf{k}_2 - \mathbf{k}_1\right)/2 \tag{2.3.3}$$

This is in the form of a sine wave moving in the $\mathbf{k}_0$ direction with phase speed $c_{\text{phase}} = \omega/|\mathbf{k}_0|$ and wavenumber $|\mathbf{k}_0|$, modulated in the $\Delta\mathbf{k}$ direction by a cosine function. The lines of destructive interference are normal to $\Delta\mathbf{k}$. The distance w between lines of destructive interference is the distance between successive zeros of the cosine function in equation (2.3.2), implying that $|\Delta\mathbf{k}|w = \pi$, which leads to

$$w = \pi/|\Delta\mathbf{k}| \tag{2.3.4}$$

Thus, the smaller $|\Delta\mathbf{k}|$ the greater is the beam diameter.

Waves of Identical Wavelength

In this section we investigate the beams produced by superimposing isotropic waves of the same frequency. Figure 2.3.7: illustrates what happens in such a superposition. Vectors $\mathbf{k}_1$ and $\mathbf{k}_2$ of equal length give rise to a mean wave vector $\mathbf{k}_0$ and half the difference, $\Delta\mathbf{k}$. As illustrated, the lines of constructive and destructive interference are perpendicular to $\Delta\mathbf{k}$. Figure 2.3.8: shows a concrete example of the beams produced by superposition of two plane waves of equal wavelength oriented as in Figure 2.3.7:. The beams are aligned vertically, since $\Delta\mathbf{k}$ is horizontal, with the lines of destructive interference separating the beams located near x = ±16. The transverse width of the beams of ≈ 32 satisfies equation (2.3.4) with $|\Delta\mathbf{k}|$ = 0.1. Each beam is made up of vertically propagating phase waves, with the crests and troughs indicated by the regions of white and black.

Waves of Differing Wavelength

In the third example of Figure 2.3.6:, the frequency of the wave depends only on the direction of the wave vector, independent of its magnitude, which is the reverse of the case for an isotropic dispersion relation. In this highly anisotropic case, different plane waves with the same frequency have wave vectors which point in the same direction, but have different lengths.

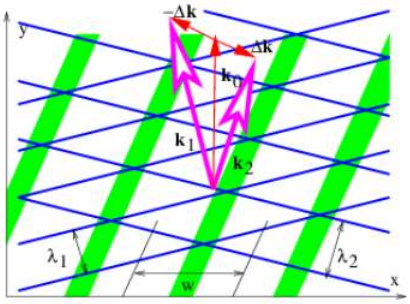


Figure 2.3.2: .9: Wave fronts and wave vectors ($\mathbf{k}_1$ and $\mathbf{k}_2$) of two plane waves with different wavelengths oriented in different directions. The slanted bands show regions of constructive interference where wave fronts coincide. The slanted regions in between have destructive interference, and as mentioned previously, define the lateral limits of the beams produced by the superposition. The quantities $\mathbf{k}_0$ and $\Delta\mathbf{k}$ are also shown.

More generally, one might have waves for which the frequency depends on both the direction and magnitude of the wave vector. In this case, two different plane waves with the same frequency would typically have wave vectors which differ both in direction and magnitude. Such an example is illustrated in figures 2.9 and 2.10.

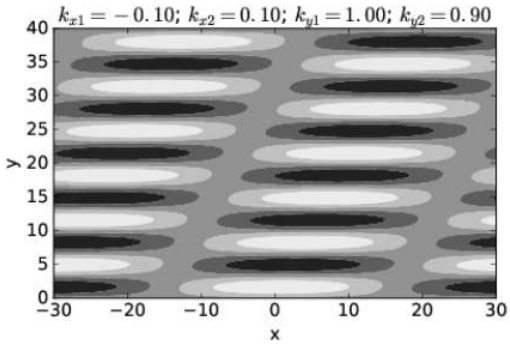


Figure 2.3.2: .10: Example of beams produced by two plane waves with wave vectors differing in both direction and magnitude. The wave vectors of the two waves are $\mathbf{k}_1$ = (-0.1, 1.0) and $\mathbf{k}_2$ = (0.1, 0.9). Regions of positive displacement are lighter, while regions of negative displacement darker.

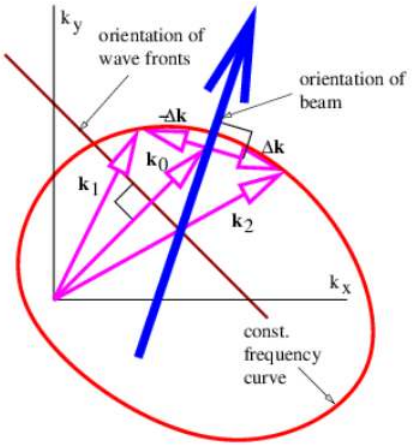


Figure 2.3.2: .11: Illustration of factors entering the addition of two plane waves with the same frequency. The wave fronts are perpendicular to the vector average of the two wave vectors, $\mathbf{k}_0 = (\mathbf{k}_1 + \mathbf{k}_2)/2$, while the lines of constructive interference, which define the beam orientation, are oriented perpendicular to the difference between these two vectors, $\Delta\mathbf{k} = (\mathbf{k}_2 - \mathbf{k}_1)/2$

Figure 2.3.11: summarizes what we have learned about adding plane waves with the same frequency. In general, the beam orientation (and the lines of constructive interference) are not perpendicular to the wave fronts. This only occurs when the wave frequency is independent of wave vector direction.

Waves with the Same Wavelength

As with wave packets in one dimension, we can add together more than two waves to produce an isolated wave packet. We will confine our attention here to the case of an isotropic dispersion relation in which all the wave vectors for a given frequency are of the same length.

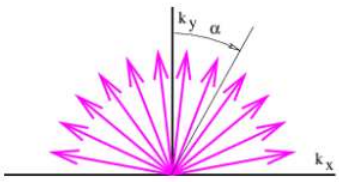


Figure 2.3.2: .12: Illustration of wave vectors of plane waves which might be added together.

Figure 2.3.12: shows an example of this in which wave vectors of the same wavelength but different directions are added together. Defining α_i as the angle of the i_{th} wave vector clockwise from the vertical, as illustrated in Figure 2.3.12:, we could write the superposition of these waves at time t = 0 as

$$\begin{aligned} h &= \sum_i h_i \sin\left(k_{xi}x + k_{yi}y\right) \\ &= \sum_i h_i \sin\left[kx\sin\left(\alpha_i\right) + ky\cos\left(\alpha_i\right)\right] \end{aligned} \tag{2.3.5}$$

where we have assumed that $k_{xi} = k\sin(\alpha_i)$ and $k_{yi} = k\cos(\alpha_i)$. The parameter $k = |\mathbf{k}|$ is the magnitude of the wave vector and is the same for all the waves. Let us also assume in this example that the amplitude of each wave component decreases with increasing $|\alpha_i|$:.

$$h_i = \exp\left[-(\alpha_i/\alpha_{\max})^2\right] \tag{2.3.6}$$

The exponential function decreases rapidly as its argument becomes more negative, and for practical purposes, only wave vectors with $|\alpha_i| \leq \alpha_{max}$ contribute significantly to the sum. We call $\alpha_{\max}$ the spreading angle.

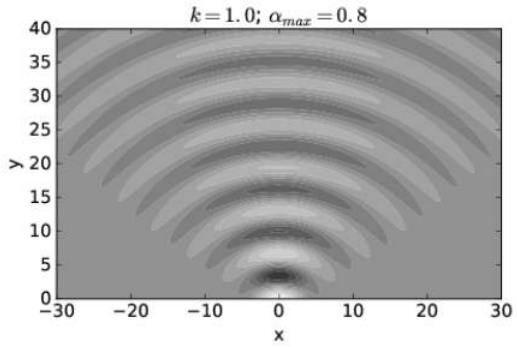


Figure 2.3.2: .13: Plot of the displacement field h (x,y) from equation (2.19) for α_{max} = 0.8 and k = 1.

Figure 2.3.13: shows what h(x,y) looks like when $\alpha_{\max} = 0.8$ radians and $k = 1$. Notice that for y = 0 the wave amplitude is only large for a small region in the range $-4 < x < 4$. However, for $y > 0$ the wave spreads into a broad, semicircular pattern.

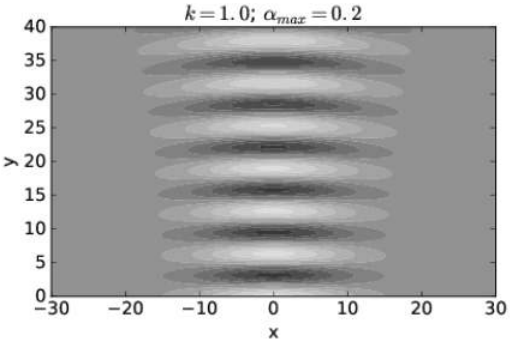


Figure 2.3.2: .14: Plot of the displacement field h (x,y) from equation (2.19) for $\alpha_{\max} = 0.2$ and $k = 1$.

Figure 2.3.14: shows the computed pattern of h(x,y) when the spreading angle $\alpha_{\max} = 0.2$ radians. The wave amplitude is large for a much broader range of x at $y = 0$ in this case, roughly $-12 < x < 12$. On the other hand, the subsequent spread of the wave is much smaller than in the case of Figure 2.3.13:.

We conclude that a superposition of plane waves with wave vectors spread narrowly about a central wave vector which points in the y direction (as in Figure 2.3.14:) produces a beam which is initially broad in x but for which the breadth increases only slightly with increasing y. However, a superposition of plane waves with wave vectors spread more broadly (as in Figure 2.3.13:) produces a beam which is initially narrow in x but which rapidly increases in width as y increases.

The relationship between the spreading angle $\alpha_{\max}$ and the initial breadth of the beam is made more understandable by comparison with the results for the two-wave superposition discussed at the beginning of this section. As indicated by equation (2.3.4), large values of k_x, and hence α, are associated with small wave packet dimensions in the x direction and vice versa. The superposition of two waves doesn't capture the subsequent spread of the beam which occurs when many waves are superimposed, but it does lead to a rough quantitative relationship between α_{max} (which is just $\tan^{-1}(k_t/k_y)$ in the two wave case) and the initial breadth of the beam. If we invoke the small angle approximation for α = α_{max} so that $\alpha_{\max} = \tan^{-1}(k_t/k_y) \approx k_t/k_y \approx k_t k$, then $k_x \approx k\alpha_{max}$ and equation (2.3.4) can be written $w = \pi k_x \approx \pi(k\alpha_{\max}) = \lambda(2\alpha_{\max})$. Thus, we can find the approximate spreading angle from the wavelength of the wave λ and the initial breadth of the beam w:

$$\alpha_{\max} \approx \lambda/(2w) \quad \text{(single slit spreading angle)} \tag{2.3.7}$$

This page titled 2.3: Superposition of Plane Waves is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

2.4: Diffraction Through a Single Slit

How does all of this apply to the passage of waves through a slit? Imagine a plane wave of wavelength λ impingent on a barrier with a slit. The barrier transforms the plane wave with infinite extent in the lateral direction into a beam with initial transverse dimensions equal to the width of the slit. The subsequent development of the beam is illustrated in figures 2.13 and 2.14, and schematically in Figure 2.4.15:. In particular, if the slit width is comparable to the wavelength, the beam spreads broadly as in Figure 2.4.13:. If the slit width is large compared to the wavelength, the beam doesn't spread as much, as Figure 2.4.14: illustrates. Equation (2.3.7) gives us an approximate quantitative result for the spreading angle if w is interpreted as the width of the slit.

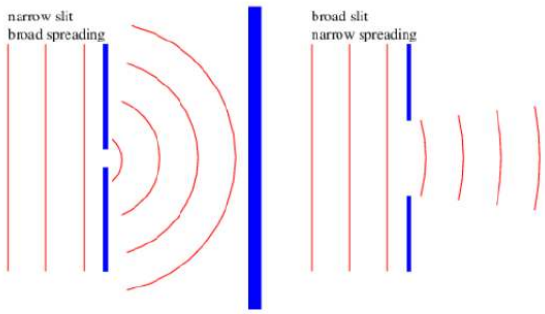


Figure 2.4.2: .15: Schematic behavior when a plane wave impinges on a narrow slit and a broad slit.

One use of the above equation is in determining the maximum angular resolution of optical instruments such as telescopes. The primary lens or mirror can be thought of as a rather large "slit". Light from a distant point source is essentially in the form of a plane wave when it arrives at the telescope. However, the light passed by the telescope is no longer a plane wave, but is a beam with a tendency to spread. The spreading angle α_{max} is given by equation (2.3.7), and the telescope cannot resolve objects with an angular separation less than α_{max}. Replacing w with the diameter of the lens or mirror in equation (2.3.7) thus yields the telescope's angular resolution. For instance, a moderate sized telescope with aperture 1 m observing red light with $\lambda \approx 6 \times 10^{-7}$ m has a maximum angular resolution of about 3×10^{-7} radians.

This page titled 2.4: Diffraction Through a Single Slit is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

2.5: Two Slits

Let us now imagine a plane sine wave normally impingent on a screen with two narrow slits spaced by a distance *d*, as shown in Figure 2.5.16:. Since the slits are narrow relative to the wavelength of the wave impingent on them, the spreading angle of the beams is large and the diffraction pattern from each slit individually is a cylindrical wave spreading out in all directions, as illustrated in Figure 2.5.13:. The cylindrical waves from the two slits interfere, resulting in oscillations in wave intensity at the screen on the right side of Figure 2.5.16:.

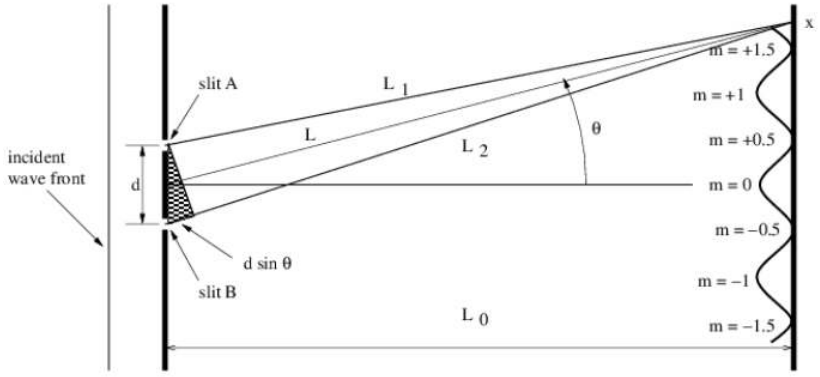


Figure 2.5.2: .16: Definition sketch for the double slit. Light passing through slit B travels an extra distance to the screen equal to $d \sin\theta$ compared to light passing through slit A.

Constructive interference occurs when the difference in the paths traveled by the two waves from their originating slits to the screen, $L_2 - L_1$ is an integer multiple *m* of the wavelength λ : $L_2 - L_1 = m\lambda$. If $L_0 \gg d$ the lines L_1 and L_2 are nearly parallel, which means that the narrow end of the dark triangle in Figure 2.5.16: has an opening angle of θ. Thus, the path difference between the beams from the two slits is $L_2 - L_1 = d \sin\theta$. Substitution of this into the above equation shows that constructive interference occurs when

$$d \sin\theta = m\lambda, \quad m = 0, \pm 1, \pm 2, \ldots \quad \text{(two slit interference)}. \tag{2.5.1}$$

Destructive interference occurs when *m* is an integer plus $1/2$. The integer *m* is called the *interference order* and is the number of wavelengths by which the two paths differ.

This page titled 2.5: Two Slits is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

2.6: Diffraction Gratings

Since the angular spacing $\Delta\theta$ of interference peaks in the two slit case depends on the wavelength of the incident wave, the two slit system can be used as a crude device to distinguish between the wavelengths of different components of a non-sinusoidal wave impingent on the slits. However, if more slits are added, maintaining a uniform spacing d between slits, we obtain a more sophisticated device for distinguishing beam components. This is called a diffraction grating.

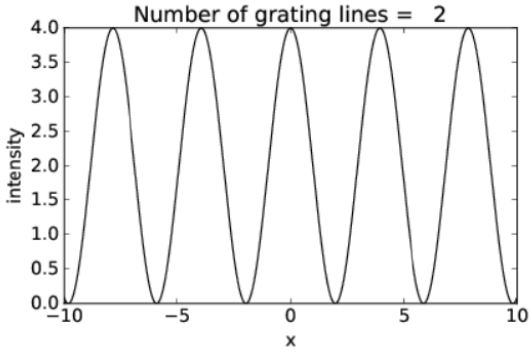


Figure 2.6.2: .17: Intensity of interference pattern from a diffraction grating with 2 slits on the screen in Figure 2.6.16:. The position x on the screen is proportional to the angle θ in the small angle approximation.

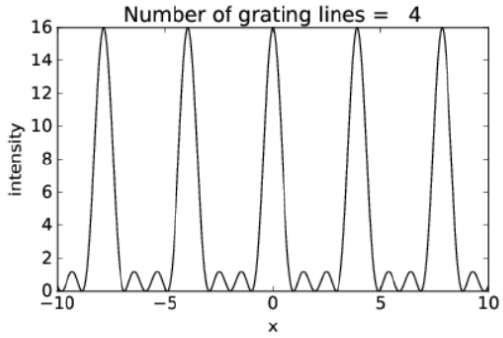


Figure 2.6.2: .18: Intensity of interference pattern from a diffraction grating with 4 slits.

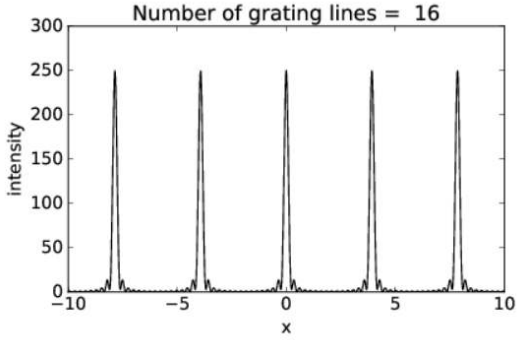


Figure 2.6.2: .19: Intensity of interference pattern from a diffraction grating with 16 slits.

Figures 2.17-2.19 show the intensity of the diffraction pattern as a function of position x on the display screen (see Figure 2.6.16:) for gratings with 2, 4, and 16 slits respectively, with the same slit spacing. Notice how the interference peaks remain in the same place but increase in sharpness as the number of slits increases.

The width of the peaks is actually related to the overall width of the grating, $w = nd$, where n is the number of slits. Thinking of this width as the dimension of a large single slit, the single slit equation, $\mathrm{a}_{\max} = \lambda/(2\mathrm{w})$, tells us the angular width of the peaks.

Whereas the angular width of the interference peaks is governed by the single slit equation, their angular positions are governed by the two slit equation. Let us assume for simplicity that $|\theta| < 1$ so that we can make the small angle approximation to the two slit

equation, $\mathrm{m}\lambda = \mathrm{d}\sin\theta \approx \mathrm{d}\theta$, and ask the following question: How different do two wavelengths differing by $\Delta\lambda$ have to be in order that the interference peaks from the two waves not overlap? In order for the peaks to be distinguishable, they should be separated in θ by an angle $\Delta\theta = m\Delta\lambda/d$, which is greater than the angular width of each peak, α_{max}:

$$\Delta\theta > \alpha_{\max} \tag{2.6.1}$$

Substituting in the above expressions for $\Delta\theta \text{ and } \alpha_{\max}$ and solving for $\Delta\lambda,\ \text{we get } \Delta\lambda > \lambda/(2mn)$, where λ is the average of the two wavelengths and $n = w\ /\ d$ is the number of slits in the diffraction grating. Thus, the fractional difference between wavelengths which can be distinguished by a diffraction grating depends solely on the interference order m and the number of slits n in the grating:

$$\frac{\Delta\lambda}{\lambda} > \frac{1}{2mn} \tag{2.6.2}$$

This page titled 2.6: Diffraction Gratings is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

2.7: Problems

1. Point A is at the origin. Point B is 3 m distant from A at 30˚ counterclockwise from the x axis. Point C is 2 m from point A at 100˚ counterclockwise from the x axis.
    1. Obtain the Cartesian components of the vector $\mathbf{D}_1$ which goes from A to B and the vector $\mathbf{D}_2$ which goes from A to C.
    2. Find the Cartesian components of the vector $\mathbf{D}_3$ which goes from B to C.

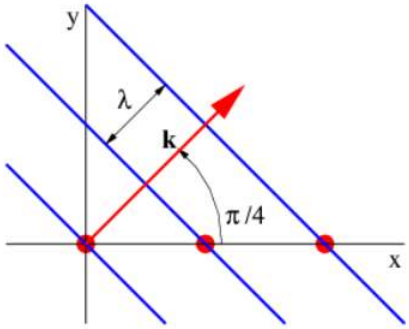
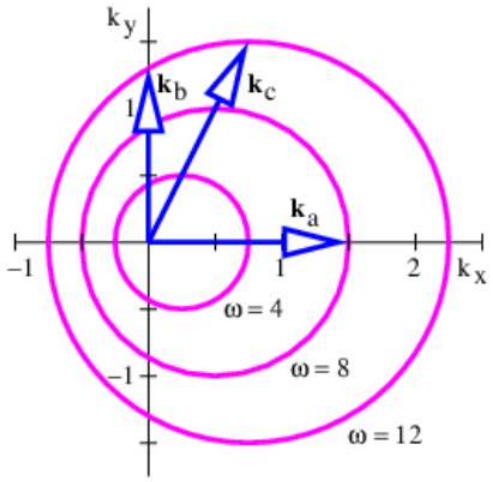


Figure 2.7.20**::** Sketch of wave moving at 45˚ to the *x*-axis.

    3. Find the direction and magnitude of $\mathbf{D}_3$.
2. For the vectors in the previous problem, find $\mathbf{D}_1 \cdot \mathbf{D}_2$ using both the cosine form of the dot product and the Cartesian form. Check to see if the two answers are the same.
3. Show graphically or otherwise that $|\mathbf{A} + \mathbf{B}| \neq |\mathbf{A}| + |\mathbf{B}|$ except when the vectors **A** and **B** are parallel.
4. A wave in the x-y plane is defined by $h = h_0 \sin(\mathbf{k} \cdot \mathbf{x})$ where $\mathbf{k} = (1, 2)$ cm^{-1}.
    1. On a piece of graph paper draw x and y axes and then plot a line passing through the origin which is parallel to the vector **k**.
    2. On the same graph plot the line defined by $\mathbf{k} \cdot \mathbf{x} = k_x x + k_y y = 0$, $\mathbf{k} \cdot \mathbf{x} = \pi$, and $\mathbf{k} \cdot \mathbf{x} = 2\pi$. Check to see if these lines are perpendicular to **k**.
5. A plane wave in two dimensions in the x - y plane moves in the direction 45˚ counterclockwise from the x-axis as shown in Figure 2.7.20:. Determine how fast the intersection between a wave front and the x-axis moves to the right in terms of the phase speed c of the wave. Hint: What is the distance between wave fronts along the x-axis compared to the wavelength?
6. Two deep plane ocean waves with the same frequency ω are moving approximately to the east. However, one wave is oriented a small angle β north of east and the other is oriented β south of east.
    1. Determine the orientation of lines of constructive interference between these two waves.
    2. Determine the spacing between lines of constructive interference.
7. An example of waves with a dispersion relation in which the frequency is a function of both wave vector magnitude and direction is shown graphically in Figure 2.7.21:.
    1. What is the phase speed of the waves for each of the three wave vectors? Hint: You may wish to obtain the length of each wave vector graphically.
    2. For each of the wave vectors, what is the orientation of the wave fronts?
    3. For each of the illustrated wave vectors, sketch two other wave vectors whose average value is approximately the illustrated vector, and whose tips lie on the same frequency contour line. Determine the orientation of lines of constructive interference produced by the superimposing pairs of plane waves for which each of the vector pairs are the wave vectors.
8. Two gravity waves have the same frequency, but slightly different wavelengths.
    1. If one wave has an orientation angle $\theta = \pi/4$ radians, what is the orientation angle of the other? (See Figure 2.7.6:.)
    2. Determine the orientation of lines of constructive interference between these two waves.
9. A plane wave impinges on a single slit, spreading out a half-angle α after the slit. If the whole apparatus is submerged in a liquid with index of refraction n = 1.5, how does the spreading angle of the light change? (Hint: Recall that the index of refraction in a transparent medium is the ratio of the speed of light in a vacuum to the speed in the medium. Furthermore, when light goes from a vacuum to a transparent medium, the light frequency doesn't change. Therefore, how does the wavelength of



the light change?) **Figure** 2.7.21**::** Graphical representation of the dispersion relation for shallow water waves in a river flowing in the *x* direction. Units of frequency are hertz, units of wavenumber are inverse meters.

10. Determine the diameter of the telescope needed to resolve a planet 2×10^8 km from a star which is 6 light years from the earth. (Assume blue light which has a wavelength $\lambda \approx 4 \times 10^{-7}$ m = 400 nm. Also, don't worry about the great difference in brightness between the two for the purposes of this problem.)
11. A laser beam from a laser on the earth is bounced back to the earth by a corner reflector on the moon.
    1. Engineers find that the returned signal is stronger if the laser beam is initially spread out by the beam expander shown in Figure 2.7.22:. Explain why this is so.
    2. The beam has a diameter of 1 m leaving the earth. How broad is it when it reaches the moon, which is 4×10^5 km away? Assume the wavelength of the light to be 5×10^{-7} m.
    3. How broad would the laser beam be at the moon if it weren't initially passed through the beam expander? Assume its initial diameter to be 1 cm.
12. Suppose that a plane wave impinges on two slits in a barrier at an angle, such that the phase of the wave at one slit lags the phase at the other slit by half a wavelength. How does the resulting interference pattern change from the case in which there is no lag?
13. Suppose that a thin piece of glass of index of refraction n = 1.33 is placed in front of one slit of a two slit diffraction setup.
    1. How thick does the glass have to be to slow down the incoming wave so that it lags the wave going through the other slit by a phase difference of π? Take the wavelength of the light to be $\lambda = 6 \times 10^{-7}$ m.
    2. For the above situation, describe qualitatively how the diffraction pattern changes from the case in which there is no glass in front of one of the slits. Explain your results.

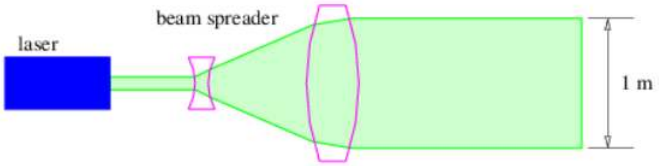


Figure 2.7.22**::** Sketch of a beam expander for a laser.

14. A light source produces two wavelengths, $\lambda_1 = 400$ nm (blue) and $\lambda_2 = 600$ nm (red).
    1. Qualitatively sketch the two slit diffraction pattern from this source. Sketch the pattern for each wavelength separately.
    2. Qualitatively sketch the 16 slit diffraction pattern from this source, where the slit spacing is the same as in the two slit case.
15. A light source produces two wavelengths, $\lambda_1 = 631$ nm and $\lambda_2 = 635$ nm. What is the minimum number of slits needed in a grating spectrometer to resolve the two wavelengths? (Assume that you are looking at the first order diffraction peak.) Sketch the diffraction peak from each wavelength and indicate how narrow the peaks must be to resolve them.

2.7: Problems is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

CHAPTER OVERVIEW

3: Geometrical Optics

As was shown previously, when a plane wave is impingent on an aperture which has dimensions much greater than the wavelength of the wave, diffraction effects are minimal and a segment of the plane wave passes through the aperture essentially unaltered. This plane wave segment can be thought of as a wave packet, called a *beam* or *ray*, consisting of a superposition of wave vectors very close in direction and magnitude to the central wave vector of the wave packet. In most cases the ray simply moves in the direction defined by the central wave vector, i. e., normal to the orientation of the wave fronts. However, this is not true when the medium through which the light propagates is optically anisotropic, i. e., light traveling in different directions moves at different phase speeds. An example of such a medium is a calcite crystal. In the anisotropic case, the orientation of the ray can be determined once the dispersion relation for the waves in question is known, by using the techniques developed in the previous chapter.

If light moves through some apparatus in which all apertures are much greater in dimension than the wavelength of light, then we can use the above rule to follow rays of light through the apparatus. This is called the *geometrical optics* approximation.

3.1: Reflection and Refraction
3.2: Total Internal Reflection
3.3: Anisotropic Media
3.4: Thin Lens Equation and Optical Instruments
3.5: Fermat's Principle
3.6: Problems

This page titled 3: Geometrical Optics is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

3.1: Reflection and Refraction

Most of what we need to know about geometrical optics can be summarized in two rules, the laws of reflection and refraction. These rules may both be inferred by considering what happens when a plane wave segment impinges on a flat surface. If the surface is polished metal, the wave is reflected, whereas if the surface is an interface between two transparent media with differing indices of refraction, the wave is partially reflected and partially refracted. Reflection means that the wave is turned back into the half-space from which it came, while refraction means that it passes through the interface, acquiring a different direction of motion from that which it had before reaching the interface.

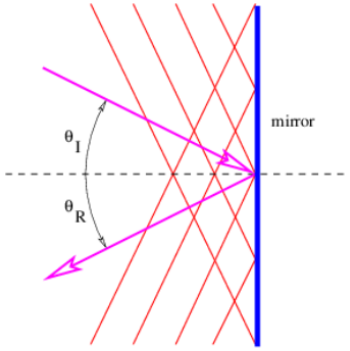


Figure 3.1.1: Sketch showing the reflection of a wave from a plane mirror. The law of reflection states that $\theta_I = \theta_R$

Figure 3.1.1: shows the wave vector and wave front of a wave being reflected from a plane mirror. The angles of incidence, θ_I, and reflection, θ_R, are defined to be the angles between the incoming and outgoing wave vectors respectively and the line normal to the mirror. The law of reflection states that $\theta_R = \theta_I$. This is a consequence of the need for the incoming and outgoing wave fronts to be in phase with each other all along the mirror surface. This plus the equality of the incoming and outgoing wavelengths is sufficient to insure the above result.

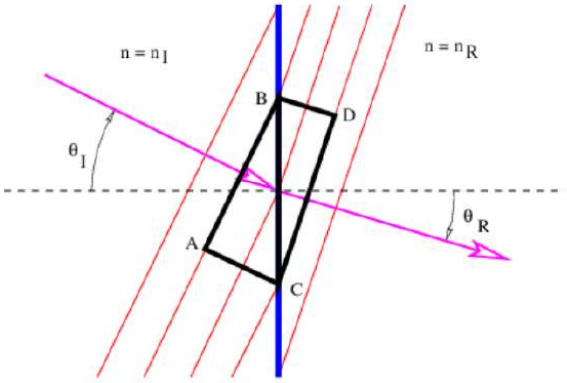


Figure 3.1.2: Sketch showing the refraction of a wave from an interface between two dielectric media with $n_2 > n_1$.

Refraction, as illustrated in Figure 3.1.2:, is slightly more complicated. Since $n_R > n_l$, the speed of light in the right-hand medium is less than in the left-hand medium. (Recall that the speed of light in a medium with refractive index n is $C_{\text{medium}} = c_{\text{Yac}} / n_.$) The frequency of the wave packet doesn't change as it passes through the interface, so the wavelength of the light on the right side is less than the wavelength on the left side.

Let us examine the triangle ABC in Figure 3.1.2:. The side AC is equal to the side BC times $\sin(\theta_{\text{I}})$. However, AC is also equal to $2\lambda_I$, or twice the wavelength of the wave to the left of the interface. Similar reasoning shows that $2\lambda_{\text{R}}$, twice the wavelength to the right of the interface, equals BC times $\sin(\theta_{\text{R}})$. Since the interval BC is common to triangles ABC and DBC, we easily see that

$$\frac{\lambda_I}{\lambda_R} = \frac{\sin(\theta_I)}{\sin(\theta_R)} \tag{3.1.1}$$

Since $\lambda_I = c_I T = c_{vac} T / n_I$ and $\lambda_R = c_R T = c_{vac} T / n_R$ where c_{I} and c_{R} are the wave speeds to the left and right of the interface, c_{vac} is the speed of light in a vacuum, and T is the (common) period, we can easily recast the above equation in the form

$$n_I \sin(\theta_I) = n_R \sin(\theta_R) \tag{3.1.2}$$

$$n_I \sin(\theta_I) = n_R \sin(\theta_R) \tag{3.1.2}$$

This is called Snell's law, and it governs how a ray of light bends as it passes through a discontinuity in the index of refraction. The angle θ_I is called the incident angle and θ_R is called the refracted angle. Notice that these angles are measured from the normal to the surface, not the tangent.

This page titled 3.1: Reflection and Refraction is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

3.2: Total Internal Reflection

When light passes from a medium of lesser index of refraction to one with greater index of refraction, Snell's law indicates that the ray bends toward the normal to the interface. The reverse occurs when the passage is in the other direction. In this latter circumstance a special situation arises when Snell's law predicts a value for the sine of the refracted angle greater than one. This is physically untenable. What actually happens is that the incident wave is reflected from the interface. This phenomenon is called total internal reflection. The minimum incident angle for which total internal reflection occurs is obtained by substituting $\theta_R = \pi/2$ into equation (3.1.2), resulting in

$$\sin(\theta_I) = n_R/n_I \quad (\text{ total internal reflection }) \tag{3.2.1}$$

This page titled 3.2: Total Internal Reflection is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

3.3: Anisotropic Media

Notice that Snell's law makes the implicit assumption that rays of light move in the direction of the light's wave vector, i. e., normal to the wave fronts. As the analysis in the previous chapter makes clear, this is valid only when the optical medium is isotropic, i. e., the wave frequency depends only on the magnitude of the wave vector, not on its direction.

Certain kinds of crystals, such as those made of calcite, are not isotropic — the speed of light in such crystals, and hence the wave frequency, depends on the orientation of the wave vector. As an example, the angular frequency in an anisotropic medium might take the form

$$\omega = \left[\frac{c_1^2 (k_x + k_y)^2}{2} + \frac{c_2^2 (k_x - k_y)^2}{2} \right]^{1/2} \tag{3.3.1}$$

where c_1 is the speed of light for waves in which $k_y = k_x$, and c_2 is its speed when $k_y = -k_x$.

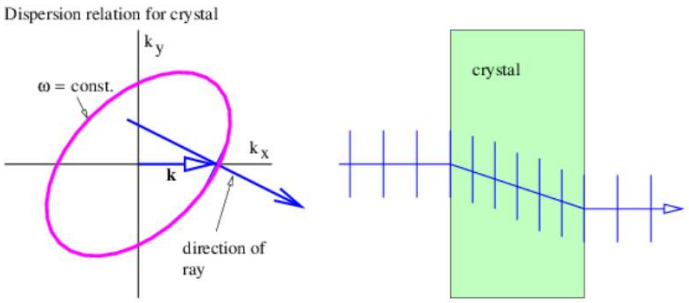


Figure 3.3.3: The right panel shows the fate of a light ray normally incident on the face of a properly cut calcite crystal. The anisotropic dispersion relation which gives rise to this behavior is shown in the left panel.

Figure 3.3.3: shows an example in which a ray hits a calcite crystal oriented so that constant frequency contours are as specified in equation (3.3.1). The wave vector is oriented normal to the surface of the crystal, so that wave fronts are parallel to this surface. Upon entering the crystal, the wave front orientation must stay the same to preserve phase continuity at the surface. However, due to the anisotropy of the dispersion relation for light in the crystal, the ray direction changes as shown in the right panel. This behavior is clearly inconsistent with the usual version of Snell's law!

It is possible to extend Snell's law to the anisotropic case. However, we will not present this here. The following discussions of optical instruments will always assume that isotropic optical media are used.

This page titled 3.3: Anisotropic Media is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

3.4: Thin Lens Equation and Optical Instruments

Given the laws of reflection and refraction, one can see in principle how the passage of light through an optical instrument could be traced. For each of a number of initial rays, the change in the direction of the ray at each mirror surface or refractive index interface can be calculated. Between these points, the ray traces out a straight line.

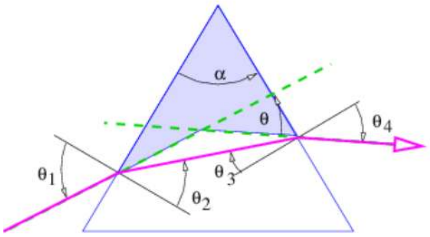


Figure 3.4.4: Bending of a ray of light as it passes through a prism.

Though simple in conception, this procedure can be quite complex in practice. However, the procedure simplifies if a number of approximations, collectively called the thin lens approximation, are valid. We begin with the calculation of the bending of a ray of light as it passes through a prism, as illustrated in Figure 3.4.4:.

The pieces of information needed to find θ, the angle through which the ray is deflected, are as follows: the geometry of the triangle defined by the entry and exit points of the ray and the upper vertex of the prism leads to

$$\alpha + (\pi/2 - \theta_2) + (\pi/2 - \theta_3) = \pi \tag{3.4.1}$$

which simplifies to

$$\alpha = \theta_2 + \theta_3 \tag{3.4.2}$$

Snell's law at the entrance and exit points of the ray tell us that

$$n = \frac{\sin(\theta_1)}{\sin(\theta_2)} \quad n = \frac{\sin(\theta_4)}{\sin(\theta_3)} \tag{3.4.3}$$

where n is the index of refraction of the prism. (The index of refraction of the surroundings is assumed to be unity.) One can also infer that

$$\theta = \theta_1 + \theta_4 - \alpha \tag{3.4.4}$$

This comes from the fact that the the sum of the internal angles of the shaded quadrangle in Figure 3.4.4: is

$$(\pi/2 - \theta_1) + \alpha + (\pi/2 - \theta_4) + (\pi + \theta) = 2\pi. \tag{3.4.5}$$

Combining equations (3.4.2), (3.4.3), and (3.4.4) allows the ray deflection θ to be determined in terms of θ_1 and α, but the resulting expression is very messy. However, great simplification occurs if the following conditions are met:

- The angle $\alpha \ll 1$.
- The angles $\theta_1, \theta_2, \theta_3, \theta_4 \ll 1$.

With these approximations it is easy to show that

$$\theta = \alpha(n-1) \quad (\text{ small angles }) \tag{3.4.6}$$

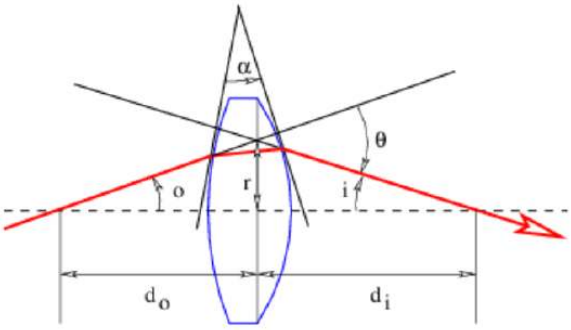


Figure 3.4.5: Light ray undergoing deflection through an angle θ by a lens. The angle α is the angle between the tangents to the entry and exit points of the ray on the lens.

Generally speaking, lenses and mirrors in optical instruments have curved rather than flat surfaces. However, we can still use the laws for reflection and refraction by plane surfaces as long as the segment of the surface on which the wave packet impinges is not curved very much on the scale of the wave packet dimensions. This condition is easy to satisfy with light impinging on ordinary optical instruments. In this case, the deflection of a ray of light is given by equation (3.4.6) if α is defined as the intersection of the tangent lines to the entry and exit points of the ray, as illustrated in Figure 3.4.5:.

A positive lens is thicker in the center than at the edges. The angle α between the tangent lines to the two surfaces of the lens at a distance r from the central axis takes the form $\alpha = Cr$, where C is a constant. The deflection angle of a beam hitting the lens a distance r from the center is therefore $\theta = Cr(n-1)$, as indicated in Figure 3.4.5:. The angles o and i sum to the deflection angle: $o + i = \theta = Cr(n-1)$. However, to the extent that the small angle approximation holds, $o = r/d_o$ and $i = r/d_i$ where d_o is the distance to the object and $d_i$is the distance to the image of the object. Putting these equations together and cancelling the r results in the thin lens equation:

$$\frac{1}{d_o} + \frac{1}{d_i} = C(n-1) \equiv \frac{1}{f} \tag{3.4.7}$$

The quantity f is called the focal length of the lens. Notice that $f = d_i$ if the object is very far from the lens, i. e., if d_o is extremely large.

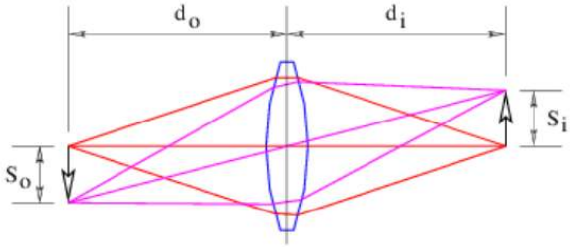


Figure 3.4.6: A positive lens producing an image on the right of the arrow on the left.

Figure 3.4.6: shows how a positive lens makes an image. The image is produced by all of the light from each point on the object falling on a corresponding point in the image. If the arrow on the left is an illuminated object, an image of the arrow will appear at right if the light coming from the lens is allowed to fall on a piece of paper or a ground glass screen. The size of the object S_o and the size of the image S_i are related by simple geometry to the distances of the object and the image from the lens:

$$\frac{S_i}{S_o} = \frac{d_i}{d_o} \tag{3.4.8}$$

Notice that a positive lens inverts the image.

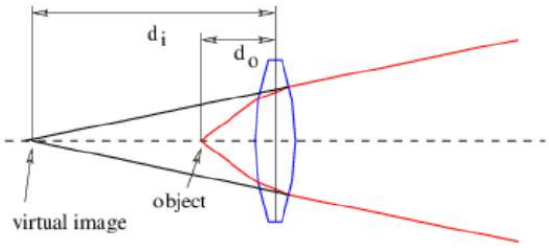


Figure 3.4.7: Production of a virtual image by a positive lens.

An image will be produced to the right of the lens only if $d_o > f$. If $d_o < f$, the lens is unable to converge the rays from the image to a point, as is seen in Figure 3.4.7:. However, in this case the backward extension of the rays converge at a point called a virtual image, which in the case of a positive lens is always farther away from the lens than the object. The image is called virtual because it does not appear on a ground glass screen placed at this point. Unlike the real image seen in Figure 3.4.6:, the virtual image is not inverted. The thin lens equation still applies if the distance from the lens to the image is taken to be negative.

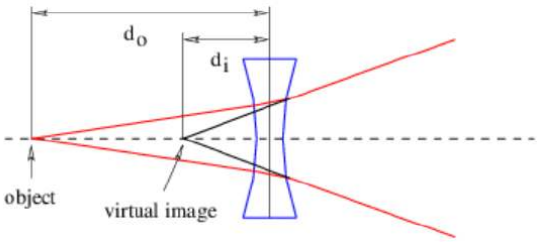


Figure 3.4.8: Production of a virtual image by a negative lens.

A *negative lens* is thinner in the center than at the edges and produces only virtual images. As seen in Figure 3.4.8:, the virtual image produced by a negative lens is closer to the lens than is the object. Again, the thin lens equation is still valid, but both the distance from the image to the lens and the focal length must be taken as negative. Only the distance to the object remains positive.

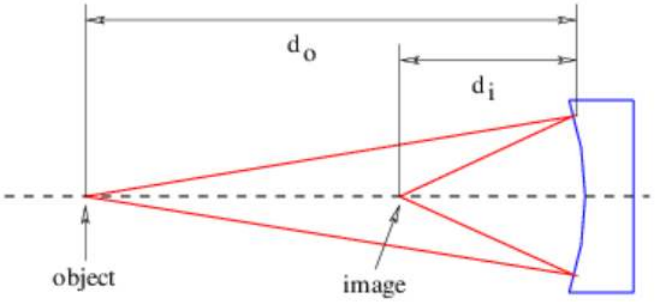


Figure 3.4.9: Production of a real image by a concave mirror.

Curved mirrors also produce images in a manner similar to a lens, as shown in Figure 3.4.9:. A concave mirror, as seen in this figure, works in analogy to a positive lens, producing a real or a virtual image depending on whether the object is farther from or closer to the mirror than the mirror's focal length. A convex mirror acts like a negative lens, always producing a virtual image. The thin lens equation works in both cases as long as the angles are small.

This page titled 3.4: Thin Lens Equation and Optical Instruments is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

3.5: Fermat's Principle

An alternate approach to geometrical optics can be developed from Fermat's principle. This principle states (in its simplest form) that light waves of a given frequency traverse the path between two points which takes the least time. The most obvious example of this is the passage of light through a homogeneous medium in which the speed of light doesn't change with position. In this case the shortest time corresponds to the shortest distance between the points, which, as we all know, is a straight line. Thus, Fermat's principle is consistent with light traveling in a straight line in a homogeneous medium.

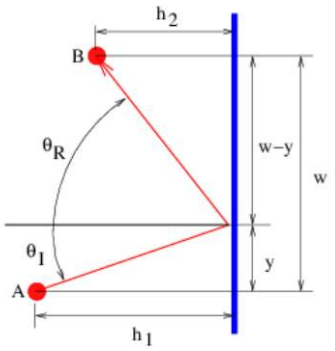


Figure 3.5.10: Definition sketch for deriving the law of reflection from Fermat's principle. θ_I is the angle of incidence and θ_R the angle of reflection as in Figure 3.5.1:.

Fermat's principle can also be used to derive the laws of reflection and refraction. For instance, Figure 3.5.10: shows a candidate ray for reflection in which the angles of incidence and reflection are not equal. The time required for the light to go from point A to point B is

$$t = \left(\left[h_1^2 + y^2 \right]^{1/2} + \left[h_2^2 + (w-y)^2 \right]^{1/2} \right) / c \tag{3.5.1}$$

where c is the speed of light. We find the minimum time by differentiating t with respect to y and setting the result to zero, with the result that

$$\frac{y}{\left[h_1^2 + y^2 \right]^{1/2}} = \frac{w-y}{\left[h_2^2 + (w-y)^2 \right]^{1/2}} \tag{3.5.2}$$

However, we note that the left side of this equation is simply sin θ_I, while the right side is sin θ_R, so that the minimum time condition reduces to sin θ_I = sin θ_R or $\theta_I = \theta_R$, which is the law of reflection.

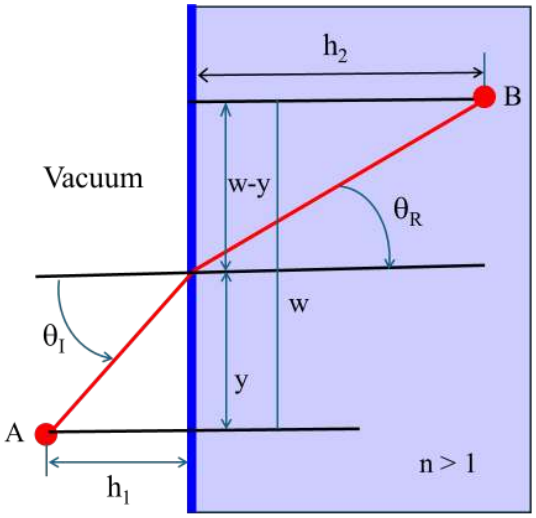


Figure 3.5.11: Definition sketch for deriving Snell's law of refraction from Fermat's principle. The shaded area has index of refraction n > 1.

A similar analysis may be done to derive Snell's law of refraction. The speed of light in a medium with refractive index n is c/n, where c is its speed in a vacuum. Thus, the time required for light to go some distance in such a medium is n times the time light takes to go the same distance in a vacuum. Referring to Figure 3.5.11:, the time required for light to go from A to B becomes

$$t = \left(\left[h_1^2 + y^2 \right]^{1/2} + n \left[h_2^2 + (w-y)^2 \right]^{1/2} \right) / c \tag{3.5.3}$$

This results in the condition

$$\sin \theta_I = n \sin \theta_R \tag{3.5.4}$$

where θ_R is now the refracted angle. We recognize this result as Snell's law.

Notice that the reflection case illustrates a point about Fermat's principle: The minimum time may actually be a local rather than a global minimum — after all, in Figure 3.5.10:, the global minimum distance from A to B is still just a straight line between the two points! In fact, light starting from point A will reach point B by both routes — the direct route and the reflected route.

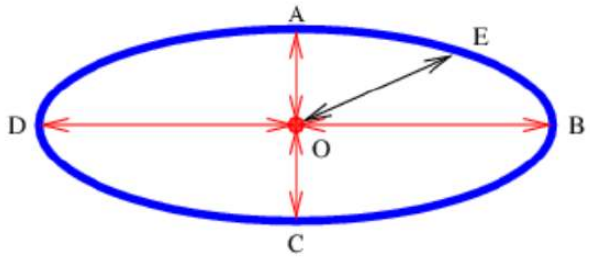


Figure 3.5.12: Ellipsoidal mirror showing minimum and maximum time rays from the center of the ellipsoid to the mirror surface and back again.

It turns out that trajectories allowed by Fermat's principle don't strictly have to be minimum time trajectories. They can also be *maximum* time trajectories, as illustrated in Figure 3.5.12:. In this case light emitted at point O can be reflected back to point O from four points on the mirror, A, B, C, and D. The trajectories O-A-O and O-C-O are minimum time trajectories while O-B-O and O-D-O are maximum time trajectories.

Fermat's principle seems rather mysterious. However, the American physicist Richard Feynman made sense out of it by invoking an even more fundamental principle, as we now see.

If a light ray originates at point O in Figure 3.5.12:, reflects off of the ellipsoidal mirror surface at point A, and returns to point O, the elapsed time isn't much different from that experienced by a ray which reflects off the mirror a slight distance from point A and returns to O. This is because at point A the beam from point O is perpendicular to the tangent to the surface of the mirror at point A. In contrast, the time experienced by a ray going from point O to point E and back would differ by a much greater amount than the time experienced by a ray reflecting off the mirror a slight distance from point E. This is because the tangent to the mirror surface at point E is not perpendicular to the beam from point O.

Technically, the change in the round trip time varies linearly with the deviation in the reflection point from point E, but quadratically with the deviation from point A. If this deviation is small in the first place, then the change in the round trip time will be much smaller for the quadratic case than for the linear case.

It seems odd that we would speak of a beam reflecting back to point O if it hit the mirror at any point except A, B, C, or D, due to the requirements of the law of reflection. However, recall that the law of reflection itself depends on Fermat's principle, so we cannot assume the validity of that law in this investigation.

Feynman postulated that light rays explore *all* possible paths from one point to another, but that the only paths realized in nature are those for which light taking closely neighboring paths experiences nearly the same elapsed time (or more generally, traverses nearly the same number of wavelengths) as the original path. If this is true, then neighboring rays interfere constructively with each other, resulting in a much brighter beam than would occur in the absence of this constructive interference. Thus, the round-trip paths O-A-O, O-B-O, O-C-O, and O-D-O in Figure 3.5.12: actually occur, but not O-E-O. Feynman explains Fermat's principle by invoking constructive and destructive interference!

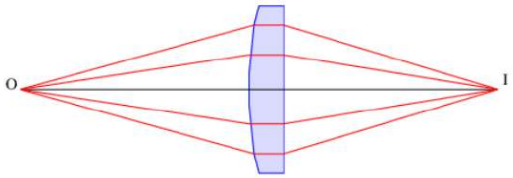


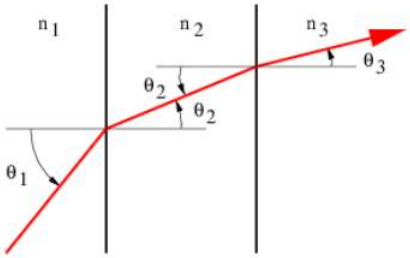
Figure 3.5.13: Ray trajectories from a point O being focused to another point I by a lens.

Figure 3.5.13: illustrates a rather peculiar situation. Notice that *all* the rays from point O which intercept the lens end up at point I. This would seem to contradict Fermat's principle, in that only the minimum (or maximum) time trajectories should occur. However, a calculation shows that all the illustrated trajectories in this particular case take the *same* time. Thus, the light cannot choose one trajectory over another using Fermat's principle and all of the trajectories are equally favored. Note that this inference applies not to just any set of trajectories, but only those going from an object point to the corresponding image point.

This page titled 3.5: Fermat's Principle is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

3.6: Problems

1. The index of refraction varies as shown in Figure 3.6.14::
    1. Given θ_1, use Snell's law to find θ_2.
    2. Given θ_2, use Snell's law to find θ_3.



    3. From the above results, find θ_3, given θ_1. Do n_2 or θ_2 matter?

Figure 3.6.14**::** Refraction through multiple parallel layers with different refractive indices.

2. A 45°-45°-90° prism is used to totally reflect light through 90° as shown in Figure 3.6.15:. What is the minimum index of refraction of the prism needed for this to work?

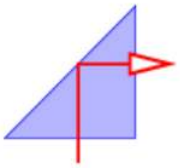


Figure 3.6.15**::** Refraction through a 45°-45°-90° prism.

3. Show graphically which way the wave vector must point inside the calcite crystal of Figure 3.6.3: for a light ray to be horizontally oriented. Sketch the orientation of the wave fronts in this case.

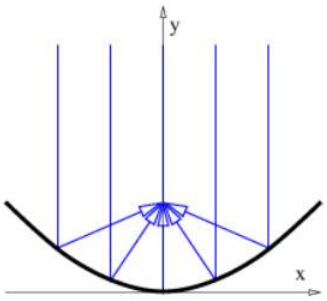


Figure 3.6.16**::** Focusing of parallel rays by a parabolic mirror.

4. The human eye is a lens which focuses images on a screen called the retina. Suppose that the normal focal length of this lens is 4 cm and that this focuses images from far away objects on the retina. Let us assume that the eye is able to focus on nearby objects by changing the shape of the lens, and thus its focal length. (The lens-retina distance remains the same.) If an object is 20 cm from the eye, what must the altered focal length of the eye be in order for the image of this object to be in focus on the retina?
5. An amoeba 0.01 cm in diameter has its image projected on a screen as shown in Figure 3.6.18: by a positive lens of diameter 0.1 cm.
    1. How big is the image of the amoeba?
    2. What is the focal length of the lens?
    3. What is the minimum distance between features on the image of the amoeba which can be resolved? Assume that the wavelength of light used is 5×10^{-7} m. (Hint: What is the spreading angle of a beam of light passing through an opening the size of the lens?)

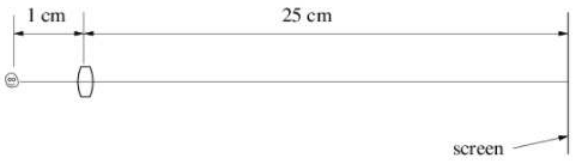


Figure 3.6.18**::** An amoeba imaged by a lens.

6. The great refractor telescope of Yerkes Observatory in Wisconsin (see Figure 3.6.19:) has primary lens D = 1.02 m in diameter with a focal length of L = 19.4 m. Use the small angle approximation in all calculations and assume that the light has wavelength 5×10^{-7} m.
    1. Jupiter has a diameter of 1.5×10^5 km and an average distance from the earth of 8×10^8 km. How big is the image of Jupiter (in cm) at the focal plane of the primary lens?
    2. Given perfect atmospheric "seeing" conditions, how far apart must two features be on Jupiter (in km) for the Yerkes telescope to be able to resolve them.
    3. What should the focal length l of the secondary lens or eyepiece be for Jupiter to subtend the same angle as the moon subtends to the naked eye? The moon's diameter is 3.5×10^3 km and its distance from the earth is 3.8×10^5 km. Hint: Imagine that a translucent sheet of ground glass is placed at the focal plane so that the image is seen projected on this ground glass, which scatters light over a broad range of angles. The eyepiece can then be thought of as a magnifying glass with which you can examine the image on the ground glass. Using this artifice, you need consider only light rays that pass through the center of each lens.

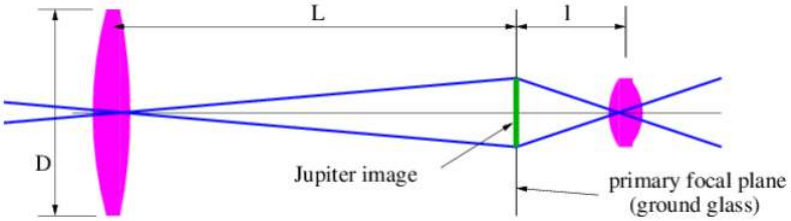


Figure 3.6.19: The great Yerkes refractor telescope observing Jupiter.

7. Show that a concave mirror that focuses incoming rays parallel to the optical axis of the mirror to a point on the optical axis, as illustrated in Figure 3.6.16:, is parabolic in shape. Hint: Since rays following different paths all move from the distant source to the focal point of the mirror, Fermat's principle implies that all of these rays take the same time to do so (why is this?), and therefore all traverse the same distance.
8. Use *Fermat's principle* to explain *qualitatively* why a ray of light follows the solid rather than the dashed line through the wedge of glass shown in Figure 3.6.17:.
9. Test your knowledge of Fermat's principle by using equation (3.5.3) to derive Snell's law.

This page titled 3.6: Problems is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

4: Special Relativity

Albert Einstein invented the special and general theories of relativity early in the 20th century, though many other people contributed to the intellectual climate which made these discoveries possible. The special theory of relativity arose out of a conflict between the ideas of mechanics as developed by Galileo and Newton, and the theory of electromagnetism. For this reason relativity is often discussed in textbooks after electromagnetism is developed. However, special relativity is actually a valid extension to the Galilean world view which is needed when objects move at very high speeds, and it is only coincidentally related to electromagnetism. For this reason we discuss relativity before electromagnetism.

The only fact from electromagnetism that we need is introduced now: There is a maximum speed at which objects can travel. This is coincidentally equal to the speed of light in a vacuum, $c = 3 \times 10^8$ m s^{-1}. Furthermore, a measurement of the speed of a particular light beam yields the same answer regardless of the speed of the light source or the speed at which the measuring instrument is moving.

This rather bizarre experimental result is in contrast to what occurs in Galilean relativity. If two cars pass a pedestrian standing on a curb, one at 20 m s^{-1} and the other at 50 m s^{-1}, the faster car appears to be moving at 30 m s^{-1} relative to the slower car. However, if a light beam moving at 3×10^8 m s^{-1} passes an interstellar spaceship moving at 2×10^8 m s^{-1}, then the light beam appears to occupants of the spaceship to be moving at 3×10^8 m s^{-1}, not 1×10^8 m s^{-1}. Furthermore, if the spaceship beams a light signal forward to its (stationary) destination planet, then the resulting beam appears to be moving at 3×10^8 m s^{-1} to instruments at the destination, not 5×10^8 m s^{-1}.

The fact that we are talking about light beams is only for convenience. Any other means of sending a signal at the maximum allowed speed would result in the same behavior. We therefore cannot seek the answer to this apparent paradox in the special properties of light. Instead we have to look to the basic nature of space and time.

4.1: Galilean Spacetime Thinking
4.2: Spacetime Thinking in Special Relativity
4.3: Postulates of Special Relativity
4.4: Time Dilation
4.5: Lorentz Contraction
4.6: Twin Paradox
4.7: Problems

This page titled 4: Special Relativity is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

4.1: Galilean Spacetime Thinking

In order to gain an understanding of both Galilean and Einsteinian relativity it is important to begin thinking of space and time as being different dimensions of a four-dimensional space called **spacetime**. Actually, since we can't visualize four dimensions very well, it is easiest to start with only one space dimension and the time dimension. Figure 4.1.1: shows a graph with time plotted on the vertical axis and the one space dimension plotted on the horizontal axis. An event is something that occurs at a particular time and a particular point in space. ("Julius X. wrecks his car in Lemitar, NM on 21 June at 6:17 PM.") A **world line** is a plot of the position of some object as a function of time on a spacetime diagram, although it is conventional to put time on the vertical axis. Thus, a world line is really a line in space time, while an event is a point in space time. A horizontal line parallel to the position axis is a line of simultaneity in Galilean relativity — i. e., all events on this line occur simultaneously.

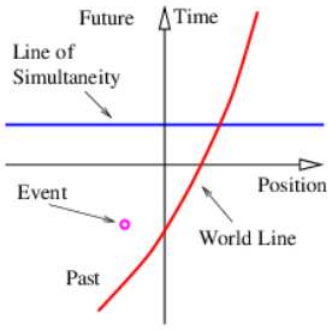


Figure 4.1.1:: Spacetime diagram showing an event, a world line, and a line of simultaneity.

In a spacetime diagram the slope of a world line has a special meaning. Notice that a vertical world line means that the object it represents does not move — the velocity is zero. If the object moves to the right, then the world line tilts to the right, and the faster it moves, the more the world line tilts. Quantitatively, we say that

$$\text{velocity} = \frac{1}{\text{slope of world line}} \tag{4.1.1}$$

in Galilean relativity. Notice that this works for negative slopes and velocities as well as positive ones. If the object changes its velocity with time, then the world line is curved, and the instantaneous velocity at any time is the inverse of the slope of the tangent to the world line at that time.

The hardest thing to realize about spacetime diagrams is that they represent the past, present, and future all in one diagram. Thus, spacetime diagrams don't change with time — the evolution of physical systems is represented by looking at successive horizontal slices in the diagram at successive times. Spacetime diagrams represent evolution, but they don't evolve themselves.

The principle of relativity states that the laws of physics are the same in all inertial reference frames. An inertial reference frame is one that is not accelerated. Reference frames attached to a car at rest and to a car moving steadily down the freeway at 30 m s^{-1} are both inertial. A reference frame attached to a car accelerating away from a stop light is not inertial.

The principle of relativity is an educated guess or hypothesis based on extensive experience. If the principle of relativity weren't true, we would have to do all our calculations in some preferred reference frame. This would be very annoying. However, the more fundamental problem is that we have no idea what the velocity of this preferred frame might be. Does it move with the earth? That would be very earth-centric. How about the velocity of the center of our galaxy or the mean velocity of all the galaxies? Rather than face the issue of a preferred reference frame, physicists have chosen to stick with the principle of relativity.

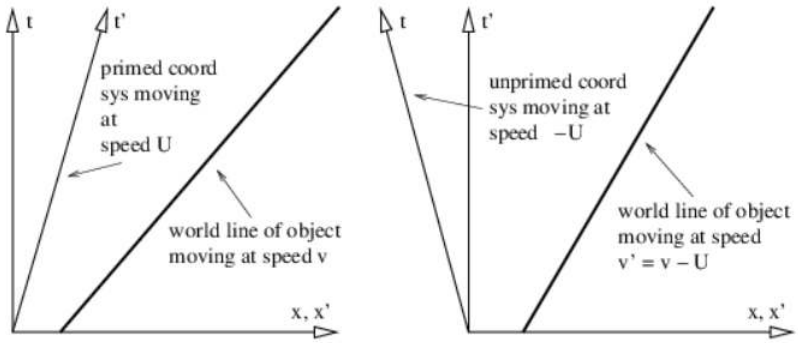


Figure 4.1.2:: The left panel shows the world line in the unprimed reference frame, while the right panel shows it in the primed frame, which moves to the right at speed U relative to the unprimed frame. (The "prime" is just a label that allows us to distinguish the axes corresponding to the two reference frames.)

If an object is moving to the left at velocity v relative to a particular reference frame, it appears to be moving at a velocity $v' = v - U$ relative to another reference frame which itself is moving at velocity U. This is the Galilean velocity transformation law, and it is based on everyday experience. If you are traveling 30 m s^{-1} down the freeway and another car passes you doing 40 m s^{-1}, then the other car moves past you at 10 m s^{-1} relative to your car.

Figure 4.1.2: shows how the world line of an object is represented differently in the unprimed (x, t) and primed (x′, t′) reference frames. The difference between the velocity of the object and the velocity of the primed frame (i. e., the difference in the inverses of the slopes of the corresponding world lines) is the same in both reference frames in this Galilean case. This illustrates the difference between a physical law independent of reference frame (the difference between velocities in Galilean relativity) and the different motion of the object in the two different reference frames.

This page titled 4.1: Galilean Spacetime Thinking is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

4.2: Spacetime Thinking in Special Relativity

In special relativity we find that space and time “mix” in a way that they don’t in Galilean relativity. This suggests that space and time are different aspects of the same “thing”, which we call **spacetime**.

If time and position are simply different dimensions of the same abstract space, then they should have the same units. The easiest way to arrange this is to multiply time by the maximum speed, *c*, resulting in the kind of spacetime diagram shown in Figure 4.2.3:. Notice that world lines of light have slope ±1 when the time axis is scaled this way. Furthermore, the relationship between speed and the slope of a world line must be revised to read

$$v = \frac{c}{\text{slope}} \quad (\text{ world line }) . \tag{4.2.1}$$

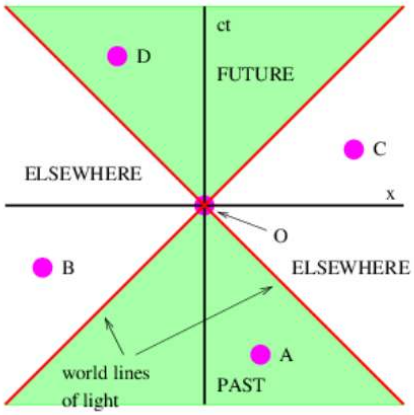


Figure 4.2.3:: Scaled spacetime diagram showing world lines of light passing left and right through the origin.

Notice that it is physically possible for an object to have a world line which connects event O at the origin and the events A and D in Figure 4.2.3:, since the slope of the resulting world line would exceed unity, and thus represent a velocity less than the speed of light. Events which can be connected by a world line are called *timelike* relative to each other. On the other hand, event O cannot be connected to events B and C by a world line, since this would imply a velocity greater than the speed of light. Events which cannot be connected by a world line are called *spacelike* relative to each other. Notice the terminology in Figure 4.2.3:: Event A is in the *past* of event O, while event D is in the *future*. Events B and C are *elsewhere* relative to event O.

This page titled 4.2: Spacetime Thinking in Special Relativity is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

4.3: Postulates of Special Relativity

As we learned previously, the principle of relativity states that the laws of physics are the same in all inertial reference frames. The principle of relativity applies to Einsteinian relativity just as it applies to Galilean relativity.

Notice that the constancy of the speed of light in all reference frames is consistent with the principle of relativity. However, as noted above, it is inconsistent with our notions as to how velocities add, or alternatively, how we think the world should look from reference frames moving at different speeds. We have called the classical way of understanding the view from different reference frames Galilean relativity. The new way that reconciles the behavior of objects moving at very high speeds is called Einsteinian relativity. Einstein's great contribution was to discover the laws that tell us how the world looks from reference frames moving at high speeds relative to each other. These laws constitute a geometry of spacetime, and from them all of special relativity can be derived.

All of the observed facts about spacetime can be derived from two postulates:

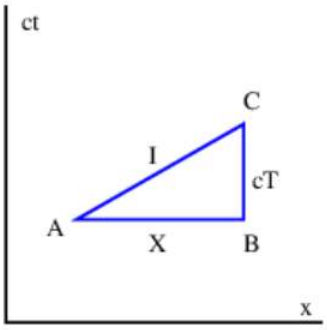


Figure 4.3.4:: Triangle for Pythagorean theorem in spacetime.

- Whether two events are simultaneous depends on the reference frame from which they are viewed.
- Spacetime obeys a modified Pythagorean theorem, which gives the distance, I, in spacetime or spacetime interval as

$$I^2 = X^2 - c^2T^2 \tag{4.3.1}$$

where X, T, and I are defined in Figure 4.3.4:.

Let us discuss these postulates in turn.

Simultaneity

The classical way of thinking about simultaneity is so ingrained in our everyday habits that we have a great deal of difficulty adjusting to what special relativity has to say about this subject. Indeed, understanding how relativity changes this concept is the single most difficult part of the theory — once you understand this, you are well on your way to mastering relativity!

Before tackling simultaneity, let us first think about collocation. Two events (such as A and E in Figure 4.3.5:) are collocated if they have the same x value. However, collocation is a concept that depends on the reference frame. For instance, George is driving from Boston to Washington, with the line passing through events A and D being his world line. Just as he passes New York he sneezes (event A in Figure 4.3.5:). As he drives by Baltimore, he sneezes again (event D). In the reference frame of the earth, these two sneezes are not collocated, since they are separated by many kilometers. However, in the reference frame of George's car, they occur in the same place — assuming that George hasn't left the driver's seat!

Notice that any two events separated by a timelike interval are collocated in some reference frame. The speed of the reference frame is given by equation (4.2.1), where the slope is simply the slope of the world line connecting the two events.

In Galilean relativity, if two events are simultaneous, we consider them to be simultaneous in all reference frames. For instance, if two clocks, one in New York and one in Los Angeles, strike the hour at the same time in the earth reference frame, then in Galilean relativity these events also appear to be simultaneous to instruments in the space shuttle as it flies over the United States. However, if the space shuttle is moving from west to east, i. e., from Los Angeles toward New York, careful measurements will show that the clock in New York strikes the hour before the clock in Los Angeles! Thus, the Galilean point of view is not accurate.

Just as collocation depends on one's reference frame, this result shows that simultaneity also depends on the reference frame. Figure 4.3.5: shows how this works. In Figure 4.3.5: events A and B are simultaneous in the rest or unprimed reference frame.

However, in the primed reference frame, events A and C are simultaneous, and event B occurs at an earlier time. If A and B correspond to the clocks striking in Los Angeles and New York respectively, then it is clear that B must occur at an earlier time in the primed frame if indeed A and C are simultaneous in that frame.

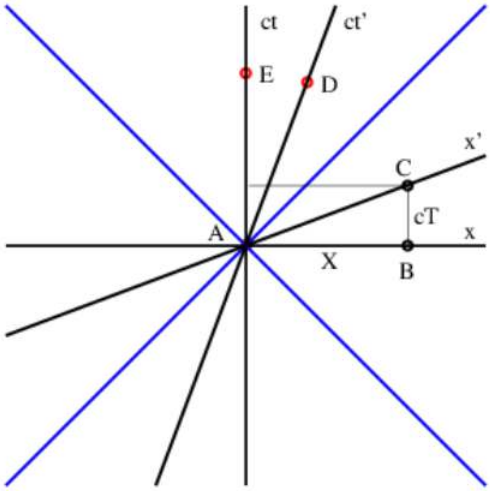


Figure 4.3.5:: Sketch of coordinate axes for a moving reference frame, x′, ct′. The meanings of the events A-E are discussed in the text. The lines tilted at ±45° are the world lines of light passing through the origin.

The tilted line passing through events A and C in Figure 4.3.5: is called the line of simultaneity for the primed reference frame. Its slope is related to the speed, U, of the reference frame by

$$\text{slope } = U/c \quad \text{(line of simultaneity)}. \tag{4.3.2}$$

Notice that this is the inverse of the slope of the world line attached to the primed reference frame. There is thus a symmetry between the world line and the line of simultaneity of a moving reference frame — as the reference frame moves faster to the right, these two lines close like the blades of a pair of scissors on the 45° line.

In Galilean relativity it is fairly obvious what we mean by two events being simultaneous — it all boils down to coordinating portable clocks which are sitting next to each other, and then moving them to the desired locations. Two events separated in space are simultaneous if they occur at the same time on clocks located near each event, assuming that the clocks have been coordinated in the above manner.

In Einsteinian relativity this doesn't work, because the very act of moving the clocks changes the rate at which the clocks run. Thus, it is more difficult to determine whether two distant events are simultaneous.

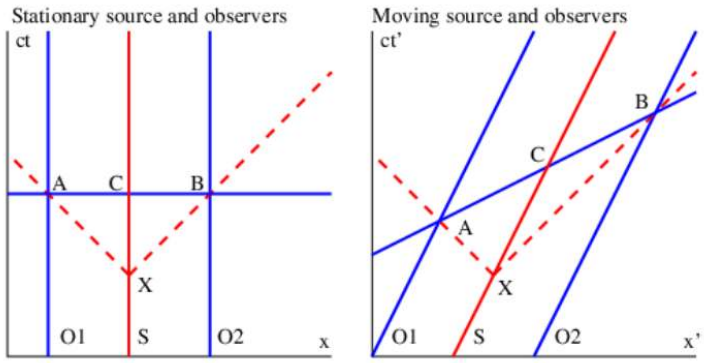


Figure 4.3.6:: World lines of two observers (O1, O2) and a pulsed light source (S) equidistant between them. In the left frame the observers and the source are all stationary. In the right frame they are all moving to the right at half the speed of light. The dashed lines show pulses of light emitted simultaneously to the left and the right.

An alternate way of experimentally determining simultaneity is shown in Figure 4.3.6:. Since we know from observation that light travels at the same speed in all reference frames, the pulses of light emitted by the light sources in Figure 4.3.6: will reach the two equidistant observers simultaneously in both cases. The line passing through these two events, A and B, defines a line of

simultaneity for both stationary and moving observers. For the stationary observers this line is horizontal, as in Galilean relativity. For the moving observers the light has to travel farther in the rest frame to reach the observer receding from the light source, and it therefore takes longer in this frame. Thus, event B in the right panel of Figure 4.3.6: occurs later than event A in the stationary reference frame and the line of simultaneity is tilted. We see that the postulate that light moves at the same speed in all reference frames leads inevitably to the dependence of simultaneity on reference frame.

Spacetime Pythagorean Theorem

The Pythagorean theorem of spacetime differs from the usual Pythagorean theorem in two ways. First, the vertical side of the triangle is multiplied by c. This is a trivial scale factor that gives time the same units as space. Second, the right side of equation (4.3.1) has a minus sign rather than a plus sign. This highlights a fundamental difference between spacetime and the ordinary xyz space in which we live. Spacetime is said to have a non-Euclidean (but not curved) geometry — in other words, the normal rules of geometry that we learn in high school don't always work for spacetime!

The main consequence of the minus sign in equation (4.3.1) is that I^2 can be negative and therefore I can be imaginary. Furthermore, in the special case where $X = \pm cT$, we actually have I = 0 even though X, $T \neq 0 -$ i. e., the "distance" between two well-separated events can be zero. Clearly, spacetime has some weird properties!

The quantity I is usually called an interval in spacetime. Generally speaking, if I^2 is positive, the interval is called spacelike, while for a negative I^2, the interval is called timelike.

A concept related to the spacetime interval is the proper time τ. The proper time between the two events A and C in Figure 4.3.4: is defined by the equation

$$\tau^2 = T^2 - X^2/c^2 \tag{4.3.3}$$

Notice that I and τ are related by

$$\tau^2 = -I^2/c^2 \tag{4.3.4}$$

so the spacetime interval and the proper time are not independent concepts. However, *I* has the dimensions of length and is real when the events defining the interval are spacelike relative to each other, whereas *τ* has the dimensions of time and is real when the events are timelike relative to each other. Both equation (4.3.1) and equation (4.3.3) express the spacetime Pythagorean theorem.

If two events defining the end points of an interval have the same *t* value, then the interval is the ordinary space distance between the two events. On the other hand, if they have the same *x* value, then the proper time is just the time interval between the events. If the interval between two events is spacelike, but the events are not simultaneous in the initial reference frame, they can always be made simultaneous by choosing a reference frame in which the events lie on the same line of simultaneity. Thus, the meaning of the interval in that case is just the distance between the events in the new reference frame. Similarly, for events separated by a timelike interval, the proper time is just the time between two events in a reference frame in which the two events are collocated.

This page titled 4.3: Postulates of Special Relativity is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

4.4: Time Dilation

Stationary and moving clocks run at different rates in relativity. This is illustrated in Figure 4.4.7:. The triangle ABC in the left panel of Figure 4.4.7: can be used to illustrate this point. Suppose that the line passing through the events A and C in this figure is the world line of a stationary observer. At zero time another observer moving with velocity V passes the stationary observer. The moving observer's world line passes through events A and B.

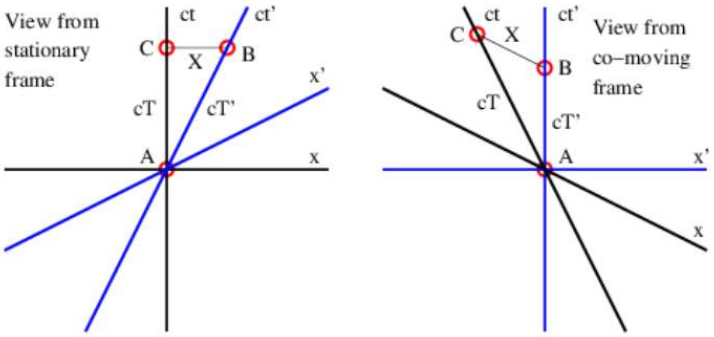


Figure 4.4.7:: Two views of the relationship between three events, A, B, and C. The left panel shows the view from the unprimed reference frame, in which A and C are collocated, while the right panel shows the view from the primed frame, in which A and B are collocated.

We assume that events B and C are simultaneous in the rest frame, so ABC is a right triangle. Application of the space time Pythagorean theorem thus yields

$$c^2T'^2 = c^2T^2 - X^2 \tag{4.4.1}$$

Since the second observer is moving at velocity V , the slope of his world line is

$$\frac{c}{V} = \frac{cT}{X} \tag{4.4.2}$$

where the right side of the above equation is the slope calculated as the rise of the world line cT over the run X between events A and B. Eliminating X between the above two equations results in a relationship between T and T′:

$$T' = T\left(1 - V^2/c^2\right)^{1/2} \equiv T/\gamma \tag{4.4.3}$$

where

$$\gamma = \frac{1}{\left(1 - V^2/c^2\right)^{1/2}} \tag{4.4.4}$$

The quantity γ occurs so often in relativistic calculations that we give it this special symbol. Note by its definition that $\gamma \geq 1$.

Equation (4.4.3) tells us that the time elapsed for the moving observer is less than that for the stationary observer, which means that the clock of the moving observer runs more slowly. This is called the time dilation effect.

Let us view this situation from the reference frame of the moving observer. In this frame the moving observer becomes stationary and the stationary observer moves in the opposite direction, as illustrated in the right panel of Figure 4.4.7:. By symmetric arguments, one infers that the clock of the initially stationary observer who is now moving to the left runs more slowly in this reference frame than the clock of the initially moving observer. One might conclude that this contradicts the previous results. However, examination of the right panel of Figure 4.4.7: shows that this is not so. The interval cT is still greater than the interval cT′, because such intervals are relativistically invariant quantities. However, events B and C are no longer simultaneous, so one cannot use these results to infer anything about the rate at which the two clocks run in this frame. Thus, the relative nature of the concept of simultaneity saves us from an incipient paradox, and we see that the relative rates at which clocks run depends on the reference frame in which these rates are observed.

This page titled 4.4: Time Dilation is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

4.5: Lorentz Contraction

A similar argument can be made to show how the postulates of relativity result in the Lorentz contraction. Figure 4.5.8: compares the length X′ of a moving object measured in its own reference frame (left panel) with its length X as measured in a stationary reference frame (right panel). The length of a moving object is measured by simultaneously measuring the positions of the front and the rear of the object and subtracting these two numbers. Events A and C correspond to these position measurements for the stationary reference frame since they are respectively on the rear and front world lines of the object. Thus, the interval AC, which is equal to X, is the length of the object as measured in the stationary frame.

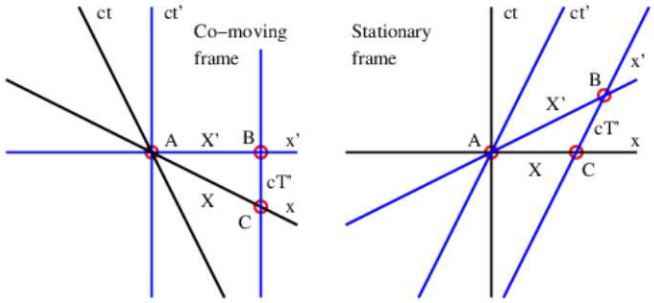


Figure 4.5.8:: Definition sketch for understanding the Lorentz contraction. The parallel lines represent the world lines of the front and the rear of a moving object. The left panel shows a reference frame moving with the object, while the right panel shows a stationary reference frame.

In the left panel, X is the hypotenuse of a right triangle. Therefore, by the Pythagorean theorem of spacetime, we have

$$X = I = \left(X'^2 - c^2T'^2\right)^{1/2} \tag{4.5.1}$$

Now, the line passing through A and C in the left panel is the line of simultaneity of the stationary reference frame. The slope of this line is -V/c, where V is the speed of the object relative to the stationary reference frame. Geometrically in Figure 4.5.8:, the slope of this line is $-\mathrm{cT}'/\mathrm{X}'$, so we find by equating these two expressions for the slope that

$$T' = VX'/c^2 \tag{4.5.2}$$

Finally, eliminating T′ between (4.5.1) and (4.5.2) results in

$$X = X'\left(1 - V^2/c^2\right)^{1/2} = X'/\gamma \tag{4.5.3}$$

This says that the length of a moving object as measured in a stationary reference frame (X) is less than the actual length of the object as measured in its own reference frame (X′). This reduction in length is called the Lorentz contraction.

Note that the Lorentz contraction only occurs in the direction of motion. The dimensions of a moving object perpendicular to the motion remain unchanged.

This page titled 4.5: Lorentz Contraction is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

4.6: Twin Paradox

An interesting application of time dilation is the so-called **twin paradox**, which turns out not to be a paradox at all. Two twins are initially the same age. One twin travels to a distant star on an interstellar spaceship which moves at speed V , which is close to the speed of light. Upon reaching the star, the traveling twin immediately turns around and heads home. When reaching home, the traveling twin has aged less than the twin that stayed home. This is easily explained by the time dilation effect, which shows that the proper time elapsed along the world line of the traveling twin is $(1 - V^2/c^2)^{1/2}$ times the proper time elapsed along the world line of the other twin.

The “paradox” part of the twin paradox arises from making the symmetric argument in which one assumes the reference frame of the traveling twin to be stationary. The frame of the earth-bound twin must then travel in the sense opposite that of the erstwhile traveling twin, which means that the earth-bound twin must age less rather than more. This substitution is justified on the basis of the principle of relativity, which states that the laws of physics must be the same in all inertial reference frames.

However, the above argument is fallacious, because the reference frame of the traveling twin is not inertial throughout the entire trip, since at various points the spaceship has to accelerate and decelerate. Thus, the principle of relativity cannot be used to assert the equivalence of the traveler’s reference frame to the stationary frame.

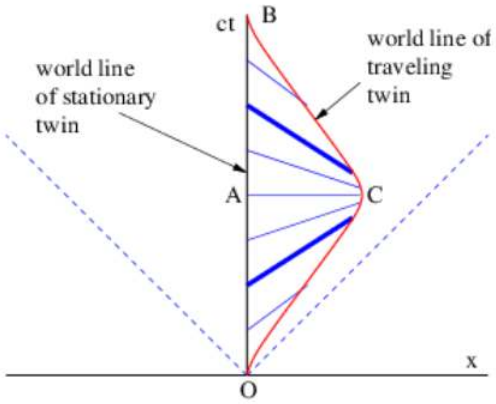


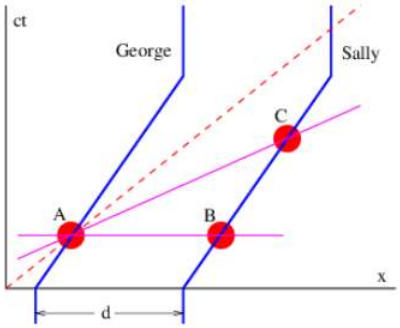
Figure 4.6.9:: Definition sketch for the twin paradox. The vertical line is the world line of the twin that stays at home, while the traveling twin has the curved world line to the right. The slanted lines between the world lines are lines of simultaneity at various times for the traveling twin. The heavy lines of simultaneity bound the period during which the traveling twin is decelerating to a stop and accelerating toward home.

Of particular importance is the period of deceleration and acceleration near the distant star. During this interval, the line of simultaneity of the traveling twin rotates, as illustrated in Figure 4.6.9:, such that the twin staying at home rapidly ages in the reference frame of the traveler. Thus, even though the acceleration of the traveling twin may occupy only a negligible segment of the twin’s world line, the overall effect is not negligible. In fact, the shorter and more intense the period of acceleration, the more rapidly the earth-bound twin ages in the traveling frame during this interval!

This page titled 4.6: Twin Paradox is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

4.7: Problems

1. Sketch your personal world line on a spacetime diagram for the last 24 hours, labeling by time and location special events such as meals, physics classes, etc. Relate the slope of the world line at various times to how fast you were walking, riding in a car, etc.
2. Spacetime conversions:
    1. What is the distance from New York to Los Angeles in seconds? From here to the moon? From here to the sun?
    2. What is one nanosecond in meters? One second? One day? One year?
3. Three events have the following spacetime coordinates: A is at $(x, ct) = (2\ \mathrm{m}, 1\ \mathrm{m})$; B is at $(x, ct) = (-2\ \mathrm{m}, 0\ \mathrm{m})$; C is at $(x, ct) = (0\ \mathrm{m}, 3\ \mathrm{m})$ (Hint: Draw a spacetime diagram with all the events plotted before trying to answer the above questions.)
    1. A world line for an object passes through events B and C. How fast and in which direction is the object moving?
    2. A line of simultaneity for a coordinate system passes through events A and B. How fast and in which direction is the coordinate system moving?
    3. What is the invariant interval between events A and B? B and C? A and C?
    4. Can a signal from event B reach event A? Can it reach event C? Explain.
4. In the following problem be sure to indicate the slope of all pertinent lines drawn.
    1. In a spacetime diagram, sketch a line of simultaneity for a reference frame moving to the left at $V = c/2$, where c is the speed of light.
    2. Sketch the world line of an object which is initially stationary, but which accelerates to a velocity of $v = c/3$.
5. If the slopes of the world lines of the observers in the right panel of Figure $4.7.6$: are both $1/\beta$, find the slope of their line of simultaneity, AB.
6. Suppose that an interstellar spaceship goes a distance X = 100 light years relative to the rest frame in $\mathrm{T}' = 10$ years of its own time.
    1. Draw a spacetime diagram in the rest frame showing X, T′, and the time T needed for this journey relative to the rest frame.
    2. Compute T, using your spacetime diagram as an aid.
    3. Compute the speed of the spaceship.
7. If an airline pilot flies 80 hr per month (in the rest frame) at 300 m s^{-1} for 30 years, how much younger will she be than her twin brother (who handles baggage) when she retires? Hint: Use $(1 + \epsilon)^{\mathrm{X}} \approx 1 + \mathrm{x}\epsilon$ for small ϵ.
8. A mu particle normally lives about 2×10^{-6} sec before it decays. However, muons created by cosmic rays 20 km up in the atmosphere reach the Earth's surface. How fast must they be going?
9. The Stanford Linear Accelerator accelerates electrons to a speed such that the 3 km long accelerator appears to be 8 cm long to the electron, due to the Lorentz contraction. How much less than the speed of light is the electron traveling? Hint: It is best to first develop an approximation for the relationship between $\gamma = (1 - v^2/c^2)^{-1/2}$ and the difference between c and v for a particle moving close to the speed of light.
10. How fast do you have to go to reach the center of our galaxy in your expected lifetime? At this speed, what does this distance appear to be? (We are about 30000 light years from the galactic center.)
11. Two identical spaceships pass each other going in the opposite direction at the same speed.
    1. Sketch a spacetime diagram showing the world lines of the front and rear of each spaceship as well as lines of simultaneity for each spaceship.
    2. Indicate an interval on the diagram corresponding to the rightward-moving spaceship's length in its own reference frame.
    3. Indicate an interval corresponding to the leftward-moving spaceship's length in the reference frame of the rightward-moving spaceship.



4. Indicate an interval equal to the length of either spaceship in the rest frame.

Figure 4.7.10**::** Sketch for moving twins. Line AC is the line of simultaneity for a reference frame moving with Sally and George.

12. George and Sally are twins initially separated by a distance d and at rest. In the rest frame they are initially the same age. At time t = 0 both George and Sally get in their spaceships and head to the right at velocity U. Both move a distance d to the right and decelerate to a halt. (See Figure 4.7.10:.) Hint: Draw the triangle ABC in a reference frame moving with George and Sally.
    1. When both are moving, how far away is Sally according to George?
    2. How much older or younger is Sally relative to George while both are moving?
    3. How much older or younger is Sally relative to George after both stop?

4.7: Problems is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

CHAPTER OVERVIEW

5: Applications of Special Relativity

In this chapter we continue the study of special relativity. Three important applications of the ideas developed in the previous chapter are made here. First, we show how to describe waves in the context of spacetime. We then see how waves which have no preferred reference frame (such as that of a medium supporting them) are constrained by special relativity to have a dispersion relation of a particular form. This dispersion relation turns out to be that of the relativistic matter waves of quantum mechanics. Second, we investigate the Doppler shift phenomenon, in which the frequency of a wave takes on different values in different coordinate systems. Third, we show how to add velocities in a relativistically consistent manner.

A new mathematical idea is presented in the context of relativistic waves, namely the spacetime vector or four-vector. Writing the laws of physics totally in terms of relativistic scalars and four-vectors insures that they will be valid in all inertial reference frames.

5.1: Waves in Spacetime
5.2: Math Tutorial – Four-Vectors
5.3: Principle of Relativity Applied
5.4: Characteristics of Relativistic Waves
5.5: The Doppler Shift
5.6: Addition of Velocities
5.7: Problems

This page titled 5: Applications of Special Relativity is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

5.1: Waves in Spacetime

We now look at the characteristics of waves in spacetime. Recall that a sine wave moving to the right in one space dimension can be represented by

$$A(x,t) = A_0 \sin(kx - \omega t) \tag{5.1.1}$$

where A_0 is the (constant) amplitude of the wave, k is the wavenumber, and ω is the angular frequency, and that the quantity $\phi = kx - \omega t$ is called the phase of the wave. For a plane wave in three space dimensions, the wave is represented in a similar way,

$$A(\mathbf{x},t) = A_0 \sin(\mathbf{k}\cdot\mathbf{x} - \omega t) \tag{5.1.2}$$

where k is now the position vector and k is the wave vector. The magnitude of the wave vector, $|\mathbf{k}| = \mathrm{k}$ is just the wavenumber of the wave and the direction of this vector indicates the direction the wave is moving. The phase of the wave in this case is $\phi = \mathbf{k}\cdot\mathbf{x} - \omega t$.

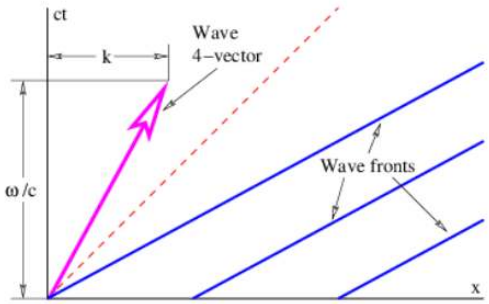


Figure 5.1.1:: Sketch of wave fronts for a wave in spacetime. The large arrow is the associated wave four-vector, which has slope ω/ck. The slope of the wave fronts is the inverse, ck/ω . . The phase speed of the wave is greater than c in this example. (Can you tell why?)

In the case of a one-dimensional wave moving to the right $\phi = kx - \omega t$. A wave front has constant phase ϕ, so solving this equation for t and multiplying by c, the speed of light in a vacuum, gives us an equation for the world line of a wave front:

$$ct = \frac{ckx}{\omega} - \frac{c\phi}{\omega} = \frac{cx}{u_p} - \frac{c\phi}{\omega} \quad (\text{ wave front }) \tag{5.1.3}$$

The slope of the world line in a spacetime diagram is the coefficient of x, or $c/\mathrm{u_p},$ where $\mathrm{u_p} = \omega/\mathrm{k}$ is the phase speed. The world lines of the wave fronts of a wave are illustrated in Figure 5.1.1:.

This page titled 5.1: Waves in Spacetime is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

5.2: Math Tutorial – Four-Vectors

Also shown in Figure 5.2.1: is a spacetime vector or four-vector which represents the frequency and wavenumber of the wave, which we refer to as the wave four-vector. It is called a four-vector because it has 3 spacelike components and one timelike component when there are 3 space dimensions. In the case shown, there is only a single space dimension. The spacelike component of the wave four-vector is just k (or **k** when there are 3 space dimensions), while the timelike component is ω/c. The c is in the denominator to give the timelike component the same dimensions as the spacelike component. From Figure 5.2.1: it is clear that the slope of the line representing the four-vector is ω/ck, which is just the inverse of the slope of the wave fronts.

Let us define some terminology. We indicate a four-vector by underlining and write the components in the following way: $\underline{k} = (k, \omega/c)$, where k is the wave four-vector, k is its spacelike component, and ω/c is its timelike component. For three space dimensions, where we have a wave vector rather than just a wavenumber, we write $\underline{\mathrm{k}} = (\mathbf{k}, \omega/\mathrm{c})$.

Another example of a four-vector is simply the position vector in spacetime, $\underline{x} = (x, ct),$ or $\underline{x} = (\mathbf{x}, \mathrm{ct})$ in three space dimensions. The c multiplies the timelike component in this case, because that is what is needed to give it the same dimensions as the spacelike component.

In three dimensions we define a vector as a quantity with magnitude and direction. Extending this to spacetime, a four-vector is a quantity with magnitude and direction in spacetime. Implicit in this definition is the notion that the vector's magnitude is a quantity independent of coordinate system or reference frame. We have seen that the invariant interval in spacetime from the origin to the point $(\mathrm{x}, \mathrm{ct})$ is $I = (x^2 - c^2t^2)^{1/2}$ so it makes sense to identify this as the magnitude of the position vector. This leads to a way of defining a dot product of four-vectors. Given two four-vectors $\underline{A} = (\mathbf{A}, A_t)$ and $\underline{B} = (\mathbf{B}, B_t)$, the dot product is

$$\underline{A} \cdot \underline{B} = \mathbf{A} \cdot \mathbf{B} - A_t B_t \quad \text{(dot product in spacetime)}. \tag{5.2.1}$$

This is consistent with the definition of invariant interval if we set $\underline{A} = \underline{B} = \underline{x},$ since then $\underline{x} \cdot \underline{x} = x^2 - c^2t^2 = I^2$

In the odd geometry of spacetime it is not obvious what "perpendicular" means. We therefore define two four-vectors A and B to be perpendicular if their dot product is zero, $\underline{A} \cdot \underline{B} = 0$, in analogy with ordinary vectors.

The dot product of two four-vectors is a scalar result, i. e., its value is independent of coordinate system. This can be used to our advantage on occasion. For instance, consider the dot product of a four-vector A which resolves into (A_x, A_t) in the unprimed frame. Let us further suppose that the spacelike component is zero in some primed frame, so that the components in this frame are $(0, A'_t)$. The fact that the dot product is independent of coordinate system means that

$$\underline{A} \cdot \underline{A} = A_x^2 - A_t^2 = -A_t'^2 \tag{5.2.2}$$

This constitutes an extension of the spacetime Pythagorean theorem to four-vectors other than the position four-vector. Thus, for instance, the wavenumber for some wave may be zero in the primed frame, which means that the wavenumber and frequency in the unprimed frame are related to the frequency in the primed frame by

$$k^2 - \omega^2/c^2 = -\omega'^2/c^2 \tag{5.2.3}$$

This page titled 5.2: Math Tutorial – Four-Vectors is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

5.3: Principle of Relativity Applied

Returning to the phase of a wave, we immediately see that

$$\phi = \mathbf{k}\cdot\mathbf{x} - \omega t = \mathbf{k}\cdot\mathbf{x} - (\omega/c)(ct) = \underline{k}\cdot\underline{x}. \tag{5.3.1}$$

Thus, a compact way to rewrite equation (5.2.2) is

$$A(\underline{x}) = A_0 \sin(\underline{k}\cdot\underline{x}) \tag{5.3.2}$$

Since x is known to be a four-vector and since the phase of a wave is known to be a scalar independent of reference frame, it follows that k is also a four-vector rather than just a set of numbers. Thus, the square of the length of the wave four-vector must also be a scalar independent of reference frame:

$$\underline{k}\cdot\underline{k} = \mathbf{k}\cdot\mathbf{k} - \omega^2/c^2 = \text{ const.} \tag{5.3.3}$$

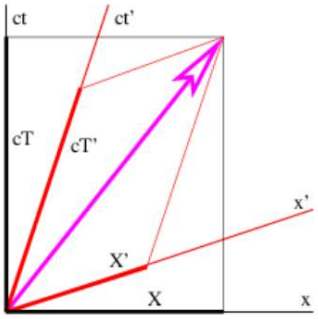


Figure 5.3.2:: Resolution of a four-vector into components in two different reference frames.

Let us review precisely what this means. As Figure 5.3.2: shows, we can resolve a position four-vector into components in two different reference frames, $\underline{x} = (X, cT) = (X', cT')$. However, even though $X \neq X'$ and $T \neq T'$, the vector lengths computed from these two sets of components are necessarily the same: $\underline{x}\cdot\underline{x} = X^2 - c^2T^2 = X'^2 - c^2T'^2$.

Applying this to the wave four-vector, we infer that

$$k^2 - \omega^2/c^2 = k^2 - \omega'^2/c^2 = \text{ const.} \tag{5.3.4}$$

where the unprimed and primed values of k and ω refer to the components of the wave four-vector in two different reference frames.

Up to now, this argument applies to any wave. However, waves can be divided into two categories, those for which a “special” reference frame exists, and those for which there is no such special frame. As an example of the former, sound waves look simplest in the reference frame in which the gas carrying the sound is stationary. The same is true of light propagating through a material medium with an index of refraction not equal to unity. In both cases the speed of the wave is the same in all directions only in the frame in which the material medium is stationary.

Suppose we have a machine that produces a wave with wavenumber k and frequency ω in its own rest frame. If we observe the wave from a moving reference frame, the wavenumber and frequency will be different, say, k' and ω'. However, these quantities will be related by equation (5.3.4).

Up to this point the argument applies to any wave whether a special reference frame exists or not; the observed changes in wavenumber and frequency have nothing to do with the wave itself, but are just consequences of how we have chosen to observe it. However, if there is no special reference frame for the type of wave under consideration, then the same result can be obtained by keeping the observer stationary and moving the wave-producing machine in the opposite direction. By moving it at various speeds, any desired value of k' can be obtained in the initial reference frame (as opposed to some other frame), and the resulting value of ω' can be computed using equation (5.3.4).

This is actually an amazing result. We have shown on the basis of the principle of relativity that any wave type for which no special reference frame exists can be made to take on a full range of frequencies and wavenumbers in any given reference frame, and furthermore that these frequencies and wavenumbers obey

$$\omega^2 = k^2c^2 + \omega^2 \tag{5.3.5}$$

Equation (5.3.5) comes from solving equation (5.10) for ω^2 and the constant μ^2 equals the constant in equation (5.3.4) times $-c^2$. Equation (5.3.5) relates frequency to wavenumber and therefore is the dispersion relation for such waves. We call waves which have no special reference frame and therefore necessarily obey equation (5.3.5) relativistic waves. The only difference in the dispersion relations between different types of relativistic waves is the value of the constant μ. The meaning of this constant will become clear later.

This page titled 5.3: Principle of Relativity Applied is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

5.4: Characteristics of Relativistic Waves

Light in a vacuum is an example of a wave for which no special reference frame exists. For light, $\mu = 0$, and we have (taking the positive root) $\omega = \mathrm{ck}$. This simply states what we know already, namely that the phase speed of light in a vacuum is c.

If $\mu \neq 0$, waves of this type are dispersive. The phase speed is

$$u_p = \frac{\omega}{k} = \left(c^2 + \mu^2/k^2\right)^{1/2} \tag{5.4.1}$$

This phase speed always exceeds c, which at first seems like an unphysical conclusion. However, the group velocity of the wave is

$$u_g = \frac{d\omega}{dk} = \frac{c^2 k}{\left(k^2 c^2 + \mu^2\right)^{1/2}} = \frac{kc^2}{\omega} = \frac{c^2}{u_p} \tag{5.4.2}$$

which is always less than c. Since wave packets and hence signals propagate at the group velocity, waves of this type are physically reasonable even though the phase speed exceeds the speed of light.

Another interesting property of such waves is that the wave four-vector is parallel to the world line of a wave packet in spacetime. This is easily shown by the following argument. As Figure $5.4.1$: shows, the spacelike component of a wave four-vector is k, while the timelike component is ω/c. The slope of the four-vector on a spacetime diagram is therefore ω/kc. However, the slope of the world line of a wave packet moving with group velocity $\mathrm{u_g}$ is $\mathrm{c/u_g} = \omega/(\mathrm{kc})$, which is the same as the slope of the $\underline{\mathrm{k}}$ four-vector.

Note that when $\mathrm{k} = 0$ we have $\omega = \mu$. In this case the group velocity of the wave is zero. For this reason we call μ the rest frequency of the wave.

This page titled 5.4: Characteristics of Relativistic Waves is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

5.5: The Doppler Shift

You have probably heard how the pitch of a train horn changes as it passes you. When the train is approaching, the pitch or frequency is higher than when it is moving away from you. This is called the Doppler shift. A similar, but distinct shift occurs if you are moving past a source of sound. If a stationary whistle is blowing, the pitch perceived from a moving car is higher while moving toward the source than when moving away. The first case thus has a moving source, while the second case has a moving observer.

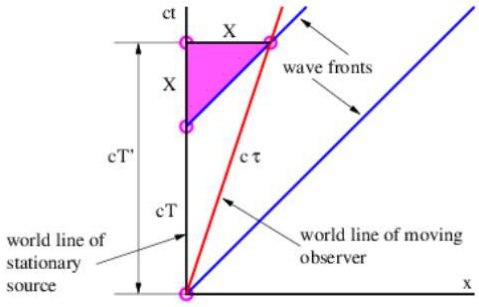


Figure 5.5.3:: Definition sketch for computing the Doppler shift for light.

In this section we compute the Doppler shift as it applies to light moving through a vacuum. Figure 5.5.3: shows the geometry for computing the time between wave fronts of light for a stationary and a moving reference frame. The time in the stationary frame is just T. Since the world lines of the wave fronts have a slope of unity, the sides of the shaded triangle have the same value, X. If the observer is moving at speed U, the slope of the observer's world line is c/U, which means that $\mathrm{c/U = (cT + X)/X}$. Solving this for X yields $\mathrm{X = UT/(1 - U/c)}$, which can then be used to compute $\mathrm{T' = T + X/c = T/(1 - U/c)}$. This formula as it stands leads to the classical Doppler shift for a moving observer. However, with relativistic velocities, one additional factor needs to be taken into account: The observer experiences time dilation since he or she is moving. The actual time measured by the observer between wave fronts is actually

$$\tau = \left(T'^2 - X^2/c^2\right)^{1/2} \tag{5.5.1}$$

$$= T\frac{\left(1 - U^2/c^2\right)^{1/2}}{1 - U/c} \tag{5.5.2}$$

$$= T\left(\frac{1 + U/c}{1 - U/c}\right)^{1/2} \tag{5.5.3}$$

where the last step uses $\mathrm{1 - U^2/c^2 = (1 - U/c)(1 + U/c)}$. From this we infer the **relativistic Doppler shift** formula for light in a vacuum:

$$\omega' = \omega\left(\frac{1 - U/c}{1 + U/c}\right)^{1/2} \tag{5.5.4}$$

where the frequency measured by the moving observer is $\omega' = 2\pi/\mathrm{T}$ and the frequency observed in the stationary frame is $\omega = 2\pi/T$.

We could go on to determine the Doppler shift resulting from a moving source. However, by the principle of relativity, the laws of physics should be the same in the reference frame in which the observer is stationary and the source is moving. Furthermore, the speed of light is still c in that frame. Therefore, the problem of a stationary observer and a moving source is conceptually the same as the problem of a moving observer and a stationary source when the wave is moving at speed c. This is unlike the case for, say, sound waves, where the stationary observer and the stationary source yield different formulas for the Doppler shift.

This page titled 5.5: The Doppler Shift is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

5.6: Addition of Velocities

Figure 5.6.4: shows the world line of a moving object from the point of view of two different reference frames, with the primed frame (left panel) moving to the right at speed U relative to the unprimed frame (right panel). The goal is to calculate the velocity of the object relative to the unprimed frame, v, assuming its velocity in the primed frame, v' is known. The classical result is simply

$$v = U + v' \quad \text{(classical result)}. \tag{5.6.1}$$

However, this is inconsistent with the speed of light being constant in all reference frames, since if we substitute c for v′, this formula predicts that the speed of light in the unprimed frame is U + c.

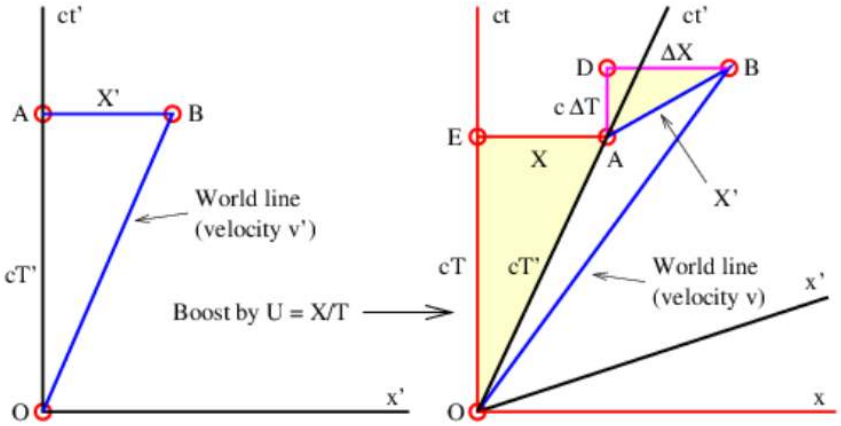


Figure 5.6.4:: Definition sketch for relativistic velocity addition. The two panels show the world line of a moving object relative to two different reference frames moving at velocity U with respect to each other. The velocity of the world line in the left panel is v' while its velocity in the right panel is v.

We can use the geometry of Figure 5.6.4: to come up with the correct relativistic formula. From the right panel of this figure we infer that

$$\frac{v}{c} = \frac{X + \Delta X}{cT + c\Delta T} = \frac{X/(cT) + \Delta X/(cT)}{1 + \Delta T/T} \tag{5.6.2}$$

This follows from the fact that the slope of the world line of the object in this frame is c/v. The slope is calculated as the ratio of the rise, $c(T + \Delta T)$ to the run, $X + \Delta X$.

From the left panel of Figure 5.6.4: we similarly see that

$$\frac{v'}{c} = \frac{X'}{cT'} \tag{5.6.3}$$

However, we can apply our calculations of Lorentz contraction and time dilation from the previous chapter to triangles ABD and OAE in the right panel. The slope of AB is U/c because AB is horizontal in the left panel, so $\mathrm{X'} = \Delta \mathrm{X}\left(1 - \mathrm{U}^2/\mathrm{c}^2\right)^{1/2}$. Similarly, the slope of OA is c/U since OA is vertical in the left panel, and $\mathrm{T'} = \mathrm{T}\left(1 - \mathrm{U}^2/\mathrm{c}^2\right)^{1/2}$. Substituting these formulas into the equation for v′/c yields

$$\frac{v'}{c} = \frac{\Delta X}{cT} \tag{5.6.4}$$

Again using what we know about the triangles ABD and OAE, we see that

$$\frac{U}{c} = \frac{c\Delta T}{\Delta X} = \frac{X}{cT} \tag{5.6.5}$$

Finally, we calculate $\Delta T/T$ by noticing that

$$\frac{\Delta T}{T} = \frac{\Delta T}{T}\frac{c\Delta X}{c\Delta X} = \left(\frac{c\Delta T}{\Delta X}\right)\left(\frac{\Delta X}{cT}\right) = \frac{U}{c}\frac{v'}{c} \tag{5.6.6}$$

Substituting equations (5.6.4), (5.6.5), and (5.6.6) into equation (5.6.2) and simplifying yields the relativistic velocity addition formula:

$$v = \frac{U + v'}{1 + Uv'/c^2} \qquad \text{(special relativity)} \tag{5.6.7}$$

Notice how this equation behaves in various limits. If $|Uv'| \ll c^2$, the denominator of equation (5.6.7) is nearly unity, and the special relativistic formula reduces to the classical case. On the other hand, if $\mathrm{v}' = \mathrm{c}$, then equation (5.6.7) reduces to $v = c$. In other words, if the object in question is moving at the speed of light in one reference frame, it is moving at the speed of light in all reference frames, i. e., for all possible values of U. Thus, we have found a velocity addition formula that 1) reduces to the classical formula for low velocities and 2) gives the observed results for very high velocities as well.

Equation (5.6.7) is valid even if v' is negative, i. e., if the object is moving to the right less rapidly than the primed reference frame, or even if it is moving to the left in the unprimed frame.

This page titled 5.6: Addition of Velocities is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

5.7: Problems

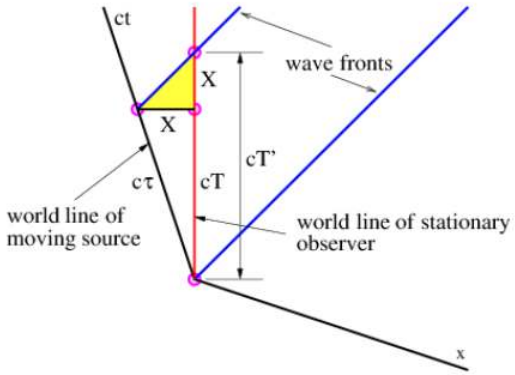


Figure 5.7.5:: Doppler shift for a moving light source.

1. Sketch the wave fronts and the k four-vector in a spacetime diagram for the case where $\omega/k = 2c$. Label your axes and space the wave fronts correctly for the case $\mathrm{k} = 4n\ \mathrm{m}^{-1}$.
2. If the four-vector $\mathrm{k} = (0, 1\mathrm{nm}^{-1})$ in the rest frame, find the space and time components of k in a frame moving to the left at speed $\mathrm{c}/2$.
3. Let's examine the four-vector $\underline{\mathrm{u}} = (\mathrm{u}_9, \mathrm{c})\,(1-\beta^2)^{1/2}$ where $\beta = \mathrm{u}_g/\mathrm{c}, \mathrm{u}_\mathrm{g}$ being the velocity of some object. The four-vector u is called the four-velocity.
    1. Show that u is parallel to the world line of the object.
    2. Show that $\mathrm{u}\cdot\mathrm{u} = -\mathrm{c}^2$.
    3. If u_g is the group velocity of a relativistic wave packet, show that $\mathrm{k} = (\mu/\mathrm{c}^2)\,\mathrm{u}$, where k is the central wave four-vector of the wave packet.
4. Find the Doppler shift for a moving source of light from Figure 5.7.5:, roughly following the procedure used in the text to find the shift for a moving observer. (Assume that the source moves to the left at speed U.) Is the result the same as for the moving observer, as demanded by the principle of relativity?
5. Find the Doppler shift for a stationary source of sound ($c \approx 380\ \mathrm{m\ s}^{-1}$ in a stationary atmosphere) as measured by a moving observer. Follow the procedure as for light in section 5.5 except note that the proper time τ is virtually the same as the time T′ in Figure 5.7.3: since speeds are much less than the speed of light.
    1. Suppose you shine a laser with frequency ω and wavenumber k on a mirror moving toward you at speed v, as seen in Figure 5.7.6:. What are the frequency ω' and wavenumber k' of the reflected beam? Hint: Find the frequency of the incident beam in the reference frame of the mirror. The frequency of the reflected beam will be the same as that of the incident beam in this frame. Then transform back to the reference frame of the laser.

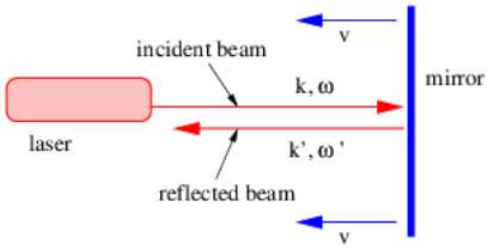


Figure 5.7.6: Laser beam reflecting off of a moving mirror.

6. Suppose the moving twin in the twin paradox has a powerful telescope so that she can watch her twin brother back on earth during the entire trip. Describe how the earthbound twin appears to age to the travelling twin compared to her own rate of aging. Use a spacetime diagram to illustrate your argument and consider separately the outbound and return legs. Remember that light travels at the speed of light! Hint: Does the concept of Doppler shift help here?
7. Find the velocity of an object with respect to the rest frame if it is moving at a velocity of $0.1c$ with respect to another frame which itself is moving in the same direction at $0.1c$ relative to the rest frame using
    1. the Galilean formula and
    2. the formula of special relativity. Determine the fractional error made in using the Galilean formula.

8. Each stage of a high performance 3 stage rocket can accelerate to a speed of 0.9c from rest. If the rocket starts from rest, how fast does the final stage eventually go?
    1. the Galilean formula and
    2. the formula of special relativity. Determine the fractional error made in using the Galilean formula.
9. An interstellar spaceship is going from Earth to Sirius with speed U = 0.8c relative to the rest frame. It passes a spaceship which is going from Sirius to Earth at a speed of 0.95c in the reference frame of the first spaceship. What is the velocity (direction and speed) of the second spaceship in the rest frame?
10. An interstellar spaceship is going from Earth to Sirius with speed $U = 0.8c$ relative to the rest frame. It passes a spaceship which is going from Sirius to Earth at a speed of $0.95c$ in the reference frame of the first spaceship. What is the velocity (direction and speed) of the second spaceship in the rest frame?

5.7: Problems is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

CHAPTER OVERVIEW

6: Acceleration and General Relativity

General relativity is Einstein's extension of special relativity to include gravity. An important aspect of general relativity is that spacetime is no longer necessarily flat, but in fact may be curved under the influence of mass. Understanding curved spacetime is an advanced topic which is not easily accessible at the level of this text. However, it turns out that some insight into general relativistic phenomena may be obtained by investigating the effects of acceleration in the flat (but non-Euclidean) space of special relativity.

The central assumption of general relativity is the equivalence principle, which states that gravity is a force which arises from being in an accelerated reference frame. To understand this we must first investigate the concept of acceleration. We then see how this leads to phenomena such as the gravitational red shift, event horizons, and black holes. We also introduce in a preliminary way the notions of force and mass.

6.1: Acceleration
6.2: Circular Motion
6.3: Acceleration, Force, and Mass
6.4: Acceleration in Special Relativity
6.5: Accelerated Reference Frames
6.6: Gravitational Red Shift
6.7: Event Horizons
6.8: Problems

This page titled 6: Acceleration and General Relativity is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

6.1: Acceleration

Imagine that you are in a powerful luxury car stopped at a stoplight. As you sit there, gravity pushes you into the comfortable leather seat. The light turns green and you “floor it”. The car accelerates and an additional force pushes you into the seat back. You round a curve, and yet another force pushes you toward the outside of the curve. (But the well designed seat and seat belt keep you from feeling discomfort!)

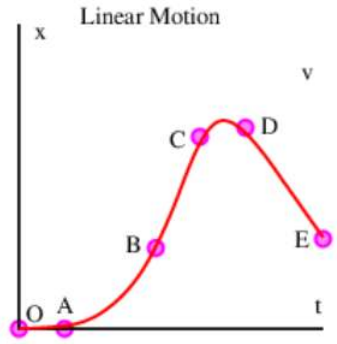


Figure 6.1.1:: Example of linear motion.

Let us examine the idea of acceleration more closely. Considering first acceleration in one dimension, Figure 6.1.1: shows the position of an object as a function of time, $\mathrm{x(t)}$. The velocity is simply the time rate of change of the position:

$$v(t) = \frac{dx(t)}{dt} \tag{6.1.1}$$

The acceleration is the time rate of change of velocity:

$$a(t) = \frac{dv(t)}{dt} = \frac{d^2x(t)}{dt^2} \tag{6.1.2}$$

In Figure 6.1.1:, only the segment OA has zero velocity. Velocity is increasing in AB, and the acceleration is positive there. Velocity is constant in BC, which means that the acceleration is zero. Velocity is decreasing in CD, and the acceleration is negative. Finally, in DE, the velocity is negative and the acceleration is zero.

In two or three dimensions, position **x**, velocity **v**, and acceleration **a** are all vectors, so that the velocity is

$$\mathbf{v}(t) = \frac{d\mathbf{x}(t)}{dt} \tag{6.1.3}$$

while the acceleration is

$$\mathbf{a}(t) = \frac{d\mathbf{v}(t)}{dt} \tag{6.1.4}$$

Thus, over some short time interval Δt, the changes in **x** and **v** can be written

$$\Delta\mathbf{x} = \mathbf{v}\Delta t \quad \Delta\mathbf{v} = \mathbf{a}\Delta t \tag{6.1.5}$$

These are vector equations, so the subtractions implied by the “delta” operations must be done vectorially. An example where the vector nature of these quantities is important is motion in a circle at constant speed, which is discussed in the next section.

This page titled 6.1: Acceleration is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

6.2: Circular Motion

Imagine an object constrained by an attached string to move in a circle at constant speed, as shown in the left panel of Figure 6.2.2:. We now demonstrate that the acceleration of the object is toward the center of the circle. The acceleration in this special case is called the **centripetal acceleration**.

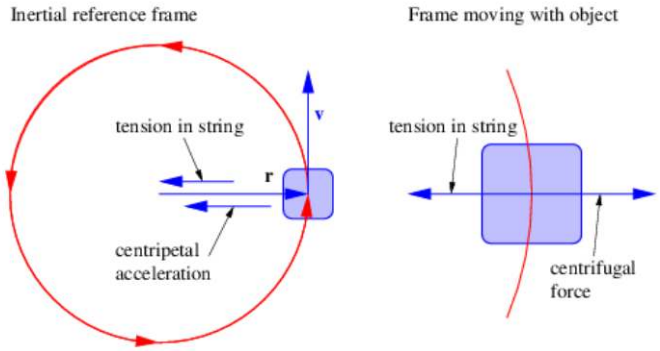


Figure 6.2.2:: Two different views of circular motion of an object. The left panel shows the view from the inertial reference frame at rest with the center of the circle. The tension in the string is the only force and it causes an acceleration toward the center of the circle. The right panel shows the view from an accelerated frame in which the object is at rest. In this frame the tension in the string balances the centrifugal force, which is the inertial force arising from being in an accelerated reference frame, leaving zero net force.

Figure 6.2.3: shows the position of the object at two times spaced by the time interval Δt. The position vector of the object relative to the center of the circle rotates through an angle $\Delta\theta$ during this interval, so the angular rate of revolution of the object about the center is $\omega = \Delta\theta/\Delta t$. The magnitude of the velocity of the object is v, so the object moves a distance $v\Delta t$ during the time interval. To the extent that this distance is small compared to the radius r of the circle, the angle $\Delta\theta = v\Delta t/r$. Solving for v and using $\omega = \Delta\theta/\Delta t$, we see that

$$v = \omega r \quad (\text{ circular motion }). \tag{6.2.1}$$

The direction of the velocity vector changes over this interval, even though the magnitude v stays the same. Figure 6.2.3: shows that this change in direction implies an acceleration a which is directed toward the center of the circle, as noted above. The magnitude of the vectoral change in velocity in the time interval Δ is $a\Delta t$ . . Since the angle between the initial and final velocities is the same as the angle $\Delta\theta$ between the initial and final radius vectors, we see from the geometry of the triangle in Figure 6.2.3: that $a\Delta t/v = \Delta\theta$. Solving for a results in

$$a = \omega v \quad (\text{ circular motion }). \tag{6.2.2}$$

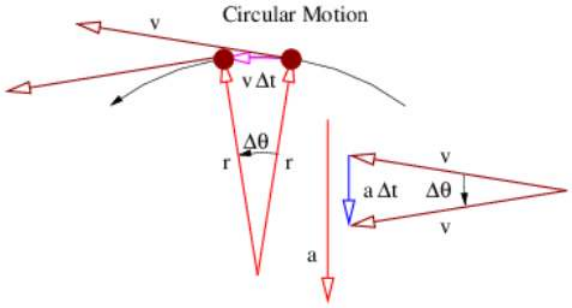


Figure 6.2.3:: Definition sketch for computing centripetal acceleration.

Combining equations (6.2.1) and (6.2.2) yields the equation for centripetal acceleration:

$$a = \omega^2 r = v^2/r \quad (\text{centripetal acceleration}). \tag{6.2.3}$$

The second form is obtained by eliminating ω from the first form using equation (6.2.1).

This page titled 6.2: Circular Motion is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

6.3: Acceleration, Force, and Mass

We have a good intuitive feel for the concepts of force and mass because they are very much a part of our everyday experience. We think of force as how hard we push on something. Mass is the resistance of an object to acceleration if it is otherwise free to move. Thus, pushing on a bicycle on a smooth, level road causes it to accelerate more readily than pushing on a car. We say that the car has greater mass. We can summarize this relationship with Newton's second law

$$F = ma \tag{6.3.1}$$

where F is the total force on an object, m is its mass, and a is the acceleration resulting from the force.

Three provisos apply to equation (6.3.1). First, it only makes sense in unmodified form when the velocity of the object is much less than the speed of light. For relativistic velocities it is best to write this equation in a slightly different form which we introduce later. Second, the force must be the total force, including all frictional and other incidental forces which might otherwise be neglected by an uncritical observer. Third, it only works in a reference frame which itself is unaccelerated, i.e., an *inertial reference frame*. We deal below with accelerated reference frames.

This page titled 6.3: Acceleration, Force, and Mass is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

6.4: Acceleration in Special Relativity

As noted above, acceleration is just the time rate of change of velocity. We use the above results to determine how acceleration transforms from one reference frame to another. Figure 6.4.4: shows the world line of an accelerated reference frame, with a time-varying velocity $\mathrm{U(t)}$ relative to the unprimed inertial rest frame. Defining $\Delta \mathrm{U} = \mathrm{U(T)} - \mathrm{U}(0)$ as the change in the velocity of the accelerated frame (relative to the unprimed frame) between events A and C, we can relate this to the change of velocity, $\Delta \mathrm{U}'$, of the accelerated frame relative to an inertial frame moving with the initial velocity, $\mathrm{U}(0)$.

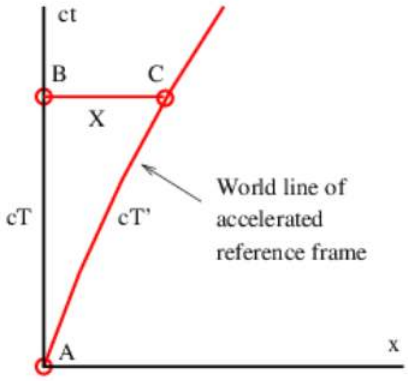


Figure 6.4.4:: World line of the origin of an accelerated reference frame.

Applying the equation for the relativistic addition of velocities, we find

$$U(T) = U(0) + \Delta U = \frac{U(0) + \Delta U'}{1 + U(0)\Delta U'/c^2} \tag{6.4.1}$$

We now note that the mean acceleration of the reference frame between events A and C in the rest frame is just $\mathrm{a} = \Delta \mathrm{U}/\mathrm{T}$, whereas the mean acceleration in the primed frame between the same two events is $\mathrm{a}' = \Delta \mathrm{U}'/\mathrm{T}'$. From equation (6.4.1) we find that

$$\Delta U = \frac{\Delta U' \left[1 - U(0)^2/c^2\right]}{1 + U(0)\Delta U'/c^2} \tag{6.4.2}$$

and the acceleration of the primed reference frame as it appears in the unprimed frame is

$$a = \frac{\Delta U}{T} = \frac{\Delta U' \left[1 - U(0)^2/c^2\right]}{T\left[1 + U(0)\Delta U'/c^2\right]} \tag{6.4.3}$$

Since we are interested in the instantaneous rather than the average acceleration, we let T become small. This has three consequences. First, $\Delta \mathrm{U}$ and $\Delta \mathrm{U}'$ become small, which means that the term $\mathrm{U}(0)\Delta \mathrm{U}'/\mathrm{c}^2$ in the denominator of Equation 6.4.3 can be ignored compared to 1. This means that

$$a \approx \frac{\Delta U' \left[1 - U(0)^2/c^2\right]}{T} \tag{6.4.4}$$

with the approximation becoming exact as $\mathrm{T} \rightarrow 0$. Second, the "triangle" with the curved side in Figure 6.4.4: becomes a true triangle, with the result that $\mathrm{T}' = \mathrm{T}\left[1 - \mathrm{U}(0)^2/\mathrm{c}^2\right]^{1/2}$. The acceleration of the primed frame with respect to an inertial frame moving at speed $\mathrm{U}(0)$ can therefore be written

$$a' = \frac{\Delta U'}{T'} = \frac{\Delta U'}{T\left[1 - U(0)^2/c^2\right]^{1/2}} \tag{6.4.5}$$

Third, we can replace U(0) with U, since the velocity of the accelerated frame doesn't change very much over a short time interval.

Dividing equation (6.4.4) by equation (6.4.5) results in a relationship between the two accelerations:

$$a = a'\left(1 - U^2/c^2\right)^{3/2} \tag{6.4.6}$$

which shows that the acceleration of a rapidly moving object, a, as observed from the rest frame, is less than its acceleration relative to an inertial reference frame in which the object is nearly stationary, a', by the factor $\left(1-U^2/c^2\right)^{3/2}$. We call this latter acceleration the intrinsic acceleration. This difference in observed acceleration between the two inertial reference frames is purely the result of the geometry of space time, but it has interesting consequences.

Identifying a with $\mathrm{dU/dt}$, we can integrate the acceleration equation assuming that the intrinsic acceleration a' is constant and that the velocity $U=0$ at time $t=0$. We get the following result (verify this by differentiating with respect to time):

$$a't = \frac{U}{\left(1-U^2/c^2\right)^{1/2}} \tag{6.4.7}$$

which may be solved for $\mathrm{U/c}$:

$$\frac{U}{c} = \frac{a't/c}{\left[1+\left(a't/c\right)^2\right]^{1/2}} \tag{6.4.8}$$

This is plotted in Figure 6.4.5:. Classically, the velocity U would reach the speed of light when $a't/c=1$. However, as figure 6.5 shows, the rate at which the velocity increases with time slows as the object moves faster, such that U approaches c asymptotically, but never reaches it.

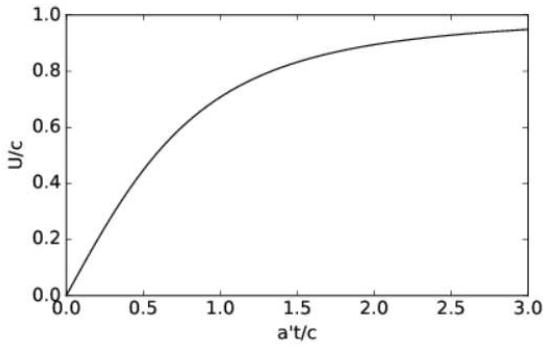


Figure 6.4.5:: Velocity divided by the speed of light as a function of the product of the time and the (constant) acceleration divided by the speed of light.

The results for this section are valid only for acceleration components in the direction of motion. The components perpendicular to this direction behave differently and are treated in more advanced texts.

This page titled 6.4: Acceleration in Special Relativity is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

6.5: Accelerated Reference Frames

Referring back to the forces being felt by the occupant of a car, it is clear that the forces associated with accelerations are directed opposite the accelerations and proportional to their magnitudes. For instance, when accelerating away from a stoplight, the acceleration is forward and the perceived force is backward. When turning a corner, the acceleration is toward the corner while the perceived force is away from the corner. Such forces are called *inertial forces*.

The origin of these forces can be understood by determining how acceleration changes when one observes it from a reference frame which is itself accelerated. Suppose that the primed reference frame is accelerating to the right with acceleration *A* relative to the unprimed frame. The position *x*′ in the primed frame can be related to the position *x* in the unprimed frame by

$$x = x' + X \tag{6.5.1}$$

where *X* is the position of the origin of the primed frame in the unprimed frame. Taking the second time derivative, we see that

$$a = a' + A \tag{6.5.2}$$

where $\mathrm{a} = \mathrm{d}^2\mathrm{x}/\mathrm{dt}^2$ is the acceleration in the unprimed frame and $a' = d^2x'/dt^2$ is the acceleration according to an observer in the primed frame.

We now substitute this into equation (6.9) and move the term involving A to the left side:

$$F - mA = ma' \tag{6.5.3}$$

This shows that Newton's second law represented by equation (6.9) is not valid in an accelerated reference frame, because the total force F and the acceleration a' in this frame don't balance as they do in the unaccelerated frame; the additional term $-mA$ messes up this balance.

We can fix this problem by considering $-mA$ to be a type of force, in which case we can include it as a part of the total force F. This is the inertial force which we mentioned above. Thus, to summarize, we can make Newton's second law work when objects are observed from accelerated reference frames if we include as part of the total force an inertial force which is equal to $-mA$, A being the acceleration of the reference frame of the observer and m the mass of the object being observed.

The right panel of Figure 6.5.2: shows the inertial force observed in the reference frame of an object moving in circular motion at constant speed. In the case of circular motion the inertial force is called the centrifugal force. It points away from the center of the circle and just balances the tension in the string. This makes the total force on the object zero in its own reference frame, which is necessary since the object cannot move (or accelerate) in this frame.

General relativity says that gravity is nothing more than an inertial force. This was called the equivalence principle by Einstein. Since the gravitational force on the Earth points downward, it follows that we must be constantly accelerating upward as we stand on the surface of the Earth! The obvious problem with this interpretation of gravity is that we don't appear to be moving away from the center of the Earth, which would seem to be a natural consequence of such an acceleration. However, relativity has a surprise in store for us here.

It follows from the above considerations that something can be learned about general relativity by examining the properties of accelerated reference frames. In particular, we can gain insight into the above apparent paradox. Equation (6.17) shows that the velocity of an object undergoing constant intrinsic acceleration a (note that we have dropped the "prime" from a to simplify the notation) is

$$v = \frac{dx}{dt} = \frac{at}{\left[1 + (at/c)^2\right]^{1/2}} \tag{6.5.4}$$

where t is the time and c is the speed of light. A function $x(t)$ which satisfies equation (6.5.4) is

$$x(t) = \left(c^2/a\right)\left[1 + (at/c)^2\right]^{1/2} \tag{6.5.5}$$

(Verify this by differentiating it.) The interval OB in Figure 6.5.6: is of length $x(0) = c^2/A$.

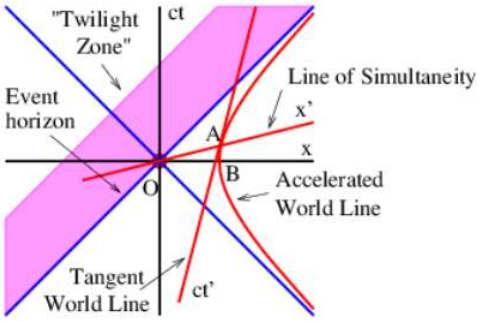


Figure 6.5.6:: Spacetime diagram showing the world line of the origin of a reference frame undergoing constant acceleration.

The slanted line OA is a line of simultaneity associated with the unaccelerated world line tangent to the accelerated world line at point A. This line of simultaneity goes through the origin, as is shown in Figure 6.5.6:. To demonstrate this, multiply equations (6.5.4) and (6.5.5) together and solve for v/c:

$$v/c = ct/x. \tag{6.5.6}$$

From Figure 6.5.6: we see that ct/x is the slope of the line OA, where (x, ct) are the coordinates of event A. Equation (6.23) shows that this line is indeed the desired line of simultaneity, since its slope is the inverse of the slope of the world line, c/v. Since there is nothing special about the event A, we infer that all lines of simultaneity associated with the accelerated world line pass through the origin.

We now inquire about the length of the invariant interval OA in Figure 6.5.6:. Recalling that $I^2 = x^2 - c^2t^2$ and using equation (6.5.5), we find that the length of OA is

$$I = \left(x^2 - c^2t^2\right)^{1/2} = \left(c^4/a^2\right)^{1/2} = c^2/a \tag{6.5.7}$$

which is the same as the length of the interval OB. By extension, *all* events on the accelerated world line are the same invariant interval from the origin. Recalling that the interval along a line of simultaneity is the distance in the associated reference frame, we reach the astonishing conclusion that even though the object associated with the curved world line in Figure 6.5.6: is accelerating away from the origin, it always remains the same distance (in its own frame) from the origin.

The analogy between this problem and the apparent paradox in which one remains a fixed distance from the center of the earth while accelerating away from it is not perfect. In particular, the earth case depends on the existence of the earth's mass.

This page titled 6.5: Accelerated Reference Frames is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

6.6: Gravitational Red Shift

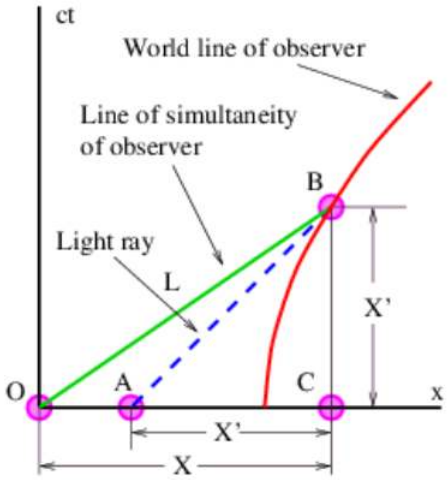


Figure 6.6.7:: Spacetime diagram for explaining the gravitational red shift. Why is the interval AC equal to the interval BC? L is the length of the invariant interval OB.

Light emitted at a lower level in a gravitational field has its frequency reduced as it travels to a higher level. This phenomenon is called the gravitational red shift. Figure 6.26 shows why this happens. Since experiencing a gravitational force is equivalent to being in an accelerated reference frame, we can use the tools of special relativity to view the process of light emission and absorption from the point of view of the unaccelerated or inertial frame. In this reference frame the observer of the light is accelerating to the right, as indicated by the curved world line in Figure 6.6.26:, which is equivalent to a gravitational force to the left. The light is emitted at point A with frequency ω by a source which is stationary at this instant. At this instant the observer is also stationary in this frame. However, by the time the light gets to the observer, he or she has a velocity to the right which means that the observer measures a Doppler shifted frequency ω′ for the light. Since the observer is moving away from the source, ω′ < ω, as indicated above.

The relativistic Doppler shift is given by

$$\frac{\omega'}{\omega} = \left(\frac{1 - U/c}{1 + U/c}\right)^{1/2} \tag{6.6.1}$$

so we need to compute $\mathrm{U/c}$. The line of simultaneity for the observer at point B goes through the origin, and is thus given by line segment OB in Figure 6.6.26:. The slope of this line is $\mathrm{U/c}$, where U is the velocity of the observer at point B. From the figure we see that this slope is also given by the ratio $\mathrm{X'/X}$. Equating these, eliminating X in favor of $\mathrm{L} = \left(\mathrm{X}^2 - \mathrm{X'}^2\right)^{1/2} = \mathrm{c}^2/\mathrm{g}$, which is the actual invariant distance of the observer from the origin, and substituting into equation (6.6.1) results in our gravitational red shift formula:

$$\frac{\omega'}{\omega} = \left(\frac{X - X'}{X + X'}\right)^{1/2} = \left(\frac{\left(L^2 + X'^2\right)^{1/2} - X'}{\left(L^2 + X'^2\right)^{1/2} + X'}\right)^{1/2} \tag{6.6.2}$$

If X′ = 0, then there is no redshift, because the source is collocated with the observer. On the other hand, if the source is located at the origin, so X′ = X, the Doppler shifted frequency is zero. In addition, the light never gets to the observer, since the world line is asymptotic to the light world line passing through the origin. If the source is at a higher level in the gravitational field than the observer, so that X′ < 0, then the frequency is shifted to a higher value, i. e., it becomes a “blue shift”.

Equation (6.6.2) works for more complex geometries than that associated with an accelerated reference frame, e.g., for the gravitational field g associated with a star, as long as $|X'| < c^2/g$. In this case L is no longer the distance to the center of the star but remains equal to c^2/g.

This page titled 6.6: Gravitational Red Shift is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

6.7: Event Horizons

The 45˚ diagonal line passing through the origin in Figure 6.7.6: is called the event horizon for the accelerated observer in this figure. Notice that light from the “twilight zone” above and to the left of the event horizon cannot reach the accelerated observer. However, the reverse is not true — a light signal emitted to the left by the observer can cross the event horizon into the “twilight zone”. The event horizon thus has a peculiar one-way character — it passes signals from right to left, but not from left to right.

6.7: Event Horizons is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

6.8: Problems

1. An object moves as described in Figure 6.8.8:, which shows its position x as a function of time t.
    1. Is the velocity positive, negative, or zero at each of the points A, B, C, D, E, and F?
    2. Is the acceleration positive, negative, or zero at each of the points A, B, C, D, E, and F?

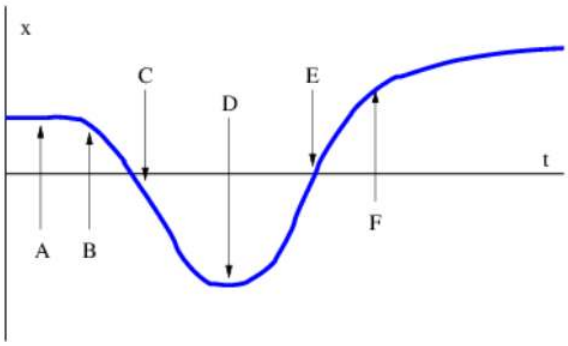
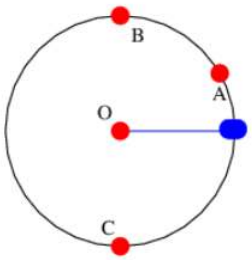


Figure 6.8.8**::** Position of an object as a function of time.

2. An object is moving counterclockwise at constant speed around the circle shown in Figure 6.8.9: due to the fact that it is attached by a string to the center of the circle at point O.
    1. Sketch the object's velocity vectors at points A, B, and C.
    2. Sketch the object's acceleration vectors at points A, B, and C.



    3. If the string breaks at point A, sketch the subsequent trajectory followed by the object.

Figure 6.8.9**::** Object in circular motion.

3. How fast are you going after accelerating from rest with intrinsic acceleration $a = 10ms^{-2}$ after the given times measured in the rest frame:
    1. 10 y?
    2. 100000 y? Express your answer as the speed of light minus your actual speed. Hint: You may have a numerical problem on the second part, which you should try to resolve using the approximation $(1+\epsilon)^{X} \approx 1 + x\epsilon$, which is valid for $|\epsilon| \ll 1$.
4. An object's world line is defined by $x(t) = (d^2 + c^2t^2)^{1/2}$ where d is a constant and c is the speed of light.
    1. Find the object's velocity as a function of time.
    2. Using the above result, find the slope of the tangent to the world line as a function of time.
    3. Find where the line of simultaneity corresponding to each tangent world line crosses the x axis.
5. A car accelerates in the positive x direction at 3 m s^{-2}.
    1. What is the net force on a 100 kg man in the car as viewed from an inertial reference frame?
    2. What is the inertial force experienced by this man in the reference frame of the car?
    3. What is the net force experienced by the man in the car's (accelerated) reference frame?
6. A person is sitting in a comfortable chair in her home in Bogotá, Colombia, which is essentially on the equator.
    1. What would the rotational period of the earth have to be to make this person weightless?
    2. What is her acceleration according to the equivalence principle in the earth frame in this situation?
7. At time t = 0 a Zork (a creature from the planet Zorkheim) accelerating to the right at a = 10^3 m s^{-2} in a spaceship accidently drops its stopwatch from the spaceship just when its velocity is zero.
    1. Describe qualitatively how the hands of the watch appear to move to the Zork as it observes the watch through a powerful telescope.

2. After a very long time what does the watch read? Hint: Draw a spacetime diagram with the world lines of the spaceship and the watch. Then send light rays from the watch to the spaceship.

8. Using a spacetime diagram, show why signals from events on the hidden side of the event horizon from an accelerating spaceship cannot reach the spaceship.
9. Approximate equation (6.26) to first order in X′ for the case in which $X' \ll L$.
10. Imagine two identical clocks, one on top of the volcano Chimborazo in Ecuador (6300 m above sea level), the other in the Ecuadorian city of Guayaquil (at sea level).
    1. From the perspective of Chimborazo, does the clock in Guayaquil appear to be running faster or slower than the Chimborazo clock? Explain.
    2. Compute the fractional frequency difference $(\omega - \omega')/\omega$ in this case, where ω is the freqency of the Guayaquil clock as observed in Guayaquil (and the frequency of the Chimborazo clock on Chimborazo) and ω' is the frequency of the Guayaquil clock as observed from Chimborazo. You may wish to use the results of the previous problem.

6.8: Problems is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

CHAPTER OVERVIEW

7: Matter Waves

We begin our study of quantum mechanics by discussing the diffraction undergone by X-rays and electrons when they interact with a crystal. X-rays are a form of electromagnetic radiation with wavelengths comparable to the distances between atoms. Scattering from atoms in a regular crystalline structure results in an interference pattern which is in many ways similar to the pattern from a diffraction grating. We first develop Bragg's law for diffraction of X-rays from a crystal. Two practical techniques for doing X-ray diffraction are then described.

It turns out that electrons have wave-like properties and also undergo Bragg diffraction by crystals. Bragg diffraction thus provides a crucial bridge between the worlds of waves and particles. With this bridge we introduce the classical ideas of momentum and energy by relating them to the wave vector and frequency of a wave. The properties of waves also give rise to the Heisenberg uncertainty principle.

Table 7.1 shows a table of the Nobel prizes associated with the ideas presented in this chapter. This gives us a feel for the chronology of these discoveries and indicates how important they were to the development of physics in the early 20th century.

Table 7.1: Selected Nobel prize winners, year of award, and contribution.

Year	Recipient	Contribution
1901	W. K. Röntgen	Discovery of X-rays
1906	J. J. Thomson	Discovery of electron
1914	M. von Laue	X-ray diffraction in crystals
1915	W. and L. Bragg	X-ray analysis of crystal structure
1918	M. Planck	Energy quantization
1921	A. Einstein	Photoelectric effect
1922	N. Bohr	Structure of atoms
1929	L.-V. de Broglie	Wave nature of electrons
1932	W. Heisenberg	Quantum mechanics
1933	Schrödinger and Dirac	Atomic theory
1937	Davisson and Thomson	Electron diffraction in crystals

7.1: Bragg's Law
7.2: X-Ray Diffraction Techniques
7.3: Meaning of Quantum Wave Function
7.4: Sense and Nonsense in Quantum Mechanics
7.5: Mass, Momentum, and Energy
7.6: Heisenberg Uncertainty Principle
7.7: Problems

This page titled 7: Matter Waves is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

7.1: Bragg's Law

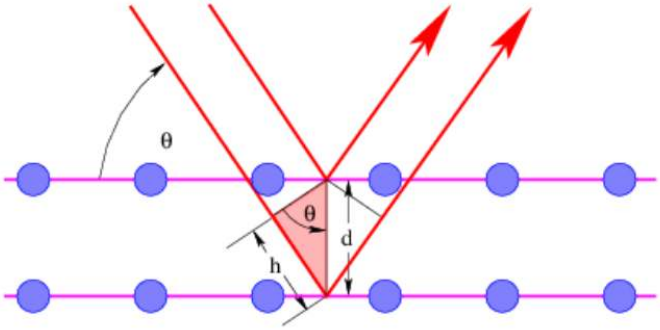


Figure 7.1.1:: Schematic diagram for determining Bragg's law.

Figure 7.1.1: schematically illustrates interference between waves scattering from two adjacent rows of atoms in a crystal. The net effect of scattering from a single row is equivalent to partial reflection from a mirror imagined to be aligned with the row. Thus, the angle of "reflection" equals the angle of incidence for each row. Interference then occurs between the beams reflecting off different rows of atoms in the crystal.

For the two adjacent rows shown in Figure 7.1.1:, the path difference between beams is $2h = 2d\sin\theta$. For constructive interference this must be an integer number of wavelengths, $m\lambda$, where the integer m is called the *order of interference*. The result is Bragg's law of diffraction:

$$m\lambda = 2d\sin\theta, \quad m = 1, 2, 3 \ldots \quad \text{(Bragg's law)} \tag{7.1.1}$$

If only two rows are involved, the transition from constructive to destructive interference as θ changes is gradual. However, if interference from many rows occurs, then the constructive interference peaks become very sharp with mostly destructive interference in between. This sharpening of the peaks as the number of rows increases is very similar to the sharpening of the diffraction peaks from a diffraction grating as the number of slits increases.

This page titled 7.1: Bragg's Law is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

7.2: X-Ray Diffraction Techniques

Two types of targets are used in Bragg diffraction experiments: single crystals and powder targets.

Single Crystal

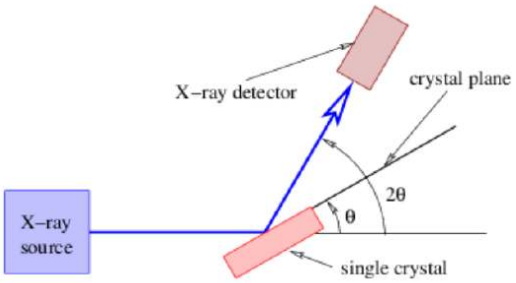


Figure 7.2.2:: Setup for single crystal Bragg diffraction.

In a single crystal setup, an X-ray detector is mounted as shown in Figure 7.2.2:. A mechanical device keeps the detector oriented so that the angle of incidence equals the angle of reflection for the desired crystal plane. Peaks in the X-ray detection rate are sought as the angle θ is varied.

The advantage of this type of apparatus is that diffraction peaks from only the selected crystal plane are observed.

Powder Target

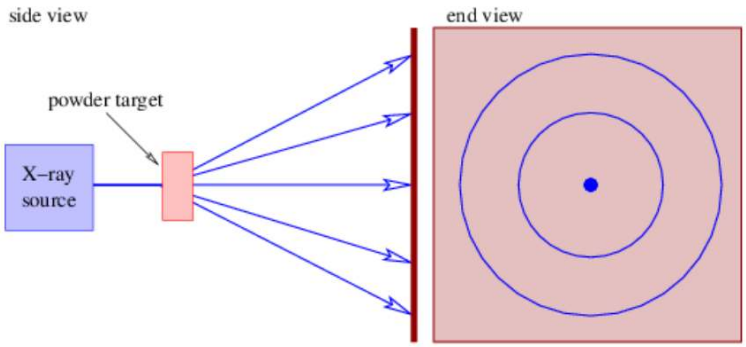


Figure 7.2.3:: Setup for powder target Bragg diffraction.

The powder in a powder target is really a conglomeration of many tiny crystals randomly oriented. Thus, for each possible Bragg diffraction angle there are crystals oriented correctly for Bragg diffraction to take place. The detector is usually a photographic plate or an equivalent electronic device as illustrated in Figure 7.2.3:. For each Bragg diffraction angle one sees a ring on the plate concentric with the axis of the incident X-ray beam.

The advantage of this type of system is that no knowledge is needed of the crystal plane orientations. Furthermore, a single large crystal is not required. However, all possible Bragg scattering angles are seen at once, which can lead to confusion in the interpretation of the results.

This page titled 7.2: X-Ray Diffraction Techniques is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

7.3: Meaning of Quantum Wave Function

Bragg diffraction illustrates the most difficult thing to understand about quantum mechanics, namely that particles can have wave-like properties and waves can have particle-like properties.

The variation of X-ray intensity with angle seen in a Bragg diffraction apparatus is very difficult to explain in any terms other than wave interference. Yet, X-rays are typically detected by a device such as a Geiger counter which produces a pulse of electricity for each X-ray particle, or photon, which hits it. Thus, X-rays sometimes act like particles and sometimes like waves.

Light isn't alone in having both particle and wave properties. Davisson and Germer and later G. P. Thomson (son of J. J. Thomson, the discoverer of the electron) showed that electrons also can act like waves. They did this by demonstrating that electrons undergo Bragg diffraction in crystals, much as X-rays do.

Most physicists (including Albert Einstein) have found quantum mechanics to be extremely bizarre, so if you feel the same way, you are in good company! However, there is a useful interpretation of quantum mechanics which at least allows us to get on with using it to solve problems, even though it may not satisfy our intuitive reservations about the theory.

The displacement of the matter wave associated with a particle is usually called the wave function, Ψ. It is not at all clear what Ψ is a displacement of, but its use is straightforward. The absolute square of the wave function, $|\Psi(x,t)|^2$, is proportional to the probability of finding the associated particle at position x and time t. The absolute square is taken because under many circumstances the wave function is actually complex, i. e., it has both real and imaginary parts. The reasons for this will be discussed later.

Due to the interpretation of the wave function, quantum mechanics is a probabilistic theory. It does not tell us with certainty what happens to a particular particle. Instead, it tells us the probability for detecting the particle in any given location. If many experiments are done, with one particle per experiment, the numbers of experiments with particles being detected in the various possible locations are in proportion to the quantum mechanical probabilities.

This page titled 7.3: Meaning of Quantum Wave Function is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

7.4: Sense and Nonsense in Quantum Mechanics

The essential mystery of quantum mechanics becomes clearer when discussed in the conceptually simpler context of two slit interference. If light and electrons can have both particle and wave properties, then one might ask through which of the two slits the particle passed. However, in physics a question simply doesn't make sense if it cannot be answered by experiment.

One can indeed perform an experiment to determine which slit the X-ray photon or electron passes through in the two slit experiment. However, by the very act of making this measurement, the form of the associated wave is altered. In particular, since the absolute square of the wave displacement represents the probability of finding the particle, once the particle has definitely been found passing through one or the other of the slits, the wave function collapses into a very small wave packet located at the observed position of the particle. Thus, the wave displacement becomes zero at the slit it didn't pass through. However, the interference pattern results from the superposition of waves emanating from two slits. If no wave comes from one of the slits (because the wave displacement is zero there), then there can be no interference pattern!

We can now make the inverse argument. If there is an interference pattern, then we know that the wave displacement is non-zero at both slits. From the probability interpretation of the wave displacement, we conclude that we cannot say, even in principle, through which slit the particle passed. It is not just that we don't know the answer to this question; there is simply no experiment which can give us an answer without destroying the interference pattern. In other words, the question "Which slit did the particle pass through?" is a nonsensical question in the case where an interference pattern is actually produced.

The American physicist Richard Feynman noticed that the above behavior can be interpreted as violating the normal laws of probability. These laws say that the probability of an event is the sum of the probabilities of alternate independent ways for that event to occur. For instance, the probability for a particle to reach point A on the detection screen of a two slit setup is just the probability P_1 for the particle to reach point A after going through slit 1, plus the probability P_2 for the particle to reach point A after going through slit 2. Thus, if $P_1 = P_2 = 0$, then the probability for the particle to reach point A irrespective of which slit it went through should be $P_{\text{total}} = P_1 + P_2 = 0.2$. However, if point A happens to be a point of destructive interference, then we know that $P_{\text{total}} = 0$

Feynman proposed that the above rule stating that alternate independent probabilities add, is simply incorrect. In its place Feynman asserted that probability amplitudes add instead, where the probability amplitude in this case is just the wave function associated with the particle. The total probability for a process is obtained by adding the alternate probability amplitudes together and taking the absolute square of the sum.

Feynman's view is a particularly compact expression of the so-called Copenhagen interpretation of quantum mechanics which evolved from the ideas of Niels Bohr, Werner Heisenberg, Max Born, and others in the 1920s. It dispenses with the wave-particle duality and other philosophical baggage by saying "Particles are real in that we can observe them, but the only theory we have is about probability amplitudes for particles." This interpretation of quantum mechanics may be weird, but it appears to be self-consistent and in agreement with experiment.

This page titled 7.4: Sense and Nonsense in Quantum Mechanics is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

7.5: Mass, Momentum, and Energy

In this section we relate the classical ideas of mass, momentum and energy to what we have done so far. Historically, these connections were first made by Max Planck and Louis de Broglie with help from Albert Einstein. Bragg diffraction of electrons is invoked as an experimental test of the Planck and de Broglie relations.

Technically, we don't need the ideas of mass, momentum, and energy to do physics – the notions of wavenumber, frequency, and group velocity are sufficient to describe and explain all observed phenomena. However, mass, momentum, and energy are so firmly embedded in physics that one couldn't talk to other physicists without an understanding of these quantities!

Planck, Einstein, and de Broglie

Max Planck was the first to develop a theory explaining the energy density of electromagnetic radiation in a box at a fixed temperature. Albert Einstein extended Planck's ideas by postulating that the energy of electromagnetic radiation is quantized into chunks called photons. The energy E of a photon is related to the frequency of the electromagnetic radiation by the equation

$$E = hf = \hbar\omega \quad \text{(Planck-Einstein relation)} \tag{7.5.1}$$

where f is the rotational frequency of the associated electromagnetic wave and ω is its angular frequency. The constant $\mathrm{h} = 6.63 \times 10^{-34}\ \mathrm{kg\ m^2\ s^{-1}}$ is called Planck's constant. The related constant $\hbar = h/2\pi = 1.06 \times 10^{-34}\ \mathrm{kg\ m^2\ s^{-1}}$ is also referred to as Planck's constant, but to avoid confusion with the original constant, we will generally refer to it as "h bar".

Notice that a new physical dimension has appeared, namely mass, with the unit kilogram, abbreviated "kg". The physical meaning of mass is much like our intuitive understanding of the concept, i. e., as a measure of the resistance of an object to its velocity being changed. The precise scientific meaning will emerge shortly.

Einstein showed that Planck's idea could be used to explain the emission of electrons which occurs when light impinges on the surface of a metal. This emission, which is called the photoelectric effect, can only occur when electrons are supplied with a certain minimum energy $\mathrm{E_B}$ required to break them loose from the metal. Experiment shows that this emission occurs only when the frequency of the light exceeds a certain minimum value. This value turns out to equal $\omega_{\min} = \mathrm{E_B}/\hbar$, which suggests that electrons gain energy by absorbing a single photon. If the photon energy, $\hbar\omega$, exceeds $\mathrm{E_B}$, then electrons are emitted, otherwise they are not. It is much more difficult to explain the photoelectric effect from the classical theory of light. The value of $\mathrm{E_B}$, called the binding energy or work function, is different for different metals.

Louis de Broglie proposed that Planck's energy-frequency relationship be extended to all kinds of particles. In addition he hypothesized that the momentum Π of the particle and the wave vector k of the corresponding wave were similarly related:

$$\mathbf{\Pi} = \hbar\mathbf{k} \quad \text{(de Broglie relation)} \tag{7.5.2}$$

Note that this can also be written in scalar form in terms of the wavelength as $\Pi = \mathrm{h}/\lambda$. (We use Π rather than the more common p for momentum, because as we shall see, there are two different kinds of momentum, one related to the wave number, the other related to the velocity of a particle. In many cases they are equal, but there are certain important situations in which they are not.)

De Broglie's hypothesis was inspired by the fact that wave frequency and wave number are components of the same four-vector according to the theory of relativity, and are therefore closely related to each other. Thus, if the energy of a particle is related to the frequency of the corresponding wave, then there ought to be some similar quantity which is correspondingly related to the wave number. It turns out that the momentum is the appropriate quantity. The physical meaning of momentum will become clear as we proceed.

We will also find that the rest frequency, μ, of a particle is related to its mass, m:

$$E_{rest} \equiv mc^2 = \hbar\mu. \tag{7.5.3}$$

The quantity $\mathrm{E_{rest}}$ is called the rest energy of the particle.

From our perspective, energy, momentum, and rest energy are just scaled versions of frequency, wave vector, and rest frequency, with a scaling factor $\hbar$. We can therefore define a four-momentum as a scaled version of the wave four-vector:

$$\underline{\Pi} = \hbar\underline{k}. \tag{7.5.4}$$

The spacelike component of $\underline{\Pi}$ is just Π, while the time like part is $\mathrm{E/c}$.

Planck, Einstein, and de Broglie had extensive backgrounds in classical mechanics, in which the concepts of energy, momentum, and mass have precise meaning. In this text we do not presuppose such a background. Perhaps the best strategy at this point is to think of these quantities as scaled versions of frequency, wave number, and rest frequency, where the scale factor is $\hbar$. The significance of these quantities to classical mechanics will emerge bit by bit.

Particle Quantities

Let us now recapitulate what we know about relativistic waves, and how this knowledge translates into knowledge about the mass, energy, and momentum of particles. In the following equations, the form on the left is expressed in wave terms, i. e., in terms of frequency, wave number, and rest frequency. The form on the right is the identical equation expressed in terms of energy, momentum, and mass. Since the latter variables are just scaled forms of the former, the two forms of each equation are equivalent.

We begin with the dispersion relation for relativistic waves:

$$\omega^2 = k^2c^2 + \mu^2 \quad E^2 = \Pi^2c^2 + m^2c^4 \tag{7.5.5}$$

Calculation of the group velocity,$\mathrm{u_g = d\omega/dk}$, from the dispersion relation yields

$$u_g = \frac{c^2k}{\omega} \quad u_g = \frac{c^2\Pi}{E} \tag{7.5.6}$$

These two sets of equations represent what we know about relativistic waves, and what this knowledge tells us about the relationships between the mass, energy, and momentum of relativistic particles. When in doubt, refer back to these equations, as they work in all cases, including for particles with zero mass!

It is useful to turn equations (7.5.5) and (7.5.6) around so as to express the frequency as a function of rest frequency and group velocity,

$$\omega = \frac{\mu}{\left(1 - u_g^2/c^2\right)^{1/2}} \quad E = \frac{mc^2}{\left(1 - u_g^2/c^2\right)^{1/2}} \tag{7.5.7}$$

and the wavenumber as a similar function of these quantities:

$$k = \frac{\mu u_g/c^2}{\left(1 - u_g^2/c^2\right)^{1/2}} \quad \Pi = \frac{mu_g}{\left(1 - u_g^2/c^2\right)^{1/2}} \tag{7.5.8}$$

Note that equations (7.5.7) and (7.5.8) work only for particles with non-zero mass! For zero mass particles both the numerators and denominators of equations (7.5.7) and (7.5.8) are zero, making these equations undefined, and you need to use equations (7.5.5) and (7.5.6) with $\mathrm{m = 0}$ and $\mu = 0$ instead.

The quantity $\omega - \mu$ indicates how much the frequency exceeds the rest frequency. Notice that if $\omega = \mu$, then from equation (7.5.5) k = 0. Thus, positive values of $\omega_k \equiv \omega - \mu$ indicate $|\mathrm{k}| > 0$, which means that the particle is moving according to equation (7.5.6). Let us call ω_k the kinetic frequency:

$$\omega_k = \left[\frac{1}{\left(1 - u_g^2/c^2\right)^{1/2}} - 1\right]\mu \quad K = \left[\frac{1}{\left(1 - u_g^2/c^2\right)^{1/2}} - 1\right]mc^2 \tag{7.5.9}$$

We call K the *kinetic energy* for similar reasons. Again, equation (7.5.9) only works for particles with non-zero mass. For zero mass particles the kinetic energy equals the total energy.

Note that the results of this section are valid only for free particles, i. e., particles to which no force is applied. Force in classical and quantum mechanics is treated in the next chapter.

Non-Relativistic Limits

When the mass is non-zero and the group velocity is much less than the speed of light, it is useful to compute approximate forms of the above equations valid in this limit. Using the approximation $(1 + \epsilon)^X \approx 1 + x\epsilon$, we find that the dispersion relation becomes

$$\omega = \mu + \frac{k^2c^2}{2\mu} \quad E = mc^2 + \frac{\Pi^2}{2m}, \tag{7.5.10}$$

and the group velocity equation takes the approximate form

$$u_g = \frac{c^2 k}{\mu} \quad u_g = \frac{\Pi}{m} \tag{7.5.11}$$

The non-relativistic limits for equations (7.5.7) and (7.5.8) become

$$\omega = \mu + \frac{\mu u_g^2}{2c^2} \quad E = mc^2 + \frac{m u_g^2}{2} \tag{7.5.12}$$

and

$$k = \mu u_g / c^2 \quad \Pi = m u_g \tag{7.5.13}$$

while the approximate kinetic energy equation is

$$\omega_k = \frac{\mu u_g^2}{2c^2} \quad K = \frac{m u_g^2}{2} \tag{7.5.14}$$

Just a reminder — the equations in this section are not valid for massless particles!

Experimental Test

How can we test the above predictions against experiment? The key point is to be able to relate the wave aspects to the particle aspects of a quantum mechanical wave-particle. Equation (7.5.8), or equation (7.5.13) in the non-relativistic case, relates a particle's wave number k to its velocity $\mathrm{u_g}$. Both of these quantities can be measured in a Bragg's law experiment with electrons. In this experiment electrons are fired at a crystal with known atomic dimensions at a known speed, which we identify with the group velocity $\mathrm{u_g}$. The Bragg angle which yields constructive interference can be used to calculate the wavelength of the corresponding electron wave, and hence the wave number and momentum. If the momentum is plotted against group velocity in the non-relativistic case, a straight line should be found, the slope of which is the particle's mass. In the fully relativistic case one needs to plot momentum versus $\mathrm{u_g} / \left(1 - \mathrm{u_g^2/c^2}\right)^{1/2}$. Again, a straight line indicates agreement with the theory and the slope of the line is the particle's mass. This particular experiment is difficult to do, but the corresponding theories verify in many other experiments.

This page titled 7.5: Mass, Momentum, and Energy is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

7.6: Heisenberg Uncertainty Principle

Classically, we consider the location of a particle to be a knowable piece of information. In quantum mechanics the position of a particle is well known if the wave packet representing it is small in size. However, quantum mechanics imposes a price on accurately knowing the position of a particle in terms of the future predictability of its position. This is because a small wave packet, which corresponds to accurate knowledge of the corresponding particle's position, implies the superposition of plane waves corresponding to a broad distribution of wavenumbers. This translates into a large uncertainty in the wavenumber, and hence the momentum of the particle. In contrast, a broad wave packet corresponds to a narrower distribution of wavenumbers, and correspondingly less uncertainty in the momentum.

Referring back to chapters 1 and 2, recall that both the longitudinal (along the direction of motion) and transverse (normal to the direction of motion) dimensions of a wave packet, $\Delta \mathrm{x_L}$ and $\Delta \mathrm{x_T}$, can be related to the spread of longitudinal and transverse wavenumbers, $\Delta \mathrm{k_L}$ and $\Delta \mathrm{k_T}$: :

$$\Delta k_L \Delta x_L \approx 1 \tag{7.6.1}$$

$$\Delta k_T \Delta x_T \approx 1 \tag{7.6.2}$$

We have omitted numerical constants which are of order unity in these approximate relations so as to show their essential similarity.

The above equations can be interpreted in the following way. Since the absolute square of the wave function represents the probability of finding a particle, $\Delta \mathrm{x_L}$ and $\Delta \mathrm{x_T}$ represent the uncertainty in the particle's position. Similarly, $\Delta \mathrm{k_L}$ and $\Delta \mathrm{k_T}$ represent the uncertainty in the particle's longitudinal and transverse wave vector components. This latter uncertainty leads to uncertainty in the particle's future motion — larger or smaller longitudinal k results respectively in larger or smaller particle speed, while uncertainty in the transverse wavenumber results in uncertainty in the particle's direction of motion. Thus uncertainties in any component of **k** result in uncertainties in the corresponding component of the particle's velocity, and hence in its future position.

The equations (7.6.1) and (7.6.2) show that uncertainty in the present and future positions of a particle are complimentary. If the present position is accurately known due to the small size of the associated wave packet, then the future position is not very predictable, because the wave packet disperses rapidly. On the other hand, a broad-scale initial wave packet means that the present position is poorly known, but the uncertainty in position, poor as it is, doesn't rapidly increase with time, since the wave packet has a small uncertainty in wave vector and thus disperses slowly. This is a statement of the *Heisenberg uncertainty principle*.

The uncertainty principle also applies between frequency and time:

$$\Delta \omega \Delta t \approx 1 \tag{7.6.3}$$

This shows up in the beat frequency equation $1/\mathrm{T_{beat}} = \Delta \mathrm{f} = \Delta\omega/2\pi$. The beat period $\mathrm{T_{beat}}$ may be thought of as the size of a "wave packet in time". The beat frequency equation may be rewritten as $\Delta\omega \mathrm{T_{beat}} = 2\pi$, which is the same as equation (7.6.3) if the factor of 2π is ignored and $\mathrm{T_{beat}}$ is identified with Δt.

The above forms of the uncertainty principle are not relativistically invariant. A useful invariant form may be obtained by transforming to the coordinate system in which a particle is stationary. In this reference frame the time t becomes the proper time τ associated with the particle. Furthermore, the frequency ω becomes the rest frequency μ. The uncertainty principle thus becomes

$$\Delta \mu \Delta \tau \approx 1 \tag{7.6.4}$$

in this reference frame. However, since $\Delta\mu$ and $\Delta \mathrm{T}$ are relativistic invariants, this expression of the uncertainty principle is valid in any reference frame.

It is more common to express the uncertainty principle in terms of the mass, momentum, and energy by multiplying equations (7.6.1) - (7.6.4) by $\hbar$. Lumping the momentum equations, we find

$$\Delta \Pi \Delta x \approx \hbar \tag{7.6.5}$$

$$\Delta E \Delta t \approx \hbar \tag{7.6.6}$$

and

$$\Delta \left(mc^2\right) \Delta \tau \approx \hbar \tag{7.6.7}$$

Classical mechanics is the realm of quantum mechanics in which the dimensions of the system of interest are much larger than the wavelengths of the waves corresponding to the particles constituting the system. In this case the uncertainties induced by the uncertainty principle are unimportant. This limit is analogous to the geometrical optics limit for light. Thus, we can say that classical mechanics is the geometrical optics limit of quantum mechanics.

This page titled 7.6: Heisenberg Uncertainty Principle is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

7.7: Problems

1. An electron with wavelength $\lambda = 1.2 \times 10^{-10}$ m undergoes Bragg diffraction from a single crystal with atomic plane spacing of $\mathrm{d} = 2 \times 10^{-10}$ m
    1. Calculate the Bragg angles (all of them!) for which constructive interference occurs.
    2. Calculate the speed of the electron.
2. Suppose that electrons impinge on two slits in a plate, resulting in a two slit diffraction pattern on a screen on the other side of the plate. The probability for an electron to pass through either one of the slits and reach point A on the screen is P, assuming that the other slit is blocked.
    1. If there are two slits open and A is a point of constructive interference, what is the probability of an electron reaching A? Hint: Remember that amplitudes, not probabilities add.
    2. If there are two slits open and A is a point of destructive interference, what is the probability of an electron reaching A?
    3. If there are two slits open, what is the probability for an electron to reach point A according to the conventional rule that probabilities add? (This is the result one would expect if, for instance, the particles were machine gun bullets and the slits were, say, 5 cm apart.)
    4. If the slit separation is very much greater than the electron wavelength, how does this affect the spacing of regions of constructive and destructive interference? Explain how the results of parts (a) and (b) become approximately consistent with those of part (c) in this case.
3. Compute the (angular) rest frequency of an electron and a neutron. (Look up their masses.)
4. How does the dispersion relation for relativistic waves simplify if the rest frequency (and hence the particle mass) is zero? What is the group velocity in this case?
5. X-rays are photons with frequencies about 2000 times the frequencies of ordinary light photons. From this information and what you know about light, infer the approximate velocity of electrons which have Bragg diffraction properties similar to X-rays. Are the electrons relativistic or non-relativistic?
6. Electrons with velocity $v = 0.6c$ are diffracted with a 0.2 radian half-angle of diffraction when they hit an object. What is the approximate size of the object? Hint: Diffraction of a wave by an object of a certain size is quite similar to diffraction by a hole in a screen of the same size.
7. Work out an approximate formula for the kinetic energy of a particle as a function of mass m and velocity $\mathrm{u_g}$ which is valid when $u_g^2 \ll c^2$. Hint: Use the approximation $(1+\epsilon)^{\mathrm{x}} \approx 1 + \mathrm{x}\epsilon$, which is valid for $|\epsilon| \ll 1$. As $\mathrm{u_g/C}$ becomes larger, how does this approximate formula deviate from the exact formula?
8. Work out an approximate formula for the momentum of a particle as a function of m and $\mathrm{u_g}$ in the case where $\mathrm{u_g^2} \ll \mathrm{c^2}$. You may wish to use the approximation mentioned in the previous problem.
9. If a photon is localized to within a distance Δx, what is the uncertainty in the photon energy?
10. If an electron is localized to within a distance Δx, what is the uncertainty in the electron kinetic energy? Hint: As long as $\Delta\Pi \ll \Pi$, $\Delta\Pi^2 \approx 2 \sqcap \Delta\Pi$. To see why, compute $\mathrm{d}\Pi^2/\mathrm{d}\Pi$.
11. A grocer dumps some pinto beans onto a scale, estimates their mass as 2 kg, and then dumps them off after 5 s. What is the quantum mechanical uncertainty in this measurement? Assume this occurs in Quantum World where the speed of light is 10 m $\mathrm{s^{-1}}$ (speed of a fast buggy) and Planck's constant $\hbar = 1$ kg $\mathrm{m^2\ s^{-1}}$.
12. Mary's physics text (mass 0.3 kg) has to be kept on a leash (length 0.5 m) to prevent it from wandering away from her in Quantum World ($\hbar = 1$ kg $\mathrm{m^2\ s^{-1}}$).
    1. If the leash suddenly breaks, what is the maximum speed at which the book is likely to move away from its initial location?
    2. In order to reduce this speed, should Mary make the new leash shorter or longer than the old one? Explain.
13. A proton (mass $M = 1.7 \times 10^{-27}$ kg) is confined to an atomic nucleus of diameter $\mathrm{D} = 2 \times 10^{-15}$ m.
    1. What is the uncertainty in the proton's momentum?
    2. Roughly what kinetic energy might you expect the proton to have? Planck's constant is $\hbar = 1.06 \times 10^{-34}$ kg $\mathrm{m^2\ s^{-1}}$. You may use the non-relativistic equation for the energy.

This page titled 7.7: Problems is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

8: Geometrical Optics and Newton’s Laws

8.1: Fundamental Principles of Dynamics
8.2: Potential Energy
8.3: Work and Power
8.4: Mechanics and Geometrical Optics
8.5: Math Tutorial – Partial Derivatives
8.6: Motion in Two and Three Dimensions
8.7: Kinetic and Total Momentum
8.8: Problems

This page titled 8: Geometrical Optics and Newton’s Laws is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

8.1: Fundamental Principles of Dynamics

Roughly speaking, there have been three eras of physics, characterized by three different answers to the question of what makes things go.

Pre-Newtonian Dynamics

Aristotle expounded a view of dynamics which agrees closely with our everyday experience of the world. Objects only move when a force is exerted upon them. As soon as the force goes away, the object stops moving. The act of pushing a box across the floor illustrates this principle — the box certainly doesn't move by itself!

Newtonian Dynamics

In contrast to earthly behavior, the motions of celestial objects seem effortless. No obvious forces act to keep the planets in motion around the sun. In fact, it appears that celestial objects simply coast along at constant velocity unless something acts on them. The Newtonian view of dynamics — objects change their *velocity* rather than their position when a force is exerted on them — is expressed by *Newton's second law*:

$$\mathbf{F} = m\mathbf{a} \quad \text{(Newton's second law)}, \tag{8.1.1}$$

where **F** is the force exerted on a body, *m* is its mass, and **a** is its acceleration. *Newton's first law*, which states that an object remains at rest or in uniform motion unless a force acts on it, is actually a special case of Newton's second law which applies when **F** = 0.

It is no wonder that the first successes of Newtonian mechanics were in the celestial realm, namely in the predictions of planetary orbits. It took Newton's genius to realize that the same principles which guided the planets also applied to the earthly realm as well. In the Newtonian view, the tendency of objects to stop when we stop pushing on them is simply a consequence of frictional forces opposing the motion. Friction, which is so important on the earth, is negligible for planetary motions, which is why Newtonian dynamics is more obviously valid for celestial bodies.

Note that the principle of relativity is closely related to Newtonian physics and is incompatible with pre-Newtonian views. After all, two reference frames moving relative to each other cannot be equivalent in the pre-Newtonian view, because objects with nothing pushing on them can only come to rest in one of the two reference frames. Newton's second law obeys the principle of relativity because the acceleration of an object is the same when viewed from two different reference frames moving at a constant velocity with respect to each other.

Einstein's relativity is often viewed as a repudiation of Newton, but this is far from the truth — Newtonian physics makes the theory of relativity possible through its invention of the principle of relativity. Compared with the differences between pre-Newtonian and Newtonian dynamics, the changes needed to go from Newtonian to Einsteinian physics constitute minor tinkering.

Quantum Dynamics

In quantum mechanics, particles are represented by matter waves, with the absolute square of the wave displacement yielding the probability of finding the particle. The behavior of particles thus follows from the reflection, refraction, diffraction, and interference of the associated waves. The connection with Newtonian dynamics comes from tracing the trajectories of matter wave packets. Changes in the speed and direction of motion of these packets correspond to the accelerations of classical mechanics. When wavelengths are small compared to the natural length scale of the problem at hand, the wave packets can be made small, thus pinpointing the position of the associated particle, without generating excessive uncertainty in the particle's momentum. This is the geometrical optics limit of quantum mechanics.

This page titled 8.1: Fundamental Principles of Dynamics is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

8.2: Potential Energy

We now address Newtonian mechanics in the case where the force on a particle is conservative. A conservative force is one that can be derived from a so-called potential energy U. We assume that the potential energy of the particle depends only on its position. The force is obtained from the potential energy by the equation

$$F = -\frac{dU}{dx} \tag{8.2.1}$$

Using this equation we write Newton's second law as

$$-\frac{dU}{dx} = ma \tag{8.2.2}$$

We then notice that the acceleration can be written in terms of the x derivative along the particle's trajectory of $v^2/2$:

$$a = \frac{dv}{dt} = \frac{dv}{dx}\frac{dx}{dt} = \frac{dv}{dx}v = \frac{1}{2}\frac{dv^2}{dx} \tag{8.2.3}$$

The last step in the above derivation can be verified by applying the product rule: $\mathrm{dv^2/dt = d(vv)/dt = v(dv/dt) + (dv/dt)v = 2v(dv/dt)}$. Putting equations (8.2.2) and (8.2.3) together, we find that $\mathrm{d\left(mv^2/2 + U\right)/dt = 0}$, which implies that $mv^2/2 + U$ is constant. We call this constant the total energy E and the quantity $\mathrm{K = mv^2/2}$ the kinetic energy. We thus have the principle of conservation of energy for conservative forces:

$$E = K + U = \text{ constant.} \tag{8.2.4}$$

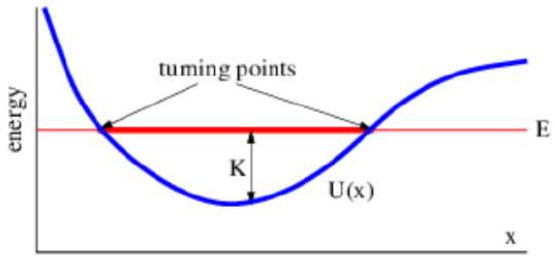


Figure 8.2.1:: Example of spatially variable potential energy $\mathrm{U(x)}$ for a particle with fixed total energy E. The kinetic energy $\mathrm{K = E - U}$ is zero where the E and U lines cross. These points are called turning points. The thick part of the horizontal line indicates the allowed range for the particle with the given energy.

Recall that in quantum mechanics the momentum is related to the group velocity $\mathrm{u_g}$ by

$$\Pi = mu_g \quad \text{(momentum)} \tag{8.2.5}$$

in the nonrelativistic case. Equating the group velocity to v and eliminating it in the kinetic energy results in an alternate expression for this quantity:

$$K \equiv \frac{1}{2}mu_g^2 = \frac{\Pi^2}{2m} \quad \text{(kinetic energy)} \tag{8.2.6}$$

Since the total energy E is constant or conserved, increases in the potential energy coincide with decreases in the kinetic energy and vice versa, as is illustrated in Figure 8.2.1:. In classical mechanics the kinetic energy cannot be negative, since it is the product of half the mass and the square of the velocity, both of which are positive. Thus, a particle with total energy E and potential energy U is forbidden to venture into regions in which the kinetic energy $\mathrm{K = E - U}$ is less than zero.

The points at which the kinetic energy is zero are called turning points. This is because a particle decreases in speed as it approaches a turning point, stops there for an instant, and reverses direction. Note also that a particle with a given total energy always has the same speed at some point x regardless of whether it approaches this point from the left or the right:

$$\text{speed } = |u_g| = \left|\pm[2(E-U)/m]^{1/2}\right| \tag{8.2.7}$$

Gravity as a Conservative Force

An example of a conservative force is gravity. An object of mass *m* near the surface of the earth has the gravitational potential energy

$$U = mgz \quad \text{(gravity near earth's surface)} \tag{8.2.8}$$

where z is the height of the object above some reference point such as the earth's surface and $g = 9.8\ \mathrm{m\ s^{-2}}$ is the local value of the gravitational field near the surface. Notice that the gravitational potential energy increases upward. The speed of the object in this case is $|u_g| = [2(E - mgz)/m]^{1/2}$. If $|\mathrm{u_g}|$ is known to equal the constant value $\mathrm{u_0}$ at elevation z = 0, then equations (8.2.7) and (8.2.8) tell us that $\mathrm{u_0 = (2E/m)^{1/2}}$ and $\mathrm{|u_g| = (u_0^2 - 2\ gz)^{1/2}}$.

There are certain types of questions which energy conservation cannot directly answer. For instance, if an object is released at elevation h with zero velocity at t = 0, at what time will it reach z = 0 under the influence of gravity? In such cases it is often easiest to return to Newton's second law. Since the force on the object is F = -dU/dz = -mg in this case, we find that the acceleration is $a = F/m = -mg/m = -g$. However, $a = du/dt = d^2z/dt^2$, so

$$u = -gt + C_1 \quad z = -gt^2/2 + C_1 t + C_2 \quad \text{(constant gravity)} \tag{8.2.9}$$

where $\mathrm{C_1\ and\ C_2}$ are constants to be determined by the initial conditions. These results can be verified by differentiating to see if the original acceleration is recovered. Since $\mathrm{u = 0\ and}\ z = \mathrm{h\ at}\ t = 0$, we have $\mathrm{C_1 = 0\ and\ C_2 = h}$. With these results it is easy to show that the object reaches $z = 0$ when $\mathrm{t = (2\ h/g)^{1/2}}$.

This page titled 8.2: Potential Energy is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

8.3: Work and Power

When a force is exerted on an object, energy is transferred to the object. The amount of energy transferred is called the *work* done on the object. However, energy is only transferred if the object moves. The work *W* done is

$$W = F\Delta x \tag{8.3.1}$$

where the distance moved by the object is Δx and the force exerted on it is F. Notice that work can either be positive or negative. The work is positive if the object being acted upon moves in the same direction as the force, with negative work occurring if the object moves opposite to the force.

Equation (8.3.1) assumes that the force remains constant over the full displacement Δx. If it is not, then it is necessary to break up the displacement into a number of smaller displacements, over each of which the force can be assumed to be constant. The total work is then the sum of the works associated with each small displacement.

If more than one force acts on an object, the works due to the different forces each add or subtract energy, depending on whether they are positive or negative. The total work is the sum of these individual works.

There are two special cases in which the work done on an object is related to other quantities. If F is the total force acting on the object, then $\mathrm{W} = \mathrm{F}\Delta\mathrm{x} = \mathrm{ma}\Delta\mathrm{x}$ by Newton's second law. However, $\mathrm{a} = \mathrm{dv}/\mathrm{dt}$ where v is the velocity of the object, and $\Delta\mathrm{x} = (\Delta\mathrm{x}/\Delta\mathrm{t})\Delta\mathrm{t} \approx \mathrm{v}\Delta\mathrm{t}$, where $\Delta\mathrm{t}$ is the time required by the object to move through distance $\Delta\mathrm{x}$. The approximation becomes exact when $\Delta\mathrm{x}$ and $\Delta\mathrm{t}$ become very small. Putting all of this together results in

$$W_{\text{total}} = m\frac{dv}{dt}v\Delta t = \frac{d}{dt}\left(\frac{mv^2}{2}\right)\Delta t = \Delta K \quad \text{(total work)} \tag{8.3.2}$$

where K is the kinetic energy of the object. Thus, when F is the only force, $W = W_{\text{total}}$ is the total work on the object, and this equals the change in kinetic energy of the object. This is called the work-energy theorem, and it demonstrates that work really is a transfer of energy to an object.

The other special case occurs when the force is conservative, but is not necessarily the total force acting on the object. In this case

$$W_{\text{cons}} = -\frac{dU}{dx}\Delta x = -\Delta U \quad \text{(conservative force)} \tag{8.3.3}$$

where ΔU is the change in the potential energy of the object associated with the force of interest.

The *power* associated with a force is simply the amount of work done by the force divided by the time interval Δt over which it is done. It is therefore the energy per unit time transferred to the object by the force of interest. From equation (8.3.1) we see that the power is

$$P = \frac{F\Delta x}{\Delta t} = Fv \quad \text{(power)} \tag{8.3.4}$$

where *v* is the velocity at which the object is moving. The *total power* is just the sum of the powers associated with each force. It equals the time rate of change of kinetic energy of the object:

$$P_{\text{total}} = \frac{W_{\text{total}}}{\Delta t} = \frac{dK}{dt} \quad \text{(total power).} \tag{8.3.5}$$

This page titled 8.3: Work and Power is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

8.4: Mechanics and Geometrical Optics

Louis de Broglie made an analogy between matter waves and light waves, pointing out that wave packets of light change their velocity as the result of spatial variations in the index of refraction of the medium in which they are traveling. This behavior comes about because the dispersion relation for light traveling through a medium with index of refraction n is $\omega = \mathrm{kc}/n$, so that the group velocity, $u_g = d\omega/dk = c/n$. Thus, when n increases, $\mathrm{u_g}$ decreases, and vice versa.

In this section we pursue de Broglie's analogy to see if we can come up with a theory of matter waves which gives the same results as classical mechanics in the geometrical optics limit of these waves. The dispersion relation for free matter waves is $\omega = \left(k^2c^2 + \mu^2\right)^{1/2}$. In the non-relativistic limit $k^2c^2 \ll \mu^2$. As done previously, we use $(1+\epsilon)^n \approx 1 + n\epsilon$ for small ϵ. In the non-relativistic limit, the dispersion relation for free waves thus becomes

$$\omega = \mu\left(1 + k^2c^2/\mu^2\right)^{1/2} \approx \mu + k^2c^2/(2\mu) \tag{8.4.1}$$

The above equation can be transformed into the total energy equation for a free, non-relativistic particle, $\mathrm{E} = \mathrm{mc}^2 + \mathrm{K}$, where mc^2 is the rest energy and K is the kinetic energy, by multiplying by $\hbar$. We convert the free particle energy equation into the equation for a particle subject to a conservative force by adding the potential energy U the right side. The analogous change to equation (8.4.1) is to add $\mathrm{S} = \mathrm{U}/\hbar$ to the right side, resulting in a modified dispersion relation:

$$\omega = S(x) + k^2c^2/(2\mu) \tag{8.4.2}$$

(Since the rest energy is just a constant, we have absorbed it into S.) This gives us the dispersion relation for one-dimensional matter waves subject to a spatially varying potential energy. The quantity S, which we see is just a scaled potential energy, plays a role for matter waves which is analogous to the role played by a spatially variable index of refraction for light waves.

Let us now imagine that all parts of the wave governed by this dispersion relation oscillate in phase. The only way this can happen is if ω is constant, i. e., it takes on the same value in all parts of the wave.

If ω is constant, the only way S can vary with x in equation (8.4.2) is if the wavenumber varies in a compensating way. Thus, constant frequency and spatially varying S together imply that $\mathrm{k} = \mathrm{k(x)}$. Solving equation (8.4.2) for k yields

$$k(x) = \pm\left[\frac{2\mu[\omega - S(x)]}{c^2}\right]^{1/2} \tag{8.4.3}$$

Since ω is constant, the wavenumber becomes smaller and the wavelength larger as the wave moves into a region of increased *S*.

In the geometrical optics limit, we assume that *S* doesn't change much over one wavelength so that the wave remains reasonably sinusoidal in shape with approximately constant wavenumber over a few wavelengths. However, over distances of many wavelengths the wavenumber and amplitude of the wave are allowed to vary considerably.

The group velocity calculated from the dispersion relation given by equation (8.4.2) is

$$u_g = \frac{d\omega}{dk} = \frac{kc^2}{\mu} = \pm\left(\frac{2c^2(\omega - S)}{\mu}\right)^{1/2} \tag{8.4.4}$$

where k is eliminated in the last step with the help of equation (8.4.3). The resulting equation tells us how the group velocity varies as a matter wave traverses a region of slowly varying S. Thus, as S increases, $\mathrm{u_g}$ decreases and vice versa.

We can now calculate the acceleration of a wave packet resulting from the spatial variation in S. We assume that x(t) represents the position of the wave packet, so that $\mathrm{u_g} = \mathrm{dx/dt}$. Using the chain rule $\mathrm{du_g/dt} = (\mathrm{du_g/dx})\,(\mathrm{dx/dt}) = (\mathrm{du_g/dx})\,\mathrm{u_g}$, we find

$$a = \frac{du_g}{dt} = \frac{du_g}{dx}u_g = \frac{du_g^2/2}{dx} = -\frac{c^2}{\mu}\frac{dS}{dx} = -\frac{\hbar}{m}\frac{dS}{dx} \tag{8.4.5}$$

The group velocity is eliminated in favor of S by squaring equation (8.4.4) and substituting the result into equation (8.4.5).

Recalling that $\mathrm{U} = \hbar\mathrm{S}$, equation (8.4.5) becomes

$$a = -\frac{1}{m}\frac{dU}{dx} = \frac{F}{m} \tag{8.4.6}$$

which is just Newton's second law! Thus, the geometrical optics approach to particle motion is completely equivalent to the classical mechanics of a particle moving under the influence of a conservative force, at least in one dimension. We therefore have two ways of solving for the motion of a particle subject to a potential energy $U(x)$. We can apply the principles of classical mechanics to get the force and the acceleration of the particle, from which we can derive the motion. Alternatively, we can apply the principles of geometrical optics to compute the spatially variable velocity of the wave packet using equation (8.4.4). The results are completely equivalent, though the methods are conceptually very different.

This page titled 8.4: Mechanics and Geometrical Optics is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

8.5: Math Tutorial – Partial Derivatives

In order to understand the generalization of Newtonian mechanics to two and three dimensions, we first need to understand a new type of derivative called the *partial derivative*. The partial derivative is used in functions of more than one variable. It is just like an ordinary derivative, except that when taking the derivative of the function with respect to one of the variables, the other variables are held constant. As an example, let us consider the function

$$f(x,y) = Ax^4 + Bx^2y^2 + Cy^4 \tag{8.5.1}$$

where *A*, *B*, and *C* are constants. The partial derivative of *f* with respect to *x* is

$$\frac{\partial f}{\partial x} = 4Ax^3 + 2Bxy^2 \tag{8.5.2}$$

and the partial derivative with respect to *y* is

$$\frac{\partial f}{\partial y} = 2Bx^2y + 4Cy^3. \tag{8.5.3}$$

That's it! Note that a special symbol " ∂" is used in place of the normal "d" for the partial derivative. This is sometimes called a "curly d".

This page titled 8.5: Math Tutorial – Partial Derivatives is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

8.6: Motion in Two and Three Dimensions

When a matter wave moves through a region of variable potential energy in one dimension, only the wavenumber changes. In two or three dimensions, the wave vector can change in both direction and magnitude. This complicates the calculation of particle movement. However, we already have an example of how to handle this situation, namely, the refraction of light. In that case Snell's law tells us how the direction of the wave vector changes, while the dispersion relation combined with the constancy of the frequency gives us information about the change in the magnitude of the wave vector. For matter waves a similar procedure works, though the details are different, because we seek the consequences of a change in potential energy rather than a change in the index of refraction.

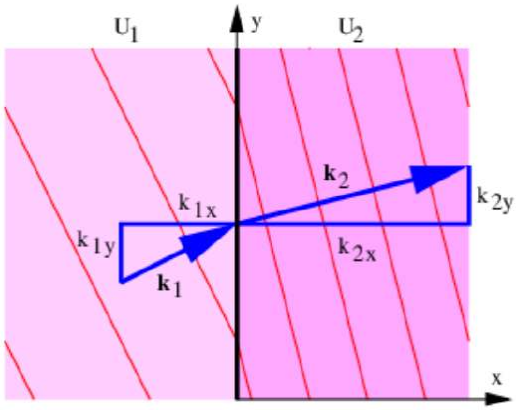


Figure 8.6.2:: Refraction of a matter wave by a discontinuity in potential energy. The component of the wave vector parallel to the discontinuity, $\mathrm{k_y}$, doesn't change, so $\mathrm{k_{1y}} = \mathrm{k_{2y}}$.

Figure 8.6.2: illustrates the refraction of matter waves at a discontinuity in the potential energy. Let us suppose that the discontinuity occurs at $x = 0$. If the matter wave to the left of the discontinuity is $\Psi_1 = \sin\left(\mathrm{k_{1x}x} + \mathrm{k_{1y}y} - \omega_1 \mathrm{t}\right)$ and to the right is $\Psi_2 = \sin\left(\mathrm{k_{2x}x} + \mathrm{k_{2y}y} - \omega_2 \mathrm{t}\right)$, then the wavefronts of the waves will match across the discontinuity for all time only if $\omega_1 = \omega_2 \equiv \omega$ and $\mathrm{k_{1y}} = \mathrm{k_{2y}} \equiv \mathrm{k_y}$. We are already familiar with the first condition from the one-dimensional problem, so the only new ingredient is the constancy of the y component of the wave vector.

In two dimensions the momentum is a vector:$\mathbf{\Pi} = \mathrm{mu\ when}\ |\mathbf{u}| \ll \mathrm{c}$, where $\mathbf{u}$ is the particle velocity. Furthermore, the kinetic energy is $K = m|\mathbf{u}|^2/2 = |\mathbf{\Pi}|^2/(2m) = \left({\Pi_x}^2 + {\Pi_y}^2\right)/(2m)$. The relationship between kinetic, potential, and total energy is unchanged from the one-dimensional case, so we have

$$E = U + \left(\Pi_x^2 + \Pi_y^2\right)/(2m) = \ \text{constant}. \tag{8.6.1}$$

The de Broglie relationship tells us that $\mathbf{\Pi} = \hbar\mathbf{k}$, so the constancy of $\mathrm{k_y}$ across the discontinuity in U tells us that

$$\Pi_y = \ \text{constant} \tag{8.6.2}$$

there

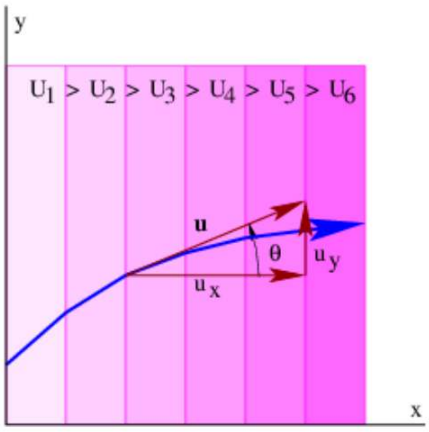


Figure 8.6.3:: Trajectory of a wave packet through a variable potential energy, U(x), which decreases to the right.

Let us now approximate a continuously variable U(x) by a series of steps of constant U oriented normal to the x axis. The above analysis can be applied at the jumps or discontinuities in U between steps, as illustrated in Figure 8.6.3:, with the result that equations (8.6.1) and (8.6.2) are valid across all discontinuities. If we now let the step width go to zero, these equations then become valid for U continuously variable in x.

An example from classical mechanics of a problem of this type is a ball rolling down an inclined ramp with an initial velocity component across the ramp, as illustrated in Figure 8.6.4:. The potential energy decreases in the down ramp direction, resulting in a force down the ramp. This accelerates the ball in that direction, but leaves the component of momentum across the ramp unchanged.

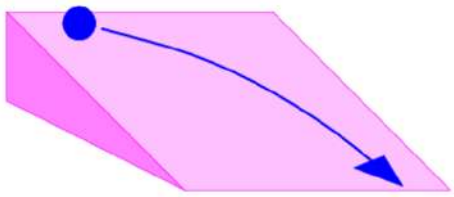


Figure 8.6.4:: Classical mechanics example of the problem illustrated in Figure 8.6.3:.

Using the procedure which we invoked before, we find the force components associated with U(x) in the x and y directions to be $F_x = -dU/dx$ and $F_y = 0$. This generalizes to

$$\mathbf{F} = -\left(\frac{\partial U}{\partial x}, \frac{\partial U}{\partial y}, \frac{\partial U}{\partial z}\right) \quad \text{(3-D conservative force)} \tag{8.6.3}$$

in the three-dimensional case where the orientation of constant U surfaces is arbitrary. It is also valid when U(x,y,z) is not limited to a simple ramp form, but takes on a completely arbitrary structure.

The definitions of work and power are slightly different in two and three dimensions. In particular, work is defined as

$$W = \mathbf{F} \cdot \Delta\mathbf{x} \tag{8.6.4}$$

where $\Delta\mathbf{x}$ is now a vector displacement. The vector character of this expression yields an additional possibility over the one dimensional case, where the work is either positive or negative depending on the direction of $\Delta\mathbf{x}$ relative to $\mathbf{F}$. If the force and the displacement of the object on which the force is acting are perpendicular to each other, the work done by the force is actually zero, even though the force and the displacement both have non-zero magnitudes. The power exhibits a similar change:

$$P = \mathbf{F} \cdot \mathbf{u} \tag{8.6.5}$$

Thus the power is zero if an object's velocity is normal to the force being exerted on it.

As in the one-dimensional case, the total work done on a particle equals the change in the particle's kinetic energy. In addition, the work done by a conservative force equals minus the change in the associated potential energy.

Energy conservation by itself is somewhat less useful for solving problems in two and three dimensions than it is in one dimension. This is because knowing the kinetic energy at some point tells us only the magnitude of the velocity, not its direction. If

conservation of energy fails to give us the information we need, then we must revert to Newton's second law, as we did in the one-dimensional case. For instance, if an object of mass m has initial velocity $\mathbf{u}_0 = (\mathrm{u}_0, 0)$ at location $(x, z) = (0, h)$ and has the gravitational potential energy $\mathrm{U} = \mathrm{mgz}$, then the force on the object is $\mathbf{F} = (0, -\mathrm{mg})$. The acceleration is therefore $\mathbf{a} = \mathbf{F}/\mathrm{m} = (0, -\mathrm{g})$. Since $\mathbf{a} = \mathrm{d}\mathbf{u}/\mathrm{dt} = \mathrm{d}^2\mathbf{x}/\mathrm{dt}^2$ where $\mathbf{x} = (\mathrm{x}, \mathrm{z})$ is the object's position, we find that

$$\mathbf{u} = (C_1, -gt + C_2) \quad \mathbf{x} = \left(C_1 t + C_3, -gt^2/2 + C_2 t + C_4\right) \tag{8.6.6}$$

where C_1, C_2, C_3, and C_4 are constants to be evaluated so that the solution reduces to the initial conditions at t = 0. The specified initial conditions tell us that $C_1 = u_0$, $C_2 = 0$, $C_3 = 0$, and $C_4 = h$ in this case. From these results we can infer the position and velocity of the object at any time.

This page titled 8.6: Motion in Two and Three Dimensions is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

8.7: Kinetic and Total Momentum

If you have previously taken a physics course then you have probably noticed that a rather odd symbol is used for momentum, namely Π, rather than the more commonly employed $\mathbf{p}$. The reason for this peculiar usage is that there are actually two kinds of momentum, **kinetic momentum** and **total momentum** (or canonical momentum), just as there are two kinds of energy, kinetic and total. The symbol Π represents total momentum while $\mathbf{p}$ represents kinetic momentum. Normally we don't need to distinguish between the two quantities, as they are generally equal to each other. However, we will find later in the course that it is crucial to make this distinction in the case of charged particles in a magnetic field. As a general rule, the total momentum is related to a particle's wave vector via the de Broglie relation,

$$\mathbf{\Pi} = \hbar \mathbf{k} \tag{8.7.1}$$

while the kinetic momentum is related to a particle's velocity,

$$\mathbf{p} = \frac{\mathrm{mu}}{\sqrt{1 - \frac{\mathrm{u}^2}{\mathrm{c}^2}}}. \tag{8.7.2}$$

This page titled 8.7: Kinetic and Total Momentum is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

8.8: Problems

1. Suppose the dispersion relation for a matter wave under certain conditions is $\omega = \mu + (\mathrm{k} - \mathrm{a})^2\mathrm{c}^2/(2\mu)$ where k is the wavenumber of the wave, $\mu = mc^2/\hbar$, m is the associated particle's mass, a is a constant, c is the speed of light, and $\hbar$ is Planck's constant divided by 2Π.
    1. Use this disperson relation and the Planck and de Broglie relations to determine the relationship between energy E, momentum Π, and mass m.
    2. Compute the group velocity of the wave and use this to determine how the group velocity depends on mass and momentum in this case.
2. A matter wave function associated with a particle of definite (constant) total energy E takes the form shown in Figure 8.8.5:. Make a sketch showing how the kinetic, potential, and total energies of the particle vary with x.

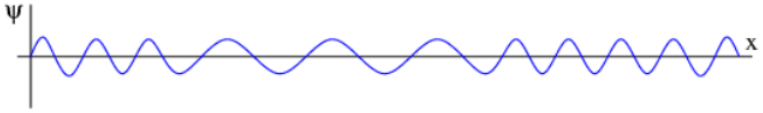


Figure 8.8.5: A wave function in which the wavelength varies with position.

3. Compute $\partial/\partial x$ and $\partial/\partial y$ of the following functions. Other symbols are constants.
    1. $f(x, y) = ax^2 + by^3$
    2. $f(x, y) = ax^2y^2$
    3. $f(x, y) = (x + a)(y + b)$
4. Given a potential energy for a particle of mass M of the form $U(x) = Ax^3 - Bx$ where A and B are positive constants:
    1. Find the force on the particle.
    2. Find the values of x where the force is zero.
    3. Sketch U(x) versus x and graphically compare the slope of U(x) to the force computed above. Do the two qualitatively match?
    4. If the total energy of the particle is zero, where are its turning points?
    5. What is the particle's speed as a function of position assuming that the total energy E is known?
5. Given a potential energy function $U(x, y) = A\left(x^2 + y^2\right)$ where A is a positive constant:
    1. Sketch lines of constant U in the x-y plane.
    2. Compute the components of force as a function of x and y and draw sample force vectors in the x-y plane on the same plot used above. Do the force vectors point "uphill" or "downhill"?
6. Do the same as in the previous question for the potential energy function $U(x, y) = Axy$
7. Suppose that the components of the force vector in the x-y plane are $\mathbf{F} = \left(2\mathrm{Axy}^3, 3\mathrm{Ax}^2\mathrm{y}^2\right)$ where A is a constant. See if you can find a potential energy function U(x,y) which gives rise to this force.
8. You are standing on top of a cliff of height H with a rock of mass M.
    1. If you throw the rock horizontally outward at speed u_0, what will its speed be when it hits the ground below?
    2. If you throw the rock upward at 45˚ to the horizontal at speed u_0, what will its speed be when it hits the ground? Hint: Can you use conservation of energy to solve this problem? Ignore air friction.
9. A car of mass 1200 kg initially moving $30\ \mathrm{m\ s^{-1}}$ brakes to a stop.
    1. What is the net work done on the car due to all the forces acting on it during the indicated period?
    2. Describe the motion of the car relative to an inertial reference frame initially moving with the car.

3. In the above reference frame, what is the net work done on the car during the indicated period?Is work a relativistically invariant quantity?

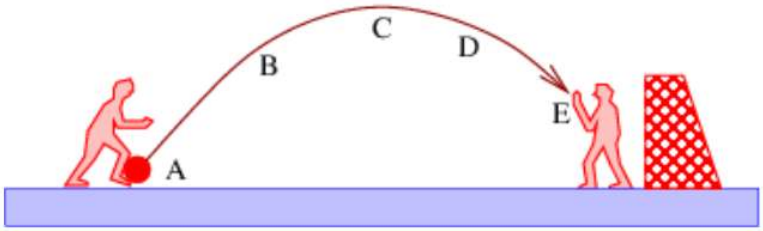


Figure 8.8.6**::** The trajectory of a soccer ball.

10. A soccer player kicks a soccer ball, which is caught by the goal keeper as shown in Figure 8.8.6:. At various points forces exerted by gravity, air friction, the foot of the offensive player, and the hands of the goal keeper act on the ball.
    1. List the forces acting on the soccer ball at each of the points A, B, C, D, and E.
    2. State whether the instantaneous power being applied to the soccer ball due to each of the forces listed above is positive, negative, or zero at each of the labeled points.
11. A cannon located at $(x, z) = (0, 0)$ shoots a cannon ball upward at an angle of θ from the horizontal at initial speed u_0. Hint: In order to solve this problem you must first obtain the x and z components of acceleration from Newton's second law. Second, you must find the velocity components as a function of time from the components of acceleration. Third, you must find x and z as a function of time from the the components of velocity. Only then should you attempt to answer the questions below.
    1. How long does it take the cannon ball to reach its peak altitude?
    2. How high does the cannon ball go?
    3. At what value of x does the cannon ball hit the ground (z = 0)?
    4. Determine what value of θ yields the maximum range.

This page titled 8.8: Problems is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

9: Symmetry and Bound States

9.1: Math Tutorial — Complex Waves
9.2: Symmetry and Quantum Mechanics
9.3: Confined Matter Waves
9.4: Problems

This page titled 9: Symmetry and Bound States is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

9.1: Math Tutorial — Complex Waves

Until now we have represented quantum mechanical plane waves by sine and cosine functions, just as with other types of waves. However, plane matter waves cannot be truly represented by sines and cosines. We need instead mathematical functions in which the wave displacement is complex rather than real. This requires the introduction of a bit of new mathematics, which we tackle first. Using our new mathematical tool, we are then able to explore two crucially important ideas in quantum mechanics; (1) the relationship between symmetry and conservation laws, and (2) the dynamics of spatially confined waves.

A complex number z is the sum of a real number and an imaginary number. An imaginary number is just a real number multiplied by $i \equiv (-1)^{1/2}$. Thus, we can write $z = a + ib$ for any complex z, where a and b are real. The quantities a and b are the real and imaginary parts of z, sometimes written $Re(z)$ and $Im(z)$.

Quantum mechanics requires wave functions to be complex, i. e., to possess real and imaginary parts. Plane waves in quantum mechanics actually take the form $\Psi = \exp[\mathrm{i}(\mathrm{kx} - \omega \mathrm{t})]$ rather than, say, $\cos(kx - \omega t)$. The reason for this is the need to distinguish between waves with positive and negative frequencies. If we replace k and ω with $-\mathrm{k}$ and $-\omega$ in the cosine form, we get

$$\cos(-kx + \omega t) = \cos[-(kx - \omega t)] = \cos(kx - \omega t). \tag{9.1.1}$$

In other words, changing the sign of k and ω results in no change in a wave expressed as a cosine function. The two quantum mechanical states, one with wavenumber and frequency k and ω and the other with $-\mathrm{k}$ and $-\omega$, yield indistinguishable wave functions and therefore would represent physically indistinguishable states. The cosine form is thus insufficiently flexible to represent quantum mechanical waves. On the other hand, if we replace k and ω with their negatives in the complex exponential form of a plane wave we get $\psi = \exp[-\mathrm{i}(\mathrm{kx} - \omega \mathrm{t})]$, which is different from $\exp[\mathrm{i}(\mathrm{kx} - \omega \mathrm{t})]$. These two wave functions are distinguishable and thus correspond to distinct physical states.

It is not immediately obvious that a complex exponential function provides the oscillatory behavior needed to represent a plane wave. However, the complex exponential can be expressed in terms of sines and cosines using Euler's equation:

$$\exp(i\phi) = \cos(\phi) + i\sin(\phi) \quad \text{(Euler's equation)}. \tag{9.1.2}$$

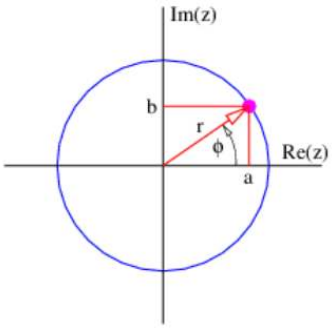


Figure 9.1.1:: Graphical representation of a complex number z as a point in the complex plane. The horizontal and vertical Cartesian components give the real and imaginary parts of z respectively.

If we define $r = (a^2 + b^2)^{1/2}$ and $\phi = \tan^{-1}(b/a)$, then an alternate way of expressing a complex number is z = r exp(iϕ), which by Euler's equation equals $r\cos(\phi) + \mathrm{ir}\sin(\phi)$. Comparison shows that $\mathrm{a} = r\cos(\phi)$ and $\mathrm{b} = r\sin(\phi)$. Thus, a complex number can be thought of as a point in the a-b plane with Cartesian coordinates a and b and polar coordinates r and ϕ. The a-b plane is called the complex plane.

We now see how the complex wave function represents an oscillation. If $\Psi = \exp[\mathrm{i}(\mathrm{kx} - \omega \mathrm{t})]$, the complex function $\psi(x,t)$ moves round and round the unit circle in the complex plane as x and t change, as illustrated in Figure 9.1.1:. This contrasts with the back and forth oscillation along the horizontal axis of the complex plane represented by $\cos(\mathrm{kx} - \omega \mathrm{t})$.

We will not present a formal proof of Euler's equation — you will eventually see it in your calculus course. However, it may be helpful to note that the ϕ derivatives of $\exp(\ \mathrm{i}\phi\ )$ and $\cos(\phi) + \mathrm{i}\sin(\phi)$ have the same behavior:

$$\frac{d}{d\phi}\exp(i\phi) = i\exp(i\phi) \tag{9.1.3}$$

$$\frac{d}{d\phi}[\cos(\phi) + i\sin(\phi)] = -\sin(\phi) + i\cos(\phi) \tag{9.1.4}$$

$$= i[\cos(\phi) + i\sin(\phi)]. \tag{9.1.5}$$

(In the second of these equations we have replaced the minus sign in front of the sine function by i^2 and then extracted a common factor of i.) The ϕ derivative of both of these functions thus yields the function back again times i. This is a strong hint that $\exp(i\phi)$ and $\cos(\phi) + i\sin(\phi)$ are different ways of representing the same function.

We indicate the complex conjugate of a complex number z by a superscripted asterisk, i. e., z^*. It is obtained by replacing i by -i. Thus, $(a + ib)^* = a - ib$. The absolute square of a complex number is the number times its complex conjugate:

$$|z|^2 = |a + ib|^2 \equiv (a + ib)(a - ib) = a^2 + b^2 = r^2 \tag{9.1.6}$$

Notice that the absolute square of a complex exponential function is one:

$$|\exp(i\phi)|^2 = \exp(i\phi)\exp(-i\phi) = \exp(i\phi - i\phi) = \exp(0) = 1 \tag{9.1.7}$$

In quantum mechanics the absolute square of the wave function at any point expresses the relative probability of finding the associated particle at that point. Thus, the probability of finding a particle represented by a plane wave is uniform in space. Contrast this with the relative probability associated with a sine wave: $|\sin(kx - \omega t)|^2 = \sin^2(kx - \omega t)$. This varies from zero to one, depending on the phase of the wave. The "waviness" in a complex exponential plane wave resides in the phase rather than in the magnitude of the wave function.

One more piece of mathematics is needed. The complex conjugate of Euler's equation is

$$\exp(-i\phi) = \cos(\phi) - i\sin(\phi) \tag{9.1.8}$$

Taking the sum and the difference of this with the original Euler's equation results in the expression of the sine and cosine in terms of complex exponentials:

$$\cos(\phi) = \frac{\exp(i\phi) + \exp(-i\phi)}{2} \quad \sin(\phi) = \frac{\exp(i\phi) - \exp(-i\phi)}{2i} \tag{9.1.9}$$

We aren't used to having complex numbers show up in physical theories and it is hard to imagine how we would measure such a number. However, everything observable comes from taking the absolute square of a wave function, so we deal only with real numbers in experiments.

This page titled 9.1: Math Tutorial — Complex Waves is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

9.2: Symmetry and Quantum Mechanics

The idea of symmetry plays a huge role in physics. We have already used symmetry arguments in the theory of relativity — applying the principle of relativity to obtain the dispersion relation for relativistic matter waves is just such an argument. In this section we begin to explore how symmetry can be used to increase our understanding of quantum mechanics.

Particle

For our first example we take the case of a free particle in quantum mechanics, i. e., a particle subject to no force. The wave function for a free particle of definite momentum Π and energy E is given by

$$\psi = \exp[i(kx - \omega t)] = \exp[i(\Pi x - Et)/\hbar] \quad \text{(free particle)} \tag{9.2.1}$$

For this wave function $|\Psi|^2 = 1$ everywhere, so the probability of finding the particle anywhere in space and time is uniform. This contrasts with the probability distribution which arises if we assume a free particle to have the wave function $\Psi = \cos[(\Pi \mathrm{x} - \mathrm{Et})/\hbar]$. In this case $|\Psi|^2 = \cos^2[(\Pi x - Et)/\hbar]$, which varies with position and time, and is inconsistent with a uniform probability distribution.

Symmetry and Definiteness

Quantum mechanics is a probabilistic theory, in the sense that the predictions it makes tell us, for instance, the probability of finding a particle somewhere in space. If we know nothing about a particle's previous history, and if there are no physical constraints that would make it more likely for a particle to be at one point along the x axis than any another, then the probability distribution must be P(x) = constant.

This is an example of a symmetry argument. Expressed more formally, it states that if the above conditions apply, then the probability distribution ought to be subject to the condition P(x + D) = P(x) for any constant value of D. The only possible P(x) in this case is P = constant. In the language of physics, if there is nothing that gives the particle a higher probability of being at one point rather than another, then the probability is independent of position and the system is invariant under displacement in the x direction.

The above argument doesn't suffice for quantum mechanics, since as we have learned, the fundamental quantity describing a particle is not the probability distribution, but the wave function $\Psi(\mathrm{x})$. Thus, the wave function rather than the probability distribution ought to be the quantity which is invariant under displacement, i. e., $\psi(x + D) = \Psi(x)$.

This condition turns out to be too restrictive, because it implies that $\Psi(\mathrm{x}) = \text{ constant }$, whereas we know that a one-dimensional plane wave, which describes a particle with a uniform probability of being found anywhere along the x axis, has the form $\psi(\mathrm{x}) = \exp(\mathrm{ikx})$. (For simplicity we temporarily ignore the time dependence.) If we make the substitution x → x + D in a plane wave, we get $\exp[\mathrm{ik(x + D)}] = \exp(\mathrm{ikx})\exp(\mathrm{ikD})$. The wave function is thus technically not invariant under displacement, in that the displaced wave function is multiplied by the factor $\exp(\mathrm{ikD})$. However, the probability distribution of the displaced wave function still equals one everywhere, so there is no change in what we observe. Thus, in determining invariance under displacement, we are allowed to ignore changes in the wave function which consist only of multiplying it by a complex constant with an absolute value of one. Such a multiplicative constant is called a phase factor.

It is easy to convince oneself by trial and error or by more sophisticated means that the only form of wave function $\psi(\mathrm{x})$ which satisfies the condition $\Psi(\mathrm{x + D}) = \Psi(\mathrm{x}) \times (\text{ phase factor })$ is $\psi(\mathrm{x}) = \mathrm{A}\exp(\mathrm{ikx})$ where A is a (possibly complex) constant. This is just in the form of a complex exponential plane wave with wave number k. Thus, not only is the complex exponential wave function invariant under displacements in the manner defined above, it is the only wave function which is invariant to displacements. Furthermore, the phase factor which appears for a displacement D of such a plane wave takes the form $\exp(\mathrm{iC}) = \exp(\mathrm{ikD})$, where k is the wave number of the plane wave.

As an experiment, let us see if a wave packet is invariant under displacement. Let's define a wave packet consisting of two plane waves:

$$\psi(x) = \exp(ik_1 x) + \exp(ik_2 x) \tag{9.2.2}$$

Making the substitution $x \to x + D$ in this case results in

$$\begin{aligned} \psi(x + D) &= \exp[ik_1(x + D)] + \exp[ik_2(x + D)] \\ &= \exp(ik_1 x)\exp(ik_1 D) + \exp(ik_2 x)\exp(ik_2 D) \end{aligned} \tag{9.2.3}$$

$$\neq [\exp(i\kappa_1 x) + \exp(i\kappa_2 x)] \times (\text{ phase factor })$$

The impossibility of writing $\psi(x+D) = \psi(x) \times (\text{ phase factor })$ lends plausibility to the assertion that a single complex exponential is the only possible form of the wave function that is invariant under displacement.

Notice that the wave packet does not have definite wavenumber, and hence, momentum. In particular, the wave packet is a sum of complex exponentials with wavenumbers k_1 and k_2, which means that the associated particle can have either momentum $\Pi_1 = \hbar k_1$ or $\Pi_2 = \hbar k_2$. This makes sense from the point of view of the uncertainty principle – for a single plane wave the uncertainty in position is complete and the uncertainty in momentum is zero. For a wave packet the uncertainty in position is reduced and the uncertainty in the momentum is non-zero. As we have seen, this idea can be carried further: A definite value of momentum must be associated with a completely indefinite probability distribution in position, i. e., with P = constant. This corresponds to a wave function which has the form of a complex exponential plane wave. However, such a plane wave is invariant under displacement D, except for the multiplicative phase factor $\exp(ikD)$, , which has no physical consequences since it disappears when the probability distribution is obtained. Thus, we see that invariance under displacement of the wave function and a definite value of the momentum are linked, in that each implies the other:

$$\text{invariance under displacement} \iff \text{definite momentum} \tag{9.2.4}$$

The idea of potential energy was introduced in the previous chapter. In particular, we found that if the total energy is constant, then the momentum cannot be constant in the presence of spatially varying potential energy. This means that the wavenumber, and hence the wavelength of the oscillations in the wave function also vary with position. The spatial inhomogeneity of the potential energy gives rise to spatial inhomogeneity in the wave function, and hence an indefinite momentum.

The above argument can be extended to other variables besides momentum. In particular since the time dependence of a complex exponential plane wave is $\exp(-i\omega t) = \exp(-iEt/\hbar)$, where E is the total energy, we have by analogy with the above argument that

$$\text{invariance under time shift} \iff \text{definite energy} \tag{9.2.5}$$

Thus, invariance of the wave function under a displacement in time implies a definite value of the energy of the associated particle.

In the previous chapter we assumed that the frequency (and hence the energy) was definite and constant for a particle passing through a region of variable potential energy. We now see that this assumption is justified only if the potential energy doesn't change with time. This is because a time-varying potential energy eliminates the possibility of invariance under time shift.

Invariance

We have seen a few examples of invariance in quantum mechanics. It is now time to define this concept more precisely. A quantum mechanical wave function is said to be invariant under some transformation if the transformed wave function is observationally indistinguishable from the original.

In the above examples, the transformation is accomplished by replacing x by $x + D$ in the case of displacement in space and similarly by replacing t by $t + T$ for displacement in time. However, the idea of a transformation is much more general; other examples will be discussed as they arise.

The idea of "observationally indistinguishable" can be tricky. For example, if some transformation results in a new wave function which is the old wave function times a constant phase factor, then the new wave function is observationally indistinguishable from the old one. This is because physical measurements capture phase differences between different parts of wave functions (think of how interferometers work), but not absolute phases. The constant phase factor disappears in this difference calculation. However, if the multiplicative phase factor created by some transformation is a function of position, then the phase difference between different parts of a wave function changes as a result of the transformation. The wave function is not invariant under this transformation.

Compatible Variables

We already know that definite values of certain pairs of variables cannot be obtained simultaneously in quantum mechanics. For instance, the indefiniteness of position and momentum are related by the uncertainty principle — a definite value of position implies an indefinite value of the momentum and vice versa. If definite values of two variables can be simultaneously obtained, then we call these variables *compatible*. If not, the variables are *incompatible*.

If the wave function of a particle is invariant under the displacements associated with both variables, then the variables are compatible. For instance, the complex exponential plane wave associated with a free particle is invariant under displacements in

both space and time. Since momentum is associated with space displacements and energy with time displacements, the momentum and energy are compatible variables for a free particle.

Compatibility and Conservation

Variables which are compatible with the energy have a special status. The wave function which corresponds to a definite value of such a variable is invariant to displacements in time. In other words, the wave function doesn't change under this displacement except for a trivial phase factor. Thus, if the wave function is also invariant to some other transformation at a particular time, it is invariant to that transformation for all time. The variable associated with that transformation therefore retains its definite value for all time — i. e., it is *conserved*.

For example, the plane wave implies a definite value of energy, and is thus invariant under time displacements. At time $t = 0$, it is also invariant under x displacements, which corresponds to the fact that it represents a particle with a known value of momentum. However, since momentum and energy are compatible for a free particle, the wave function will represent the *same* value of momentum at all other times. In other words, if the momentum is definite at $t = 0$, it will be definite at all later times, and furthermore will have the same value. This is how the conservation of momentum (and by extension, the conservation of any other variable compatible with energy) is expressed in quantum mechanics.

Symmetries and Variables

In modern quantum physics, the discovery of new symmetries leads to new dynamical variables. In the problems we show how that comes about for the symmetries of parity $(x \rightarrow -x)$, time reversal $\mathrm{t} \rightarrow -\mathrm{t})$, and charge conjugation (the interchange of particles with antiparticles). One of the key examples of this was the development of the quark theory of matter, which came from the observation that the interchange of certain groups of elementary particles left the universe approximately unchanged, meaning that the universe was (approximately) symmetric under these interchanges.

This page titled 9.2: Symmetry and Quantum Mechanics is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

9.3: Confined Matter Waves

Confinement of a wave to a limited spatial region results in rather peculiar behavior — the wave can only fit comfortably into the confined region if the wave frequency, and hence the associated particle energy, takes on a limited set of possible values. This is the origin of the famous quantization of energy, from which the "quantum" in quantum mechanics comes. We will explore two types of confinement, position confinement due to a potential energy well, and rotational confinement due to the fact that rotation of an object through 2π radians returns the object to its original orientation.

Particle in a Box

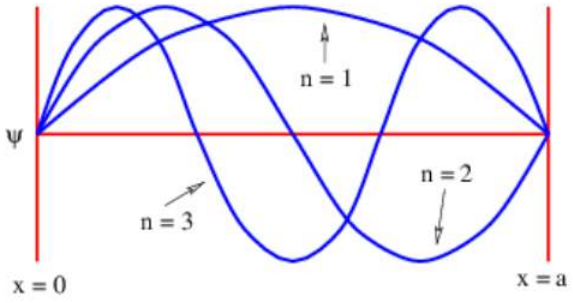


Figure 9.3.2:: First three modes for wave function of a particle in a box.

We now imagine how a particle confined to a region $0 \leq x \leq a$ on the x axis must behave. As with the displacement of a guitar string, the wave function must be zero at $x = 0$ and a, i. e., at the ends of the guitar string. A single complex exponential plane wave cannot satisfy this condition, since $|\exp[\mathrm{i}(\mathrm{kx} - \omega \mathrm{t})]|^2 = 1$ everywhere. However, a superposition (with a minus sign) of leftward and rightward traveling waves creates a standing wave, in which the the wave function separates into a function of space alone times a function of time alone.

$$\psi = \exp[i(kx - \omega t)] - \exp[i(-kx - \omega t)] = 2i \exp(-i\omega t) \sin(kx) \tag{9.3.1}$$

Notice that the time dependence is still a complex exponential, which means that $|\Psi|^2$ is independent of time. This insures that the probability of finding the particle somewhere in the box remains constant with time. It also means that the wave packet corresponds to a definite energy $\mathrm{E} = \hbar\omega$.

Because we took a difference rather than a sum of plane waves, the condition $\psi = 0$ is already satisfied at $x = 0$. To satisfy it at $x = a$, we must have $\mathrm{ka} = \mathrm{n}\pi$, where $\mathrm{n} = 1, 2, 3, \ldots$. Thus, the absolute value of the wave number must take on the discrete values

$$k_n = \frac{n\pi}{a}, \quad n = 1, 2, 3, \ldots \tag{9.3.2}$$

(The wavenumbers of the two plane waves equal plus or minus this absolute value respectively.) This implies that the absolute value of the particle momentum is $\Pi_n = \hbar \mathrm{k_n} = \mathrm{n}\hbar\mathrm{a}$, which in turn means that the energy of the particle must be

$$E_n = \left(\Pi_n^2 c^2 + m^2 c^4\right)^{1/2} = \left(n^2 \pi^2 \hbar^2 c^2 / a^2 + m^2 c^4\right)^{1/2} \tag{9.3.3}$$

where *m* is the particle mass. In the non-relativistic limit this becomes

$$E_n = \frac{\Pi_n^2}{2m} = \frac{n^2 \pi^2 \hbar^2}{2ma^2} \quad \text{(non-relativistic)} \tag{9.3.4}$$

where we have dropped the rest energy mc^2 since it is a constant offset. In the ultra-relativistic case where we can ignore the particle mass, we find

$$E_n = |\Pi_n| c = \frac{n\pi\hbar c}{a} \quad \text{(zero mass)} \tag{9.3.5}$$

The shapes of the wave functions for the first three values of n for the particle in the box are illustrated in Figure 9.3.2:.

In both limits the energy takes on only a certain set of possible values. This is called *energy quantization* and the integer n is called the *energy quantum number*. In the non-relativistic limit the energy is proportional to n^2, while in the ultra-relativistic case the energy is proportional to n.

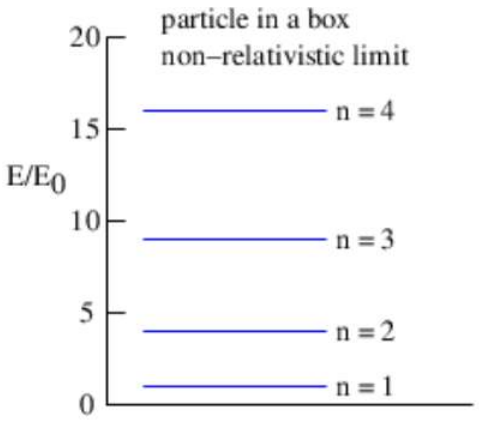


Figure 9.3.3:: Allowed energy levels for the non-relativistic particle in a box. The constant $\mathrm{E}_0 = \Pi^2\hbar^2/(2\mathrm{ma}^2)$. See text for the meanings of symbols.

We can graphically represent the allowed energy levels for the particle in a box by an energy level diagram. Such a diagram is shown in Figure 9.3.3: for the non-relativistic case.

One aspect of this problem deserves a closer look. Equation (9.3.1) shows that the wave function for this problem is a superposition of two plane waves corresponding to momenta $\Pi_1 = +\hbar\mathrm{k}$ and $\Pi_2 = -\hbar\mathrm{k}$ and is therefore a kind of wave packet. Thus, the wave function is not invariant under displacement and does not correspond to a definite value of the momentum — the momentum's absolute value is definite, but its sign is not. Following Feynman's prescription, equation (9.3.1) tells us that the amplitude for the particle in the box to have momentum $+\hbar\mathrm{k}$ is $\exp[\mathrm{i}(\mathrm{kx} - \omega\mathrm{t})]$, while the amplitude for it to have momentum $-\hbar k$ is $-\exp[\mathrm{i}(-\mathrm{kx} - \omega\mathrm{t})]$. The absolute square of the sum of these amplitudes gives us the relative probability of finding the particle at position x:

$$P(x) = |2i\exp(-i\omega t)\sin(kx)|^2 = 4\sin^2(kx) \tag{9.3.6}$$

Which of the two possible values of the momentum the particle takes on is unknowable, just as it is impossible in principle to know which slit a particle passes through in two slit interference. If an experiment is done to measure the momentum, then the wave function is irreversibly changed, just as the interference pattern in the two slit problem is destroyed if the slit through which the particle passes is unambiguously determined.

Barrier Penetration

Unlike the situation in classical mechanics, quantum mechanics allows the kinetic energy K to be negative. This makes the momentum Π (equal to $(2\mathrm{mK})^{1/2}$ in the nonrelativistic case) imaginary, which in turn gives rise to an imaginary wavenumber.

Let us investigate the nature of a wave with an imaginary wavenumber. Let us assume that $\mathrm{k} = \mathrm{i}\kappa$ in a complex exponential plane wave, where κ is real:

$$\psi = \exp[i(kx - \omega t)] = \exp(-\kappa x - i\omega t) = \exp(-\kappa x)\exp(-i\omega t) \tag{9.3.7}$$

The wave function doesn't oscillate in space when $\mathrm{K} = \mathrm{E} - \mathrm{U} < 0$, but grows or decays exponentially with x, depending on the sign of κ.

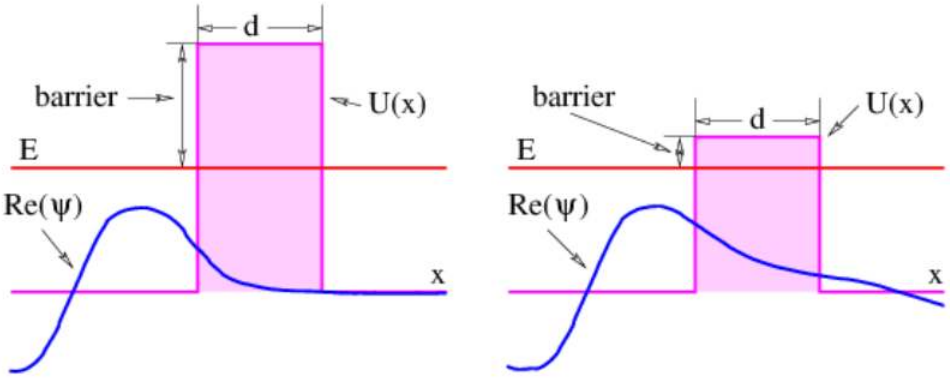


Figure 9.3.4:: Real part of wave function $\mathrm{Re}[\Psi(x)]$ for barrier penetration. The left panel shows weak penetration occurring for a large potential energy barrier, while the right panel shows stronger penetration which occurs when the barrier is small.

For a particle moving to the right, with positive k in the allowed region, κ turns out to be positive, and the solution decays to the right. Thus, a particle impingent on a potential energy barrier from the left (i. e., while moving to the right) will have its wave amplitude decay in the classically forbidden region, as illustrated in Figure 9.3.4:. If this decay is very rapid, then the result is almost indistinguishable from the classical result — the particle cannot penetrate into the forbidden region to any great extent. However, if the decay is slow, then there is a reasonable chance of finding the particle in the forbidden region. If the forbidden region is finite in extent, then the wave amplitude will be small, but non-zero at its right boundary, implying that the particle has a finite chance of completely passing through the classical forbidden region. This process is called *barrier penetration*.

The probability for a particle to penetrate a barrier is the absolute square of the amplitude after the barrier divided by the square of the amplitude before the barrier. Thus, in the case of the wave function illustrated in equation (9.3.7), the probability of penetration is

$$P = |\psi(d)|^2/|\psi(0)|^2 = \exp(-2\kappa d) \tag{9.3.8}$$

where *d* is the thickness of the barrier.

The rate of exponential decay with *x* in the forbidden region is related to how negative *K* is in this region. Since

$$-K = U - E = -\frac{\Pi^2}{2m} = -\frac{\hbar^2 k^2}{2m} = \frac{\hbar^2 \kappa^2}{2m} \tag{9.3.9}$$

we find that

$$\kappa = \left(\frac{2mB}{\hbar^2}\right)^{1/2} \tag{9.3.10}$$

where the potential energy barrier is $B \equiv -K = U - E$. The smaller B is, the smaller is κ, resulting in less rapid decay of the wave function with x. This corresponds to stronger barrier penetration. (Note that the way B is defined, it is positive in forbidden regions.)

If the energy barrier is very high, then the exponential decay of the wave function is very rapid. In this case the wave function goes nearly to zero at the boundary between the allowed and forbidden regions. This is why we specify the wave function to be zero at the walls for the particle in a box. These walls act in effect as infinitely high potential barriers.

Barrier penetration is important in a number of natural phenomena. Certain types of radioactive decay and the fissioning of heavy nuclei are governed by this process.

Orbital Angular Momentum

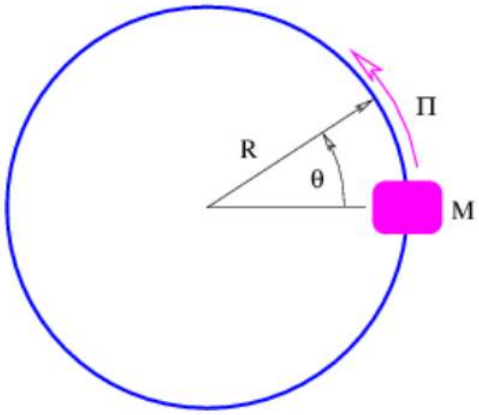


Figure 9.3.5:: Illustration of a bead of mass *M* sliding (without friction) on a circular loop of wire of radius *R* with momentum Π.

Another type of bound state motion occurs when a particle is constrained to move in a circle. (Imagine a bead sliding on a circular loop of wire, as illustrated in Figure 9.3.5:.) We can define x in this case as the path length around the wire and relate it to the angle $\theta : x = \mathrm{R}\theta$. For a plane wave we have

$$\psi = \exp[i(kx - \omega t)] = \exp[i(kR\theta - \omega t)] \tag{9.3.11}$$

This plane wave differs from the normal plane wave for motion along a Cartesian axis in that we must have $\psi(\theta) = \psi(\theta + 2\pi)$. This can only happen if the circumference of the loop, $2\pi\mathrm{R}$, is an integral number of wavelengths, i. e., if $2\pi\mathrm{R}/\lambda = \mathrm{m}$ where m is an integer. However, since $2\Pi/\lambda = \mathrm{k}$, this condition becomes $\mathrm{kR} = \mathrm{m}$.

Since $\Pi = \hbar k$, the above condition can be written $\Pi_m R = m\hbar$. The quantity

$$L_m \equiv \Pi_m R \tag{9.3.12}$$

is called the *angular momentum*, leading to our final result,

$$L_m = m\hbar, \quad m = 0, \pm 1, \pm 2, ... \tag{9.3.13}$$

We see that the angular momentum can only take on values which are integer multiples of $\hbar$. This represents the *quantization of angular momentum*, and *m* in this case is called the *angular momentum quantum number*. Note that this quantum number differs from the energy quantum number for the particle in the box in that zero and negative values are allowed.

The energy of our bead on a loop of wire can be expressed in terms of the angular momentum:

$$E_m = \frac{\Pi_m^2}{2M} = \frac{L_m^2}{2MR^2} \tag{9.3.14}$$

This means that angular momentum and energy are compatible variables in this case, which further means that angular momentum is a conserved variable. Just as definite values of linear momentum are related to invariance under translations, definite values of angular momentum are related to invariance under rotations. Thus, we have

$$\text{invariance under rotation} \iff \text{definite angular momentum} \tag{9.3.15}$$

for angular momentum.

We need to briefly address the issue of angular momentum in three dimensions. Angular momentum is actually a vector oriented perpendicular to the wire loop in the example we are discussing. The direction of the vector is defined using a variation on the right-hand rule: Curl your fingers in the direction of motion of the bead around the loop (using your right hand!). The orientation of the angular momentum vector is defined by the direction in which your thumb points. This tells you, for instance, that the angular momentum in Figure 9.3.5: points out of the page.

In quantum mechanics it is only possible to measure simultaneously the square of the length of the angular momentum vector and one component of this vector. Two different components of angular momentum cannot be simultaneously measured because of the uncertainty principle. However, the *length* of the angular momentum vector may be measured simultaneously with one component. Thus, in quantum mechanics, the angular momentum is completely specified if the length and one component of the angular momentum vector are known.

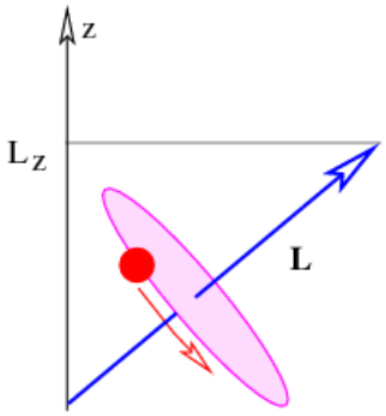


Figure 9.3.6:: Illustration of the angular momentum vector **L** for a tilted loop and its z component L_z.

Figure 9.6 illustrates the angular momentum vector associated with a bead moving on a wire loop which is tilted from the horizontal. One component (taken to be the z component) is shown as well. For reasons we cannot explore here, the square of the length of the angular momentum vector L^2 is quantized with the following values:

$$L_l^2 = \hbar^2 l(l+1), \quad l = 0, 1, 2, ... \tag{9.3.16}$$

One component (say, the z component) of angular momentum is quantized just like angular momentum in the two-dimensional case, except that l acts as an upper bound on the possible values of |m|. In other words, if the square of the length of the angular momentum vector is $\hbar^2 \mid (\mid +1)$, then the z component can take on the values

$$L_{zm} = \hbar m, \quad m = -l, -l+1, ..., l-1, l \tag{9.3.17}$$

The quantity *l* is called the *angular momentum quantum number*, while *m* is called the *orientation* or *magnetic quantum number*, the latter for historical reasons.

Angular Momentum

The type of angular momentum discussed above is associated with the movement of particles in orbits. However, it turns out that even stationary particles can possess angular momentum. This is called *spin* angular momentum. The *spin quantum number s* plays a role analogous to l for spin angular momentum, i. e., the square of the spin angular momentum vector of a particle is

$$L_s^2 = \hbar^2 s(s+1) \tag{9.3.18}$$

The spin orientation quantum number m_s is similarly related to s:

$$L_{zs} = \hbar m_s, \quad m_s = -s, -s+1, ..., s-1, s \tag{9.3.19}$$

The spin angular momentum for an elementary particle is absolutely conserved, i. e., it can never change. Thus, the value of s is an intrinsic property of a particle. The major difference between spin and orbital angular momentum is that the spin quantum number can take on more values, i. e., $s = 0, 1/2, 1, 3/2, 2, 5/2, ...$

Particles with integer spin values $s = 0, 1, 2, ...$ are called bosons after the Indian physicist Satyendra Nath Bose. Particles with half-integer spin values $s = 1/2, 3/2, 5/2, ...$ are called fermions after the Italian physicist Enrico Fermi. As we shall see later in the course, bosons and fermions play very different roles in the universe.

This page titled 9.3: Confined Matter Waves is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

9.4: Problems

1. Suppose that a particle is represented by the wave function $\Psi = \sin(\mathrm{kx} - \omega \mathrm{t}) + \sin(-\mathrm{kx} - \omega \mathrm{t})$.
    1. Use trigonometry to simplify this wave function.
    2. Compute the *x* and *t* dependence of the probability of finding the particle by squaring the wave function.
    3. Explain what this result says about the time dependence of the probability of finding the particle. Does this make sense?
2. Repeat the above problem for a particle represented by the wave function $\Psi = \exp[\mathrm{i}(\mathrm{kx} - \omega \mathrm{t})] + \exp[\mathrm{i}(-\mathrm{kx} - \omega \mathrm{t})]$.
3. Determine if the wavefunction $\psi(x) = \exp\left(iCx^2\right)$ is invariant under displacement in the sense that the displaced wave function differs from the original wave function by just a phase factor.
4. Just as invariance under the substitution $x \rightarrow x + D$ is associated with momentum, invariance under the substitution $x \rightarrow -x$ is associated with a quantum mechanical variable called parity, denoted P. However, unlike momentum, which can take on any numerical value, parity can take on only two possible values, ± 1. The parity of a wave function $\Psi(x)$ is $+1$ if $\Psi(-x) = \psi(x)$, while the parity is -1 if $\Psi(-\mathrm{x}) = -\Psi(\mathrm{x})$. If $\Psi(\mathrm{x})$ satisfies neither of these conditions, then it has no definite value of parity.
    1. What is the parity of $\Psi = \sin(\mathrm{kx})$? Of $\Psi = \cos(\mathrm{kx})$? The quantity k is a constant.
    2. Is $\psi(x) = \cos(kx)$ invariant under the substitution x = x + D for all possible values of D? Does this wave function have a definite value of the momentum?
    3. Show that a wave function with a definite value of the momentum does not have a definite value of parity. Are momentum and parity compatible variables?
5. Realizing that $\cos(kx - \omega t)$ can be written in terms of complex exponential functions, give a physical interpretation of the meaning of the above cosine wave function. In particular, what are the possible values of the associated particle's momentum and energy?
6. The time reversal operation T makes the substitution $\mathrm{t} \rightarrow -\mathrm{t}$. Similar to parity, time reversal can only take on values ± 1. Is symmetry of a wave function under time reversal, i. e., i. e., $\psi(-t) = \Psi(t)$, consistent with a definite value of the energy? Hint: Any wave function corresponding to a definite value of energy E must have the form $\Psi = A\exp(-\mathrm{iEt}/\hbar)$ where A is not a function of time t. (Why?)
7. The operation C takes the complex conjugate of the wave function, i. e., it makes the substitution $\mathrm{i} \rightarrow -\mathrm{i}$. In modern quantum mechanics this corresponds to interchanging particles and antiparticles, and is called charge conjugation. What does the combined operation CPT do to a complex plane wave, i. e., one with definite wave vector and frequency?
8. Make an energy level diagram for the case of a massless particle in a box.
9. Compare $|\Pi|$ for the ground state of a non-relativistic particle in a box of size a with $\Delta\Pi$ obtained from the uncertainty principle in this situation. Hint: What should you take for $\Delta \mathrm{x}$?
10. Imagine that a billiard table has an infinitely high rim around it. For this problem assume that $\hbar = 1\ \mathrm{kg\ m^2\ s^{-1}}$.
    1. If the table is 1.5 m long and if the mass of a billiard ball is M = 0.5 kg, what is the billiard ball's lowest or ground state energy? Hint: Even though the billiard table is two dimensional, treat this as a one-dimensional problem. Also, treat the problem nonrelativistically and ignore the contribution of the rest energy to the total energy.
    2. The energy required to lift the ball over a rim of height H against gravity is $\mathrm{U = MgH}$ where $\mathrm{g} = 9.8\ \mathrm{m\ s^{-2}}$. What rim height makes the gravitational potential energy equal to the ground state energy of the billiard ball calculated above.
    3. If the rim is actually twice as high as calculated above but is only 0.1 m thick, determine the probability of the ball penetrating the rim.

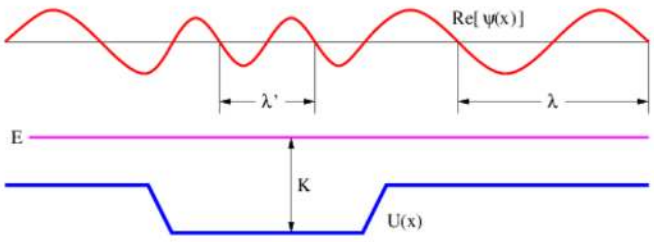


Figure 9.4.7**::** Real part of the wave function ψ, corresponding to a fixed total energy E, occurring in a region of spatially variable potential energy $\mathrm{U(x)}$. Notice how the wavelength λ changes as the kinetic energy $\mathrm{K = E - U}$ changes.

11. The real part of the wave function of a particle with positive energy E passing through a region of negative potential energy is shown in Figure 9.4.7:.

1. If the total energy is definitely *E*, what is the dependence of this wave function on time?
2. Is the wave function invariant under displacement in space in this case? Why or why not?
3. Does this wave function correspond to a definite value of momentum? Why or why not?
4. Is the momentum compatible with the energy in this case? Why or why not?

12. Assuming again that $\hbar = 1\ \mathrm{kg\ m^2\ s^{-1}}$, what are the possible speeds of a toy train of mass 3 kg running around a circular track of radius 0.8 m?
13. If a particle of zero mass sliding around a circular loop of radius R can take on angular momenta $\mathrm{L_m = m\hbar}$ where m is an integer, what are the possible kinetic energies of the particle? Hint: Remember that $\mathrm{L = IIR}$.

This page titled 9.4: Problems is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

10: Dynamics of Multiple Particles

10.1: Momentum and Newton’s Second Law
10.2: Newton’s Third Law
10.3: Conservation of Momentum
10.4: Collisions
10.5: Rockets and Conveyor Belts
10.6: Problems

This page titled 10: Dynamics of Multiple Particles is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

10.1: Momentum and Newton's Second Law

Up to this point we have stated Newton's second law in its conventional form, $\mathbf{F} = \mathrm{ma}$. However, in the non-relativistic case $\mathrm{ma} = \mathrm{md}\mathbf{u}/\mathrm{dt} = \mathrm{d}(\mathrm{mu})/\mathrm{dt}$, so we can also write Newton's second law as

$$\mathbf{F} = \frac{d\mathbf{p}}{dt} \quad \text{(Newton's second law)} \tag{10.1.1}$$

where $\mathbf{p} = m\mathbf{u}$ is the non-relativistic kinetic momentum. This form of Newton's second law is actually closer to Newton's original statement of the law. It also has the advantage that it is correct even in the relativistic case when the relativistic definition of kinetic momentum, $\mathbf{p} = m\mathbf{w}\left(1 - u^2/c^2\right)^{1/2}$ (as defined earlier), is substituted.

This page titled 10.1: Momentum and Newton's Second Law is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

10.2: Newton's Third Law

Newton's third law states that if particle A exerts a force **F** on particle B, then particle B exerts a force **-F** on particle A. Newton's third law makes it possible to apply Newton's second law to systems of particles without considering the detailed interactions between particles within the system.

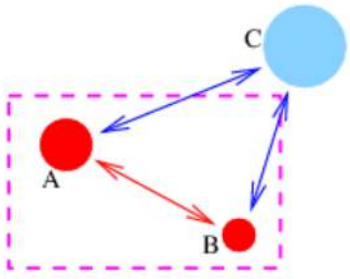


Figure 10.2.1:: Interactions between three particles, A, B, and C. A and B are considered to be part of the system defined by the dashed line.

For instance, if we (arbitrarily) define the system in Figure 10.2.1: to be the particles A and B inside the dashed lines, then we can divide the forces acting on these particles into internal and external parts:

$$\mathbf{F}_A = \mathbf{F}_{A-\text{ internal}} + \mathbf{F}_{A-\text{ external}} = \frac{d\mathbf{p}_A}{dt} \tag{10.2.1}$$

$$\mathbf{F}_B = \mathbf{F}_{B-\text{ internal}} + \mathbf{F}_{B-\text{ external}} = \frac{d\mathbf{p}_B}{dt} \tag{10.2.2}$$

Adding these equations together results in the net force $\mathbf{F}_{net}$ being equal to

$$\mathbf{F}_{A-\text{ internal}} + \mathbf{F}_{A-\text{ external}} + \mathbf{F}_{B-\text{ internal}} + \mathbf{F}_{B-\text{ external}} = \frac{d}{dt}(\mathbf{p}_A + \mathbf{p}_B) \tag{10.2.3}$$

However, the internal interactions in this case are A acting on B and B acting on A. These forces are equal in magnitude but opposite in direction, so they cancel out, leaving us with the net force equal to the sum of the external parts, $\mathbf{F}_{\text{net}} = \mathbf{F}_{\text{A-external}} + \mathbf{F}_{\text{B-external}}$. The external forces in Figure 10.2.1: are the force of C on A and the force of C on B. Defining the total kinetic momentum of the system as the sum of the A and B momenta, $\mathbf{p}_{\text{tot}} = \mathbf{p}_{\text{A}} + \mathbf{p}_{\text{B}}$, the above equation becomes

$$\mathbf{F}_{net} = \frac{d\mathbf{p}_{tot}}{dt} \tag{10.2.4}$$

which looks just like Newton's second law for a single particle, except that it now applies to the system of particles (A and B in the present case) as a whole. This argument easily generalizes to any number of particles inside and outside the system. Thus, for instance, even though a soccer ball consists of billions of atoms, we are sure that the forces between atoms within the soccer ball cancel out, and the trajectory of the ball as a whole is determined solely by external forces such as gravity, wind drag, friction with the ground, and the kicks of soccer players.

Remember that for two forces to be a third law pair, they have to be acting on different particles. Furthermore, if one member of the pair is the force of particle A acting on particle B, then the other must be the force of particle B acting on particle A. A counterexample would be gravity and the upward normal force acting on a mass sitting on a table; these forces are equal and opposite in the stationary case, but act on the same object, and therefore *do not* constitute a third law pair. However, the upward normal force of the table on the mass and the downward normal force of the mass on the table *would be* a third law pair.

This page titled 10.2: Newton's Third Law is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

10.3: Conservation of Momentum

If all external forces on a system are zero, then equation (10.5) reduces to

$$\mathbf{p}_{tot} = \text{ const} \qquad (\text{isolated system}). \tag{10.3.1}$$

A system of particles with no external forces acting on it is called *isolated*. Newton's third law thus tells us that the kinetic momentum of an isolated system doesn't change with time. This law is called the *conservation of momentum*.

This page titled 10.3: Conservation of Momentum is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

10.4: Collisions

Let us now consider the situation in which two particles collide with each other. There can be several outcomes to this collision, of which we will study two:

- The two particles collide *elastically*, in essence bouncing off of each other.
- The two particles stick together, resulting in the production of a single particle, or a single particle breaks apart into two particles. These are *inelastic processes*.

In both of the above cases energy and momentum are conserved. We assume that the forces acting between the particles are short range, so that except in the instant of collision, we need not worry about potential energy or potential momentum — all energy is in the form of rest plus kinetic energy except in this short interval, and all momenta are kinetic momenta.

Because of the principle of relativity, we are free to consider collisions in any convenient reference frame. We can then transform the results to any reference frame we please. Generally speaking, the most convenient reference frame to consider is the one in which the total momentum of the two particles is zero. For the sake of simplicity we only consider collisions in one dimension.

Elastic Collisions

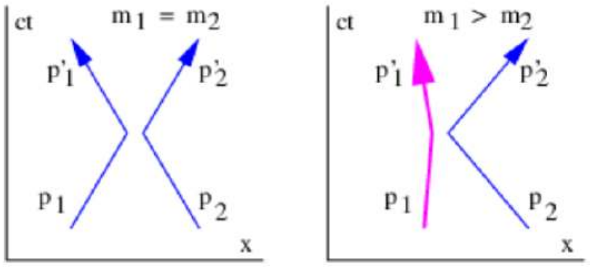


Figure 10.4.2:: One-dimensional elastic collisions of two particles in the center of momentum frame as seen in spacetime diagrams.

Suppose a particle with mass m_1 and initial velocity u_1 in the center of momentum frame, i. e., the reference frame in which the total momentum is zero, collides elastically with another particle of mass m_2 with initial velocity u_2. The momenta of the two particles are

$$p_1 = \frac{m_1 u_1}{\left(1 - u_1^2/c^2\right)^{1/2}} \quad p_2 = \frac{m_2 u_2}{\left(1 - u_2^2/c^2\right)^{1/2}} \tag{10.4.1}$$

In the center of momentum frame we must have

$$p_1 = -p_2 \tag{10.4.2}$$

Figure 10.4.2: shows what happens when these two particles collide. The first particle acquires momentum p_1' while the second acquires momentum p_2'. The conservation of momentum tells us that the total momentum after the collision is the same as before the collision, namely zero, so

$$p_1' = -p_2' \tag{10.4.3}$$

In the center of momentum frame we know that $|p_1| = |p_2|$ and we know that the two momentum vectors point in opposite directions. Similarly, $|p_1'| = |p_2'|$. However, we as yet don't know how p_1' is related to p_1. Conservation of energy,

$$E_1 + E_2 = E_1' + E_2' \tag{10.4.4}$$

gives us this information. Notice that if $p_1' = -p_1$, then $E_1'^2 = p_1'^2c^2 + m_1^2c^4 = p_1^2c^2 + m_1^2c^4 = E_1^2$. Assuming positive energies, we therefore have $E_1' = E_1$. If $p_2' = -p_2$, then we can similiarly infer that $E_2' = E_2$. If these conditions are satisfied, then so is equation (10.4.4). Therefore, a complete solution to the problem is

$$p_1 = -p_1' = -p_2 = p_2' \equiv p \tag{10.4.5}$$

and

$$E_1 = E_1' = \left(p^2c^2 + m_1^2c^4\right)^{1/2} \quad E_2 = E_2' = \left(p^2c^2 + m_2^2c^4\right)^{1/2} \tag{10.4.6}$$

In other words, the particles just exchange momenta.

The left panel of Figure 10.4.2: shows what happens in a collision when the masses of the two colliding particles are equal. If $m_1 = m_2$, then the incoming and outgoing velocities of the two particles are the same, as indicated by the inverse slopes of the world lines. On the other hand, if $m_1 > m_2$, then the velocity of particle 2 is greater than the velocity of particle 1, as is illustrated in the right panel of Figure 10.4.2:.

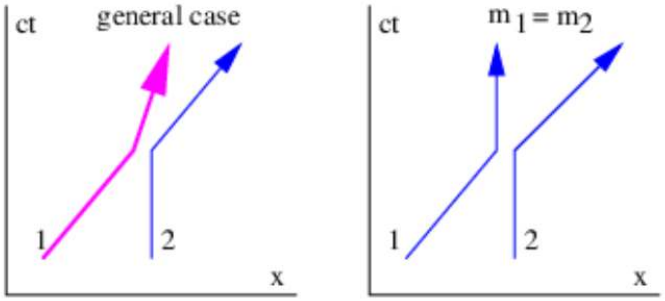


Figure 10.4.3:: Elastic collisions viewed from a reference frame in which one particle is initially stationary.

Suppose we wish to view the results of an elastic collision in a reference frame in which particle 2 is initially stationary. All we have to do is to transform the velocities into a reference frame moving with the initial velocity of particle 2, as illustrated in Figure 10.4.3:. We do this by relativistically adding $U = -u_2$ to each velocity. (Note that the velocity U of the moving frame is positive since u_2 is negative.) Using the relativistic velocity translation formula, we find that

$$v_1 = \frac{u_1 + U}{1 + u_1 U/c^2} \quad v_1' = \frac{u_1' + U}{1 + u_1' U/c^2} \quad v_2' = \frac{u_2' + U}{1 + u_2' U/c^2} \tag{10.4.7}$$

where $u_1, u_1', u_2,$ and u_2' indicate velocities in the original, center of momentum reference frame and v_1, v_1', etc., indicate velocities in the transformed frame.

In the special case where the masses of the two particles are equal to each other, we have $v_1 = 2U/\left(1 + U^2/c^2\right), v_1' = 0,$ and $v_2' = 2U/\left(1 + U^2/c^2\right) = v_1$. Thus, when the masses are equal, the particles simply exchange velocities.

If the velocities are nonrelativistic, then the simpler Galilean transformation law $v = u + U$ can be used in place of the relativistic equations invoked above.

Inelastic Collisions

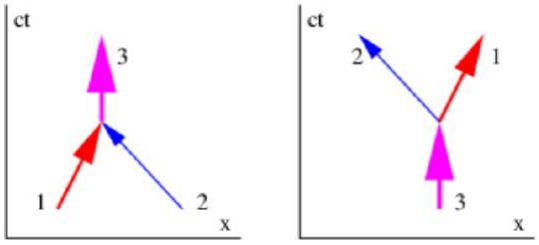


Figure 10.4.4:: Building blocks of inelastic collisions. In the left panel two particles collide to form a third particle. In the right panel a particle breaks up, forming two particles.

An inelastic collision is one in which the particles coming out of the collision are not the same as the particles going into it. Inelastic collisions conserve both total momentum and energy just as elastic collisions do. However, unlike elastic collisions, inelastic collisions generally do not conserve the total kinetic energy of the particles, as some rest energy is generally created or destroyed.

Figure 10.4.4: shows the fundamental building blocks of inelastic collisions. We can consider even the most complex inelastic collisions to be made up of composites of only two processes, the creation of one particle from two, and the disintegration of one particle into two.

Let us consider each of these in the center of momentum frame. In both cases the single particle must be stationary in this frame since it carries the total momentum of the system, which has to be zero. By conservation of momentum, if particle 1 in the left

panel of Figure 10.4.4: has momentum p, then the momentum of particle 2 is -p. If the two particles have masses m_1 and m_2, then their energies are $E_1 = \left(p^2c^2 + m_1^2c^4\right)^{1/2}$ and $E_2 = \left(p^2c^2 + m_2^2c^4\right)^{1/2}$. The energy of particle 3 is therefore $E_3 = E_1 + E_2$, and since it is at rest, all of its energy is in the form of "mc^2" or rest energy, and so the mass of this particle is

$$\begin{aligned} m_3 &= \left(E_1 + E_2\right)/c^2 \quad (\text{ center of momentum frame }) \\ &= \left(p^2/c^2 + m_1^2\right)^{1/2} + \left(p^2/c^2 + m_2^2\right)^{1/2} \\ &= m_1\left[1 + p^2/\left(m_1^2c^2\right)\right]^{1/2} + m_2\left[1 + p^2/\left(m_2^2c^2\right)\right]^{1/2} \end{aligned} \tag{10.4.8}$$

The last line in the above equation shows that $m_3 > m_1 + m_2$ because it is in the form $m_1A + m_2B$ where both A and B are greater than one. Thus, rest energy is created in the amount $\Delta E_{\text{rest}} = \left(m_3 - m_1 - m_2\right)c^2$.

Actually, it is easy to calculate the mass of particle 3 in the above case from any reference frame as long as the momenta and energies of particles 1 and 2 are known in this frame. By conservation of energy and momentum, $E_3 = E_1 + E_2$ and $\mathbf{p}_3 = \mathbf{p}_1 + \mathbf{p}_2$. Furthermore, $E_3^2 = p_3^2c^2 + m_3^2c^4$, so we can solve for m_3:

$$m_3 = \left[\left(E_1 + E_2\right)^2/c^4 - \left(\mathbf{p}_1 + \mathbf{p}_2\right)\cdot\left(\mathbf{p}_1 + \mathbf{p}_2\right)/c^2\right]^{1/2} \quad (\text{ any frame }) \tag{10.4.9}$$

The right panel of Figure 10.4.4: shows the process of particle decay. This is just the inverse of the particle creation process, and all of the analysis we have done for creation is valid for particle decay except that rest energy is converted to kinetic energy rather than vice versa.

This page titled 10.4: Collisions is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

10.5: Rockets and Conveyor Belts

Normally when we define a system to which Newton's second law is to be applied, the system is *closed* in the sense that mass cannot enter or exit the system. However, sometimes it is convenient to work with *open* systems for which this is not true. The classic example is the rocket, where exhaust gases leave the system, thus decreasing the mass of the rocket with time.

Open systems can be analyzed if momentum is considered to be a quantity which is accounted for much as money is accounted for in a bank account. The bank account can change in three ways: money can be deposited in the account, it can be withdrawn from the account, and the amount can grow or shrink as a consequence of interest payments or fees. Similarly, the amount of momentum in a system can change as the result of mass entering the system, mass leaving the system, and forces acting on the system. The time rate of change of momentum in a system is therefore

$$\frac{dp}{dt} = F + \left(\frac{dp}{dt}\right)_{in} - \left(\frac{dp}{dt}\right)_{out} \tag{10.5.1}$$

where F is the net force on the system, $(\mathrm{dp/dt})_{\mathrm{in}}$ is the momentum per unit time added by mass entering the system, and $(\mathrm{dp}\ /\mathrm{dt})_{\ \mathrm{out}}$ is the amount lost per unit time by mass exiting the system. In the non-relativistic case, $(\mathrm{dp/dt})_{\mathrm{in}} = \mathrm{u_{in}}(\mathrm{dm/dt})_{\mathrm{in}}$ and $(\mathrm{dp/dt})_{\mathrm{out}} = \mathrm{u_{out}}(\mathrm{dm/dt})_{\mathrm{out}}$, where $(\mathrm{dm/dt})_{\mathrm{in}}$ is the mass entering the system per unit time with velocity $\mathrm{u_{in}}$ and $(\mathrm{dm/dt})_{\mathrm{out}}$ is the mass per unit time exiting the system with velocity $\mathrm{u_{out}}$.

For non-relativistic velocities, the momentum of the system can be written as $p = mu$ so that

$$\frac{dp}{dt} = \frac{dm}{dt}u + m\frac{du}{dt} \tag{10.5.2}$$

To complete the analysis, we need an accounting of the mass entering and leaving the system:

$$\frac{dm}{dt} = \left(\frac{dm}{dt}\right)_{in} - \left(\frac{dm}{dt}\right)_{out} \tag{10.5.3}$$

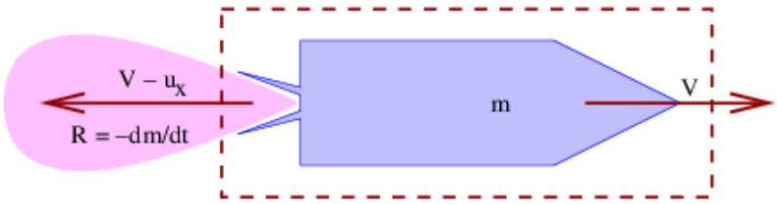


Figure 10.5.5:: Rocket moving with velocity V while expelling gas at a rate R with velocity V - $\mathrm{u_x}$.

Let us see how to apply this to a rocket for which all velocities are non-relativistic. As Figure 10.5.5: indicates, a rocket spews out a stream of exhaust gas. The system is defined by the dashed box and includes the rocket and the part of the exhaust gas inside the box. The reaction to the momentum carried off in this stream of gas is what causes the rocket to accelerate. We note that $(\mathrm{dm/dt})_{\mathrm{in}} = 0$ since no mass is entering the system, and $(\mathrm{dm/dt})_{\mathrm{out}} = \mathrm{R}$, the rate at which mass is ejected by the rocket in the form of exhaust gas. The rocket is assumed to be moving to the right at speed V and the gas is ejected at a speed $\mathrm{u_x}$ relative to the rocket, which means that its actual velocity after ejection is V - $\mathrm{u_x}$. We call $\mathrm{u_x}$ the exhaust velocity. Notice that V - $\mathrm{u_x}$ may be either positive or negative, depending on how big V is.

Equating the mass of the rocket to the system mass, we find that $\mathrm{R} = -\mathrm{dm/dt}$. The momentum balance equation (10.5.1) becomes $\mathrm{dp/dt} = -(\mathrm{V} - \mathrm{u_x})\,\mathrm{R}$. The force on the rocket is actually zero, so the force term does not enter the momentum balance equation. This is non-intuitive, because we are used to acceleration being the result of a force. However, nothing, including the ejected gas, is actually pushing on the system, so we must indeed conclude that there is no force — all of the change in the system's momentum arises from the ejection of gas with the opposite momentum.

Finally, we see that $\mathrm{dp/dt} = (\mathrm{dm/dt})\mathrm{V} + \mathrm{m}(\mathrm{dV/dt}) = -\mathrm{RV} + \mathrm{m}(\mathrm{dV/dt})$. Equating this to the results of the momentum balance calculation gives us $-\mathrm{RV} + \mathrm{m}(\mathrm{dV/dt}) = -(\mathrm{V} - \mathrm{u_X})\,\mathrm{R}$. Solving for the acceleration $\mathrm{dV/dt}$ results in

$$\frac{dV}{dt} = \frac{u_x R}{m} \quad \text{(rocket acceleration)} \tag{10.5.4}$$

Thus, the acceleration of the rocket depends on the exhaust velocity of the ejected gas, the rate at which the gas is being ejected, and the mass of the rocket.

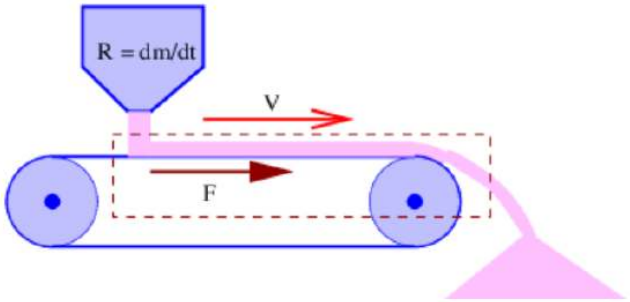


Figure 10.5.6:: Sand is dumped on a conveyor belt and in turn is dumped off the end of the belt.

Figure 10.5.6: illustrates another type of open system problem. A hopper dumps sand on a conveyor belt at a rate of R kilograms per second. The conveyor belt is moving to the right at (non-relativistic) speed V and the sand is dumped off at the end. What force F is needed to keep the conveyor belt moving at a constant speed, assuming that the conveyor belt mechanism itself is frictionless? In this case $(\mathrm{dm/dt})_{\mathrm{in}} = (\mathrm{dm/dt})_{\mathrm{out}} = \mathrm{R}$. Furthermore, since the system outlined by the dashed line is in a steady state, $\mathrm{dp/dt} = 0$.

The key to understanding this problem is that the sand enters the system with zero horizontal velocity, but exits the system with the horizontal velocity of the conveyor belt, V . The momentum balance equation is thus

$$0 = F - VR \tag{10.5.5}$$

and the force is

$$F = VR \text{ (force on conveyor belt).} \tag{10.5.6}$$

This force serves to accelerate the sand up to the velocity of the conveyor belt.

10.5: Rockets and Conveyor Belts is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

10.6: Problems

1. Imagine a block of mass M resting on a plate under the influence of gravity, as shown in Figure 10.6.7:.
    1. Determine the force of the plate on the block, $\mathbf{N}_b$, and the force of the block on the plate, $\mathbf{N}_p$.
    2. State which of the three forces, $Mg, \mathbf{N}_b, \text{ and } \mathbf{N}_p$, form a Newton's third law pair.

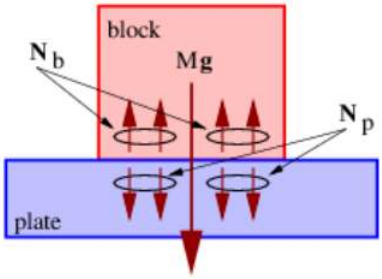


Figure 10.6.7**::** Block of mass *M* subject to gravitational force Mgwhile resting on a plate. The force of the block on the plate is Np while the force of the plate on the block is Nb.

2. Repeat the previous problem assuming that the block and the plate are in an elevator accelerating upward with acceleration **a**.
3. Straighten out the misunderstanding of Newton's third law implicit in the question "If the force of the horse on the cart equals the force of the cart on the horse, why does anything ever go anywhere"? Examine in particular the conditions under which the horse-cart system accelerates.
4. A pusher boat (mass M) on the Mississippi is pushing two barges (each mass m) at a steady speed as shown in Figure 10.6.8:. Each barge is subject to a drag force by the water of F_B. Consider only horizontal force components in the following.
    1. What is the total horizontal force of the water on the barge-boat system? Explain.
    2. What is the direction and magnitude of the force of the pusher boat on barge 1? Explain.

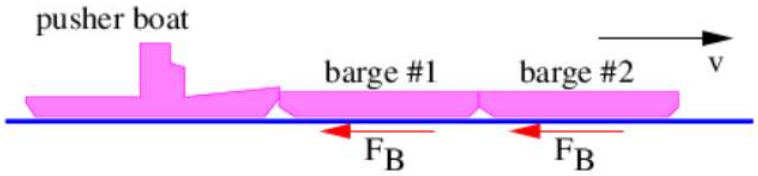


Figure 10.6.8**::** Barges being pushed by a pusher boat on the Mississippi. Each barge experiences a drag force F_b.

5. A train with an engine of mass M and 2 freight cars, each of mass m, is accelerating to the right with acceleration a on a horizontal track as shown in Figure 10.6.9:. Assume that the two freight cars roll with negligible friction. Consider only horizontal force components below.
    1. Find the direction and magnitude of the force of the rails on the engine and specify the system to which Newton's second law is applied.
    2. Find the direction and magnitude of the force of the engine on the first car and specify the system to which Newton's second law is applied.
    3. Find the direction and magnitude of the force of the first car on the second car and specify the system to which Newton's second law is applied.
    4. Find the direction and magnitude of the force of the second car on the first car and specify the law used to obtain this force.



Figure 10.6.9**::** An engine and two freight cars accelerating to the right.

6. A car and trailer are descending a hill as shown in Figure 10.6.10:. Assume that the trailer rolls without friction and that air friction can be ignored. Consider only forces parallel to the road surface.

1. Compute the force of the road on the car if the car-trailer system shown in Figure 10.6.10: is moving down the hill at constant speed.
2. Compute the force of the trailer on the car in the above conditions.
3. If the driver takes his foot off the brake and lets the car coast frictionlessly, recompute the force of the trailer on the car.

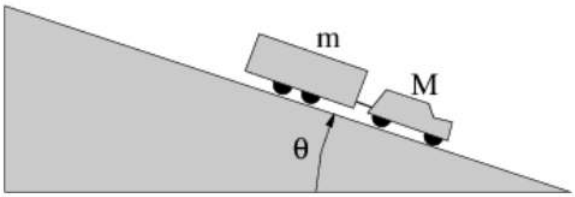


Figure 10.6.10**::** A car and a trailer going down a hill.

7. Consider a one-dimensional elastic collision between particles of masses m_1 and m_2. If particle 2 is initially stationary, what range of values must m_1/m_2 have for the initial particle to rebound backwards along its initial track after the collision? (Do this problem non-relativistically.)
8. A stationary pion (mass M) decays into a muon (mass $m < M$)) and a neutrino (massless).
    1. What is the (fully relativistic) momentum of the muon after the decay?
    2. What is the energy of the neutrino?
9. In an elastic collision viewed in the center of momentum frame, the energy of each particle is conserved individually. Is this true for the same process viewed from a reference frame in which one of the particles is initially stationary?
10. A space probe approaches a planet in the -x direction, curves around it under the influence of the planet's powerful gravity (a conservative force) and recedes from the planet in the +x direction, as seen in Figure 10.6.11:. The planet is moving in the +x direction at speed V , while the space probe is initially moving in the -x direction at speed u_1. What is its speed u_2 in the +x direction after this close approach to the planet? Treat this problem nonrelativistically. Hint: First transform to the center of mass frame in which the planet is essentially stationary. Work out the interaction between the probe and the planet in this frame. Then transform back to the original reference frame. Assume that the mass of the probe is negligible compared to that of the planet.

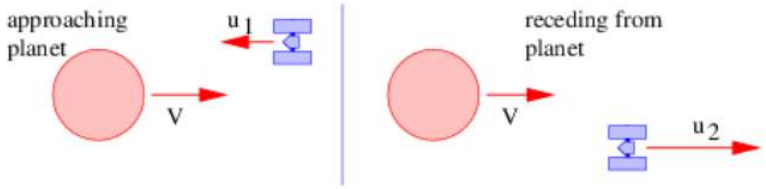
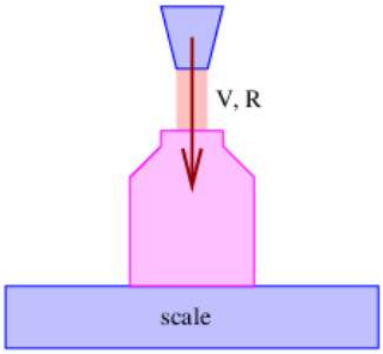


Figure 10.6.11**::** A space probe approaches a planet, curves around it, and heads off in the opposite direction.

11. Two asteroids, each with mass 10^{10} kg and initial speed 10^5 m s^{-1}, collide head on. The whole mess congeals into one large mass. How much rest mass (rest energy divided by c^2) is created?
12. Two equal objects, both with mass m, collide and stick together. Before the collision, one mass is stationary and the other is moving at speed v. In the following, assume that velocities are fully relativistic.
    1. Compute the total momentum and energy (including rest energy) of the two masses before the collision.
    2. Compute the mass *M* of the combined system after the collision, taking the conversion of energy into mass into account.
13. Explain qualitatively why a fireman needs to push forward on a firehose to keep it stationary. Hint: The water is flowing faster after it comes out of the nozzle of the hose than before.
14. Solve equation (10.19) for V as a function of m, assuming that $\mathrm{V} = 0 \text{ and } \mathrm{m} = \mathrm{m}_0$ at t = 0. Hint: Since $\mathrm{R} = -\mathrm{dm}/\mathrm{dt}$, we have $\mathrm{R}/\mathrm{m} = -\mathrm{d}\ln(\mathrm{m})/\mathrm{dt}$.
15. Bottles are filled with soft drink at a bottling plant as shown in Figure 10.6.12:. The bottles sit on a scale which is used to determine when to shut off the flow of soft drink. If the desired mass of the bottle plus soft drink after filling is M, what weight should the scale read when the bottle is full? The rate at which mass is being added to the bottle is R and its velocity entering



the bottle is V . **Figure** 10.6.12**::** A bottle being filled with soft drink at a rate *R*. The liquid enters the bottle with velocity *V* .

16. An interstellar space probe has frontal area A, initial mass M_0, and initial velocity V_0, which is non-relativistic. The tenuous gas between the stars has mass density ρ. These gas molecules stick to the probe when they hit it. Find the probe's acceleration. Hint: In a frame of reference in which the gas is stationary, does the momentum of the space probe change with time? Does its mass?
17. A light beam with power J hits a plate which is oriented normally to the beam. Compute the force required to hold the plate in place if
    1. the plate completely absorbs the light, and
    2. the plate completely reflects the light. (Hint: Photons are massless, so the momentum of a photon with energy E is E/c. Thus, the momentum per unit time hitting the plate is J/c. .)
18. Find the acceleration of a rocket when the exhaust "gas" is actually a laser beam of power J. Assume that the rocket moves at non-relativistic velocities and that the decrease in mass due to the loss of energy in the laser beam is negligible.

10.6: Problems is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

CHAPTER OVERVIEW

11: Rotational Dynamics

We have already seen the quantum mechanical treatment of angular momentum and rotational dynamics. In this section we study these subjects in a classical, non-relativistic context. We first define the concepts of torque and angular momentum in order to understand the orbital motion of a single particle. Next we examine two particles in arbitrary motion and learn how kinetic energy and angular momentum are partitioned between orbital and internal components. Two particles fixed to the ends of a light rod constitute a dumbbell, which serves as a prototype for the rotation of rigid bodies. We then see how what we learned for two particles extends to an arbitrary number of particles. Finally, we explore the physics of structures in static equilibrium.

Before we begin, we need to extend our knowledge of vectors to the cross product.

11.1: Math Tutorial — Cross Product
11.2: Torque and Angular Momentum
11.3: Two Particles
11.4: The Uneven Dumbbell
11.5: Many Particles
11.6: New Page
11.7: Statics
11.8: Problem

This page titled 11: Rotational Dynamics is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

11.1: Math Tutorial — Cross Product

There are two ways to multiply two vectors together, the dot product and the cross product. We have already studied the dot product of two vectors, which results in a scalar or single number.

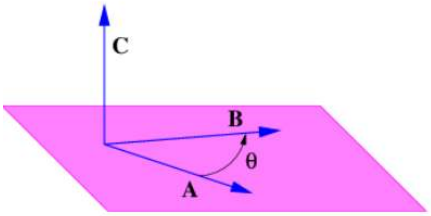


Figure 11.1.1:: Illustration of the cross product of two vectors **A** and **B**. The resulting vector **C** is perpendicular to the plane defined by **A** and **B**.

The cross product of two vectors results in a third vector, and is written symbolically as follows:

$$\mathbf{C} = \mathbf{A} \times \mathbf{B} \tag{11.1.1}$$

As illustrated in Figure 11.1.1:, the cross product of two vectors is perpendicular to the plane defined by these vectors. However, this doesn't tell us whether the resulting vector in Figure 11.1.1: points upward out of the plane or downward. This ambiguity is resolved using the *right-hand rule*:

1. Point the uncurled fingers of your right hand along the direction of the first vector **A**.
2. Rotate your arm until you can curl your fingers in the direction of the second vector **B**.
3. Your stretched out thumb now points in the direction of the cross product vector **C**

The magnitude of the cross product is given by

$$|\mathbf{C}| = |\mathbf{A}||\mathbf{B}| \sin(\theta) \tag{11.1.2}$$

where |**A**| and |**B**| are the magnitudes of **A** and **B**, and θ is the angle between these two vectors. Note that the magnitude of the cross product is zero when the vectors are parallel or anti-parallel, and maximum when they are perpendicular. This contrasts with the dot product, which is maximum for parallel vectors and zero for perpendicular vectors.

Notice that the cross product does not commute, i. e., the order of the vectors is important. In particular, it is easy to show using the right-hand rule that

$$\mathbf{A} \times \mathbf{B} = -\mathbf{B} \times \mathbf{A} \tag{11.1.3}$$

An alternate way to compute the cross product is most useful when the two vectors are expressed in terms of components, i. e., $A = (A_x, A_y \cdot A_z)$ and $B = (B_x, B_y, B_z)$

$$\begin{aligned} C_x &= A_y B_z - A_z B_y \\ C_y &= A_z B_x - A_x B_z \\ C_z &= A_x B_y - A_y B_x \end{aligned} \tag{11.1.4}$$

Notice that once you have the first of these equations, the other two can be obtained by cyclically permuting the indices, i. e., $x \rightarrow y$, $y \rightarrow z$, and $z \rightarrow x$. This is useful as a memory aid.

This page titled 11.1: Math Tutorial — Cross Product is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

11.2: Torque and Angular Momentum

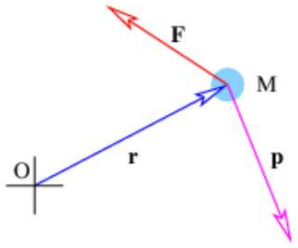


Figure 11.2.2:: A mass M located at r relative to the origin O has momentum p and has a force F applied to it. By the right-hand rule the torque $\tau = \mathbf{r} \times \mathbf{F}$ points out of the page, while the angular momentum $\mathbf{L} = \mathbf{r} \times \mathbf{p}$ points into the page.

Torque is the action of a force **F** on a mass *M* which induces it to revolve about some point, called the *origin*. It is defined

$$\tau = \mathbf{r} \times \mathbf{F} \tag{11.2.1}$$

where **r** is the position of the mass relative to the origin, as illustrated in Figure 11.2.2:.

Notice that the torque is zero in a number of circumstances. If the force points directly toward or away from the origin, the cross product is zero, resulting in zero torque, even though the force is non-zero. Likewise, if **r** = 0, the torque is zero. Thus, a force acting at the origin produces no torque. Both of these limits make sense intuitively, since neither induces the mass to revolve around the origin.

The *angular momentum* of a mass *M* relative to a point O is

$$\mathbf{L} = \mathbf{r} \times \mathbf{p} \tag{11.2.2}$$

where **p** is the ordinary kinetic momentum of the mass. The angular momentum is zero if the motion of the object is directly towards or away from the origin, or if it is located at the origin.

If we take the cross product of the position vector and Newton's second law, we obtain an equation that relates torque and angular momentum:

$$\mathbf{r} \times \mathbf{F} = \mathbf{r} \times \frac{d\mathbf{p}}{dt} = \frac{d}{dt}(\mathbf{r} \times \mathbf{p}) - \frac{d\mathbf{r}}{dt} \times \mathbf{p} \tag{11.2.3}$$

The second term on the right side of the above equation is zero because $d\mathbf{r}/dt$ equals the velocity of the mass, which is parallel to its momentum and the cross product of two parallel vectors is zero. This equation can therefore be written

$$\tau = \frac{d\mathbf{L}}{dt} \quad \text{(Newton's second law for rotation).} \tag{11.2.4}$$

It is the rotational version of Newton's second law.

For both torque and angular momentum the location of the origin is arbitrary, and is generally chosen for maximum convenience. However, it is necessary to choose the same origin for both the torque and the angular momentum.

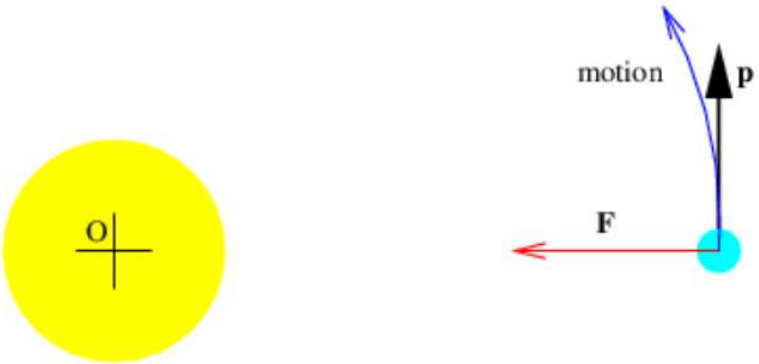


Figure 11.2.3:: A convenient choice of origin for a planet (right-hand sphere) revolving around the sun is simply the center of the sun. In this case the torque of the sun's gravitational force on the planet is zero.

For the case of a *central force*, i. e., one which acts along the line of centers between two objects (such as gravity), there often exists a particularly convenient choice of origin. Imagine a planet revolving around the sun, as illustrated in Figure 11.2.3:. If the origin is placed at the center of the sun (which is assumed not to move under the influence of the planet's gravity), then the torque exerted on the planet by the sun's gravity is zero, which means that the angular momentum of the planet about the center of the sun is constant in time. No other choice of origin would yield this convenient result.

We already know about two fundamental conservation laws — those of energy and linear momentum. We believe that angular momentum is similarly conserved in isolated systems. In other words, particles can exchange angular momentum between themselves, but the vector sum of the angular momentum of all the particles in a system isolated from outside influences must remain constant.

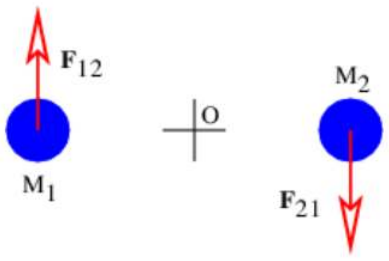


Figure 11.2.4:: Scenario for the non-conservation of angular momentum. $\mathbf{F}_{12}$ is the force of mass M_2 on mass M_1 and vice versa.

Conservation of angular momentum is not an automatic consequence of the conservation of linear momentum, even though the governing equation (11.2.4) for angular momentum is derived from Newton's second law. As an example, Figure 11.2.4: shows a hypothetical situation in which the force $\mathbf{F}_{21}$ of M_1 on M_2 is equal in magnitude but opposite in sign to the force $\mathbf{F}_{12}$ of M_2 on M_1, i. e., Newton's third law holds, and the sum of the momenta of the two masses is conserved. However, because the forces are non-central, the angular momentum of the masses is not conserved. This scenario is impossible if the forces are central.

This page titled 11.2: Torque and Angular Momentum is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

11.3: Two Particles

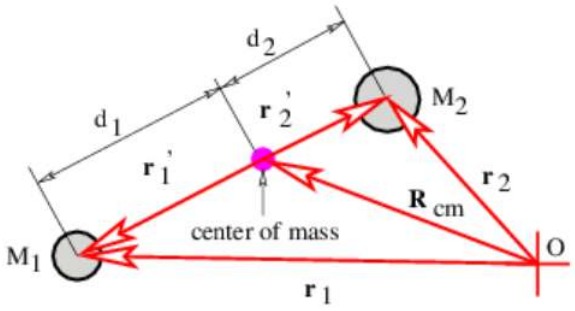


Figure 11.3.5:: Two particles of mass M_1 and M_2 with $M_2 > M_1$.

Suppose we wish to apply Newton's second law to two particles considered together as a single system. As we showed previously, only external forces act on the total momentum, $\mathbf{p}_{\text{total}} = \mathbf{p}_1 + \mathbf{p}_2$, of the two particles:

$$\mathbf{F}_{\text{external}} = \frac{d\mathbf{p}_{\text{total}}}{dt} \tag{11.3.1}$$

Let's write the total non-relativistic momentum of the two particles in a special way:

$$\mathbf{p}_{\text{total}} = M_1\mathbf{v}_1 + M_2\mathbf{v}_2 = M_{\text{total}}\left(\frac{M_1\mathbf{v}_1 + M_2\mathbf{v}_2}{M_{\text{total}}}\right) \equiv M_{\text{total}}\,\mathbf{V}_{cm} \tag{11.3.2}$$

where $\mathrm{M_{total}} = \mathrm{M_1 + M_2}$. The quantity $\mathbf{V}_{\text{cm}}$ is the velocity of the center of mass and can be expressed as the time derivative of the position of the center of mass, $\mathbf{R}_{\text{Cm}}$,

$$\mathbf{V}_{cm} = \frac{d\mathbf{R}_{cm}}{dt} \tag{11.3.3}$$

where

$$\mathbf{R}_{cm} = \frac{M_1\mathbf{r}_1 + M_2\mathbf{r}_2}{M_{\text{total}}} \tag{11.3.4}$$

We now see how the kinetic energy and the angular momentum of the two particles may be split into two parts, one having to do with the motion of the center of mass of the two particles, the other having to do with the motion of the two particles relative their center of mass. Figure 11.3.5: shows graphically how the vectors \(are defined. These vectors $\mathbf{r}_1' = \mathbf{r}_1 - \mathbf{R}_{\text{cm}}$ and $\mathbf{r}_2' = \mathbf{r}_2 - \mathbf{R}_{\text{cm}}$ represent the positions of the two particles relative to the center of mass. Substitution into equation (11.3.4) shows that $\mathrm{M_1}\mathbf{r}_1' + \mathrm{M_2}\mathbf{r}_2' = 0$. This leads to the conclusion that $\mathrm{M_1\ d_1 = M_2\ d_2}$ in Figure 11.3.5:. We also define the velocity of each mass relative to the center of mass as $\mathbf{v}_1' = \mathrm{d\mathbf{r}'_1/dt}$ and $\mathbf{v}_2' = \mathrm{d\mathbf{r}'_2/dt}$, and we therefore have $\mathrm{M_1}\mathbf{v}_1' + \mathrm{M_2}\mathbf{v}_2' = 0$.

The total kinetic energy is just the sum of the kinetic energies of the two particles, $\mathrm{K = M_1 v_1^2/2 + M_2 v_2^2/2}$, where $\mathrm{v_1}$ and $\mathrm{v_2}$ are the magnitudes of the corresponding velocity vectors. Substitution of $\mathbf{v}_1 = \mathbf{V}_{\text{cm}} + \mathbf{v}_1'$ etc., into the kinetic energy formula and rearranging yields

$$K_{\text{total}} = K_{\text{trans}} + K_{\text{intern}} = \left[M_{\text{total}}\,V_{cm}^2/2\right] + \left[M_1 v_1'^2/2 + M_2 v_2'^2/2\right]. \tag{11.3.5}$$

Terms like $\mathbf{V}_{cm}\cdot\mathbf{v}_1'$ cancel out because $M_1\mathbf{v}_1' + M_2\mathbf{v}_2' = 0$.

The first term on the right side of equation (11.3.5) in square brackets is the kinetic energy the two particles would have if all of the mass were concentrated at the center of mass. The second term is the kinetic energy computed relative to the motion of the center of mass. The first is called the *translational kinetic energy* of the system while the second is called the *internal kinetic energy*.

The angular momentum of the system is just the sum of the angular momenta of the two particles: $\mathbf{L}_{\text{total}} = \mathbf{M_1 r}_1 \times \mathbf{v}_1 + \mathbf{M_2 r}_2 \times \mathbf{v}_2$. By reasoning similar to the case of kinetic energy, we can rewrite this as

$$\mathbf{L}_{\text{total}} = \mathbf{L}_{\text{orb}} + \mathbf{L}_{\text{spin}} = [M_{\text{total}}\,\mathbf{R}_{cm}\times\mathbf{V}_{cm}] + [M_1\mathbf{r}_1'\times\mathbf{v}_1' + M_2\mathbf{r}_2'\times\mathbf{v}_2'] \tag{11.3.6}$$

The first term in square brackets on the right is called the *orbital angular momentum* while the second term is called the *spin angular momentum*. The former is the angular momentum the system would have if all the mass were concentrated at the center of

mass, while the latter is the angular momentum of motion about the center of mass.

Interestingly, the idea of center of mass and the corresponding split of kinetic energy and angular momentum into orbital and spin parts has no useful relativistic generalization. This is due to the factor of $\gamma \equiv \left(1 - v^2/c^2\right)^{-1/2}$ in the relativistic definition of momentum, $\mathbf{p} = m\mathbf{v}\gamma$, which means that

$$\frac{dm\mathbf{r}}{dt} \neq \mathbf{p} \quad \text{(relativistic case)}. \tag{11.3.7}$$

This page titled 11.3: Two Particles is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

11.4: The Uneven Dumbbell

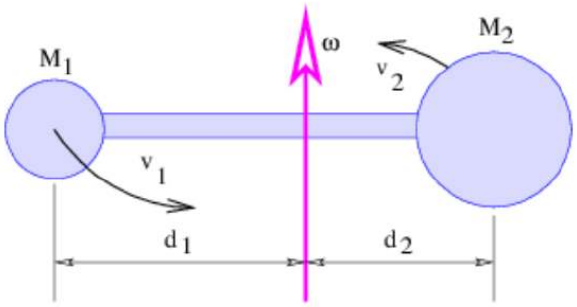


Figure 11.4.6:: Perspective view of the rotating dumbbell attached to an axle labeled ω. The axle attaches to the crossbar at the center of mass. The velocity v_1 is out of the page and v_2 is into the page.

So far we have put no restrictions on the movements of the two particles. An interesting special case occurs when the particles are connected by a lightweight, rigid rod, giving us a dumbbell. In order to further simplify things, we assume that the rod is connected rigidly to a fixed axle at the center of mass of the two particles, as shown in figure 11.6. The masses constituting the ends of the dumbbell are therefore free to revolve in circles about the axle, but they are prevented from executing any other motion. The key effect of this constraint is that both masses rotate about the axle with the same angular frequency ω.

If the particles are respectively distances d_1 and d_2 from the axle, then their speeds are $v_1 = d_1\omega$ and $v_2 = d_2\omega$. Thus the kinetic energy of the rotating dumbbell is

$$K_{\text{intern}} = \frac{1}{2}M_1v_1^2 + \frac{1}{2}M_2v_2^2 = \frac{1}{2}I\omega^2 \quad \text{(fixed axle)} \tag{11.4.1}$$

where $I = M_1d_1^2 + M_2d_2^2$ is called the *moment of inertia*. Similarly, the magnitude of the spin angular momentum, which is a vector parallel to the axle, is

$$L_{spin} = M_1d_1v_1 + M_2d_2v_2 = I\omega \quad \text{(fixed axle)} \tag{11.4.2}$$

Finally, Newton's second law for rotation becomes

$$\tau = \frac{dL_{spin}}{dt} = \frac{dI\omega}{dt} = I\frac{d\omega}{dt} \quad \text{(fixed axle)} \tag{11.4.3}$$

where τ is the component of torque along the rotation axis.

Note that the rightmost expression in equation (11.4.3) assumes that I is constant, which only is true if d_1 and d_2 are constant – i. e., the dumbbell must truly be rigid.

This page titled 11.4: The Uneven Dumbbell is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

11.5: Many Particles

The generalization from two particles to many particles is quite easy in principle. If a subscripted *i* indicates the value of a quantity for the *i*th particle, then the center of mass is given by

$$\mathbf{R}_{cm} = \frac{1}{M_{\text{total}}} \sum_i M_i \mathbf{r}_i \tag{11.5.1}$$

where

$$M_{\text{total}} = \sum_i M_i \tag{11.5.2}$$

Furthermore, if we define $\mathbf{r}'_{\text{i}} = \mathbf{r}_{\text{i}} - \mathbf{R}_{\text{cm}}$, etc., then the kinetic energy is just

$$K_{\text{total}} = M_{\text{total}}\, V_{\text{cm}}^2/2 + \sum_i M_i v_i'^2/2 \tag{11.5.3}$$

and the angular momentum is

$$\mathbf{L}_{\text{total}} = M_{\text{total}}\, \mathbf{R}_{cm} \times \mathbf{V}_{cm} + \sum_i M_i \mathbf{r}'_i \times \mathbf{v}'_i \tag{11.5.4}$$

In other words, both the kinetic energy and the angular momentum can be separated into two parts: one part is related to the overall motion of the system and the other is due to motions of system components relative to the center of mass, just as for the case of the dumbbell.

This page titled 11.5: Many Particles is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

11.6: New Page

For a rigid body rotating about a fixed axle, the moment of inertia is

$$I = \sum_i M_i d_i^2 \tag{11.6.1}$$

where $\mathrm{d_i}$ is the perpendicular distance of the ith particle from the axle. Equations (11.16)-(11.18) are valid for a rigid body consisting of many particles. Furthermore, the moment of inertia is constant in this case, so it can be taken out of the time derivative:

$$\tau = \frac{dI\omega}{dt} = I\frac{d\omega}{dt} = I\alpha \quad (\text{ fixed axle, constant } I) \tag{11.6.2}$$

The quantity $a = d\omega/dt$ is called the angular acceleration.

The sum in the equation for the moment of inertia can be converted to an integral for a continuous distribution of mass. We shall not pursue this here, but simply quote the results for a number of solid objects of uniform density:

- For rotation of a sphere of mass M and radius R about an axis piercing its center: $I = 2MR^2/5$.
- For rotation of a cylinder of mass M and radius R about its axis of symmetry: $I = MR^2/2$.
- For rotation of a thin rod of mass M and length L about an axis perpendicular to the rod passing through its center: $\mathrm{I = ML^2/12}$.
- For rotation of an annulus of mass M, inner radius $\mathrm{R_a}$, and outer radius $\mathrm{R_b}$ about its axis of symmetry: $I = M\left(R_a^2 + R_b^2\right)/2$.

This page titled 11.6: New Page is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

11.7: Statics

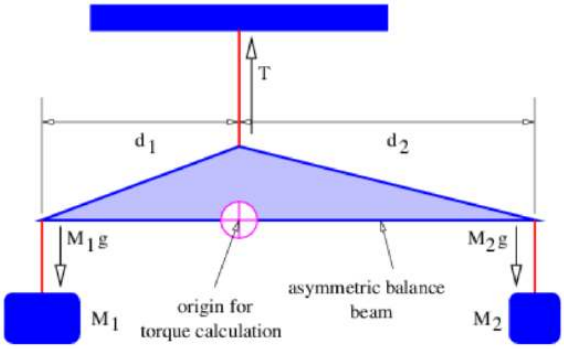


Figure 11.7.7:: Asymmetric mass balance. We assume that the balance beam is massless.

If a rigid body is initially at rest, it will remain at rest if and only if the sum of all the forces and the sum of all the torques acting on the body are zero. As an example, a mass balance with arms of differing length is shown in Figure 11.7.7:. The balance beam is subject to three forces pointing upward or downward, the tension T in the string from which the beam is suspended and the weights M_1 g and M_2 g exerted on the beam by the two suspended masses. The parameter g is the local gravitational field and the balance beam itself is assumed to have negligible mass. Taking upward as positive, the force condition for static equilibrium is

$$T - M_1 g - M_2 g = 0 \quad \text{(zero net force)} \tag{11.7.1}$$

Defining a counterclockwise torque to be positive, the torque balance computed about the pivot point in Figure 11.7.7: is

$$\tau = M_1 g d_1 - M_2 g d_2 = 0 \quad \text{(zero torque)} \tag{11.7.2}$$

where d_1 and d_2 are the lengths of the beam arms.

The first of the above equations shows that the tension in the string must be

$$T = (M_1 + M_2)\, g \tag{11.7.3}$$

while the second shows that

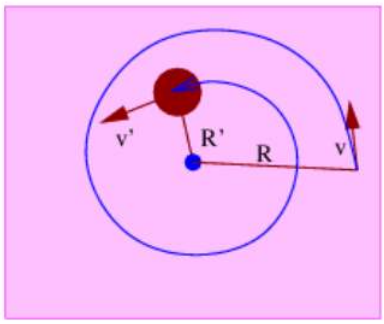
$$\frac{M_1}{M_2} = \frac{d_2}{d_1} \tag{11.7.4}$$

Thus, the tension in the string is just equal to the weight of the masses attached to the balance beam, while the ratio of the two masses equals the inverse ratio of the associated beam arm lengths.

This page titled 11.7: Statics is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

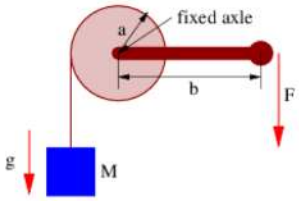
11.8: Problem

1. Show using the component form of the cross product given by equation (11.4) that $\mathbf{A} \times \mathbf{B} = -\mathbf{B} \times \mathbf{A}$.
2. A mass M is sliding on a frictionless table, but is attached to a string which passes through a hole in the center of the table as shown in Figure 11.8.8:. The string is gradually drawn in so the mass traces out a spiral pattern as shown in Figure 11.8.8:. The initial distance of the mass from the hole in the table is R and its initial tangential velocity is v. After the string is drawn in, the mass is a distance R′ from the hole and its tangential velocity is v′.
    1. Given R, v, and R′, find v′.
    2. Compute the change in the kinetic energy of the mass in going from radius R to radius R′.



    3. If the above change is non-zero, determine where the extra energy came from.

    Figure 11.8.8**::** Trajectory of a mass on a frictionless table attached to a string which passes through a hole in the table. The string is drawing the mass in.
3. A car of mass 1000 kg is heading north on a road at 30 m s^{-1} which passes 2 km east of the center of town.
    1. Compute the angular momentum of the car about the center of town when the car is directly east of the town.
    2. Compute the angular momentum of the car about the center of town when it is 3 km north of the above point.
4. The apparatus illustrated in Figure 11.8.9: is used to raise a bucket of mass M out of a well.
    1. What force F must be exerted to keep the bucket from falling back into the well?
    2. If the bucket is slowly raised a distance d, what work is done on the bucket by the rope attached to it?



    3. What work is done by the force F on the handle in the above case?

    Figure 11.8.9**::** A crank on a fixed axle turns a drum, thus winding the rope around the drum and raising the mass.
5. Derive equations below.
    1. $K_{\text{total}} = K_{\text{trans}} + K_{\text{intern}} = [M_{\text{total}}\, V_{cm}^2/2] + [M_1 v_1'^2/2 + M_2 v_2'^2/2]$.
    2. $\mathbf{L}_{\text{total}} = \mathbf{L}_{\text{orb}} + \mathbf{L}_{\text{spin}} = [M_{\text{total}}\, \mathbf{R}_{cm} \times \mathbf{V}_{cm}] + [M_1 \mathbf{r}_1' \times \mathbf{v}_1' + M_2 \mathbf{r}_2' \times \mathbf{v}_2']$
6. A mass M is held up by the structure shown in Figure 11.8.10:. The support beam has negligible mass. Find the tension T in the diagonal wire. Hint: Compute the net torque on the support beam about point A due to the tension T and the weight of the mass M.

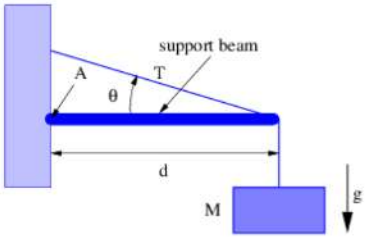


    Figure 11.8.10**::** A mass is supported by the tension in the diagonal wire. The support beam is free to pivot at point A.
7. A system consists of two stars, one of mass M moving with velocity $\mathbf{v}_1$ = (0,v, 0) at position $\mathbf{r}_1$ = (d, 0, 0), the other of mass 2M with zero velocity at the origin.

1. Find the center of mass position and velocity of the system of two stars.
2. Find the spin angular momentum of the system.

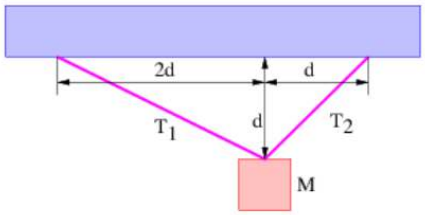


Figure 11.8.11**::** A mass is

3. Find the internal kinetic energy of the system.

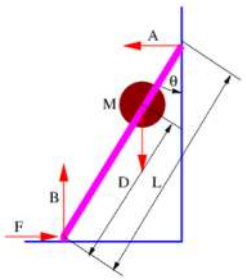


Figure 11.8.12**::** A ladder leaning against a wall is held in place the force F acting on the base of the ladder.

supported by two strings.

8. A solid disk is rolling down a ramp tilted an angle θ from the horizontal. Compute the acceleration of the disk down the ramp and compare it with the acceleration of a block sliding down the ramp without friction.
9. A mass M is suspended from the ceiling by two strings as shown in Figure 11.8.11:. Find the tensions in the strings.
10. A man of mass M is a distance D up a ladder of length L which makes an angle θ with respect to the vertical wall as shown in Figure 11.8.12:. Take the mass of the ladder to be negligible. Find the force F needed to keep the ladder from sliding if the wall and floor are frictionless and therefore can only exert normal forces A and B on the ladder.

This page titled 11.8: Problem is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

12: Harmonic Oscillator

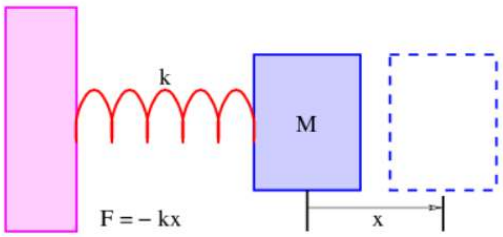


Figure 12.1:: Illustration of a mass-spring system.

Figure 12.1: illustrates the prototypical harmonic oscillator, the mass-spring system. A mass M is attached to one end of a spring. The other end of the spring is attached to something rigid such as a wall. The spring exerts a restoring force $F = -kx$ on the mass when it is stretched by an amount x, i. e., it acts to return the mass to its initial position. This is called Hooke's law and k is called the spring constant.

12.1: Energy Analysis
12.2: Analysis Using Newton's Laws
12.3: Forced Oscillator
12.4: Complex Exponential Solutions
12.5: Quantum Mechanical Harmonic Oscillator
12.6: Problems

This page titled 12: Harmonic Oscillator is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

12.1: Energy Analysis

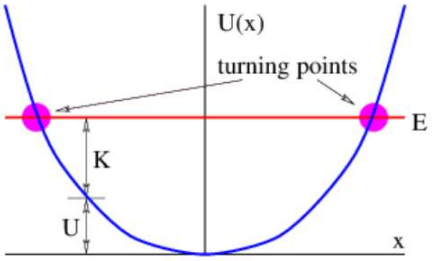


Figure 12.1.2:: Potential, kinetic, and total energy of a harmonic oscillator plotted as a function of spring displacement x.

The potential energy of the mass-spring system is

$$U(x) = kx^2/2 \tag{12.1.1}$$

which may be verified by noting that the Hooke's law force is derived from this potential energy: $F = -d\,(kx^2/2)\,dx = -kx$. This is shown in Figure 12.1.2:. Since a potential energy exists, the total energy $\mathrm{E} = \mathrm{K} + \mathrm{U}$ is conserved, i. e., is constant in time. If the total energy is known, this provides a useful tool for determining how the kinetic energy varies with the position x of the mass M : $K(x) = E - U(x)$. Since the kinetic energy is expressed (non-relativistically) in terms of the velocity u as $\mathrm{K} = \mathrm{Mu}^2/2$, the velocity at any point on the graph in Figure 12.1.2: is

$$u = \pm\left(\frac{2(E-U)}{M}\right)^{1/2} \tag{12.1.2}$$

Given all this, it is fairly evident how the mass moves. From Hooke's law, the mass is always accelerating toward the equilibrium position, x = 0. However, at any point the velocity can be either to the left or the right. At the points where $U(x) = E$, the kinetic energy is zero. This occurs at the turning points

$$x_{TP} = \pm\left(\frac{2E}{k}\right)^{1/2} \tag{12.1.3}$$

If the mass is moving to the left, it slows down as it approaches the left turning point. It stops when it reaches this point and begins to move to the right. It accelerates until it passes the equilibrium position and then begins to decelerate, stopping at the right turning point, accelerating toward the left, etc. The mass thus oscillates between the left and right turning points. (Note that equations (12.1.2) and (12.1.3) are only true for the harmonic oscillator.)

How does the period of the oscillation depend on the total energy of the system? Notice that from equation (12.1.2) the maximum speed of the mass (i. e., the speed at x = 0) is equal to $\mathrm{u_{max}} = (2\mathrm{E}M)^{1/2}$. The average speed must be some fraction of this maximum value. Let us guess here that it is half the maximum speed:

$$u_{\text{average}} \approx \frac{u_{\max}}{2} = \left(\frac{E}{2M}\right)^{1/2} \quad (\text{ approximate }). \tag{12.1.4}$$

However, the distance d the mass has to travel for one full oscillation is twice the distance between turning points, or $\mathrm{d} = 4(2\mathrm{E/k})^{1/2}$. Therefore, the period of oscillation must be approximately

$$T = \frac{d}{u_{\text{average}}} \approx 4\left(\frac{2E}{k}\right)^{1/2}\left(\frac{2M}{E}\right)^{1/2} = 8\left(\frac{M}{k}\right)^{1/2} \quad (\text{approximate}) \tag{12.1.5}$$

This page titled 12.1: Energy Analysis is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

12.2: Analysis Using Newton's Laws

The acceleration of the mass at any time is given by Newton's second law:

$$a = \frac{d^2x}{dt^2} = \frac{F}{M} = -\frac{kx}{M} \tag{12.2.1}$$

An equation of this type is known as a differential equation since it involves a derivative of the dependent variable x. Equations of this type are generally more difficult to solve than algebraic equations, as there are no universal techniques for solving all forms of such equations. In fact, it is fair to say that the solutions of most differential equations were originally obtained by guessing!

We already have the basis on which to make an intelligent guess for the solution to equation (12.2.1) since we know that the mass oscillates back and forth with a period that is independent of the amplitude of the oscillation. A function which might fill the bill is the cosine function. Let us try substituting $x = \cos(\omega t)$, where ω is a constant, into this equation. The second derivative of x with respect to t is $-\omega^2 \cos(\omega t)$, so performing this substitution results in

$$-\omega^2 \cos(\omega t) = -\frac{k}{M}\cos(\omega t) \tag{12.2.2}$$

Notice that the cosine function cancels out, leaving us with $-\omega^2 = -k/M$. The guess thus works if we set

$$\omega = \left(\frac{k}{M}\right)^{1/2} \tag{12.2.3}$$

The constant ω is the angular oscillation frequency for the oscillator, from which we infer the period of oscillation to be $\mathrm{T} = 2\Pi(\mathrm{Mk})^{1/2}$. This agrees with the earlier approximate result of equation 12.5, except that the approximation has a numerical factor of 8 rather than $2\Pi \approx 6$. Thus, the earlier guess is only off by about 30%!

It is easy to show that $x = B\cos(\omega t)$ is also a solution of equation (12.2.1), where B is any constant and $\omega = (\mathrm{k/M})^{1/2}$. This confirms that the oscillation frequency and period are independent of amplitude. Furthermore, the sine function is equally valid as a solution: $x = A\sin(\omega t)$, where A is another constant. In fact, the most general possible solution is just a combination of these two, i. e.,

$$x = A\sin(\omega t) + B\cos(\omega t) = C\cos(\omega t - \phi) \tag{12.2.4}$$

The values of A and B depend on the position and velocity of the mass at time t = 0. The right side of equation (12.2.4) shows an alternate way of writing the general harmonic oscillator solution that uses a cosine function with a phase factor ϕ.

This page titled 12.2: Analysis Using Newton's Laws is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

12.3: Forced Oscillator

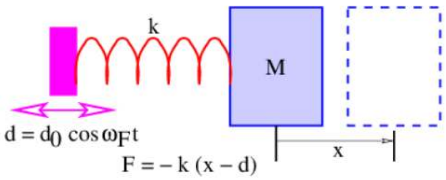


Figure 12.3.3:: Illustration of a forced mass-spring oscillator. The left end of the spring is wiggled back and forth with an angular frequency ω_F and a maximum amplitude d_0.

If we wiggle the left end of the spring by the amount $d = d_0 \cos(\omega_F t)$, as in Figure 12.3.3:, rather than rigidly fixing it as in Figure 12.3.1:, we have a forced harmonic oscillator. The constant d_0 is the amplitude of the imposed wiggling motion. The forcing frequency ω_F is not necessarily equal to the natural or resonant frequency $\omega = (k/M)^{1/2}$ of the mass-spring system. Very different behavior occurs depending on whether ω_F is less than, equal to, or greater than ω.

Given the above wiggling, the force of the spring on the mass becomes $F = -k(x-d) = -k[x - d_0 \cos(\omega_F t)]$ since the length of the spring is the difference between the positions of the left and right ends. Proceeding as for the unforced mass-spring system, we arrive at the differential equation

$$\frac{d^2x}{dt^2} + \frac{kx}{M} = \frac{kd_0}{M}\cos(\omega_F t) \tag{12.3.1}$$

The solution to this equation turns out to be the sum of a forced part in which x is proportional to $\cos(\omega_F t)$ and a free part which is the same as the solution to the unforced equation (12.2.4). We are primarily interested in the forced part of the solution, so let us set $x = x_0 \cos(\omega_F t)$ and substitute this into equation (12.3.1):

$$-\omega_F^2 x_0 \cos(\omega_F t) + \frac{kx_0}{M}\cos(\omega_F t) = \frac{kd_0}{M}\cos(\omega_F t) \tag{12.3.2}$$

Again the cosine factor cancels and we are left with an algebraic equation for x_0, the amplitude of the oscillatory motion of the mass.

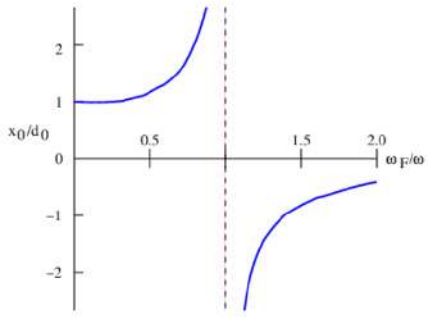


Figure 12.3.4:: Plot of the ratio of response to forcing vs. the ratio of forced to free oscillator frequency for the mass-spring system.

Solving for the ratio of the oscillation amplitude of the mass to the amplitude of the wiggling motion, x_0/d_0, we find

$$\frac{x_0}{d_0} = \frac{1}{1 - \omega_F^2/\omega^2} \tag{12.3.3}$$

where we have recognized that $k/M = \omega^2$, the square of the frequency of the free oscillation. This function is plotted in Figure 12.3.4:.

Notice that if $\omega_F < \omega_1$, the motion of the mass is in phase with the wiggling motion and the amplitude of the mass oscillation is greater than the amplitude of the wiggling. As the forcing frequency approaches the natural frequency of the oscillator, the response of the mass grows in amplitude. When the forcing is at the resonant frequency, the response is technically infinite, though practical limits on the amplitude of the oscillation will intervene in this case — for instance, the spring cannot stretch or shrink an infinite amount. In many cases friction will act to limit the response of the mass to forcing near the resonant frequency. When the forcing frequency is greater than the natural frequency, the mass actually moves in the opposite direction of the wiggling motion — i. e.,

the response is out of phase with the forcing. The amplitude of the response decreases as the forcing frequency increases above the resonant frequency.

Forced and free harmonic oscillators form an important part of many physical systems. For instance, any elastic material body such as a bridge or an airplane wing has harmonic oscillatory modes. A common engineering problem is to ensure that such modes are damped by friction or some other physical mechanism when there is a possibility of exitation of these modes by naturally occurring processes. A number of disasters can be traced to a failure to properly account for oscillatory forcing in engineered structures.

This page titled 12.3: Forced Oscillator is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

12.4: Complex Exponential Solutions

Complex exponential functions of the form $x = \exp(\pm i\omega t)$ also constitute solutions to the free harmonic oscillator governed by equation (12.2.1). This makes sense, as the complex exponential is the sum of sines and cosines. However, for the frictionless harmonic oscillator, the exponential solutions provide no particular advantage over sines and cosines. Furthermore, oscillator displacements are real, not complex quantities.

The superposition principle solves the problem of complex versus real solutions. For an equation like (12.2.1) which has real coefficients, if $\exp(i\omega t)$ is a solution, then so is $\exp(-i\omega t)$, so the superposition of these two solutions is also a solution. Furthermore

$$\exp(i\omega t) + \exp(-i\omega t) = 2\cos(\omega t) = 2\,\mathrm{Re}[\exp(i\omega t)] \tag{12.4.1}$$

This shows a shortcut for getting the physical part of a complex exponential solution to equations like the harmonic oscillator equation; simply take the real part.

Complex exponential solutions come into their own for more complicated equations. For instance, suppose the force on the mass in the mass-spring system takes the form

$$F = -kx - b\frac{dx}{dt} \tag{12.4.2}$$

The term containing b represents a frictional damping effect on the harmonic oscillator and the governing differential equation becomes

$$\frac{d^2x}{dt^2} + \frac{b}{M}\frac{dx}{dt} + \frac{k}{M}x = 0 \tag{12.4.3}$$

Trying the exponential function $\exp(\sigma t)$ in this equation results in

$$\sigma = \frac{1}{2}\left[-\frac{b}{M} \pm \left(\frac{b^2}{M^2} - \frac{4k}{M}\right)^{1/2}\right] = -\beta \pm i\left(\omega_0^2 - \beta^2\right)^{1/2} \tag{12.4.4}$$

where we have set

$$\beta = \frac{b}{2M} \quad \omega_0 = \left(\frac{k}{M}\right)^{1/2} \tag{12.4.5}$$

The quantity $\omega \equiv (\omega_0^2 - \beta^2)^{1/2}$ is the actual frequency of oscillation of the damped oscillator, which one can see is less than the oscillation frequency ω_0 that occurs with the damping turned off. The physical solution to the damped oscillator is thus

$$x(t) = \mathrm{Re}[\exp(\sigma t)] = \mathrm{Re}[\exp(i\omega t)\exp(-\beta t)] = \cos(\omega t)\exp(-\beta t) \tag{12.4.6}$$

as long as $\omega_0^2 > \beta^2$. Notice that this solution is in the form of an oscillation $\cos(\omega t)$ multiplied by a decaying exponential $\exp(-\beta t)$. This confirms that the b term decreases the amplitude of the oscillation with time.

This page titled 12.4: Complex Exponential Solutions is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

12.5: Quantum Mechanical Harmonic Oscillator

The quantum mechanical harmonic oscillator shares the characteristic of other quantum mechanical bound state problems in that the total energy can take on only discrete values. Calculation of these values is too difficult for this book, but the problem is sufficiently important to warrant reporting the results here. The energies accessible to a quantum mechanical mass-spring system are given by the formula

$$E_n = (n + 1/2)\hbar(k/M)^{1/2}, \quad n = 0, 1, 2, ... \tag{12.5.1}$$

In other words, the energy difference between successive quantum mechanical energy levels in this case is constant and equals the classical resonant frequency for the oscillator, $\omega = (\mathrm{kM})^{1/2},\ \text{times}\ \hbar$.

This page titled 12.5: Quantum Mechanical Harmonic Oscillator is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

12.6: Problems

1. An oscillator (non-harmonic) has the potential energy function $\mathrm{U(x)} = \mathrm{Cx}^4$, where C is a constant. How does the oscillation frequency depend on energy? Explain your reasoning.
2. Show that $\mathrm{C}\cos(\omega \mathrm{t} - \phi)$ is an alternate way of writing $A\sin(\omega t) + B\cos(\omega t)$ by finding the values of A and B in terms of the constants C and ϕ. Hint: Expand $\cos(\omega t - \phi)$ by using the trigonometric identity for the cosine of the sum of two angles.
3. If a mass-spring harmonic oscillator has displacement $x = 0$ and velocity $\mathrm{dx/dt} = \mathrm{V}$ at time $\mathrm{t} = 0$, determine the values of A and B as well as those of C and ϕ in equation (12.2.4).
4. A mass M is suspended against gravity by a spring of spring constant k. The unstretched length of the spring is x_0 and under the influence of gravity the spring is stretched to a resting length $x_1 > x_0$.
    1. Compute the length of the spring x_1 in the steady, resting case.
    2. Set up the equation of motion for the mass moving under the influence of the two forces, gravity and spring. Solve the equation for the frequency of the oscillation and the position of the spring as a function of time x(t). Does the oscillation frequency change from the case without gravity?
5. Determine the two real solutions to the damped harmonic oscillator problem in the case in which $\omega_0^2 < \beta^2$.

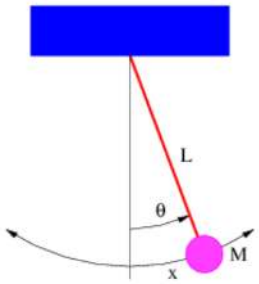


Figure 12.6.5**::** The pendulum as a harmonic oscillator.

6. Consider the pendulum in Figure 12.6.5:. The mass M moves along an arc with x denoting the distance along the arc from the equilibrium point.
    1. Find the component of the gravitational force tangent to the arc (and thus in the direction of motion of the mass) as a function of the angle $theta$. Use the small angle approximation on $\sin(\theta)$ to simplify this answer.
    2. Get the force in terms of x rather than θ. (Recall that $\theta = x/L$.)
    3. Use Newton's second law for motion in the x direction (i. e., along the arc followed by the mass) to get the equation of motion for the mass.
    4. Solve the equation of motion using the solution to the mass-spring problem as a guide.
7. Forced damped oscillator:
    1. Add a damping term to the forced harmonic oscillator equation (12.3.1) and solve for the forced part of the solution using complex exponential methods. Hint: Change the cosine on the right side of this equation to $\exp(i\omega_F t)$ to convert the equation to complex form and then try the solution $x = x_0 \exp(i\omega_F t)$ where x_0 will depend of ω_F . Also, write the equation in terms of $\beta = b/(2M)$ and $\omega^2 = k/M$.
    2. Find the physical part of this solution by taking the real part of x(t). Hint: While taking the real part of x, it may be helpful to recall that the inverse of any complex number can be written $1/(a+ib) = (a-ib)/(a^2+b^2)$.
    3. Determine how x_0 differs from that in the undamped case when ω_F is near the resonant frequency of the unforced oscillator. In particular, show how the phase and amplitude of the oscillation change as the forcing frequency changes from less than to greater than the resonant frequency.
8. A massless particle is confined to a box of length a. (Think of a photon between two mirrors.) Treating the particle classically, compute the period of one round trip from one end of the box to the other and back again. From this compute an angular frequency for the oscillation of this particle in the box. Does this frequency depend on the particle's energy?
9. Compute the ground state energy $\mathrm{E}_{\mathrm{ground}}$ of a massless particle in a box of length a using quantum mechanics. Compare $\mathrm{E}_{\mathrm{ground}}/\hbar$ with the angular frequency computed in the previous problem.

This page titled 12.6: Problems is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by David J. Raymond (The New Mexico Tech Press) via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

13: Appendix A, Constants, Units, and Conversions

13.1: Appendix A Constants, Units, and Conversions

13: Appendix A, Constants, Units, and Conversions is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

13.1: Appendix A Constants, Units, and Conversions

Appendix A Constants, Units, and Conversions

This appendix contains various useful constants and conversion factors as well as information on the International System of Units.

Units

"SI" is the French abbreviation for the International System of Units, the system used universally in science. See http://physics.nist.gov/cuu/Units/ for the last word on this subject. This treatment is derived from the National Institute of Science and Technology (NIST) website.

The most fundamental units of measure are length (meters; m), mass (kilograms; kg), time (seconds; s), electric current (ampere; A), temperature (kelvin; K), amount of a substance (mole; mol), and the luminous intensity (candela; cd). The candela is a rather specialized unit related to the perceived brightness of a light source by a "standard" human eye. As such, it is rather anthropocentric and hardly seems to merit the designation "fundamental". The mole is also less fundamental than the other units, as it is simply a convenient way to refer to a multiple of Avogadro's number of atoms or molecules.

Fundamental units can be combined to form derived units with special names. Some of these derived units are listed below.

Fundamental and derived SI units can have multipliers expressed as prefixes, e. g., 1 km = 1000 m. The NIST website points out a minor irregularity with the fundamental unit of mass, the kilogram. This already has the multiplier "kilo" prefixed to the unit "gram". In this case 1000 kg is written 1 Mg, not 1 kkg, etc. SI multipliers are listed below as well.

Derived Units

Name	Abbrev.	Units	Meaning
hertz	Hz	s^{-1}	frequency (cycles/sec)
(unnamed)		s^{-1}	angular frequency (radians/sec)
newton	N	kg m s^{-2}	force
pascal	Pa	N m^{-2}	pressure
joule	J	N m	energy
watt	W	J s^{-1}	power
coulomb	C	A s	electric charge
volt	V	N m C^{-1}	scalar potential
(unnamed)		N s C^{-1}	vector potential
(unnamed)		V m^{-1}	electric field
tesla	T	N s C^{-1} m^{-1}	magnetic field
(unnamed)		V m	electric flux
weber	Wb	T m^2	magnetic flux
volt	V	V	electric circulation (EMF)
(unnamed)		T m	magnetic circulation
farad	F	C V^{-1}	capacitance
ohm	Ω	V A^{-1}	resistance
henry	H	V s^2 C^{-1}	inductance

Multipliers

Multiplier	Name	Prefix
10^{24}	yotta	Y
10^{21}	zetta	Z
10^{18}	exa	E
10^{15}	peta	P
10^{12}	tera	T
10^{9}	giga	G
10^{6}	mega	M
10^{3}	kilo	k
10^{2}	hecto	h
10^{1}	deka	da
10^{-1}	deci	d
10^{-2}	centi	c
10^{-3}	milli	m
10^{-6}	micro	μ
10^{-9}	nano	n
10^{-12}	pico	p
10^{-15}	femto	f
10^{-18}	atto	a
10^{-21}	zepto	z
10^{-24}	yocto	y

Centimeter-Gram-Second Units

An older system of scientific units is the CGS system. This system is still used widely in certain areas of physics. The fundamental units of length, mass, and time are as implied by the title given above. The most common CGS derived units are those for force (1 dyne = 10^{-5} N) and energy (1 erg = 10^{-7} J).

Electromagnetism is expressed in several different ways in CGS units. Electromagnetic quantities in CGS not only have different units than in SI, they also have different physical dimensions, with different versions differing among themselves. The most common variant of CGS electromagnetic units is called “Gaussian” units. This variant is advocated by some physicists, though many others consider the whole subject of CGS electromagnetic units to be a terrible mess! SI units for electromagnetism are used in this text and CGS units will not be discussed further here.

Miscellaneous Conversions

1 lb = 4. 448 N
1 ft = 0. 3048 m
1 mph = 0. 4470 m s^{-1}
1 eV = 1. 60 × 10^{-19} J

1 mol = 6. 022 × 10^{23} molecules
1 gauss = 10^{-4} T (CGS unit of magnetic field)
1 millibar = 1 mb = 100 Pa (Old unit of pressure)

Advice on Calculations

Substituting Numbers

When faced with solving an algebraic equation to obtain a numerical answer, solve the equation symbolically first and then substitute numbers. For example, given the equation

$$ax^2 - b = 0 \tag{13.1.1}$$ (A.1)

where a = 2 and b = 8, first solve for x,

$$1 / 2x = \pm(b / a), \tag{13.1.2}$$ (A.2)

and then substitute the numerical values:

$$1 / 21 / 2x = \pm(8 / 2) = \pm 4 = \pm 2. \tag{13.1.3}$$ (A.3)

This procedure is far better than substituting numbers first,

$$2x^2 - 8 = 0 \tag{13.1.4}$$ (A.4)

and then solving for x. Solving first and then substituting has two advantages: (1) It is easier to make algebraic manipulations with symbols than it is with numbers. (2) If you decide later that numerical values should be different, then the entire solution procedure doesn't have to be repeated, only the substitutions at the end.

Significant Digits

In numerical calculations, keep only one additional digit beyond those present in the least accurate input number. For instance, if you are taking the square root of 3. 4, your calculator might tell you that the answer is 1. 843908891. The answer you write down should be 1. 84. Keeping all ten digits of the calculator's answer gives a false sense of the accuracy of the result.

Round the result up if the digit following the last significant digit is 5 or greater and round it down if it is less than 5. Thus, the square root of 4. 1, which the calculator tells us is 2. 049390153, should be represented as 2. 05 rather than 2. 04.

Changing Units

It is easy to make mistakes when changing the units of a quantity. Adopting a systematic approach to changing units greatly reduces the chance of error. We illustrate a systematic approach to this problem with an example in which we change the units of acceleration from meters per second squared to kilometers per minute squared:

$$\begin{aligned} 5\ \mathrm{m/s^2} &\rightarrow 5\ \mathrm{m/s^2} \times (0.001\ \mathrm{km/m}) \times (60\ \mathrm{s/min})^2 \\ &= 5 \times 0.001 \times 60^2\ \mathrm{km/min^2} \\ &= 18\ \mathrm{km/min^2} \end{aligned} \tag{13.1.5}$$

The trick is to multiply by the conversion factor for each unit to the power that makes the original unit cancel out. The conversion factors to the proper powers are then multiplied by the original number and the proper cancellations of the old units are double checked. If done with care, this yields the correct result every time!

Constants of Nature

Symbol	Value	Meaning
h	6.63×10^{-34} J s	Planck's constant
$\hbar$	1.06×10^{-34} J s	$h/(2\pi)$
c	2.998×10^{8} m s^{-1}	speed of light
G	6.67×10^{-11} m^{3} s^{-2} kg^{-1}	universal gravitational constant
k_B	1.38×10^{-23} J K^{-1}	Boltzmann's constant
σ	5.67×10^{-8} W m^{-2} K^{-4}	Stefan-Boltzmann constant
K	3.67×10^{11} s^{-1} K^{-1}	thermal frequency constant
ϵ_0	8.85×10^{-12} C^{2} N^{-1} m^{-2}	permittivity of free space
μ_0	$4\pi \times 10^{-7}$ N s^{2} C^{-2}	permeability of free space (= $1/(\epsilon_0 c^2)$).

Properties of Stable Particles

Symbol	Value	Meaning
e	1.60×10^{-19} C	fundamental unit of charge
m_e	9.11×10^{-31} kg = 0.511 MeV	mass of electron
m_p	1.672648×10^{-27} kg = 938.280 MeV	mass of proton
m_n	1.674954×10^{-27} kg = 939.573 MeV	mass of neutron

Properties of Solar System Objects

Symbol	Value	Meaning
M_e	5.98×10^{24} kg	mass of earth
M_m	7.36×10^{22} kg	mass of moon
M_s	1.99×10^{30} kg	mass of sun
R_e	6.37×10^{6} m	radius of earth
R_m	1.74×10^{6} m	radius of moon
R_s	6.96×10^{8} m	radius of sun
D_m	3.82×10^{8} m	earth-moon distance
D_s	1.50×10^{11} m	earth-sun distance
g	9.81 m s^{-2}	earth's surface gravity

13.1: Appendix A Constants, Units, and Conversions is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

Index

A

accelerated reference frames
6.5: Accelerated Reference Frames

B

beats
1.5: Beats

Bragg's Law
7.1: Bragg’s Law

C

canonical momentum
8.7: Kinetic and Total Momentum

centripetal acceleration
6.2: Circular Motion

complex exponentials
12.4: Complex Exponential Solutions

complex waves
9.1: Math Tutorial — Complex Waves

D

de Broglie wavelength
7.5: Mass, Momentum, and Energy

derivatives
1.8: Math Review — Derivatives

diffraction fringes
2.4: Diffraction Through a Single Slit

diffraction grating
2.6: Diffraction Gratings

dispersion relation
1.9: Group Velocity
2.2: Plane Waves

E

elastic collisions
10.4: Collisions

Euler’s equation
9.1: Math Tutorial — Complex Waves

F

Fermat’s Principle of Least Action
3.5: Fermat’s Principle

forced oscillator
12.3: Forced Oscillator

G

gravity waves
2.2: Plane Waves

group velocity
1.9: Group Velocity

I

inelastic collisions
10.4: Collisions

K

kinetic momentum
8.7: Kinetic and Total Momentum

L

line of simultaneity
4.1: Galilean Spacetime Thinking

Lorentz contraction
4.5: Lorentz Contraction

M

Michelson interferometer
1.6: Interferometers

N

Newton’s third law
10.2: Newton’s Third Law

O

orbital angular momentum
9.3: Confined Matter Waves

P

partial derivatives
8.5: Math Tutorial – Partial Derivatives

particle in box
9.3: Confined Matter Waves

R

reflection
3.1: Reflection and Refraction

refraction
3.1: Reflection and Refraction

relativistic Doppler effect
5.5: The Doppler Shift

rocket propulsion
10.5: Rockets and Conveyor Belts

S

simultaneity
4.3: Postulates of Special Relativity

Snell’s law of refraction
3.1: Reflection and Refraction

spacetime
4.1: Galilean Spacetime Thinking
4.2: Spacetime Thinking in Special Relativity

spin angular momentum
9.3: Confined Matter Waves

superposition
2.3: Superposition of Plane Waves

T

thin film interference
1.7: Thin Films

time dilation
4.4: Time Dilation

total momentum
8.7: Kinetic and Total Momentum

twin paradox
4.6: Twin Paradox

V

vector
2.1: Math Tutorial — Vectors

vector addition
2.1: Math Tutorial — Vectors

W

wave vector
2.2: Plane Waves

world line
4.1: Galilean Spacetime Thinking
4.2: Spacetime Thinking in Special Relativity

Glossary

absolute pressure | sum of gauge pressure and atmospheric pressure [OpenStax]

acceleration due to gravity | acceleration of an object as a result of gravity [OpenStax]

acceleration vector | instantaneous acceleration found by taking the derivative of the velocity function with respect to time in unit vector notation [OpenStax]

accuracy | the degree to which a measured value agrees with an accepted reference value for that measurement [OpenStax]

Achimedes' principle | buoyant force on an object equals the weight of the fluid it displaces [OpenStax]

action-at-a-distance force | type of force exerted without physical contact [OpenStax]

amplitude (A) | maximum displacement from the equilibrium position of an object oscillating around the equilibrium position [OpenStax]

angular acceleration | time rate of change of angular velocity [OpenStax]

angular frequency | ω, rate of change of an angle with which an object that is moving on a circular path [OpenStax]

angular momentum | rotational analog of linear momentum, found by taking the product of moment of inertia and angular velocity [OpenStax]

angular position | angle a body has rotated through in a fixed coordinate system [OpenStax]

angular velocity | time rate of change of angular position [OpenStax]

anticommutative property | change in the order of operation introduces the minus sign [OpenStax]

antinode | location of maximum amplitude in standing waves [OpenStax]

antiparallel vectors | two vectors with directions that differ by 180° [OpenStax]

aphelion | farthest point from the Sun of an orbiting body; the corresponding term for the Moon's farthest point from Earth is the apogee [OpenStax]

apparent weight | reading of the weight of an object on a scale that does not account for acceleration [OpenStax]

associative | terms can be grouped in any fashion [OpenStax]

average acceleration | the rate of change in velocity; the change in velocity over time [OpenStax]

average power | work done in a time interval divided by the time interval [OpenStax]

average speed | the total distance traveled divided by elapsed time [OpenStax]

average velocity | the displacement divided by the time over which displacement occurs [OpenStax]

banked curve | curve in a road that is sloping in a manner that helps a vehicle negotiate the curve [OpenStax]

base quantity | physical quantity chosen by convention and practical considerations such that all other physical quantities can be expressed as algebraic combinations of them [OpenStax]

base unit | standard for expressing the measurement of a base quantity within a particular system of units; defined by a particular procedure used to measure the corresponding base quantity [OpenStax]

beat frequency | frequency of beats produced by sound waves that differ in frequency [OpenStax]

beats | constructive and destructive interference of two or more frequencies of sound [OpenStax]

Bernoulli's equation | equation resulting from applying conservation of energy to an incompressible frictionless fluid:

$$p + \frac{1}{2}\rho v^2 + \rho g h = constant, \quad (1)$$

throughout the fluid [OpenStax]

Bernoulli's principle | Bernoulli's equation applied at constant depth:

$$p_1 + \frac{1}{2}\rho v_1^2 = p_2 + \frac{1}{2}\rho v_2^2 \quad (2)$$

[OpenStax]

black hole | mass that becomes so dense, that it collapses in on itself, creating a singularity at the center surround by an event horizon [OpenStax]

bow wake | v-shaped disturbance created when the wave source moves faster than the wave propagation speed [OpenStax]

breaking stress (ultimate stress) | value of stress at the fracture point [OpenStax]

bulk modulus | elastic modulus for the bulk stress [OpenStax]

bulk strain (or volume strain) | strain under the bulk stress, given as fractional change in volume [OpenStax]

bulk stress (or volume stress) | stress caused by compressive forces, in all directions [OpenStax]

buoyant force | net upward force on any object in any fluid due to the pressure difference at different depths [OpenStax]

center of gravity | point where the weight vector is attached [OpenStax]

center of mass | weighted average position of the mass [OpenStax]

centripetal acceleration | component of acceleration of an object moving in a circle that is directed radially inward toward the center of the circle [OpenStax]

centripetal force | any net force causing uniform circular motion [OpenStax]

closed system | system for which the mass is constant and the net external force on the system is zero [OpenStax]

commutative | operations can be performed in any order [OpenStax]

component form of a vector | a vector written as the vector sum of its components in terms of unit vectors [OpenStax]

compressibility | reciprocal of the bulk modulus [OpenStax]

compressive strain | strain that occurs when forces are contracting an object, causing its shortening [OpenStax]

compressive stress | stress caused by compressive forces, only in one direction [OpenStax]

conservative force | force that does work independent of path [OpenStax]

conserved quantity | one that cannot be created or destroyed, but may be transformed between different forms of itself [OpenStax]

constructive interference | when two waves arrive at the same point exactly in phase; that is, the crests of the two waves are precisely aligned, as are the troughs [OpenStax]

conversion factor | a ratio that expresses how many of one unit are equal to another unit [OpenStax]

Coriolis force | inertial force causing the apparent deflection of moving objects when viewed in a rotating frame of reference [OpenStax]

corkscrew right-hand rule | a rule used to determine the direction of the vector product [OpenStax]

critically damped | condition in which the damping of an oscillator causes it to return as quickly as possible to its equilibrium position without oscillating back and forth about this position [OpenStax]

cross product | the result of the vector multiplication of vectors is a vector called a cross product; also called a vector product [OpenStax]

density | mass per unit volume of a substance or object [OpenStax]

derived quantity | physical quantity defined using algebraic combinations of base quantities [OpenStax]

derived units | units that can be calculated using algebraic combinations of the fundamental units [OpenStax]

destructive interference | when two identical waves arrive at the same point exactly out of phase; that is, precisely aligned crest to trough [OpenStax]

difference of two vectors | vector sum of the first vector with the vector antiparallel to the second [OpenStax]

dimension | expression of the dependence of a physical quantity on the base quantities as a product of powers of symbols representing the base quantities; in general, the dimension of a quantity has the form $L^a M^b T^c I^d \Theta^e N^f J^g$ for some powers a, b, c, d, e, f, and g [OpenStax]

dimensionally consistent | equation in which every term has the same dimensions and the arguments of any mathematical functions appearing in the equation are dimensionless [OpenStax]

dimensionless | quantity with a dimension of $L^0 M^0 T^0 I^0 \Theta^e N^0 J^0 = 1$; also called quantity of dimension 1 or a pure number [OpenStax]

direction angle | in a plane, an angle between the positive direction of the x-axis and the vector, measured counterclockwise from the axis to the vector [OpenStax]

discrepancy | the difference between the measured value and a given standard or expected value [OpenStax]

displacement | the change in position of an object [OpenStax]

displacement | change in position [OpenStax]

displacement vector | vector from the initial position to a final position on a trajectory of a particle [OpenStax]

distance traveled | the total length of the path traveled between two positions [OpenStax]

distributive | multiplication can be distributed over terms in summation [OpenStax]

Doppler effect | alteration in the observed frequency of a sound due to motion of either the source or the observer [OpenStax]

Doppler shift | actual change in frequency due to relative motion of source and observer [OpenStax]

dot product | the result of the scalar multiplication of two vectors is a scalar called a dot product; also called a scalar product [OpenStax]

drag force | force that always opposes the motion of an object in a fluid; unlike simple friction, the drag force is proportional to some function of the velocity of the object in that fluid [OpenStax]

dynamics | study of how forces affect the motion of objects and systems [OpenStax]

elapsed time | the difference between the ending time and the beginning time [OpenStax]

elastic | object that comes back to its original size and shape when the load is no longer present [OpenStax]

elastic | collision that conserves kinetic energy [OpenStax]

elastic limit | stress value beyond which material no longer behaves elastically and becomes permanently deformed [OpenStax]

elastic modulus | proportionality constant in linear relation between stress and strain, in SI pascals [OpenStax]

elastic potential energy | potential energy stored as a result of deformation of an elastic object, such as the stretching of a spring [OpenStax]

energy conservation | total energy of an isolated system is constant [OpenStax]

English units | system of measurement used in the United States; includes units of measure such as feet, gallons, and pounds [OpenStax]

equal vectors | two vectors are equal if and only if all their corresponding components are equal; alternately, two parallel vectors of equal magnitudes [OpenStax]

equilibrium | body is in equilibrium when its linear and angular accelerations are both zero relative to an inertial frame of reference [OpenStax]

equilibrium point | position where the assumed conservative, net force on a particle, given by the slope of its potential energy curve, is zero [OpenStax]

equilibrium position | position where the spring is neither stretched nor compressed [OpenStax]

escape velocity | initial velocity an object needs to escape the gravitational pull of another; it is more accurately defined as the velocity of an object with zero total mechanical energy [OpenStax]

estimation | using prior experience and sound physical reasoning to arrive at a rough idea of a quantity's value; sometimes called an "order-of-magnitude approximation," a "guesstimate," a "back-of-the-envelope calculation", or a "Fermi calculation" [OpenStax]

event horizon | location of the Schwarzschild radius and is the location near a black hole from within which no object, even light, can escape [OpenStax]

exact differential | is the total differential of a function and requires the use of partial derivatives if the function involves more than one dimension [OpenStax]

explosion | single object breaks up into multiple objects; kinetic energy is not conserved in explosions [OpenStax]

external force | force applied to an extended object that changes the momentum of the extended object as a whole [OpenStax]

external force | force acting on an object or system that originates outside of the object or system [OpenStax]

first equilibrium condition | expresses translational equilibrium; all external forces acting on the body balance out and their vector sum is zero [OpenStax]

fixed boundary condition | when the medium at a boundary is fixed in place so it cannot move [OpenStax]

flow rate | abbreviated Q, it is the volume V that flows past a particular point during a time t, or $Q = \frac{dV}{dt}$ [OpenStax]

fluids | liquids and gases; a fluid is a state of matter that yields to shearing forces [OpenStax]

force | push or pull on an object with a specific magnitude and direction; can be represented by vectors or expressed as a multiple of a standard force [OpenStax]

force constant (k) | characteristic of a spring which is defined as the ratio of the force applied to the spring to the displacement caused by the force [OpenStax]

free boundary condition | exists when the medium at the boundary is free to move [OpenStax]

free fall | situation in which the only force acting on an object is gravity [OpenStax]

free fall | the state of movement that results from gravitational force only [OpenStax]

free-body diagram | sketch showing all external forces acting on an object or system; the system is represented by a single isolated point, and the forces are represented by vectors extending outward from that point [OpenStax]

frequency (f) | number of events per unit of time [OpenStax]

friction | force that opposes relative motion or attempts at motion between systems in contact [OpenStax]

fundamental | the lowest-frequency resonance [OpenStax]

fundamental frequency | lowest frequency that will produce a standing wave [OpenStax]

gauge pressure | pressure relative to atmospheric pressure [OpenStax]

gravitational field | vector field that surrounds the mass creating the field; the field is represented by field lines, in which the direction of the field is tangent to the lines, and the magnitude (or field strength) is inversely proportional to the spacing of the lines; other masses respond to this field [OpenStax]

gravitational torque | torque on the body caused by its weight; it occurs when the center of gravity of the body is not located on the axis of rotation [OpenStax]

gravitationally bound | two object are gravitationally bound if their orbits are closed; gravitationally bound systems have a negative total mechanical energy [OpenStax]

harmonics | the term used to refer collectively to the fundamental and its overtones [OpenStax]

hearing | perception of sound [OpenStax]

Hooke's law | in a spring, a restoring force proportional to and in the opposite direction of the imposed displacement [OpenStax]

hydraulic jack | simple machine that uses cylinders of different diameters to distribute force [OpenStax]

hydrostatic equilibrium | state at which water is not flowing, or is static [OpenStax]

ideal banking | sloping of a curve in a road, where the angle of the slope allows the vehicle to negotiate the curve at a certain speed without the aid of friction between the tires and the road; the net external force on the vehicle equals the horizontal centripetal force in the absence of friction [OpenStax]

ideal fluid | fluid with negligible viscosity [OpenStax]

impulse | effect of applying a force on a system for a time interval; this time interval is usually small, but does not have to be [OpenStax]

impulse-momentum theorem | change of momentum of a system is equal to the impulse applied to the system [OpenStax]

inelastic | collision that does not conserve kinetic energy [OpenStax]

inertia | ability of an object to resist changes in its motion [OpenStax]

inertial force | force that has no physical origin [OpenStax]

inertial reference frame | reference frame moving at constant velocity relative to an inertial frame is also inertial; a reference frame accelerating relative to an inertial frame is not inertial [OpenStax]

instantaneous acceleration | acceleration at a specific point in time [OpenStax]

instantaneous angular acceleration | derivative of angular velocity with respect to time [OpenStax]

instantaneous angular velocity | derivative of angular position with respect to time [OpenStax]

instantaneous speed | the absolute value of the instantaneous velocity [OpenStax]

instantaneous velocity | the velocity at a specific instant or time point [OpenStax]

intensity (I) | power per unit area [OpenStax]

interference | overlap of two or more waves at the same point and time [OpenStax]

internal force | force that the simple particles that make up an extended object exert on each other. Internal forces can be attractive or repulsive [OpenStax]

Kepler's first law | law stating that every planet moves along an ellipse, with the Sun located at a focus of the ellipse [OpenStax]

Kepler's second law | law stating that a planet sweeps out equal areas in equal times, meaning it has a constant areal velocity [OpenStax]

Kepler's third law | law stating that the square of the period is proportional to the cube of the semi-major axis of the orbit [OpenStax]

kilogram | SI unit for mass, abbreviated kg [OpenStax]

kinematics | the description of motion through properties such as position, time, velocity, and acceleration [OpenStax]

kinematics of rotational motion | describes the relationships among rotation angle, angular velocity, angular acceleration, and time [OpenStax]

kinetic energy | energy of motion, one-half an object's mass times the square of its speed [OpenStax]

kinetic friction | force that opposes the motion of two systems that are in contact and moving relative to each other [OpenStax]

laminar flow | type of fluid flow in which layers do not mix [OpenStax]

law | description, using concise language or a mathematical formula, of a generalized pattern in nature supported by scientific evidence and repeated experiments [OpenStax]

law of conservation of angular momentum | angular momentum is conserved, that is, the initial angular momentum is equal to the final angular momentum when no external torque is applied to the system [OpenStax]

Law of Conservation of Momentum | total momentum of a closed system cannot change [OpenStax]

law of inertia | see Newton's first law of motion [OpenStax]

lever arm | perpendicular distance from the line that the force vector lies on to a given axis [OpenStax]

linear mass density | the mass per unit length λ of a one dimensional object [OpenStax]

linear mass density | λ, expressed as the number of kilograms of material per meter [OpenStax]

linear wave equation | equation describing waves that result from a linear restoring force of the medium; any function that is a solution to the wave equation describes a wave moving in the positive x-direction or the negative x-direction with a constant wave speed v [OpenStax]

linearity limit (proportionality limit) | largest stress value beyond which stress is no longer proportional to strain [OpenStax]

longitudinal wave | wave in which the disturbance is parallel to the direction of propagation [OpenStax]

loudness | perception of sound intensity [OpenStax]

magnitude | length of a vector [OpenStax]

mechanical energy | sum of the kinetic and potential energies [OpenStax]

mechanical wave | wave that is governed by Newton's laws and requires a medium [OpenStax]

meter | SI unit for length, abbreviated m [OpenStax]

method of adding percents | the percent uncertainty in a quantity calculated by multiplication or division is the sum of the percent uncertainties in the items used to make the calculation [OpenStax]

metric system | system in which values can be calculated in factors of 10 [OpenStax]

model | representation of something often too difficult (or impossible) to display directly [OpenStax]

moment of inertia | rotational mass of rigid bodies that relates to how easy or hard it will be to change the angular velocity of the rotating rigid body [OpenStax]

momentum | measure of the quantity of motion that an object has; it takes into account both how fast the object is moving, and its mass; specifically, it is the product of mass and velocity; it is a vector quantity [OpenStax]

natural angular frequency | angular frequency of a system oscillating in SHM [OpenStax]

neap tide | low tide created when the Moon and the Sun form a right triangle with Earth [OpenStax]

net external force | vector sum of all external forces acting on an object or system; causes a mass to accelerate [OpenStax]

net work | work done by all the forces acting on an object [OpenStax]

neutron star | most compact object known—outside of a black hole itself [OpenStax]

newton | SI unit of force; 1 N is the force needed to accelerate an object with a mass of 1 kg at a rate of 1 m/s2 [OpenStax]

Newton's first law of motion | body at rest remains at rest or, if in motion, remains in motion at constant velocity unless acted on by a net external force; also known as the law of inertia [OpenStax]

Newton's law of gravitation | every mass attracts every other mass with a force proportional to the product of their masses, inversely proportional to the square of the distance between them, and with direction along the line connecting the center of mass of each [OpenStax]

Newton's second law for rotation | sum of the torques on a rotating system equals its moment of inertia times its angular acceleration [OpenStax]

Newton's second law of motion | acceleration of a system is directly proportional to and in the same direction as the net external force acting on the system and is inversely proportional to its mass [OpenStax]

Newton's third law of motion | whenever one body exerts a force on a second body, the first body experiences a force that is equal in magnitude and opposite in direction to the force that it exerts [OpenStax]

node | point where the string does not move; more generally, nodes are where the wave disturbance is zero in a standing wave [OpenStax]

non-conservative force | force that does work that depends on path [OpenStax]

non-Euclidean geometry | geometry of curved space, describing the relationships among angles and lines on the surface of a sphere, hyperboloid, etc. [OpenStax]

non-renewable | energy source that is not renewable, but is depleted by human consumption [OpenStax]

noninertial frame of reference | accelerated frame of reference [OpenStax]

normal force | force supporting the weight of an object, or a load, that is perpendicular to the surface of contact between the load and its support; the surface applies this force to an object to support the weight of the object [OpenStax]

normal mode | possible standing wave pattern for a standing wave on a string [OpenStax]

normal pressure | pressure of one atmosphere, serves as a reference level for pressure [OpenStax]

notes | basic unit of music with specific names, combined to generate tunes [OpenStax]

null vector | a vector with all its components equal to zero [OpenStax]

orbital period | time required for a satellite to complete one orbit [OpenStax]

orbital speed | speed of a satellite in a circular orbit; it can be also be used for the instantaneous speed for noncircular orbits in which the speed is not constant [OpenStax]

order of magnitude | the size of a quantity as it relates to a power of 10 [OpenStax]

orthogonal vectors | two vectors with directions that differ by exactly 90°, synonymous with perpendicular vectors [OpenStax]

oscillation | single fluctuation of a quantity, or repeated and regular fluctuations of a quantity, between two extreme values around an equilibrium or average value [OpenStax]

overdamped | condition in which damping of an oscillator causes it to return to equilibrium without oscillating; oscillator moves more slowly toward equilibrium than in the critically damped system [OpenStax]

overtone | frequency that produces standing waves and is higher than the fundamental frequency [OpenStax]

overtones | all resonant frequencies higher than the fundamental [OpenStax]

parallel axis | axis of rotation that is parallel to an axis about which the moment of inertia of an object is known [OpenStax]

parallel vectors | two vectors with exactly the same direction angles [OpenStax]

parallel-axis theorem | if the moment of inertia is known for a given axis, it can be found for any axis parallel to it [OpenStax]

parallelogram rule | geometric construction of the vector sum in a plane [OpenStax]

pascal (Pa) | SI unit of stress, SI unit of pressure [OpenStax]

Pascal's principle | change in pressure applied to an enclosed fluid is transmitted undiminished to all portions of the fluid and to the walls of its container [OpenStax]

percent uncertainty | the ratio of the uncertainty of a measurement to the measured value, expressed as a percentage [OpenStax]

perfectly inelastic | collision after which all objects are motionless, the final kinetic energy is zero, and the loss of kinetic energy is a maximum [OpenStax]

perihelion | point of closest approach to the Sun of an orbiting body; the corresponding term for the Moon's closest approach to Earth is the perigee [OpenStax]

period (T) | time taken to complete one oscillation [OpenStax]

periodic motion | motion that repeats itself at regular time intervals [OpenStax]

phase shift | angle, in radians, that is used in a cosine or sine function to shift the function left or right, used to match up the function with the initial conditions of data [OpenStax]

phon | numerical unit of loudness [OpenStax]

physical pendulum | any extended object that swings like a pendulum [OpenStax]

physical quantity | characteristic or property of an object that can be measured or calculated from other measurements [OpenStax]

physics | science concerned with describing the interactions of energy, matter, space, and time; especially interested in what fundamental mechanisms underlie every phenomenon [OpenStax]

pitch | perception of the frequency of a sound [OpenStax]

plastic behavior | material deforms irreversibly, does not go back to its original shape and size when load is removed and stress vanishes [OpenStax]

Poiseuille's law | rate of laminar flow of an incompressible fluid in a tube:

$$Q = \frac{(p_2 - p_1)\pi r^4}{8\eta l}. \quad (3)$$

[OpenStax]

Poiseuille's law for resistance | resistance to laminar flow of an incompressible fluid in a tube:

$$R = \frac{8\eta l}{\pi r^4} \quad (4)$$

[OpenStax]

polar coordinate system | an orthogonal coordinate system where location in a plane is given by polar coordinates [OpenStax]

polar coordinates | a radial coordinate and an angle [OpenStax]

position | the location of an object at a particular time [OpenStax]

position vector | vector from the origin of a chosen coordinate system to the position of a particle in two- or threedimensional space [OpenStax]

potential energy | function of position, energy possessed by an object relative to the system considered [OpenStax]

potential energy diagram | graph of a particle's potential energy as a function of position [OpenStax]

potential energy difference | negative of the work done acting between two points in space [OpenStax]

power | (or instantaneous power) rate of doing work [OpenStax]

precession | circular motion of the pole of the axis of a spinning object around another axis due to a torque [OpenStax]

precision | the degree to which repeated measurements agree with each other [OpenStax]

pressure | force per unit area exerted perpendicular to the area over which the force acts [OpenStax]

pressure | force pressing in normal direction on a surface per the surface area, the bulk stress in fluids [OpenStax]

principle of equivalence | part of the general theory of relativity, it states that there no difference between free fall and being weightless, or a uniform gravitational field and uniform acceleration [OpenStax]

projectile motion | motion of an object subject only to the acceleration of gravity [OpenStax]

pulse | single disturbance that moves through a medium, transferring energy but not mass [OpenStax]

radical coordinate | distance to the origin in a polar coordinate system [OpenStax]

range | maximum horizontal distance a projectile travels [OpenStax]

reference frame | coordinate system in which the position, velocity, and acceleration of an object at rest or moving is measured [OpenStax]

relative velocity | velocity of an object as observed from a particular reference frame, or the velocity of one reference frame with respect to another reference frame [OpenStax]

renewable | energy source that is replenished by natural processes, over human time scales [OpenStax]

resonance | large amplitude oscillations in a system produced by a small amplitude driving force, which has a frequency equal to the natural frequency [OpenStax]

restoring force | force acting in opposition to the force caused by a deformation [OpenStax]

resultant vector | vector sum of two (or more) vectors [OpenStax]

Reynolds number | dimensionless parameter that can reveal whether a particular flow is laminar or turbulent [OpenStax]

rocket equation | derived by the Soviet physicist Konstantin Tsiolkovsky in 1897, it gives us the change of velocity that the rocket obtains from burning a mass of fuel that decreases the total rocket mass from mi down to m [OpenStax]

rolling motion | combination of rotational and translational motion with or without slipping [OpenStax]

rotational dynamics | analysis of rotational motion using the net torque and moment of inertia to find the angular acceleration [OpenStax]

rotational kinetic energy | kinetic energy due to the rotation of an object; this is part of its total kinetic energy [OpenStax]

rotational work | work done on a rigid body due to the sum of the torques integrated over the angle through with the body rotates [OpenStax]

scalar | a number, synonymous with a scalar quantity in physics [OpenStax]

scalar component | a number that multiplies a unit vector in a vector component of a vector [OpenStax]

scalar equation | equation in which the left-hand and right-hand sides are numbers [OpenStax]

scalar product | the result of the scalar multiplication of two vectors is a scalar called a scalar product; also called a dot product [OpenStax]

scalar quantity | quantity that can be specified completely by a single number with an appropriate physical unit [OpenStax]

Schwarzschild radius | critical radius (RS) such that if a mass were compressed to the extent that its radius becomes less than the Schwarzschild radius, then the mass will collapse to a singularity, and anything that passes inside that radius cannot escape [OpenStax]

second | the SI unit for time, abbreviated s [OpenStax]

second equilibrium condition | expresses rotational equilibrium; all torques due to external forces acting on the body balance out and their vector sum is zero [OpenStax]

shear modulus | elastic modulus for shear stress [OpenStax]

shear strain | strain caused by shear stress [OpenStax]

shear stress | stress caused by shearing forces [OpenStax]

shock wave | wave front that is produced when a sound source moves faster than the speed of sound [OpenStax]

SI units | the international system of units that scientists in most countries have agreed to use; includes units such as meters, liters, and grams [OpenStax]

significant figures | used to express the precision of a measuring tool used to measure a value [OpenStax]

simple harmonic motion (SHM) | oscillatory motion in a system where the restoring force is proportional to the displacement, which acts in the direction opposite to the displacement [OpenStax]

simple harmonic oscillator | a device that oscillates in SHM where the restoring force is proportional to the displacement and acts in the direction opposite to the displacement [OpenStax]

simple pendulum | point mass, called a pendulum bob, attached to a near massless string [OpenStax]

sonic boom | loud noise that occurs as a shock wave as it sweeps along the ground [OpenStax]

sound | traveling pressure wave that may be periodic; the wave can be modeled as a pressure wave or as an oscillation of molecules [OpenStax]

sound intensity level | unitless quantity telling you the level of the sound relative to a fixed standard [OpenStax]

sound pressure level | ratio of the pressure amplitude to a reference pressure [OpenStax]

space-time | concept of space-time is that time is essentially another coordinate that is treated the same way as any individual spatial coordinate; in the equations that represent both special and general relativity, time appears in the same context as do the spatial coordinates [OpenStax]

specific gravity | ratio of the density of an object to a fluid (usually water) [OpenStax]

spring tide | high tide created when the Moon, the Sun, and Earth are along one line [OpenStax]

stable equilibrium point | point where the net force on a system is zero, but a small displacement of the mass will cause a restoring force that points toward the equilibrium point [OpenStax]

standing wave | wave that can bounce back and forth through a particular region, effectively becoming stationary [OpenStax]

static equilibrium | body is in static equilibrium when it is at rest in our selected inertial frame of reference [OpenStax]

static friction | force that opposes the motion of two systems that are in contact and are not moving relative to each other [OpenStax]

strain | dimensionless quantity that gives the amount of deformation of an object or medium under stress [OpenStax]

stress | quantity that contains information about the magnitude of force causing deformation, defined as force per unit area [OpenStax]

stress-strain diagram | graph showing the relationship between stress and strain, characteristic of a material [OpenStax]

superposition | phenomenon that occurs when two or more waves arrive at the same point [OpenStax]

surface mass density | mass per unit area σ of a two dimensional object [OpenStax]

system | object or collection of objects whose motion is currently under investigation; however, your system is defined at the start of the problem, you must keep that definition for the entire problem [OpenStax]

tail-to-head geometric construction | geometric construction for drawing the resultant vector of many vectors [OpenStax]

tangential accleration | magnitude of which is the time rate of change of speed. Its direction is tangent to the circle. [OpenStax]

tensile strain | strain under tensile stress, given as fractional change in length, which occurs when forces are stretching an object, causing its elongation [OpenStax]

tensile stress | stress caused by tensile forces, only in one direction, which occurs when forces are stretching an object, causing its elongation [OpenStax]

tension | pulling force that acts along a stretched flexible connector, such as a rope or cable [OpenStax]

terminal velocity | constant velocity achieved by a falling object, which occurs when the weight of the object is balanced by the upward drag force [OpenStax]

theory | testable explanation for patterns in nature supported by scientific evidence and verified multiple times by various groups of researchers [OpenStax]

theory of general relativity | Einstein's theory for gravitation and accelerated reference frames; in this theory, gravitation is the result of mass and energy distorting the space-time around it; it is also often referred to as Einstein's theory of gravity [OpenStax]

thrust | reaction force that pushes a body forward in response to a backward force [OpenStax]

tidal force | difference between the gravitational force at the center of a body and that at any other location on the body; the tidal force stretches the body [OpenStax]

timbre | number and relative intensity of multiple sound frequencies [OpenStax]

time of flight | elapsed time a projectile is in the air [OpenStax]

torque | cross product of a force and a lever arm to a given axis [OpenStax]

torsional pendulum | any suspended object that oscillates by twisting its suspension [OpenStax]

total accleration | vector sum of centripetal and tangential accelerations [OpenStax]

total displacement | the sum of individual displacements over a given time period [OpenStax]

total linear acceleration | vector sum of the centripetal acceleration vector and the tangential acceleration vector [OpenStax]

trajectory | path of a projectile through the air [OpenStax]

transducer | device that converts energy of a signal into measurable energy form, for example, a microphone converts sound waves into an electrical signal [OpenStax]

transverse wave | wave in which the disturbance is perpendicular to the direction of propagation [OpenStax]

turbulence | fluid flow in which layers mix together via eddies and swirls [OpenStax]

turbulent flow | type of fluid flow in which layers mix together via eddies and swirls [OpenStax]

turning point | position where the velocity of a particle, in one-dimensional motion, changes sign [OpenStax]

two-body pursuit problem | a kinematics problem in which the unknowns are calculated by solving the kinematic equations simultaneously for two moving objects [OpenStax]

uncertainty | a quantitative measure of how much measured values deviate from one another [OpenStax]

underdamped | condition in which damping of an oscillator causes the amplitude of oscillations of a damped harmonic oscillator to decrease over time, eventually approaching zero [OpenStax]

unit vector | vector of a unit magnitude that specifies direction; has no physical unit [OpenStax]

unit vectors of the axes | unit vectors that define orthogonal directions in a plane or in space [OpenStax]

units | standards used for expressing and comparing measurements [OpenStax]

universal gravitational constant | constant representing the strength of the gravitational force, that is believed to be the same throughout the universe [OpenStax]

vector | mathematical object with magnitude and direction [OpenStax]

vector components | orthogonal components of a vector; a vector is the vector sum of its vector components [OpenStax]

vector equation | equation in which the left-hand and right-hand sides are vectors [OpenStax]

vector product | the result of the vector multiplication of vectors is a vector called a vector product; also called a cross product [OpenStax]

vector quantity | physical quantity described by a mathematical vector—that is, by specifying both its magnitude and its direction; synonymous with a vector in physics [OpenStax]

vector sum | resultant of the combination of two (or more) vectors [OpenStax]

velocity vector | vector that gives the instantaneous speed and direction of a particle; tangent to the trajectory [OpenStax]

viscosity | measure of the internal friction in a fluid [OpenStax]

wave | disturbance that moves from its source and carries energy [OpenStax]

wave function | mathematical model of the position of particles of the medium [OpenStax]

wave number | $\frac{2\pi}{\lambda}$ [OpenStax]

wave speed | magnitude of the wave velocity [OpenStax]

wave velocity | velocity at which the disturbance moves; also called the propagation velocity [OpenStax]

wavelength | distance between adjacent identical parts of a wave [OpenStax]

weight | force $\vec{w}$ due to gravity acting on an object of mass m [OpenStax]

work | done when a force acts on something that undergoes a displacement from one position to another [OpenStax]

work done by a force | integral, from the initial position to the final position, of the dot product of the force and the infinitesimal displacement along the path over which the force acts [OpenStax]

work-energy theorem | net work done on a particle is equal to the change in its kinetic energy [OpenStax]

work-energy theorem for rotation | the total rotational work done on a rigid body is equal to the change in rotational kinetic energy of the body [OpenStax]

Young's modulus | elastic modulus for tensile or compressive stress [OpenStax]

Detailed Licensing

Overview

Title: Radically Modern Introductory Physics Text I (Raymond)

Webpages: 108

Applicable Restrictions: Noncommercial

All licenses found:

- CC BY-NC-SA 3.0: 79.6% (86 pages)
- Undeclared: 20.4% (22 pages)

By Page

- Radically Modern Introductory Physics Text I (Raymond) - *CC BY-NC-SA 3.0*
  - Front Matter - *Undeclared*
    - TitlePage - *Undeclared*
    - InfoPage - *Undeclared*
    - Table of Contents - *Undeclared*
    - Licensing - *Undeclared*
  - 1: Waves in One Dimension - *CC BY-NC-SA 3.0*
    - 1.1: Transverse and Longitudinal Waves - *CC BY-NC-SA 3.0*
    - 1.2: Sine Waves - *CC BY-NC-SA 3.0*
    - 1.3: Types of Waves - *CC BY-NC-SA 3.0*
    - 1.4: Superposition Principle - *CC BY-NC-SA 3.0*
    - 1.5: Beats - *CC BY-NC-SA 3.0*
    - 1.6: Interferometers - *CC BY-NC-SA 3.0*
    - 1.7: Thin Films - *Undeclared*
    - 1.8: Math Review — Derivatives - *Undeclared*
    - 1.9: Group Velocity - *Undeclared*
    - 1.10: Problems - *Undeclared*
  - 2: Waves in Two and Three Dimensions - *CC BY-NC-SA 3.0*
    - 2.1: Math Tutorial — Vectors - *CC BY-NC-SA 3.0*
    - 2.2: Plane Waves - *CC BY-NC-SA 3.0*
    - 2.3: Superposition of Plane Waves - *CC BY-NC-SA 3.0*
    - 2.4: Diffraction Through a Single Slit - *CC BY-NC-SA 3.0*
    - 2.5: Two Slits - *CC BY-NC-SA 3.0*
    - 2.6: Diffraction Gratings - *CC BY-NC-SA 3.0*
    - 2.7: Problems - *Undeclared*
  - 3: Geometrical Optics - *CC BY-NC-SA 3.0*
    - 3.1: Reflection and Refraction - *CC BY-NC-SA 3.0*
    - 3.2: Total Internal Reflection - *CC BY-NC-SA 3.0*
    - 3.3: Anisotropic Media - *CC BY-NC-SA 3.0*
    - 3.4: Thin Lens Equation and Optical Instruments - *CC BY-NC-SA 3.0*
    - 3.5: Fermat's Principle - *CC BY-NC-SA 3.0*
    - 3.6: Problems - *CC BY-NC-SA 3.0*
  - 4: Special Relativity - *CC BY-NC-SA 3.0*
    - 4.1: Galilean Spacetime Thinking - *CC BY-NC-SA 3.0*
    - 4.2: Spacetime Thinking in Special Relativity - *CC BY-NC-SA 3.0*
    - 4.3: Postulates of Special Relativity - *CC BY-NC-SA 3.0*
    - 4.4: Time Dilation - *CC BY-NC-SA 3.0*
    - 4.5: Lorentz Contraction - *CC BY-NC-SA 3.0*
    - 4.6: Twin Paradox - *CC BY-NC-SA 3.0*
    - 4.7: Problems - *Undeclared*
  - 5: Applications of Special Relativity - *CC BY-NC-SA 3.0*
    - 5.1: Waves in Spacetime - *CC BY-NC-SA 3.0*
    - 5.2: Math Tutorial – Four-Vectors - *CC BY-NC-SA 3.0*
    - 5.3: Principle of Relativity Applied - *CC BY-NC-SA 3.0*
    - 5.4: Characteristics of Relativistic Waves - *CC BY-NC-SA 3.0*
    - 5.5: The Doppler Shift - *CC BY-NC-SA 3.0*
    - 5.6: Addition of Velocities - *CC BY-NC-SA 3.0*
    - 5.7: Problems - *Undeclared*
  - 6: Acceleration and General Relativity - *CC BY-NC-SA 3.0*
    - 6.1: Acceleration - *CC BY-NC-SA 3.0*
    - 6.2: Circular Motion - *CC BY-NC-SA 3.0*
    - 6.3: Acceleration, Force, and Mass - *CC BY-NC-SA 3.0*
    - 6.4: Acceleration in Special Relativity - *CC BY-NC-SA 3.0*
    - 6.5: Accelerated Reference Frames - *CC BY-NC-SA 3.0*
    - 6.6: Gravitational Red Shift - *CC BY-NC-SA 3.0*
    - 6.7: Event Horizons - *Undeclared*
    - 6.8: Problems - *Undeclared*
  - 7: Matter Waves - *CC BY-NC-SA 3.0*
    - 7.1: Bragg's Law - *CC BY-NC-SA 3.0*

- 7.2: X-Ray Diffraction Techniques - *CC BY-NC-SA 3.0*
- 7.3: Meaning of Quantum Wave Function - *CC BY-NC-SA 3.0*
- 7.4: Sense and Nonsense in Quantum Mechanics - *CC BY-NC-SA 3.0*
- 7.5: Mass, Momentum, and Energy - *CC BY-NC-SA 3.0*
- 7.6: Heisenberg Uncertainty Principle - *CC BY-NC-SA 3.0*
- 7.7: Problems - *CC BY-NC-SA 3.0*

8: Geometrical Optics and Newton's Laws - *CC BY-NC-SA 3.0*
- 8.1: Fundamental Principles of Dynamics - *CC BY-NC-SA 3.0*
- 8.2: Potential Energy - *CC BY-NC-SA 3.0*
- 8.3: Work and Power - *CC BY-NC-SA 3.0*
- 8.4: Mechanics and Geometrical Optics - *CC BY-NC-SA 3.0*
- 8.5: Math Tutorial – Partial Derivatives - *CC BY-NC-SA 3.0*
- 8.6: Motion in Two and Three Dimensions - *CC BY-NC-SA 3.0*
- 8.7: Kinetic and Total Momentum - *CC BY-NC-SA 3.0*
- 8.8: Problems - *CC BY-NC-SA 3.0*

9: Symmetry and Bound States - *CC BY-NC-SA 3.0*
- 9.1: Math Tutorial — Complex Waves - *CC BY-NC-SA 3.0*
- 9.2: Symmetry and Quantum Mechanics - *CC BY-NC-SA 3.0*
- 9.3: Confined Matter Waves - *CC BY-NC-SA 3.0*
- 9.4: Problems - *CC BY-NC-SA 3.0*

10: Dynamics of Multiple Particles - *CC BY-NC-SA 3.0*
- 10.1: Momentum and Newton's Second Law - *CC BY-NC-SA 3.0*
- 10.2: Newton's Third Law - *CC BY-NC-SA 3.0*
- 10.3: Conservation of Momentum - *CC BY-NC-SA 3.0*
- 10.4: Collisions - *CC BY-NC-SA 3.0*
- 10.5: Rockets and Conveyor Belts - *Undeclared*
- 10.6: Problems - *Undeclared*

11: Rotational Dynamics - *CC BY-NC-SA 3.0*
- 11.1: Math Tutorial — Cross Product - *CC BY-NC-SA 3.0*
- 11.2: Torque and Angular Momentum - *CC BY-NC-SA 3.0*
- 11.3: Two Particles - *CC BY-NC-SA 3.0*
- 11.4: The Uneven Dumbbell - *CC BY-NC-SA 3.0*
- 11.5: Many Particles - *CC BY-NC-SA 3.0*
- 11.6: New Page - *CC BY-NC-SA 3.0*
- 11.7: Statics - *CC BY-NC-SA 3.0*
- 11.8: Problem - *CC BY-NC-SA 3.0*

12: Harmonic Oscillator - *CC BY-NC-SA 3.0*
- 12.1: Energy Analysis - *CC BY-NC-SA 3.0*
- 12.2: Analysis Using Newton's Laws - *CC BY-NC-SA 3.0*
- 12.3: Forced Oscillator - *CC BY-NC-SA 3.0*
- 12.4: Complex Exponential Solutions - *CC BY-NC-SA 3.0*
- 12.5: Quantum Mechanical Harmonic Oscillator - *CC BY-NC-SA 3.0*
- 12.6: Problems - *CC BY-NC-SA 3.0*

13: Appendix A, Constants, Units, and Conversions - *Undeclared*
- 13.1: Appendix A Constants, Units, and Conversions - *Undeclared*

Back Matter - *Undeclared*
- Index - *Undeclared*
- Glossary - *Undeclared*
- Detailed Licensing - *Undeclared*