

Symmetries and the fundamental forces of Nature II

Luis A. Anchordoqui

Department of Physics and Astronomy, Lehman College, City University of New York, NY 10468, USA

Department of Physics, Graduate Center, City University of New York, 365 Fifth Avenue, NY 10016, USA

Department of Astrophysics, American Museum of Natural History, Central Park West 79 St., NY 10024, USA

(Dated: Fall 2015)

Lecture script of a one-semester course that aims to develop an understanding and appreciation of fundamental concepts in modern physics for students who are comfortable with calculus. This document contains the last part of the course where we introduce the concept of the *quantum unit* to describe the dynamic properties of subatomic particles and the interactions of matter and radiation.

IX. ORIGINS OF QUANTUM MECHANICS

Quantum mechanics was born in the beginning of the 20th century due to an apparent collapse of the deterministic classical mechanics driven by Euler-Lagrange equations [1, 2]. The collapse resulted from the discovery of various phenomena which are inexplicable with classical physics [3–7]. The historical path to quantum mechanics invariably begins with Planck and his analysis of blackbody spectral data [8–10]. It is this that we now turn to study.

A. Blackbody radiation

The rate at which an object radiates energy has been found to be proportional to the fourth power of the Kelvin temperature T and to the area A of the emitting object, i.e., $L \propto AT^4$. At normal temperatures (≈ 300 K) we are not aware of this electromagnetic radiation because of its low intensity. At higher temperatures, there is sufficient infrared radiation that we can feel heat if we are close to the object. At still higher temperatures (on the order of 1000 K), objects actually glow, such as a red-hot electric stove burner. At temperatures above 2000 K, objects glow with a yellow or whitish color, such as the filament of a lightbulb.

For an idealized object that absorbs all incident electromagnetic radiation regardless of frequency or angle of incidence (a.k.a. blackbody [11]), the bolometric luminosity becomes

$$L = \sigma A T^4, \quad (206)$$

where $\sigma = 5.67 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$ is the Stefan-Boltzmann constant [12, 13]. The total power leaving 1 m^2 of a blackbody surface – that is the radiant flux F measured in W/m^2 – at absolute temperature T is then

$$F(T) = L/A = \sigma T^4. \quad (207)$$

Another important characteristic of the blackbody radiation is the experimentally obtained Wien’s displacement law [14], which states that the wavelength λ_{max} at which the spectral emittance reaches its maximum value

decreases as the temperature is increased, in inverse proportion to the temperature

$$\lambda_{\text{max}} T = 2.90 \times 10^{-3} \text{ m K}. \quad (208)$$

The term “displacement” refers to the way the peak is moved or displaced as the temperature is varied. Wien’s law is qualitatively consistent with our previous observation that heated objects first begin to glow with a red color, and at higher temperatures the color becomes more yellow. As the temperature is increased, the wavelength at which most of the radiation is emitted moves from the longer-wavelength (red) part of the visible region toward medium (yellow or whitish) wavelengths.

An approximate realization of a blackbody surface is as follows. Consider a hollow metal box whose walls are in thermal equilibrium at temperature T . The cavity is filled with electromagnetic radiation forming standing waves (normal modes) with nodes at the walls, as shown in Fig. 29. Suppose there is a small hole in one wall of the box which allows some of the radiation to escape. It is the hole, and not the box itself, that is the blackbody. Radiation from outside that is incident on the hole gets lost inside the box and has a negligible chance of reemerging from the hole; thus no reflections occur from the blackbody (the hole). The radiation that emerges from the hole is just a sample of the radiation inside the box, so understanding the nature of the radiation inside the box allows us to understand the radiation that leaves through the hole.

Let us consider the radiation inside the box. It has an energy density (that is energy per unit volume) $u(T)$ measured in J m^{-3} , and a spectral energy density $du/d\lambda \equiv u_\lambda(\lambda, T)$, measured in $\text{J m}^{-3} \text{ nm}^{-1}$. That is, if we could look into the interior of the box and measure the spectral energy density of the electromagnetic radiation with wavelengths between λ and $\lambda + d\lambda$ in a small volume element, the result would be $u_\lambda(\lambda, T) d\lambda$. The surface brightness (or spectral emittance) $B_\lambda(\lambda, T)$ is defined as the spectral radiant flux per steradian emitted from a unit surface that lies normal to the view direction. The surface element is thus different for each view direction. Because photons of all wavelengths travel at speed c , the wavelength dependence of the spectral energy density is the same as that of the surface brightness. It does not matter whether the radiation sampled is that

in 1 m^3 at fixed time, or that impinging on a 1 m^2 in 1 s. The relation between the two quantities, B_λ and u_λ , follows from their definitions. Consider first the spectral energy density u_λ . It includes photons moving at speed c isotropically in all directions into all $4\pi \text{ sr}$. Divide by 4π to obtain the energy per unit volume flowing into 1 sr, i.e. $u_\lambda/4\pi$. Multiply this by the speed c of the photons to obtain the spectral radiant flux per steradian passing through 1 m^2 in 1 s, which is the surface brightness,

$$B_\lambda(\lambda, T) = \frac{c}{4\pi} u_\lambda(\lambda, T). \quad (209)$$

It is useful to know the total radiant flux passing through a unit area of a fixed surface immersed within the black-body cavity. Consider the radiation field we have inside the box (i.e. a bundle of rays flowing in all directions) and construct a small element of area dA at some arbitrary orientation \hat{n} as shown in Fig. 30. The differential amount of flux from the solid angle $d\Omega$ is (reduced by the lowered effective area $\cos \theta dA$),

$$dF_\lambda(\lambda, T) = B_\lambda(\lambda, T) \cos \theta d\Omega, \quad (210)$$

where the solid angle element is $d\Omega = \sin \theta d\theta d\phi$. The spectral radiant flux in the direction \hat{n} , is obtained by integrating (210) over the upper hemisphere shown in Fig. 30. Note that if B_λ is an isotropic radiation field (not a function of angle), then the *net* flux is zero, since $\int \cos \theta d\Omega = 0$ over all solid angles. That is, there is as much energy crossing dA in the \hat{n} direction as in the $-\hat{n}$ direction. The total power leaving 1 m^2 of the surface (i.e. the radiant flux) is the surface brightness integrated over all frequencies and all angles of the upper hemisphere

$$F(T) = \int_0^\infty \int_{\theta=0}^{\pi/2} \int_{\phi=0}^{2\pi} B_\lambda(\lambda, T) \frac{dA \cos \theta}{dA} d\Omega d\lambda. \quad (211)$$

The integration limits for θ specify that the integration is only over the upper hemisphere, yielding

$$\int_{\theta=0}^{\pi/2} \int_{\phi=0}^{2\pi} \cos \theta \sin \theta d\theta d\phi = \pi. \quad (212)$$

On the other hand, the total energy $u(T)$ in a unit volume is simply $u_\lambda(\lambda, T)$ integrated over all wavelengths. From (207) it follows that

$$u(T) = \int_0^\infty u_\lambda(\lambda, T) d\lambda = aT^4, \quad (213)$$

where $a = 4\sigma/c = 7.566 \times 10^{-16} \text{ J m}^{-3} \text{ K}^{-4}$. Our next goal is to derive an analytic expression for the spectral energy density that satisfies (213).

We have seen that any wave can be characterized by: the wavelength λ , the speed c , the period $T = \lambda/c$, the frequency $\nu = 1/T = c/\lambda$, or else by the angular frequency

$$\omega = 2\pi\nu = 2\pi c/\lambda. \quad (214)$$

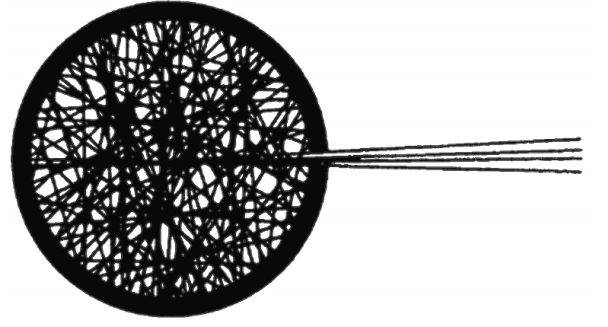


FIG. 29: A cavity filled with electromagnetic radiation in thermal equilibrium with its walls at temperature T . Some radiation escapes through the hole, which represents an ideal blackbody.

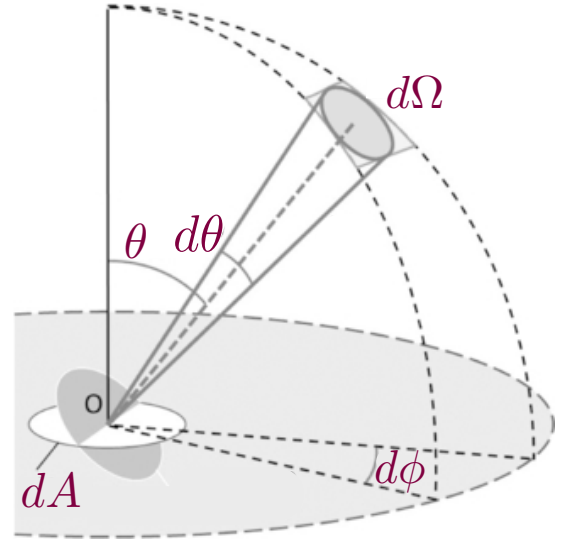


FIG. 30: Geometry for an obliquely emerging solid angle $d\Omega$ from a unit surface area dA , with unitary normal vector \hat{n} .

In what follows we will refer to both the frequency ν and the angular frequency ω as simply the frequency, since this will not lead to a confusion of the two. We start by relating $u_\omega(\omega, T)$ with $u_\lambda(\lambda, T)$. This is done as follows. The spectral energy density within the wavelength interval $(\lambda, \lambda + \Delta\lambda)$ can be written in two ways:

$$u_\lambda(\lambda, T) d\lambda = u_\omega(\omega, T) d\omega. \quad (215)$$

The left-hand side of (215) is merely the definition of the spectral energy density. The right-hand side expresses the fact that λ and ω are related by a one-to-one function (214). Then, (214) and (215) provide the following relation

$$u_\lambda(\lambda, T) = u_\omega(\omega, T) \left| \frac{d\omega}{d\lambda} \right| = u_\omega(\omega, T) \frac{2\pi c}{\lambda^2}. \quad (216)$$

To begin the derivation of $u_\omega(\omega, T)$ consider a rectangular box of volume $V = L_x L_y L_z$, which is at thermal

equilibrium at temperature T . The spectral energy density of the electromagnetic radiation with frequencies between ω and $\omega + d\omega$ in a small volume element is

$$u_\omega(\omega, T)d\omega = N(\omega, T) \langle E \rangle, \quad (217)$$

where $N(\omega, T)$ is the number of electromagnetic modes (standing waves) within a given frequency interval $(\omega, \omega + d\omega)$ allowed inside the box, $\langle E \rangle$ is the average energy of one radiation mode of frequency ω , and $u_\omega(\omega, T)$ is the average energy per ω -interval per volume; the qualifier “average” is used because the radiation is in equilibrium with its source on average over some macroscopic interval of time. As we show below, the difference between the (classical) Rayleigh-Jeans prediction [15–17] and the (quantum) prediction by Planck [8–10] depends on how $\langle E \rangle$ is calculated. Before proceeding we estimate the first term on the right-hand side of (217). To this end we note that

$$\left(\begin{array}{c} \text{number of states inside the box} \\ \text{within the interval } (\omega, \omega + d\omega) \end{array} \right) = \frac{dZ}{d\omega} d\omega, \quad (218)$$

where $Z(\omega)$ is the number of standing waves up to ω that can exist in this box. We assume that the allowed frequencies of the radiation propagating in any one direction are spaced evenly. With this in mind, we can write

$$Z(\omega) = \varkappa \left(\frac{\omega}{\omega_{\min, x}} \right) \left(\frac{\omega}{\omega_{\min, y}} \right) \left(\frac{\omega}{\omega_{\min, z}} \right) \quad (219)$$

where $\omega_{\min, x}$ is the minimum frequency of radiation that can propagate in the box in the x -direction; and similarly for $\omega_{\min, y}$ and $\omega_{\min, z}$. This minimum frequency exists because there is the maximum wavelength,

$$\lambda_{\max, x} = 2L_x \quad (220)$$

that can exist between the walls located L_x units apart. (The illustrating Fig. 31 assumes that the wave is zero at the walls, but a similar result can also be obtained for other boundary conditions.) Now, from (214) and (220) we have

$$\omega_{\min, j} = \frac{2\pi c}{2L_j} = \frac{\pi c}{L_j}, \quad (221)$$

with $j = \{x, y, z\}$. From Fig. 31 we can conclude that the next three largest wavelengths are $2L_x/2$, $2L_x/3$, and $2L_x/4$ where two, three, and four semi-periods of the wave fit between the walls, respectively. The corresponding frequencies, in analogy with (221), are $2\omega_{\min, x}$, $3\omega_{\min, x}$, and $4\omega_{\min, x}$. This justifies our assumption that the frequencies of the radiation in a box are spaced evenly. Next, from (219) and (221) one has:

$$Z(\omega) = \varkappa \frac{\omega^3}{(\pi c)^3 / (L_x L_y L_z)} = \varkappa \frac{\omega^3}{(\pi c)^3} L_x L_y L_z. \quad (222)$$

Using the fact that photons have two independent polarizations (that is there are two states per wave vector $\vec{k} = (\omega/c) \hat{n}$ of photons of frequency ω propagating

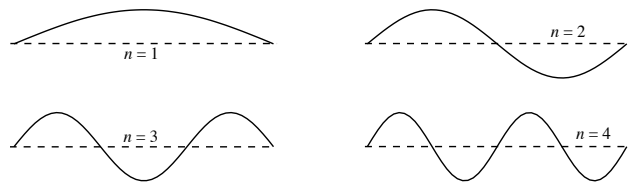


FIG. 31: Standing waves corresponding to $n\lambda/2 = L_x$ for $n = 1, 2, 3, 4$. Abscissa: x -axis of the cavity, bounded by walls at $x = 0$ and $x = L_x$. Ordinate: Electric field strengths of the longest standing waves satisfying the boundary condition $\vec{E} = 0$ at the walls.

in the direction \hat{n} inside the box) it is easily seen that $\varkappa = \pi/3$ [18]. Then according to (222)

$$\left(\begin{array}{c} \text{number of states inside the box} \\ \text{within the interval } (\omega, \omega + d\omega) \end{array} \right) = \frac{\omega^2}{\pi^2 c^3} L_x L_y L_z,$$

and finally

$$\frac{\left(\begin{array}{c} \text{number of states inside the box} \\ \text{within the interval } (\omega, \omega + d\omega) \end{array} \right)}{\text{volume of the box}} = \frac{\omega^2 d\omega}{\pi^2 c^3}. \quad (223)$$

The energy of each standing wave (normal mode) is distributed according to the Maxwell-Boltzmann distribution [19–21]

$$P(E) dE = \frac{e^{-E/kT}}{kT} dE, \quad (224)$$

where $k = 1.38 \times 10^{-23} \text{ J K}^{-1}$ is the Boltzmann’s constant and T the absolute temperature; (224) reflects Boltzmann’s connection between entropy and probability [22]. The classical Rayleigh-Jeans prediction is arrived at assuming E to be a *continuous* variable from 0 to ∞ , i.e. all energies are possible. Then,

$$\langle E \rangle = \frac{\int_0^\infty P(E) E dE}{\int_0^\infty P(E) dE} = \dots = kT. \quad (225)$$

It is worthwhile to point out that $P(E)$ is already normalized, i.e. the denominator of (225) equals one. We are keeping it so that we will be consistent when considering the discrete case later.

Putting all these ingredients together, we arrive at the spectral energy density of radiation in the wavelength interval $d\lambda$ inside the cavity

$$u_\lambda(\lambda, T) d\lambda = \frac{N(\lambda) d\lambda}{V} kT = \frac{8\pi}{\lambda^4} kT d\lambda. \quad (226)$$

The corresponding spectral emittance per unit wavelength interval $d\lambda$ is the so-called “Rayleigh-Jeans formula”

$$B_\lambda(\lambda, T) = \frac{c}{4\pi} u_\lambda(\lambda, T) = \frac{2c}{\lambda^4} kT. \quad (227)$$

This result, based firmly on the theories of electromagnetism and thermodynamics, represents our best attempt

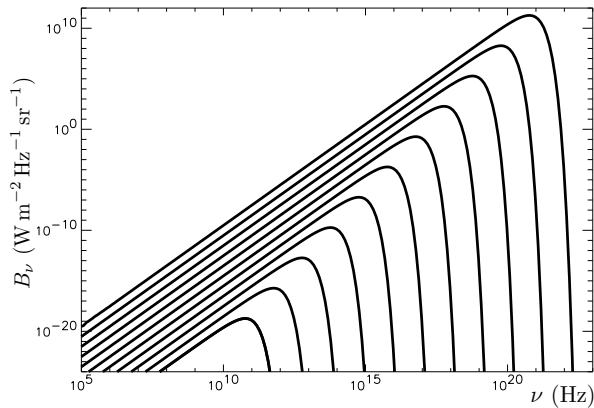


FIG. 32: Planck spectrum of blackbody radiation. Each curve corresponds to a certain absolute thermodynamic temperature value. Shown is the surface brightness for 10^0 K, 10^1 K, 10^2 K, \dots , 10^{10} K.

to apply classical physics to understanding the problem of blackbody radiation. The surface brightness calculated with (227) is in agreement with experimental data at long wavelengths. At short wavelengths, however, the classical theory (which predicts $u_\lambda \rightarrow \infty$ as $\lambda \rightarrow 0$) is absolutely not physical. The surface brightness of the emitted light (radiated energy) must remain finite! The failure of the Rayleigh-Jeans formula at short wavelengths is known as the *ultraviolet catastrophe* and represents a serious problem for classical physics, because the theories of thermodynamics and electromagnetism on which the Rayleigh-Jeans formula is based have been carefully tested in many other circumstances and found to give extremely good agreement with experiment. It is apparent in the case of blackbody radiation that the classical theories do not work, and that a new kind of physical theory is needed.

In 1901, Planck proposed a solution to this problem by considering a different way to calculate $\langle E \rangle$. In Planck's theory, the energy is *not* a continuous variable. Each oscillator in the cavity can emit or absorb energy only in quantities that are integer multiples of a certain basic quantity ΔE ,

$$E_n = n \Delta E, \quad \text{with } n = 0, 1, 2, 3, \dots \quad (228)$$

Furthermore, the discrete energy increment is proportional to the frequency of the oscillator

$$\Delta E = h\nu, \quad (229)$$

where $h = 6.626 \times 10^{-34}$ J s = 4.136×10^{-15} eV s is the Planck's constant. The average energy of an oscillator is then given by the discrete sum

$$\langle E \rangle = \frac{\sum_{n=0}^{\infty} E_n P(E_n)}{\sum_{n=0}^{\infty} P(E_n)} = \dots = \frac{hc/\lambda}{e^{hc/(\lambda kT)} - 1}. \quad (230)$$

Multiplying this result by the number of oscillators per unit volume in the interval $d\lambda$ given by (223), we obtain the spectral emittance distribution function of the

radiation inside the cavity

$$B_\lambda(\lambda, T) = \frac{2c}{\lambda^4} \langle E \rangle = \frac{2c}{\lambda^4} \frac{hc/\lambda}{e^{hc/(\lambda kT)} - 1}. \quad (231)$$

Planck's spectra for a range of values of T is shown in Fig. 32. Note that Planck's result reduces to the classical limit if $h = \Delta E = 0$. For $h \rightarrow 0$, the exponential in (231) can be expanded using $e^x \approx 1 + x + \dots$ for $x \ll 1$, where $x = hc/(\lambda kT)$. Then

$$e^{hc/(\lambda kT)} - 1 \approx \frac{hc}{\lambda kT} \quad (232)$$

and so

$$\langle E \rangle = \frac{hc/\lambda}{e^{hc/(\lambda kT)} - 1} = kT. \quad (233)$$

For long wavelength (low photon energy) again we use (232) to show

$$\lim_{\lambda \rightarrow \infty} u_\lambda \rightarrow \frac{8\pi}{\lambda^4} kT, \quad (234)$$

which is the Rayleigh-Jeans formula (226). Finally, in the quantum regime $\lambda \rightarrow 0$ (i.e. high photon energy) $e^{hc/(\lambda kT)} \rightarrow \infty$ exponentially faster than $\lambda^5 \rightarrow 0$ so

$$\lim_{\lambda \rightarrow 0} \frac{1}{\lambda^5 (e^{hc/(\lambda kT)} - 1)} \rightarrow 0. \quad (235)$$

There is no ultraviolet catastrophe in the quantum limit.

EXERCISE 9.1 Calculate the integrals in (225) and the sum of the series in (230).

EXERCISE 9.2 Verify that if you integrate (231) over all wavelengths and solid angles you can reproduce Stefan-Boltzmann law (207).

EXERCISE 9.3 Show that Wien's displacement law can be derived by determining the maximum of (231).

EXERCISE 9.4 (i) Stars behave approximately like blackbodies. Use Wien's displacement formula to obtain a rough estimate of the surface temperature of the Sun, assuming that it is an ideal blackbody as suggested by the ASTM data shown in Fig. 33 and that evolution on Earth worked well (i.e., that the human eye uses optimal the light from the Sun). (ii) The solar constant (radiant flux at the surface of the Earth) is about 1.365 kW/m^2 . Find the effective surface temperature of the Sun (iii) Assuming that the surface of Neptune and the thermodynamics of its atmosphere are similar to those of the Earth estimate the surface temperature of Neptune. Neglect any possible internal source of heat. [Hint: Astronomical data which may be helpful: radius of Sun $R_\odot = 7 \times 10^5 \text{ km}$; radius of Neptune $R_N = 2.2 \times 10^4 \text{ km}$; mean Sun-Earth distance $r_{SE} = 1 \text{ AU} = 1.5 \times 10^8 \text{ km}$;

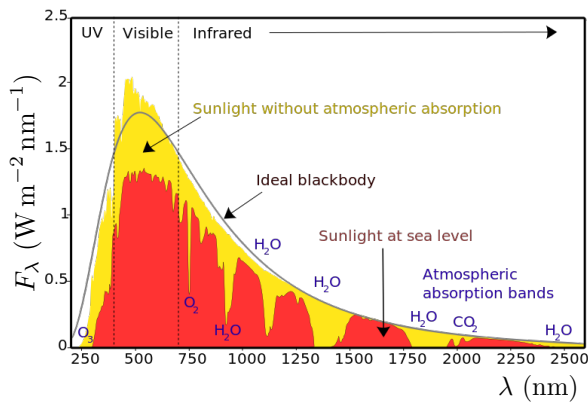


FIG. 33: Solar energy incident at Earth's atmosphere and surface. The yellow band is the radiation incident at the top of the atmosphere, while the red band is the radiation at Earth's surface, diminished by the atmospheric absorbers shown. The radiation approximates a blackbody curve. These data are from the American Society for Testing and Materials (ASTM) Terrestrial Reference Spectra.

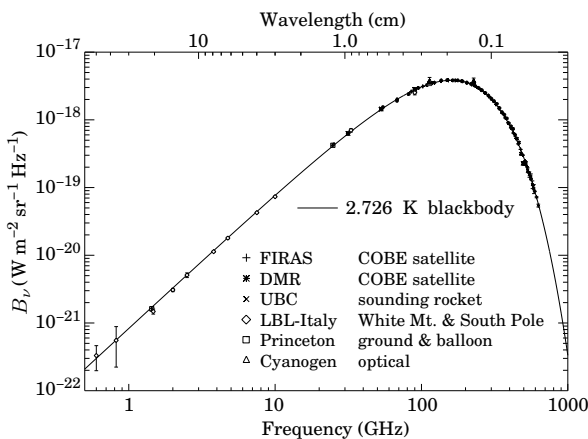


FIG. 34: The CMB blackbody spectrum as confirmed by measurements over a broad range of wavelengths [23].

mean Sun-Neptune distance $r_{\text{SN}} = 4.5 \times 10^9$ km.]

EXERCISE 9.5 A compilation of experimental measurements of the CMB reveals an accurate blackbody spectrum, see Fig. 34. Actually, according to the FIRAS (Far InfraRed Absolute Spectrometer) instrument aboard the COBE (*Cosmic Background Explorer*) satellite, which measured a temperature of $T = 2.726 \pm 0.010$ K, the CMB is the most perfect blackbody ever seen [24]. (i) Write down an integral which determines how many photons per cubic centimeter are contained in the CMB. Estimate the result within an order of magnitude and show that it agrees with the fiducial value adopted in exercise 8.9. (ii) Convince yourself that the average energy of a CMB photon is $\langle E_\gamma^{\text{CMB}} \rangle \approx 6 \times 10^{-4}$ eV (iii) Show that a freely expanding blackbody radiation remains described by the Planck formula, but

with a temperature that drops in proportion to the scale expansion.

B. Photoelectric effect

The success of Planck's idea immediately raises the question: why is it that oscillators in the walls can only emit and absorb energies in multiples of $h\nu$? The reason for this was supplied by Einstein in 1905, in connection with his explanation of the photoelectric effect: light is composed of particles called photons, and each photon has an energy $E_\gamma = h\nu$ [25].

The photoelectric effect is the observation that a beam of light can knock electrons out of the surface of a metal. The electrons emitted from the surface are called photoelectrons. The phenomenon was discovered by Hertz [4] and further studied by Lenard [6, 7]. What is surprising about the photoelectric effect is that the energy of the photoelectrons is independent of the intensity of the incident light. If the frequency of the light is swept, we find that there is a minimum frequency ν_0 below which no electrons are emitted. The energy $\varphi = h\nu_0$ corresponding to this frequency is called the work function of the surface. In experiments used to investigate the photoelectric effect, the photoelectrons are usually collected on a metal plate detector, which forms part of an electrical circuit; see Fig. 35. The current measured in the circuit is proportional to the number of electrons striking the detector plate. To measure the kinetic energy of the electrons, we can apply a static potential V (often known as a retarding potential) to the plate. Only electrons with a kinetic energy greater than eV , where e is the electron charge, will reach the plate; any electrons with a kinetic energy less than eV will be repelled, and will not be detected. The stopping potential, V_0 , is the retarding potential at which no more electrons are detected, and tells us the kinetic energy of the fastest electrons, i.e. $K_{\text{max}} = eV_0$. One of the important early observations about the photoelectric effect is that for a particular metal surface and frequency of light, the stopping potential is independent of the light intensity.

We can now explain why classical electromagnetism fails to explain the photoelectric effect, and why all the problems are solved by quantum mechanics. (i) In classical electromagnetism, increasing the intensity of a beam of light increases the amplitude of the oscillating electric field \vec{E} . Since the force that the incident beam exerts on an electron is $e\vec{E}$, the theory would predict that the energy of the photoelectrons would increase with increasing light intensity. However, this is not the case: V_0 is independent of light intensity. In quantum mechanics, however, doubling the intensity simply doubles the number of photons, it does not change their energy. If the photoelectric effect is interpreted as a collision between a single photon and an electron in the metal surface, then the total number of photons striking the surface is immaterial in determining the energy of the ejected

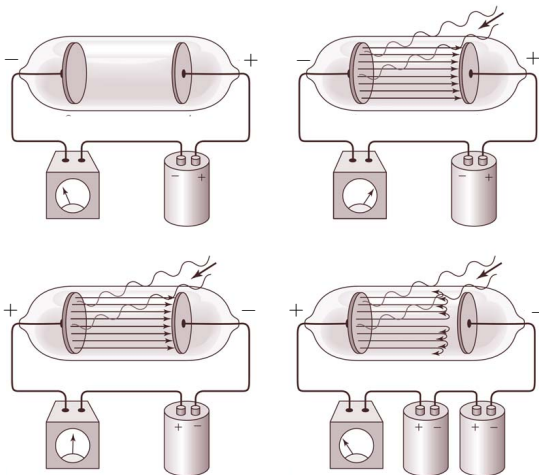


FIG. 35: Schematic diagram of the apparatus used to study the photoelectric effect.

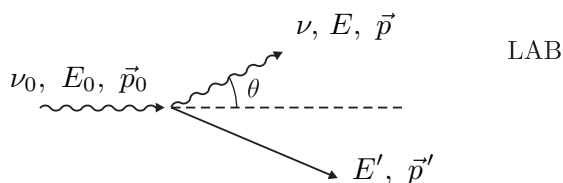


FIG. 36: The geometry of Compton scattering.

electron. (ii) According to classical electromagnetism, as long as the intensity of the light is large enough, the photoelectric effect should occur at any frequency, a direct contradiction of the experimental evidence, which shows a clear cutoff frequency below which no electrons are ejected. In quantum mechanics the frequency of the light determines the photon energy, and since it takes a certain minimum amount of energy to knock an electron out of the surface (defined by the work function φ of the surface), photons with energy $h\nu < \varphi$ simply do not have enough energy to achieve this. (iii) In classical electromagnetism the energy imparted to the electron must somehow be “soaked up” from the incident wave. Since this takes some time, if very weak light is used it would be expected that there should be a measurable time delay between the light striking the surface and the electron being emitted. This has never been observed: measurements have shown that if there is a time lag, it is less than 10^{-9} s. In quantum mechanics, the photoelectric effect is viewed as a single collisional event and no time delay is predicted. (iv) When K_{\max} is plotted as a function of frequency $\nu > \nu_0$, the experimental data fit a straight line, whose slope equals Planck’s constant: $K_{\max} = h\nu - \varphi$ [26, 27].

Einstein’s theory of the photon composition of light immediately explains Planck’s condition that the energy of electromagnetic radiation of frequency ν , in a box, is restricted to the values $E = nh\nu$. Since there can only

be an integer number of photons n at any given frequency, each of energy $h\nu$, the energy of the field at that frequency can only be $nh\nu$. Planck’s restriction on energies is thereby explained in a very natural, appealing way.

EXERCISE 9.6 For a typical case of photoemission from sodium, show that classical theory predicts that: (i) K_{\max} depends on the incident light intensity I ; (ii) K_{\max} does not depend on the frequency of the incident light; (iii) there is a long time lag between the start of illumination and the beginning of the photocurrent. The work function for sodium is $\varphi = 2.28$ eV and an absorbed power per unit area of 1.00×10^{-7} mW/cm² produces a measurable photocurrent in sodium.

EXERCISE 9.7 A metal surface has a photoelectric cutoff wavelength of 325.6 nm. It is illuminated with light of wavelength 259.8 nm. What is the stopping potential?

EXERCISE 9.8 Another effect that revealed the quantized nature of radiation is the (elastic) scattering of light on particles shown in Fig. 36, called the Compton effect [28]. (i) Using conservation of energy and momentum, derive the Compton shift formula,

$$\lambda - \lambda_0 = \lambda_c(1 - \cos \theta), \quad (236)$$

where λ_c is the Compton wavelength of the particle, which is equivalent to the wavelength of a photon whose energy is the same as the mass of the particle. (ii) In the experiment by Compton, X-rays are scattered by nearly free electrons ($\lambda_c = 2.43 \times 10^{-10}$ cm) in carbon (graphite). (Although no scattering target contains actual “free” electrons, the outer or valence electrons in many materials are very weakly attached to the atom and behave like nearly free electrons. The binding energies of these electrons in the atom are so small compared with the energies of the incident X-ray photons that they can be regarded as nearly “free” electrons.) A movable detector measured the energy of the scattered X rays at various angles θ . At each angle, two peaks appear, corresponding to scattered X-ray photons with two different energies or wavelengths. The wavelength of one peak does not change as the angle is varied; this peak corresponds to scattering that involves “inner” electrons of the atom, which are more tightly bound to the atom so that the photon can scatter with no loss of energy. The wavelength of the other peak, however, varies strongly with angle. This variation is exactly as the Compton formula predicts. Show that for a maximal scattering angle the fractional change $\Delta\lambda/\lambda_0$ is about 7%. In summary, the particle character of light is confirmed in Compton’s experiment and we assign the energy of $E = \hbar\omega$ and the momentum $\vec{p} = \hbar\vec{k}$ to the (undivisible) photon, where $\hbar = h/(2\pi)$. The Compton shift formula (236) reveals a proportionality to \hbar , a quantum mechanical property that is confirmed by experiment. Classically no change of the wavelength is to be expected.

C. Line spectra of atoms

When a gaseous element is energetically excited so that it emits radiation, the emitted radiation – when passed through a prism – is found to consist of a series of well-defined lines (called the spectrum of the element), each associated with a different wavelength. When the excitation is carried out by heating to incandescence in a flame, the spectra are found to be associated with neutral atoms, but if the excitation is more energetic, e.g., due to a high-voltage electrical discharge or spark, the resulting spectrum (spark spectrum) is found to be associated with ionized atoms.

When hydrogen in a glass tube is excited by a 5000 V electrical discharge, four lines are observed in the visible part of the emission spectrum: red at 656.3 nm, blue-green at 486.1 nm, blue violet at 434.1 nm and violet at 410.2 nm. These can be explained by Balmer's empirical formula $\lambda = 364.56 \frac{n^2}{(n^2 - 4)} \text{ nm}$, where n is a variable integer that takes on the values $n = 3, 4, 5, \dots$ [29]. Balmer's formula was generalized subsequently by Rydberg [30, 31] and Ritz [32] to accommodate newly discovered spectral lines in the ultraviolet [33, 34] and infrared [35] regions; see Fig. 37. The modern form, called the Balmer-Rydberg-Ritz formula, gives the reciprocal wavelength as

$$\frac{1}{\lambda} = \mathcal{R} \left(\frac{1}{n_1^2} - \frac{1}{n_2^2} \right), \quad \text{for } n_2 > n_1, \quad (237)$$

where n_1 and n_2 are integers and \mathcal{R} , the Rydberg constant, is the same for all series of spectral lines of the same element and varies only slightly, and in a regular way, from element to element. For hydrogen, the value of \mathcal{R} is $\mathcal{R}_H = 1.096776 \times 10^7 \text{ m}^{-1}$. The Balmer series of spectral lines in the visible region correspond to the values $n_1 = 2$, $n_2 = 3, 4, 5$ and 6. The lines with $n_1 = 1$ in the ultraviolet make up the Lyman series. The line with $n_2 = 2$, designated the Lyman alpha, has the longest wavelength (lowest wavenumber) in this series, with $1/\lambda = 82.258 \text{ cm}^{-1}$ or $\lambda = 121.57 \text{ nm}$. For very heavy elements, \mathcal{R} approaches the value of $\mathcal{R}_\infty = 1.097373 \times 10^7 \text{ m}^{-1}$.

Many attempts were made to construct a model of the atom that yielded the Balmer-Rydberg-Ritz formula. It was known that an atom was about 10^{-10} m in diameter, that it contained electrons much lighter than the atom, and that it was electrically neutral. Thomson attempted various models consisting of electrons embedded in a fluid that contained most of the mass of the atom and had enough positive charge to make the atom electrically neutral [37]. He then searched for configurations that were stable and had normal modes of vibration corresponding to the known frequencies of the spectral lines. One difficulty with all such models was that electrostatic forces alone cannot produce stable equilibrium. The nuclear model proposed by Rutherford pictures the atom as a heavy, positively-charged nucleus, around which much lighter, negatively-charged electrons circulate, much like

planets in the Solar system [38]. This model is however completely untenable from the standpoint of classical electromagnetic theory, for an accelerating electron (circular motion represents an acceleration) should radiate away its energy. In fact, a hydrogen atom should exist for no longer than $5 \times 10^{-11} \text{ s}$, time enough for the electron's death spiral into the nucleus. Bohr sought to avoid such an atomic catastrophe by proposing that certain orbits of the electron around the nucleus could be exempted from classical electrodynamics and remain stable [39, 40]. The Bohr model was quantitatively successful for the hydrogen atom, as we shall now show.

In applications to atomic phenomena we prefer to use the Gaussian unit system. We recall that the attraction between two opposite charges, such as the electron and proton, is given by Coulomb's law

$$\vec{F} = \frac{e^2}{r^2} \hat{i}_r, \quad (238)$$

where \hat{i}_r denotes the unit radial vector [41]. Since the Coulomb attraction is a central force (dependent only on r), the potential energy is related by

$$|\vec{F}| = -\frac{dV(r)}{dr}. \quad (239)$$

We find therefore, for the mutual potential energy of a proton and electron,

$$V(r) = -\frac{e^2}{r}. \quad (240)$$

Bohr considered an electron in a circular orbit of radius r around the proton. To remain in this orbit, the electron must be experiencing a centripetal acceleration

$$a = v^2/r \quad (241)$$

where v is the speed of the electron. Using (239) and (241) in Newton's second law, we find

$$\frac{e^2}{r} = \frac{m_e v^2}{r} \quad (242)$$

where m_e is the mass of the electron. For simplicity, we assume that the proton mass is infinite so that the proton's position remains fixed (actually $m_p \approx 1836m_e$). The energy of the hydrogen atom is the sum of the kinetic and potential energies:

$$E = K + V = \frac{1}{2}m_e v^2 - \frac{e^2}{r} \quad (243)$$

Using (242), we see that

$$K = -\frac{1}{2}V \quad \text{and} \quad E = \frac{1}{2}V = -K. \quad (244)$$

This is the form of the virial theorem for a force law varying as r^{-2} . Note that the energy of a bound atom

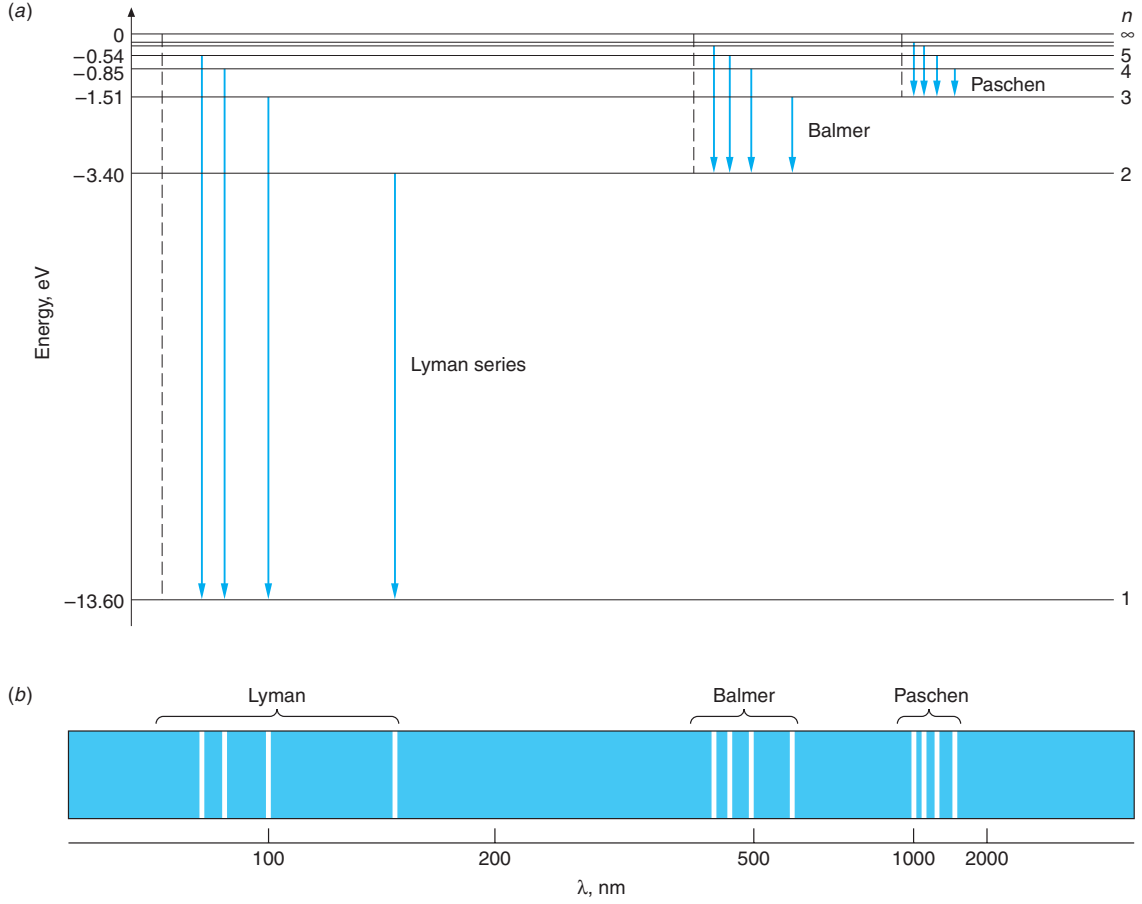


FIG. 37: (a) Energy-level diagram for hydrogen showing the seven lowest stationary states and the four lowest energy transitions each for the Lyman, Balmer, and Paschen series. There are an infinite number of levels. The dashed line shown for each series is the series limit, corresponding to the energy n that would be radiated by an electron at rest far from the nucleus ($n_2 \rightarrow \infty$) in a transition to the state n_1 for that series. The horizontal spacing between the transitions shown for each series is proportional to the wavelength spacing between the lines of the spectrum. (b) The spectral lines corresponding to the transitions shown for the three series. Notice the regularities within each series, particularly the short-wavelength limit and the successively smaller separation between adjacent lines as the limit is approached. The wavelength scale in the diagram is not linear [36].

is negative, since it is lower than the energy of the separated electron and proton, which is taken to be zero. For further progress, we need some restriction on the possible values of r or v . This is where we can introduce the quantization of angular momentum $\vec{L} = \vec{r} \times \vec{p}$. Since \vec{p} is perpendicular to \vec{r} (see Fig. 38), we can write simply

$$L = rp = mvr. \quad (245)$$

Using (244), we find also that

$$r = \frac{L^2}{me^2} \quad (246)$$

We introduce angular momentum quantization, writing

$$L = n\hbar, \quad \text{with } n = 1, 2, \dots \quad (247)$$

excluding $n = 0$, since the electron would then not be in a circular orbit. The allowed orbital radii are then given

by

$$r_n = n^2 a_0 \quad (248)$$

where

$$a_0 \equiv \frac{\hbar^2}{m_e e^2} = 5.29 \times 10^{-11} \text{ m} \simeq 0.529 \text{ \AA}, \quad (249)$$

which is known as the Bohr radius. The corresponding energy is

$$E_n = -\frac{e^2}{2a_0 n^2} = -\frac{m_e e^4}{2\hbar^2 n^2}, \quad n = 1, 2, \dots \quad (250)$$

Balmer-Rydberg-Ritz formula (237) can now be deduced from the Bohr model. We have

$$\frac{hc}{\lambda} = E_{n_2} - E_{n_1} = \frac{2\pi^2 m_e e^4}{h^2} \left(\frac{1}{n_1^2} - \frac{1}{n_2^2} \right) \quad (251)$$

and the Rydberg constant can be identified as

$$\mathcal{R} = \frac{2\pi m_e e^4}{h^3 c} \approx 1.09737 \times 10^7 \text{ m}^{-1}. \quad (252)$$

The slight discrepancy with the experimental value for hydrogen is due to the finite proton mass.

The Bohr model can be readily extended to hydrogen-like ions, systems in which a single electron orbits a nucleus of arbitrary atomic number Z . Thus $Z = 1$ for hydrogen, $Z = 2$ for He^+ , $Z = 3$ for Li^{++} , and so on. The Coulomb potential (240) generalizes to

$$V(r) = -\frac{Ze^2}{r}, \quad (253)$$

the radius of the orbit (248) becomes

$$r_n = \frac{n^2 a_0}{Z}, \quad (254)$$

and the energy (250) becomes

$$E_n = -\frac{Z^2 e^2}{2a_0 n^2}. \quad (255)$$

Wilson [43] and Sommerfeld [44] generalized Bohr's formula for the allowed orbits to

$$\oint p \, dr = nh, \quad \text{with } n = 1, 2, \dots \quad (256)$$

The Sommerfeld-Wilson quantum condition (23) reduce to Bohr's results for circular orbits, but allow, in addition, elliptical orbits along which the momentum p is variable. According to Kepler's first law of planetary motion, the orbits of planets are ellipses with the Sun at one focus. Some of these generalized orbits are drawn schematically in Fig. 39. The lowest energy state $n = 1$ is still a circular orbit. However, $n = 2$ allows an elliptical orbit in addition to the circular one; $n = 3$ has three possible orbits, and so on. The energy still depends on n alone, so that the elliptical orbits represent degenerate states.

The Bohr model was a major step in the development of quantum mechanics. It introduced the quantization of atomic energy levels (allowing transition between states n_1 and n_2 via absorption or emission of photons [45]) and gave quantitative agreement with the atomic hydrogen spectrum. With the Sommerfeld-Wilson generalization, it accounted as well for the degeneracy of hydrogen energy levels. The Bohr model was able to sidestep the atomic catastrophe. However, as we will discuss next, the assumption of well-defined electron orbits around a nucleus is completely contrary to the basic premises of quantum mechanics. Another flaw in the Bohr picture is that the angular momenta are all too large by one unit, for example, the ground state actually has zero orbital angular momentum (rather than \hbar).

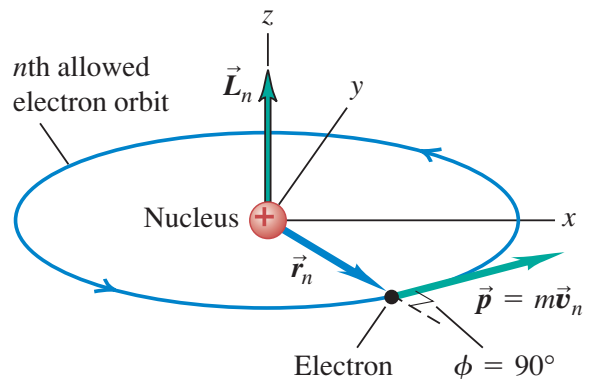


FIG. 38: Angular momentum of an electron in a circular orbit around an atomic nucleus. Angular momentum \vec{L}_n of orbiting electron is perpendicular to plane of orbit and has magnitude $L = mv_n r_n$ [42].

EXERCISE 9.9 According to general principles of classical electrodynamics, accelerated charged particles always radiate electromagnetic waves. This is the basic rule upon which all radiation sources are based. At the end of the last century Larmor calculated the total power radiated by an accelerated non-relativistic electron ($v \ll c$) [46]. His well known result is (in Gaussian units)

$$P = \frac{2}{3} \frac{q^2 a^2}{c^3}, \quad (257)$$

where a is the acceleration and q is the charge; see Appendix C. In the Rutherford model of the hydrogen atom's ground state, the electron moves in a circular orbit of radius $a_0 = 0.529 \text{ \AA}$ around the proton, which is assumed to be rigidly fixed in space. Since the electron is accelerating, a classical analysis suggests that it will continuously radiate energy, and therefore the radius of the orbit would shrink with time. Assuming that the electron is always in a nearly circular orbit and that the rate of radiation of energy is sufficiently well approximated by classical, nonrelativistic electrodynamics, how long is the fall time of the electron, i.e., the time for the electron to spiral into the origin?

D. Wave-particle duality and uncertainty principle

In view of the particle properties for light waves – photons – de Broglie ventured to consider the reverse phenomenon, he proposed to assign wave properties to matter, which we will formulate here in the following way: To every particle with mass m and momentum \vec{p} we associate

$$\lambda = h/|\vec{p}|, \quad (258)$$

called the de Broglie wavelength of the particle [47, 48]. The above statement can be easily understood when assigning energy and momentum to matter in (reversed)

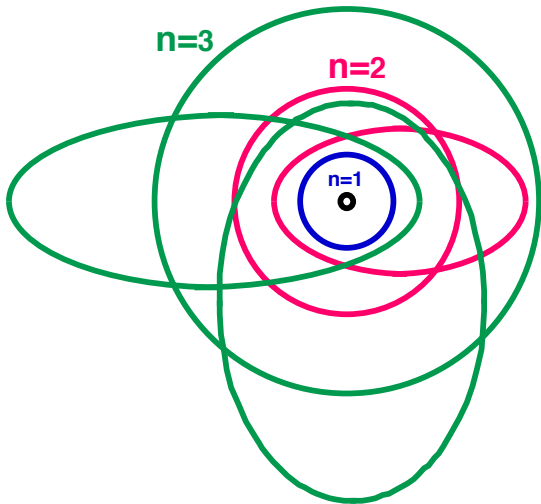


FIG. 39: Bohr-Sommerfeld-Wilson orbits for $n = 1, 2, 3$ (not to scale).

analogy to photons

$$E = \hbar\omega \quad \text{and} \quad |\vec{p}| = \hbar|\vec{k}| = h/\lambda. \quad (259)$$

The definitive evidence for the wave nature of light was deduced from the double-slit experiment discussed in Sec. III [49]. In principle, it should be possible to do double-slit experiments with particles and thereby directly observe their wavelike behavior. However, the technological difficulties of producing double slits for particles are formidable, and such experiments did not become possible until long after the time de Broglie hypothesized the wave-particle duality. The first double-slit experiment with electrons was done in 1961 [50].

Perhaps the best way to crystallize our ideas about the wave-particle duality is to consider the “simple” double-slit experiment for neutrons sketched in Fig. 40. A parallel beam of neutrons falls on a double slit, which has individual openings much smaller than the center-to-center distance between the two slits d . At a distance from the slits $D \gg d$ is a neutron detector capable of detecting individual neutrons. It is important to note that the detector always registers discrete particles localized in space and time. This can be achieved if the neutron source is weak enough. If the detector collects neutrons at different positions for a long enough time, a typical wave interference pattern for the counts per minute or probability of arrival of electrons is found. If one imagines a single neutron to produce in-phase “wavelets” at the slits, standard wave theory can be used to find the angular separation θ of the central probability maximum from its neighboring minimum. As we discuss now in detail, the similarity of the observed intensity pattern with the double-slit interference pattern for light is striking.

In a nuclear reactor, the neutrons set free by the fission process have energies around 2 MeV, which corre-

sponds to speeds around 20000 km/s. These neutrons are slowed down to a room-temperature “thermal” energy by collisions with the moderator nuclei. Their energy distribution is thus Maxwellian with a temperature around 300 K. Beams of neutrons with larger wavelengths are available at the Institute Laue-Langevin (ILL) high-flux reactor as emerging from a cold source. The ILL cold source neutron moderator is liquid deuterium at temperature ≈ 25 K (average $K \approx kT \approx 2.5 \times 10^{-2}$ eV). The neutron beam from the source enters a prism refraction monochromator system. Because of the dispersive properties of the prism medium (quartz glass), for neutrons the radiation emerging behind the prism is fanned out exhibiting a correlation between wavelength and direction very much analogous to the rainbow-colored light radiation field emerging from a glass prism. In contrast to light, the neutron refractive index is smaller than unity for most materials. As a consequence the prism orientation is reversed as compared to the customary light case. The 20- μm -wide optical bench entrance slit then selects neutrons of a specific wavelength out of the radiation field emerging from the prism. The selected neutron beam, with kinetic energy 2.4×10^{-4} eV and de Broglie wavelength 1.85 nm, passed through a gap of diameter 148 μm in a material that absorbs virtually all of the neutrons incident on it. In the center of the gap was a boron wire (also highly absorptive for neutrons) of diameter 104 μm . The neutrons could pass on either side of the wire through slits of width 22 μm . (Actually, due to difficulties in mounting the boron wire exactly centered, the two resulting slits were slightly unequal, 21.5 μm and 22.3 μm .) The intensity of neutrons that pass through this double slit is observed by sliding another slit across the beam and measuring the intensity of neutrons passing through this “scanning slit.” In Fig. 41 we show the resulting pattern of intensity maxima and minima, which leaves no doubt that interference is occurring and that the neutrons have a corresponding wave nature. The wavelength can be deduced from the slit separation using (26) to obtain the spacing between adjacent maxima, $y_{n+1} - y_n$. Estimating the spacing from Fig. 41 to be about 75 μm , it follows that

$$\lambda = \frac{d (y_{n+1} - y_n)}{D} \simeq 1.89 \text{ nm}, \quad (260)$$

where we have taken $d \simeq 126 \mu\text{m}$. This result agrees very well with the de Broglie wavelength selected for the neutron beam.

Furthermore, as first emphasized by Heisenberg, there are fundamental *uncertainties* in both the position and the momentum of an individual neutron, and these uncertainties are related inseparably [53]. To clarify this point, we adopt the reasoning introduced by Feynman [54]. We use θ_1 to denote the angle between the central maximum and the first minimum, see Fig. 42. Using (26) with $m = 1$, we find that

$$\sin \theta_1 = \lambda/d. \quad (261)$$

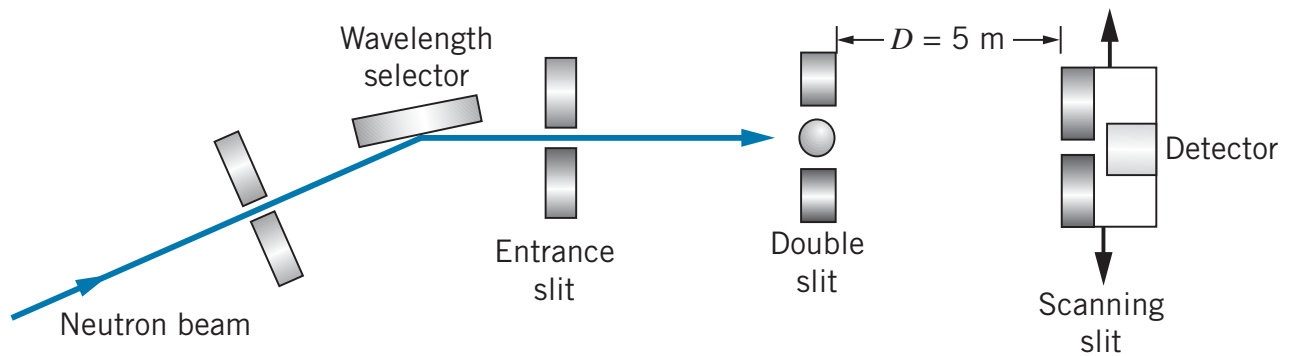


FIG. 40: Double-slit apparatus for neutrons. Thermal neutrons from a reactor are incident on a prism which together with the entrance slit select the energy of the neutrons. After passing through the double slit, the neutrons are counted by the scanning slit assembly, which moves laterally [51].

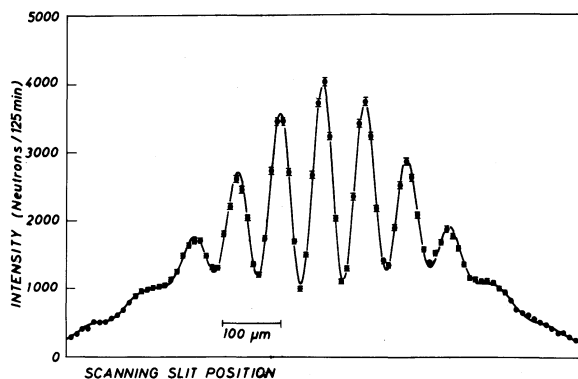


FIG. 41: Measured neutron distribution after interference at a double slit where a boron wire was used to define the two individual slits. The boron wire was opaque for the neutrons used in the experiment. The solid line represents the first-principles theoretical calculation. [52].

A neutron that strikes the screen at the outer edge of the central maximum, at angle θ_1 , must have a component of momentum p_y in the y -direction, as well as a component p_x in the x -direction, despite the fact that initially the beam was directed along the x -axis. From the geometry of the situation the two components are related by $p_y/p_x = \tan \theta_1$. Since θ_1 is small, we may use the approximation $\tan \theta_1 = \theta_1$, and

$$p_y = p_x \theta_1. \quad (262)$$

Substituting (261) into (262) gives

$$p_y = p_x \lambda / d. \quad (263)$$

Now, (263) says that for neutrons that strike the detector within the central maximum (that is, at angles between $-\lambda/d$ and λ/d), the y -component of momentum is spread out over a range from $-p_x \lambda / d$ to $+p_x \lambda / d$. Next, we consider all the neutrons that pass through the slits and strike the screen. Again, they may hit above or below the center of the pattern, so their component p_y may

be positive or negative. However, the symmetry of the interference pattern shows us the average value $\langle p_y \rangle = 0$. There will be an *uncertainty* Δp_y in the y -component of momentum at least as great as $p_x \lambda / d$. That is,

$$\Delta p_y \geq p_x \lambda / d. \quad (264)$$

The narrower the separation between slits d , the broader is the interference pattern and the greater is the uncertainty in the y -component of momentum p_y . The neutron wavelength λ is related to the momentum p_x by (258), which we can rewrite as $\lambda = h/p_x$. Substituting this expression into (264) and simplifying, we find

$$\Delta p_y \geq p_x \frac{h}{p_x d} = \frac{h}{d} \quad (265)$$

What does (265) mean? The center-to-center distance between the two slits $d \equiv \Delta y$ represents an uncertainty in the y -component of the position of a neutron as it passes through the double-slit gap. We don't know exactly where in the gap each neutron passes through. So both the y -position and the y -component of momentum have uncertainties, and the two uncertainties are related by (265), or equivalently

$$\Delta p_y \Delta y \geq h. \quad (266)$$

We can reduce the momentum uncertainty Δp_y only by reducing the width of the interference pattern. To do this, we have to increase the gap width d , which increases the position uncertainty Δy . Conversely, when we decrease the position uncertainty by narrowing the double-slit gap, the interference pattern broadens and the corresponding momentum uncertainty increases.

EXERCISE 9.10 The neutrons produced in a reactor are known as thermal neutrons, because their kinetic energies have been reduced (by collisions) until $K = \frac{3}{2} kT$, where T is room temperature (about 293 K). (i) What is the kinetic energy of such neutrons? (ii) What is their de Broglie wavelength?

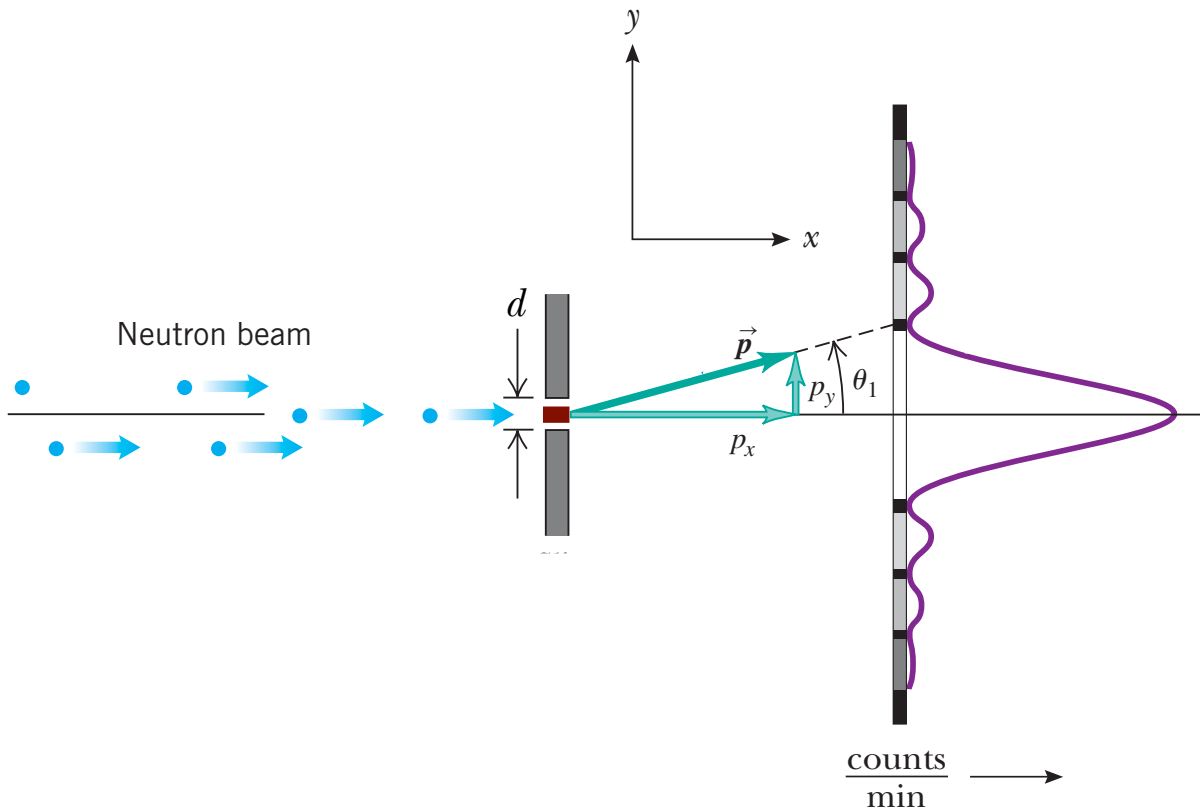


FIG. 42: Interpreting double-slit experiment in terms of neutron momentum [42].

X. SCHRÖDINGER EQUATION

In the Fall of 1925 Schrödinger was invited by Debye to give a talk at a seminar in Zurich on de Broglie's thesis. During the discussion that followed, Debye commented that he thought this approach to wave-particle duality to be somewhat “childish.” After all, said Debye, “to deal properly with waves one had to have a wave equation...” Perhaps stimulated by this comment, Schrödinger left for holiday in the Swiss Alps just before Christmas 1925, and when he returned on 9 January 1926, he had discovered wave mechanics and the equation that governs the evolution of de Broglie waves [55–59].

For non-relativistic quantum physics the basic equation to be solved is the Schrödinger equation. Like Newton's laws, the Schrödinger equation must be written down for a given situation of a quantum particle moving under the influence of some external forces, although it turns out to be easier to frame this in terms of potential energies instead of forces. However, unlike Newton's laws, the Schrödinger equation does not give the trajectory of a particle, but rather the wave function of the quantum system, which carries information about the wave nature of the particle, which allows us to only discuss the probability of finding the particle in different regions of space at a given moment in time. In this section, we introduce the Schrödinger equation, obtain solu-

tions in a few situations, and learn how to interpret these solutions.

A. Motivation and derivation

It is not possible to derive the Schrödinger equation in any rigorous fashion from classical physics. However, it had to come from somewhere, and it is indeed possible to “derive” the Schrödinger equation using somewhat less rigorous means. If we first start by considering a particle with mass m , momentum p_x moving in one dimension in a potential $V(x)$, we can express the total energy as

$$E = \frac{p_x^2}{2m} + V(x), \quad (267)$$

Multiplying both sides of (267) by the wave function $\psi(x, t)$ should not change the equality

$$E \psi(x, t) = \left[\frac{p_x^2}{2m} + V(x) \right] \psi(x, t). \quad (268)$$

Now, recall the de Broglie relations between energy and frequency as well as momentum and wave vector (259). Assume the wave function in (268) is a plane wave traveling in the x direction with a well defined energy and momentum, that is,

$$\psi(x, t) = A_0 e^{i(k_x x - \omega t)}, \quad (269)$$

where $p_x = \hbar k_x$ and $E = \hbar\omega$. If we now combine (268) and (269), using de Broglie's relations, we obtain

$$\hbar\omega A_0 e^{i(k_x x - \omega t)} = E A_0 e^{i(k_x x - \omega t)} \quad (270)$$

and

$$\left[\frac{\hbar^2 k_x^2}{2m} + V(x) \right] A_0 e^{i(k_x x - \omega t)} = \left[\frac{p_x^2}{2m} + V(x) \right] \times A_0 e^{i(k_x x - \omega t)}. \quad (271)$$

From (270) we see that for the equality to hold, the product of energy times the wave function $E\psi(x, t)$ must be equal to the first derivative of the wave function with respect to time multiplied by $i\hbar$, that is

$$E\psi(x, t) = i\hbar \frac{\partial}{\partial t} \psi(x, t). \quad (272)$$

Similarly, by examining (271), we see that the product of momentum times the wave function $p_x \psi(x, t)$ must be equal to the first derivative of the wave function with respect to position x multiplied by $-i\hbar$, that is

$$p_x \psi(x, t) = -i\hbar \frac{\partial}{\partial x} \psi(x, t). \quad (273)$$

Combining (268), (272) and (273), we arrive at the time-dependent Schrödinger equation

$$i\hbar \frac{\partial}{\partial t} \psi(x, t) = \left[-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V(x) \right] \psi(x, t). \quad (274)$$

This is a second-order linear differential equation. The term on the left-hand side of (274) represents the total energy of the particle. The first term on the right-hand side represents the kinetic energy of the particle, while the second term represents the potential energy of the particle. The time-dependent Schrödinger equation has three important properties: (i) it is consistent with energy conservation; (ii) it is linear and singular value, which implies that solutions can be constructed by superposition of two or more independent solutions; (iii) the free-particle solution, $V(x) = 0$, is consistent with a single de Broglie wave.¹⁰

If the potential energy is independent of time, as we have written above, we can separate (274) into a time-independent form using the mathematical technique known as separation of variables. Here, we assume that

our wave function can be written as a product of a temporal and spatial function

$$\psi(x, t) = \psi(x) \chi(t). \quad (275)$$

Substituting (275) into (274) we find

$$i\hbar \psi(x) \dot{\chi}(t) = -\frac{\hbar^2}{2m} \chi(t) [\psi''(x) + V(x) \psi(x)] \quad (276)$$

and so

$$i\hbar \frac{\dot{\chi}(t)}{\chi(t)} = -\frac{\hbar^2}{2m} \left[\frac{\psi''(x)}{\psi(x)} + V(x) \right] \quad (277)$$

Since the left-hand side does not depend on x while the right-hand side does not depend on t it follows that

$$i\hbar \frac{\partial}{\partial t} \chi(t) = E \chi(t) = \hbar\omega \chi(t) \quad (278)$$

and

$$\left[-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V(x) \right] \psi(x) = E \psi(x), \quad (279)$$

where E is the separation constant. (279) is called the time-independent Schrödinger equation. The solution to (278) can be easily verified to be an oscillating complex exponential

$$\chi(t) = e^{-iEt/\hbar} = e^{-i\omega t}. \quad (280)$$

The next steps involve solving (279) for a given potential energy $V(x)$. The techniques involved in solving this equation are similar regardless of the functional form of the potential and can thus be summarized in a set of steps. We assume that the potential $V(x)$ is known and we wish to determine the wave function $\psi(x)$ and its corresponding energy E for that potential. This differential equation problem is known as an eigenvalue problem. There are only particular values of E that satisfy the differential equation, which are called eigenvalues. We will not go into the general theory of solving such equations, but simply go through a few examples. However, before moving on to that, we note three further properties of the solutions of (279). (i) *Continuity*: The solutions $\psi(x)$ to (279) and its first derivative $\psi'(x)$ must be continuous for all values of x (the latter holds for finite potential $V(x)$). (ii) *Normalizable*: The solutions $\psi(x)$ to (279) must be square integrable, i.e. the integral of the modulus squared of the wave function over all space must be a finite constant so that the wave function can be normalized to give $\int |\psi(x)|^2 dx = 1$. (iii) *Linearity*: Given two independent solutions $\psi_1(x)$ and $\psi_2(x)$, we can construct other solutions by taking an appropriate superposition of these $\psi(x) = \alpha_1 \psi_1(x) + \alpha_2 \psi_2(x)$, where $\alpha_i \in \mathbb{C}$, satisfying $|\alpha_1|^2 + |\alpha_2|^2 = 1$ to ensure normalization.

¹⁰ An *eigenvalue* and *eigenvector* of a square matrix A are a scalar λ and a nonzero vector \vec{x} so that $A\vec{x} = \lambda\vec{x}$. A singular value and pair of singular vectors of a square or rectangular matrix A are a nonnegative scalar σ and two nonzero vectors \vec{u} and \vec{v} so that $A\vec{v} = \sigma\vec{u}$ and $A^\dagger\vec{u} = \sigma\vec{v}$. The superscript on A^\dagger stands for Hermitian adjoint and denotes the complex conjugate transpose of a complex matrix. If the matrix is real, then A^T denotes the same matrix.

B. Expectation value, observables, and operators

In quantum mechanics, a probability amplitude ψ is a complex function used to describe the behaviour of systems. The modulus squared of this quantity represents a probability or probability density. Probability amplitudes provide a relationship between the wave function (or, more generally, a quantum state vector) of a system and the results of observations of that system, a link first proposed by Born [60]. Indeed, the wave function gives the probability density (probability per unit length in one dimension)

$$P(x) dx = |\psi(x)|^2 dx. \quad (281)$$

This interpretation helps us understand the continuity constraint on the wave function. We do not want the probability of the particle to be zero at a point x and jump to a non-zero value infinitesimally close by. From this interpretation, we see that we can calculate the probability to find a particle between two points x_1 and x_2 from the wave function $\psi(x)$,

$$P(x_1 < x < x_2) = \int_{x_1}^{x_2} |\psi(x)|^2 dx. \quad (282)$$

Related to this is the concept of normalization of the wave function. We require that the particle must be found somewhere in space, and thus the probability to find the particle between $(-\infty, +\infty)$ should be equal to one, i.e.

$$\int_{-\infty}^{+\infty} |\psi(x)|^2 dx = 1. \quad (283)$$

EXERCISE 10.1 Starting with Schrödinger equation (274) derive the one-dimensional *continuity equation*

$$\frac{\partial \rho}{\partial t} + \frac{\partial j}{\partial x} = 0, \quad (284)$$

where

$$\rho \equiv |\psi(x, t)|^2 \quad (285)$$

and

$$j \equiv -\frac{i\hbar}{2m} \left[\psi^* \frac{\partial \psi}{\partial x} - \frac{\partial \psi^*}{\partial x} \psi \right]. \quad (286)$$

What does this equation imply in terms of conservation of probability?

Since we can no longer speak with certainty about the position of the particle, we can no longer guarantee the outcome of a single measurement of any physical quantity that depends on position. However, we can calculate the most probable outcome for a single measurement (also known as the expectation value), which is equivalent to the average outcome for many measurements. For example, suppose we wish to determine the expected location

of a particle by measuring its x coordinate. Performing a large number of measurements, we find the value x_1 a certain number of times n_1 , x_2 a number of times n_2 , etcetera, and in the usual way, we can calculate the average position

$$\langle x \rangle = \frac{n_1 x_1 + n_2 x_2 + \cdots}{n_1 + n_2 + \cdots} = \frac{\sum_i n_i x_i}{\sum_i n_i}. \quad (287)$$

Here we use the notation $\langle x \rangle$ to represent the average value of the quantity within brackets. The number of times n_i that we measure each position x_i is proportional to the probability $P(x_i) dx$ to find the particle in the interval dx at x_i . Making this substitution and changing sums to integrals, we have

$$\langle x \rangle = \frac{\int_{-\infty}^{+\infty} P(x) x dx}{\int_{-\infty}^{+\infty} P(x) dx}, \quad (288)$$

and thus

$$\langle x \rangle = \int_{-\infty}^{+\infty} x |\psi(x)|^2 dx, \quad (289)$$

where in the last step we assume that the wave function is normalized so that the integral in the denominator of (288) is equal to one. The expectation value of any function of x can be found in a similar way, by replacing x with $f(x)$ in (289)

$$\langle f(x) \rangle = \int_{-\infty}^{+\infty} f(x) |\psi(x)|^2 dx. \quad (290)$$

As a common application of (290) we can calculate the variance, denote by Δx^2 (or equivalently the standard deviation, which is the square root of the variance $\Delta x = \sqrt{\Delta x^2}$) in the position of a particle. The variance in position is given by

$$\begin{aligned} \Delta x^2 &= \langle (x - \langle x \rangle)^2 \rangle = \langle (x^2 - 2x\langle x \rangle + \langle x \rangle^2) \rangle \\ &= \langle x^2 \rangle - \langle x \rangle^2, \end{aligned} \quad (291)$$

where we used the fact that $\langle x\langle x \rangle \rangle = \langle x \rangle^2$. Hence, we see that we can express the variance as

$$\Delta x^2 = \int_{-\infty}^{+\infty} x^2 |\psi(x)|^2 dx - \left(\int_{-\infty}^{+\infty} x |\psi(x)|^2 dx \right)^2.$$

In the *parlons* of mathematics, square integrable functions (such as wave-functions) are said to form an infinite dimensional vector space, much like the familiar n -dimensional vector spaces. In the Dirac notation, a state vector or wave-function, ψ , is represented as a “ket”, $|\psi\rangle$. Just as we can express any n -dimensional vector in terms of the basis vectors, so we can expand any wave function as a superposition of basis state vectors,

$$|\psi\rangle = \lambda_1 |\psi_1\rangle + \lambda_2 |\psi_2\rangle + \cdots. \quad (292)$$

Alongside the ket, we can define the “bra”, $\langle\psi|$. Together, the bra and ket define the *scalar product*

$$\langle\phi|\psi\rangle \equiv \int_{-\infty}^{+\infty} dx \phi^*(x) \psi(x), \quad (293)$$

from which follows the identity, $\langle\phi|\psi\rangle^* = \langle\psi|\phi\rangle$. As with an n -dimensional vector, the magnitude of the scalar product is limited by the magnitude of the vectors

$$|\langle\psi|\phi\rangle| \leq \sqrt{\langle\psi|\psi\rangle\langle\phi|\phi\rangle}, \quad (294)$$

a relation known as the Schwartz inequality.

An operator \hat{A} is a “mathematical object” that maps one state vector, $|\psi\rangle$, into another, $|\phi\rangle$, i.e. $\hat{A}|\psi\rangle = |\phi\rangle$. If $\hat{A}|\psi\rangle = a|\psi\rangle$, with a real, then $|\psi\rangle$ is said to be an eigenstate (or eigenfunction) of \hat{A} with eigenvalue a . An observable is any particle property that can be measured. The position, momentum, and energy of a particle are observables.¹¹ In quantum mechanics, for any observable A , there is an operator \hat{A} which acts on the wave function so that, if a system is in a state described by $|\psi\rangle$, the expectation value of A is

$$\langle A \rangle = \langle\psi|\hat{A}|\psi\rangle = \int_{-\infty}^{+\infty} dx \psi^*(x) \hat{A}\psi(x). \quad (295)$$

An operator \hat{A} is said to be *linear* if $\hat{A}(c\psi(x)) = c\hat{A}\psi(x)$ and $\hat{A}(\psi(x) + \phi(x)) = \hat{A}\psi(x) + \hat{A}\phi(x)$, where $\psi(x)$ and $\phi(x)$ are any two functions on the domain of \hat{A} and c is a complex number. It is straightforward to check by explicit calculation that the operators \hat{x} and $\hat{p} = -i\hbar\partial/\partial x$ are linear operators. The operator \hat{A}^\dagger is called the *hermitian conjugate* (or adjoint) of \hat{A} if

$$\int_{-\infty}^{+\infty} (\hat{A}^\dagger\phi)^*\psi dx = \int_{-\infty}^{+\infty} \phi^*\hat{A}\psi dx; \quad (296)$$

in bra-ket notation (296) becomes $\langle A^\dagger\phi|\psi\rangle = \langle\phi|A\psi\rangle$. It is easy to show that $(c\hat{A})^\dagger = c^*\hat{A}^\dagger$ and $(\hat{A} + \hat{B})^\dagger = \hat{A}^\dagger + \hat{B}^\dagger$ just from the properties of the dot product.

EXERCISE 10.2 Using (296) show that $(\hat{A}^\dagger)^\dagger = \hat{A}$ and $(\hat{A}\hat{B})^\dagger = \hat{B}^\dagger\hat{A}^\dagger$.

The operator \hat{A} is called hermitian if $\hat{A}^\dagger = \hat{A}$, i.e.

$$\int_{-\infty}^{+\infty} (\hat{A}\phi)^*\psi dx = \int_{-\infty}^{+\infty} \phi^*\hat{A}\psi dx; \quad (297)$$

in bra-ket notation (297) becomes $\langle A\phi|\psi\rangle = \langle\phi|A\psi\rangle$.

EXERCISE 10.3 Convince yourself that \hat{x} and $\hat{p} = -i\hbar\partial/\partial x$ are hermitian operators.

EXERCISE 10.4 A physical variable must have real expectation values (and eigenvalues). By computing the complex conjugate of the expectation value of a physical variable show that every operator corresponding to an observable is hermitian.

Note that operators are associative but not (in general) commutative,

$$\hat{A}\hat{B}|\psi\rangle = \hat{A}(\hat{B}|\psi\rangle) = (\hat{A}\hat{B})|\psi\rangle \neq \hat{B}\hat{A}|\psi\rangle. \quad (298)$$

For example, take $\hat{A} = \hat{x}$ and $\hat{B} = \hat{p}$, then we have

$$(\hat{x}\hat{p} - \hat{p}\hat{x})\psi(x) = -i\hbar \left\{ x \frac{\partial\psi}{\partial x} - \frac{\partial}{\partial x}[x\psi(x)] \right\} \quad (299)$$

and hence by the product rule of differentiation:

$$(\hat{x}\hat{p} - \hat{p}\hat{x})\psi(x) = i\hbar\psi(x). \quad (300)$$

and since this must hold for any differentiable function $\psi(x)$, we can write this as an operator equation:

$$\hat{x}\hat{p} - \hat{p}\hat{x} = i\hbar \quad (301)$$

Therefore, we have shown that the operator product of \hat{x} and \hat{p} is non-commuting. Because combinations of operators of the form $\hat{A}\hat{B} - \hat{B}\hat{A}$ do frequently arise in quantum mechanical calculations, it is customary to use a short-hand notation: $[\hat{A}, \hat{B}] \equiv \hat{A}\hat{B} - \hat{B}\hat{A}$ and this is called the commutator of \hat{A} and \hat{B} (in that order!). If $[\hat{A}, \hat{B}] \neq 0$, then one says that \hat{A} and \hat{B} do not commute, if $[\hat{A}, \hat{B}] = 0$, then \hat{A} and \hat{B} are said to commute with each other. An operator equation of the form $[\hat{A}, \hat{B}] = \text{something}$ is called a commutation relation; e.g. $[\hat{x}, \hat{p}] = i\hbar$.

EXERCISE 10.5 Let \hat{A} and \hat{B} be Hermitian operators (i) Show that $\hat{A}\hat{B} = [\hat{A}, \hat{B}]/2 + \{\hat{A}, \hat{B}\}/2$ where $\{\hat{A}, \hat{B}\} = \hat{A}\hat{B} + \hat{B}\hat{A}$ is the anticommutator. (ii) Show that the anticommutator is Hermitian and the commutator is anti-Hermitian (that is, $[\hat{A}, \hat{B}]^\dagger = -[\hat{A}, \hat{B}]$). (iv) We know that expectation values of Hermitian operators are real. What can you say about the expectation value of an anti-Hermitian operator?

EXERCISE 10.6 We have seen that the Heisenberg uncertainty principle [53] captures the difference between classical and quantum states, and sets a limit on the precision of incompatible quantum measurements. It has been introduced in the early days of quantum mechanics, but its form has evolved with the understanding and formulation of quantum physics throughout the years. In particular, Robertson [61] and Schrödinger [62] presented a lower bound for the product of the dispersion of two non-commuting observables. This lower bound is more

¹¹ By contrast, the wavefunction ψ although clearly indispensable to the quantum description, is not directly measurable and so is not an observable.

general than the one given in (266). Let A and B be Hermitian operators. Define the “uncertainty” in A by the square root of the mean square deviation from the mean:

$$\Delta A = \sqrt{\langle (\hat{A} - \langle A \rangle)^2 \rangle} \quad (302)$$

Show that

$$\Delta A \Delta B \geq \frac{1}{2} |[\hat{A}, \hat{B}]|. \quad (303)$$

Recall that the commutator of \hat{x} and \hat{p}_x is $i\hbar$, so from (303) we have $\Delta x \Delta p_x \geq \hbar/2$. [Hint: Use the Schwarz inequality and the results from exercise 10.1.]

EXERCISE 10.7 Show that $\Delta E \Delta t \geq \hbar/2$, where Δt is the shortest time, during which the average value of a certain quantity is changed by an amount equal to the standard deviation of this quantity [63].

EXERCISE 10.8 It's often the case that we want to find the component of a function parallel to another function. We just take the dot product with the second function, but then we also need to multiply by the second function. A handy notation is $|\psi\rangle\langle\psi|$. This *projector operator* projects onto ψ . Operating on $|\phi\rangle$, we get $|\psi\rangle\langle\psi|\phi\rangle$, which is what we want. Remember $\langle\psi|\phi\rangle$ is just a number and $|\psi\rangle$ is the vector. Similarly, operating on $\langle\phi|$ we get $\langle\phi|\psi\rangle\langle\psi|$ which is the desired expression for the adjoint vector. Suppose you have a complete set of orthonormal basis vectors $|\psi_n\rangle$. What is a compact expression for transforming an arbitrary vector $|\phi\rangle$ into this basis set? (This is much easier to write down than to ask!)

C. Free particle solution

A “free” particle refers to a particle that has no external forces acting upon it, in other words the potential energy is constant V_0 . The state of such a free particle is represented by its wave function $\psi(x)$. Starting with (279), and proposing a solution of the form (this is known as an educated-guess)

$$\psi(x) = Ae^{ikx} \quad (304)$$

we find four possible solutions of the Schrödinger equation that satisfy

$$\begin{aligned} \frac{2m}{\hbar^2}(E - V_0)\psi(x) &= -\frac{\partial^2}{\partial x^2}\psi(x) \\ &= k^2\psi(x). \end{aligned} \quad (305)$$

The values $\pm k$ can take on real or imaginary values depending on the particle's energy and the potential

$$k = \pm \frac{1}{\hbar} \sqrt{2m(E - V_0)}, \quad (E > V_0), \quad (306)$$

with corresponding wave functions of the form

$$\psi(x) = Ae^{ikx} + Be^{-ikx} \quad (307)$$

which represent travelling waves solutions, and

$$i\kappa = \pm i \frac{1}{\hbar} \sqrt{2m(V_0 - E)}, \quad (E < V_0), \quad (308)$$

with corresponding wave functions of the form

$$\psi(x) = Ae^{\kappa x} + Be^{-\kappa x}, \quad (309)$$

which represent exponentially decaying solutions. The allowed energies are given by

$$E = \frac{\hbar^2 k^2}{2m} + V_0. \quad (310)$$

The case in which $E > V_0$ is classically allowed, whereas the situation in which $E < V_0$ is classically forbidden. To understand this, imagine our particle rolling on a potential surface described by $V(x)$. If it has total energy E , it can only exist in a region of space in which $V(x) < E$, and once $V(x) \geq E$, the particle must turnaround (this is the classical turning point). However, in quantum physics, the particle has a non-zero probability to be found in this classically forbidden region. We will see how this manifests itself in another section to allow quantum tunneling, in which a particle can penetrate a barrier and emerge on the other side of the barrier.

For the traveling wave solutions, consider the time evolution of the probability density, $P(x, t)$, given by

$$\begin{aligned} P(x, t) &= \psi^*(x, t)\psi(x, t) = \psi^*(x)e^{i\omega t}\psi(x)e^{-i\omega t} \\ &= \psi^*(x)\psi(x). \end{aligned} \quad (311)$$

This is independent of time. If we consider a particle traveling in only one direction, say the $+x$ direction, then the probability density is

$$P(x, t) = \psi^*(x)\psi(x) = A^*e^{-ikx}Ae^{ikx} = A^*A, \quad (312)$$

which is independent of position. This implies that the particle is equally likely to be anywhere in space. It is completely delocalized. For a superposition of both positive and negative going waves, we have

$$\begin{aligned} P(x, t) &= (Ae^{ikx} + Be^{-ikx})^*(Ae^{ikx} + Be^{-ikx}) \\ &= A^*A + B^*B + 2\Re\{A^*Be^{-2ikx} + B^*Ae^{2ikx}\}, \end{aligned}$$

where $\Re(z)$ gives the real value of z . For real-valued coefficients A and B , this simplifies to

$$P(x, t) = A^2 + B^2 + 2AB \cos(2kx). \quad (313)$$

This is the equation for a standing wave.

D. Step potential

In this section we examine the behavior of a particle initially traveling in a region of space of constant potential suddenly moves into a region of different, but also constant potential. At $x = 0$, we have the transition between potential V_1 and V_2 . The potential can thus be expressed as a piecewise function

$$V(x) = \begin{cases} V_1 & \text{for } x < 0 \\ V_2 & \text{for } x \geq 0 \end{cases} . \quad (314)$$

The particle has fixed energy E . There are two situations of interest, first when $E > V_1$ and $E > V_2$ and second when $E > V_1$ and $E < V_2$.

1. Case1: $E > V_1$ and $E > V_2$

Here we break the problem down into two sections and consider a piece-wise wave function defined on either side of the step. For region 1, $x < 0$, the wave function is given by

$$\psi_1(x) = Ae^{ik_1x} + Be^{-ik_1x}, \quad (315)$$

where A and B are coefficients to be determined, and the wave vector for region 1 is

$$k_1 = \sqrt{2m(E - V_1)}/\hbar. \quad (316)$$

Similarly, in region 2 the wave function is given by

$$\psi_2(x) = Ce^{ik_2x} + De^{-ik_2x}, \quad (317)$$

where C and D are coefficients to be determined, and the wave vector for region 2 is

$$k_2 = \sqrt{2m(E - V_2)}/\hbar. \quad (318)$$

The ratio of wave vectors is thus given by

$$\frac{k_2}{k_1} = \sqrt{\frac{1 - V_2/E}{1 - V_1/E}}. \quad (319)$$

We can set $D = 0$ since we assume the particle initially comes from the $-x$ direction. The A coefficient corresponds to the incident wave, while the B is related to the reflected wave. Now, to determine the remaining coefficients, we use the continuity of the wave function and its first derivative at the origin to give

$$\psi_1(0) = \psi_2(0) \Rightarrow A + B = C, \quad (320)$$

and

$$\psi'_1(0) = \psi'_2(0) \Rightarrow ik_1(A - B) = ik_2C, \quad (321)$$

where the prime implies the first derivative with respect to x . Combining these and eliminating C , we find the ratio of B/A

$$\frac{B}{A} = \frac{k_1 - k_2}{k_1 + k_2} = \frac{1 - k_2/k_1}{1 + k_2/k_1}, \quad (322)$$

which is the reflection coefficient of the barrier (the “coefficient” corresponds to the ratio of amplitudes). The reflectivity of the barrier, corresponding to the ratio of probabilities, or flux, is thus given by

$$R = \left| \frac{B}{A} \right|^2 = \left| \frac{1 - k_2/k_1}{1 + k_2/k_1} \right|^2. \quad (323)$$

Due to the conservation of particle number (or probability depending on how you want to think about the wave function), the transmissivity is simply given by

$$T = 1 - R = 1 - \left| \frac{1 - k_2/k_1}{1 + k_2/k_1} \right|^2. \quad (324)$$

Note that in going from region 1 to region 2, the de Broglie wavelength of the particle with energy E changes and becomes longer for an increased potential step. The specific values of A and B are typically determined by considering the normalization of the wave function. However, since plane wave solutions are infinite in extent, they are not normalizable.

2. Case2: $E > V_1$ and $E < V_2$

In this case, we follow a similar approach to above. However, in region 2, we now have decaying solutions since $E < V_2$. Again we have the following wave function in region 1

$$\psi_1(x) = Ae^{ik_1x} + Be^{-ik_1x}, \quad (325)$$

where A and B are coefficients to be determined, and the wave vector for region 1 is

$$k_1 = \sqrt{2m(E - V_1)}/\hbar. \quad (326)$$

In region 2 the wave function is now given by exponentials

$$\psi_2(x) = Ce^{\kappa_2x} + De^{-\kappa_2x}, \quad (327)$$

where C and D are coefficients to be determined, and κ_2 is given by

$$\kappa_2 = \sqrt{2m(V_2 - E)}/\hbar. \quad (328)$$

We can set $C = 0$ since we cannot have the probability amplitude growing infinitely large as $x \rightarrow \infty$. The A coefficient corresponds to the incident wave, while the B is related to the reflected wave. Now, to determine the remaining coefficients, we use the continuity of the wave function and its first derivative at the origin to give

$$\psi_1(0) = \psi_2(0) \Rightarrow A + B = D, \quad (329)$$

and

$$\psi'_1(0) = \psi'_2(0) \Rightarrow ik_1(A - B) = -\kappa_2D. \quad (330)$$

Combining these and eliminating D , we find the ratio of B/A

$$\frac{B}{A} = \frac{k_1 - i\kappa_2}{k_1 + i\kappa_2}, \quad (331)$$

which is the reflection coefficient of the barrier. The reflectivity of the barrier is thus given by

$$R = \left| \frac{B}{A} \right|^2 = \left(\frac{k_1 - i\kappa_2}{k_1 + i\kappa_2} \right) \left(\frac{k_1 + i\kappa_2}{k_1 - i\kappa_2} \right) = 1. \quad (332)$$

Thus, we see that even though the particle has non-zero probability to penetrate into the classically forbidden region, it will always be reflected (eventually). We could have also obtained the same result in (331) from (323) by allowing $k_2 = -i\kappa$. By allowing k_1 and k_2 to take on complex values, we do not have to consider multiple cases.

Note that the depth at which the particle penetrates into the classically forbidden region is given by the distance from $x = 0$ at which the probability drops by $1/e$,

$$P(\Delta x) = e^{-2\kappa_2 \Delta x} = e^{-1}, \quad (333)$$

which gives

$$\Delta x = \frac{1}{2\kappa_2} = \frac{1}{2} \frac{\hbar}{\sqrt{2m(V_2 - E)}}. \quad (334)$$

E. Potential barrier and tunneling

Another useful example is the scattering of a particle from a potential barrier of width L and height V_0 . The functional form of the potential can be expressed as a piecewise function

$$V(x) = \begin{cases} V_0 & \text{for } -L/2 < x < L/2 \\ 0 & \text{otherwise} \end{cases}. \quad (335)$$

Again, there are two types of behavior that occur. The first, in which the particle energy is greater than the barrier $E > V_0$, in which the particle will have some reflection and some transmission as expected classically. The second situation, in which the particle energy is less than the barrier, is of more interest. Here, the classical prediction would be that the particle should be reflected completely and there is no transmission. There should be zero probability to find the particle on the right hand side of the barrier. However, due to the wave nature of quantum systems, there is a finite, non-zero probability to find the particle on the right-hand side of the barrier, and it will continue to propagate to $+\infty$. This simple model is the precursor to discussing the decay of atomic nuclei as well as other quantum tunneling effects such as those associated with a scanning electron microscope.

On the left-hand and right-hand sides of the barrier, regions 1 and 3, the wave vector is given by

$$k = \sqrt{2mE}/\hbar, \quad (336)$$

while within the barrier, the decay is governed by

$$\kappa = \sqrt{2m(V_0 - E)}/\hbar. \quad (337)$$

Hence, we can express the wave function in each region in terms of either plane waves or decaying exponentials with unknown coefficients to be determined

$$\psi_1(x) = Ae^{ikx} + Be^{-ikx}, \quad (338)$$

$$\psi_2(x) = Ce^{\kappa x} + De^{-\kappa x}, \quad (339)$$

$$\psi_3(x) = Fe^{ikx} + Ge^{ikx}. \quad (340)$$

Assuming that the particle initially starts on the left-hand side of the barrier (i.e. $x \leq -L/2$), then we can set $G = 0$, since there is no way to obtain a solution on the right-hand side propagating in the negative x direction. Next we apply the boundary conditions on the wave function at the edges of the barrier, which leads to the following four equations

$$\begin{aligned} e^{-ikL/2} + \frac{B}{A}e^{ikL/2} &= \frac{C}{A}e^{-\kappa L/2} + \frac{D}{A}e^{\kappa L/2} \\ ik \left(e^{-ikL/2} - \frac{B}{A}e^{ikL/2} \right) &= \kappa \left(\frac{C}{A}e^{-\kappa L/2} - \frac{D}{A}e^{\kappa L/2} \right), \\ ik \left(\frac{F}{A}e^{ikL/2} \right) &= \kappa \left(\frac{C}{A}e^{\kappa L/2} - \frac{D}{A}e^{-\kappa L/2} \right), \\ \frac{F}{A}e^{ikL/2} &= \frac{C}{A}e^{\kappa L/2} + \frac{D}{A}e^{-\kappa L/2}, \end{aligned} \quad (341)$$

where we have divided through by A , since we want to solve for the transmission and reflection coefficients given by the ratios of F/A and B/A respectively. By combining the last two equations in (341), we can solve for C/A and D/A in terms of F/A . For example, $1/\kappa$ times the 3rd equation added to the 4th gives

$$\frac{C}{A} = \frac{F}{A} \frac{e^{ikL/2}e^{-\kappa L/2}}{2\kappa}(\kappa + ik), \quad (342)$$

and subtracting $1/\kappa$ times the 3rd equation from 4th gives

$$\frac{D}{A} = \frac{F}{A} \frac{e^{ikL/2}e^{\kappa L/2}}{2\kappa}(\kappa - ik). \quad (343)$$

We can now solve for the transmission and reflection coefficients by combining the first two equations in (341) and substituting in for C/A and D/A to give

$$\begin{aligned} \frac{F}{A} &= \frac{4ik\kappa e^{-ikL}}{(\kappa + ik)^2 e^{-\kappa L} - (\kappa - ik)^2 e^{\kappa L}} \\ \frac{B}{A} &= \frac{F}{A} \frac{\kappa^2 + k^2}{2ik\kappa} \cosh(\kappa L). \end{aligned} \quad (344)$$

The transmittance $T = |F/A|^2$ is thus given by

$$\begin{aligned} T &= \left| \frac{F}{A} \right|^2 \\ &= \frac{8k^2\kappa^2}{(k^2 + \kappa^2)^2 \cosh(2\kappa L) - (k^4 - \kappa^4 - 6k^2\kappa^2)}. \end{aligned} \quad (345)$$

In the limit of weak transmission, i.e. $\kappa L \ll 1$, the transmittance can be approximated by setting $\cosh(2\kappa L) \approx e^{2\kappa L/2}$, and dropping the smaller terms in the denominator, leading to

$$T \approx \frac{16k^2\kappa^2}{(k^2 + \kappa^2)^2} e^{-2\kappa L}. \quad (346)$$

In the last few sections we found the allowed solutions to the Schrödinger equation for a particle of mass m traveling in free space, and through different potentials – a step and a barrier. We found that the allowed energy and momentum the particle may take on have a continuous range of values, which is in line with our classical notions of energy and momentum for a particle. A key signature of non-classical behavior is the non-zero probability to find the particle in a region of space where it is classically forbidden to go, in which its total energy is less than the potential in that region, i.e. $E < V_0$. This quantum penetration is due to the wave nature of quantum objects and can be shown to be consistent with the uncertainty principle, in which the particle is allowed to penetrate the potential for a finite period of time Δt , which is on the order of $\hbar/\Delta E$ where $\Delta E = V_0 - E$ is the difference between the total particle energy and the potential energy of the barrier. The depth into the barrier can be approximated by multiplying this allowed time by the “particle velocity” which we approximate as $v = p/m = \sqrt{2m(V_0 - E)}/m$. The reason to take $p = \sqrt{2m(V_0 - E)}$ as the value of the momentum is that $K = V_0 - E$ is the maximum value of kinetic energy the particle could have and still not propagate in the potential barrier region, it is also the only relevant energy quantity by dimensional analysis. This gives us a penetration depth from the uncertainty principle of

$$\begin{aligned} \Delta x &= \frac{1}{2} v \Delta t = \frac{1}{2} \frac{\hbar}{V_0 - E} \frac{\sqrt{2m(V_0 - E)}}{m} \\ &= \frac{\hbar}{\sqrt{2m(V_0 - E)}}, \end{aligned} \quad (347)$$

where the factor of $1/2$ arises from the fact that the particle must travel into the forbidden region and return in the time Δt . This is consistent with the result we obtained in (334) (to within a factor of two), and thus in good agreement.

F. Particle in a box

The final example that we will look at is that of a particle confined to an infinite potential well. Here a particle of mass m is trapped (or bound) to the well that has a width L , and finite potential V_0 inside the infinite walls. We want to determine what the allowed energies and wave functions that the particle can have. We choose to center the well (or potential “box”) on the origin of our coordinate axes, so that the walls of the well are located

at $x = -L/2$ and $x = +L/2$. The potential can thus be written as

$$V(x) = \begin{cases} \infty & \text{for } x < -L/2 \\ V_0 & \text{for } -L/2 \leq x \leq L/2 \\ \infty & \text{for } x > L/2 \end{cases}. \quad (348)$$

We again break up the problem into different regions of space as dictated by the boundaries of the potential. In regions 1 and 3 outside the box where the potential is infinite, we must have

$$\psi(x) = 0, \quad x < -L/2 \wedge x > L/2, \quad (349)$$

so that the Schrödinger equation can be satisfied (otherwise the term $V(x)\psi(x)$ would be infinite, which makes no sense). Inside the box, we have a 1-dimensional plane wave again, similar to the situation of the finite potential in Sec. XD,

$$\psi(x) = Ae^{ikx} + Be^{-ikx}, \quad -L/2 \leq x \leq L/2, \quad (350)$$

where the magnitude of the wave vector comes from the Schrödinger equation, in which the energy is

$$E = \frac{\hbar^2 k^2}{2m} + V_0, \quad (351)$$

which gives the square of the wave vector

$$k^2 = \frac{2m(E - V_0)}{\hbar^2}, \quad (352)$$

while the expansion coefficients A and B , and the allowed energy values E are still to be determined. These unknown values of A , B and E are set by the initial conditions of the system and the boundary conditions on the wave function respectively. The boundary conditions for the wave function at the left-hand side of the box, $x = -L/2$ implies that

$$\psi(-L/2) = Ae^{-ikL/2} + Be^{ikL/2} = 0, \quad (353)$$

and similarly on the right hand side of the box $x = L/2$

$$\psi(L/2) = Ae^{ikL/2} + Be^{-ikL/2} = 0, \quad (354)$$

Adding (353) to (354) gives

$$2(A + B) \cos(kL/2) = 0 \quad (355)$$

while subtracting (353) from (354) gives

$$2i(A - B) \sin(kL/2) = 0 \quad (356)$$

Both conditions in (355) and (356) must be met. There are two cases we can consider. First, when $A = B$ (356) is met, and to satisfy (355), we see that the wave vector can only take on discrete values

$$k = \frac{2\pi n_1}{L} + \frac{\pi}{L}, \quad (357)$$

where $n_1 = 0, 1, 2, 3, \dots$. The second case is when $A = -B$ in which (355) is met, and to satisfy (356) the wave vector can only take on discrete values

$$k = \frac{2\pi n_2}{L}, \quad (358)$$

where $n_2 = 1, 2, 3, \dots$. We can consolidate these conditions and rewrite the solutions by noting that both (358) and (357) are satisfied by

$$k = \frac{\pi n}{L}, \quad (359)$$

where $n = 1, 2, 3, \dots$ and the solution to the time-independent Schrödinger equation is

$$\begin{aligned} \psi_n(x) &= A \begin{cases} \cos(n\pi x/L) & \text{for } n \text{ odd} \\ \sin(n\pi x/L) & \text{for } n \text{ even} \end{cases} \\ &= A \sin \left[\frac{n\pi}{L} \left(x + \frac{L}{2} \right) \right]. \end{aligned} \quad (360)$$

Note that this implies that not only is the wave vector quantized, but also the particle momentum and energy

$$p = \hbar k = \hbar \pi n / L, \quad (361)$$

and

$$E = V_0 + \frac{\hbar^2 k^2}{2m} = V_0 + \frac{\hbar^2 \pi^2 n^2}{2mL^2}. \quad (362)$$

The amplitude A can be found by considering the normalization condition of the wave function

$$\begin{aligned} \int_{-\infty}^{+\infty} |\psi_n(x)|^2 dx &= \int_{-L/2}^{+L/2} \left| A \sin \left[\frac{n\pi}{L} \left(x + \frac{L}{2} \right) \right] \right|^2 dx \\ &= |A|^2 \frac{L}{2}, \end{aligned} \quad (363)$$

where we have made use of the fact that the wave function is zero outside the box, and the following definite integral

$$\int_{-L/2}^{+L/2} \left| \sin \left[\frac{n\pi}{L} \left(x + \frac{L}{2} \right) \right] \right|^2 dx = \frac{L}{2}. \quad (364)$$

For the normalization integral in (363) to equal unity, we see that the normalization amplitude must be

$$A = \sqrt{\frac{2}{L}}, \quad (365)$$

since we require $|A|^2 L / 2 = 1$. The solutions for the time independent Schrödinger equation can thus be expressed as

$$\psi_n(x) = \sqrt{\frac{2}{L}} \sin \left[\frac{n\pi}{L} \left(x + \frac{L}{2} \right) \right]. \quad (366)$$

Note that the normalization in (365) can be met for a range of complex amplitudes

$$A = e^{i\phi} \sqrt{\frac{2}{L}}, \quad (367)$$

in which the phase ϕ is arbitrary. This implies that the outcome of a measurement about the particle position, which is proportional to $|\psi(x)|^2$ is invariant under a *global* phase factor. However, as we will see shortly, if there are two or more amplitudes that contribute to a measurement outcome, it is important to keep track of the relative phase between amplitudes, as in the double slit experiment, where the fringe pattern arises from the interference of two probability amplitudes with different phases.

Each solution $\psi_n(x)$, labelled by the quantum number n , has quantized wave number, momentum, and energy. The wave function $\psi_n(x)$ is often referred to as an energy eigenstate with corresponding eigenvalue (or eigenenergy E_n), because it satisfies the eigenvalue problem

$$\hat{H}\psi_n(x) = E_n\psi_n(x) \quad (368)$$

where we have introduced the Hamiltonian operator,

$$\hat{H} = \left[-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V(x) \right], \quad (369)$$

which in this case is the sum of kinetic energy and potential energy operators. The wave vector and momentum scale linearly with the quantum number ($k_n = k_1 n$ and $p_n = p_1 n$), while the energy scales quadratically ($E_n = E_1 n^2$). The ground state, labelled by $n = 1$ has the lowest allowed energy, followed by the first excited state $n = 2$, which has four times the energy.

We often speak of the state of the system, which implies specification of the wave function at some initial time, although this is not always explicitly written. Furthermore, the solutions are orthogonal to one another, that is, they have zero overlap across the entire box

$$\int_{-L/2}^{+L/2} \psi_m^*(x) \psi_n(x) dx = \delta_{mn} \quad (370)$$

where δ_{mn} is the Kronecker delta, which is one for $m = n$ and zero otherwise.

G. Superposition and time dependence

The time dependence of quantum states is governed by the time dependent Schrödinger equation (274). We can separate out the time and spatial degrees of freedom if the potential energy is time-independent ($V(x)$ is independent of time) by using the method of separation of variables, as described in Sec. X A. In this case the temporal evolution of the wave function is described by an oscillating exponential phase factor

$$\psi_n(x, t) = \psi_n e^{-iE_n t / \hbar}, \quad (371)$$

where $\psi_n(x)$ is the solution to the time independent Schrödinger equation with quantum number n and E_n is its corresponding energy eigenvalue. The states, $\psi_n(x)$

are called energy eigenstates and have special properties in terms of their temporal evolution. Owing to the linearity of the Schrödinger equation, a general solution for the “particle in a box” can be expressed as a sum of different solutions

$$\Psi(x, t) = \sum_{n=1}^{\infty} c_n \psi_n(x, t), \quad (372)$$

where the amplitudes c_n that weight the superposition can be complex, and must obey the normalization condition

$$\sum_{n=1}^{\infty} |c_n|^2 = 1. \quad (373)$$

This follows from considering the normalization condition, and the orthogonality of the states as in (370). The set of amplitudes $\{c_n\}$, are determined by the specific initial conditions of the problem. The modulus squared of each coefficient gives the probability to find the particle in that state

$$P_n = |c_n|^2. \quad (374)$$

For example, we can consider a particle initially prepared in the symmetric superposition of the ground and first excited states, i.e. an equally weighted superposition of the ground and first excited states, at time $t = 0$

$$\Psi^{(+)}(x, t = 0) = \frac{1}{\sqrt{2}} [\psi_1(x) + \psi_2(x)]. \quad (375)$$

The probability to find the particle in state 1 or 2 is 1/2. The state will then evolve in time, with each amplitude having a different time-dependent phase

$$\begin{aligned} \Psi^{(+)}(x, t) &= \frac{1}{\sqrt{2}} [\psi_1(x)e^{-i\omega_1 t} + \psi_2(x)e^{-i\omega_2 t}] \\ &= e^{-i\omega_1 t} \frac{1}{\sqrt{2}} [\psi_1(x) + \psi_2(x)e^{-i\Delta\omega t}] \end{aligned} \quad (376)$$

where we have introduced the angular frequency associated with the energy eigenvalues $\omega_n = E_n/\hbar$ and the frequency difference $\Delta\omega = \omega_2 - \omega_1$. The probability to find the particle in state 1, is given by the modulus squared of the corresponding amplitude coefficient, which for the state in (376) is just 1/2. The same holds true for the probability to find the particle in state 2. However, the probability to find the particle in the initial superposition state is not time independent.

EXERCISE 10.9 Consider the bound state problem $E < V_0$ in the asymmetric infinite well potential

$$V(x) = \begin{cases} \infty & x \leq 0 \\ 0 & 0 < x < L \\ V_0 & L < x < 2L \\ \infty & x > 2L \end{cases}. \quad (377)$$

(i) Solve the time-dependent Schrödinger equations in regions I ($0 < x < L$) and II ($L < x < 2L$) and impose appropriate boundary conditions. (ii) Use the results in (i) to derive an equation in terms of E , V_0 , L whose solution will determine the possible energy eigenvalues E . (iii) Duplicate the procedure to determine an equation in terms of E , V_0 , L whose solution will resolve the possible energy eigenvalues for $E > V_0$.

EXERCISE 10.10 Consider the square well of width a which is infinite on the left side and finite on the right,

$$V(x) = \begin{cases} \infty & x < 0 \\ 0 & 0 < x < L \\ V_0 & x > L \end{cases}. \quad (378)$$

(i) Apply appropriate boundary conditions and derive the transcendental equation that determines the energy eigenvalues for the bound states. (ii) Show that if the parameter $2mL^2V_0/\hbar^2$ is smaller than a critical value, there are no bound states. What is this critical value?

EXERCISE 10.11 Consider a “downstep” potential, which drops at $x = 0$ as one goes from left to right.

$$V(x) = \begin{cases} 0 & x < 0 \\ -V_0 & x > 0 \end{cases}. \quad (379)$$

A particle of mass m and kinetic energy $E > 0$ approaches the abrupt potential drop. (i) Derive the reflection and transmission coefficients in terms of E and V_0 . (ii) When a free neutron enters a nucleus, it experiences a sudden drop in potential energy, from $V = 0$ outside to around -12 MeV inside. Suppose a neutron, emitted with kinetic energy 4 MeV by a fission event, strikes such a nucleus. What is the probability it will be absorbed, thereby initiating another fission? [Hint: The transmission coefficient expresses the probability of transmission through the surface.]

H. Particle in a central potential

We begin by recalling that a prescription for obtaining the (three-dimensional) Schrödinger equation for a free particle of mass m is to substitute into the classical energy momentum relation

$$E = \frac{|\vec{p}|^2}{2m} \quad (380)$$

the differential operators

$$E \rightarrow i\hbar \frac{\partial}{\partial t} \quad \text{and} \quad \vec{p} \rightarrow -i\hbar \vec{\nabla}. \quad (381)$$

The resulting operator equation is understood to act on a complex wave function $\psi(\vec{x}, t)$. That is

$$-\frac{\hbar^2}{2m} \nabla^2 \psi = i\hbar \frac{\partial}{\partial t} \psi, \quad (382)$$

where we interpret $\rho = |\psi|^2$ as the probability density ($|\psi|^2 d^3x$ gives the probability of finding the particle in a volume element d^3x .)

We are often concerned with moving particles, for example the collision of one particle with another. We therefore need to be able to calculate the density flux of a beam of particles \vec{j} . Now, from the conservation of probability, the rate of decrease of the number of particles in a given volume is equal to the total flux of particles out of that volume, i.e

$$-\frac{\partial}{\partial t} \int_V \rho dV = \int_S \vec{j} \cdot \hat{n} dS = \int_V \vec{\nabla} \cdot \vec{j} dV, \quad (383)$$

where the last equality is Gauss' theorem and \hat{n} is a unit vector along the outward normal to the element dS of the surface S enclosing the volume V . The probability and the flux densities are therefore related by the continuity equation

$$\frac{\partial \rho}{\partial t} + \vec{\nabla} \cdot \vec{j} = 0. \quad (384)$$

To determine the flux, we first form $\partial \rho / \partial t$ by subtracting the wave equation (382), multiplied by $-i\psi^*$ from the

complex conjugate equation multiplied by $-i\psi$. We then obtain

$$\frac{\partial \rho}{\partial t} - \frac{\hbar}{2m} (\psi^* \nabla^2 \psi - \psi \nabla^2 \psi^*) = 0. \quad (385)$$

Comparing this with (384), we identify the probability flux density as

$$\vec{j} = -\frac{i\hbar}{2m} (\psi^* \nabla \psi - \psi \nabla \psi^*). \quad (386)$$

For example, a solution of (382),

$$\psi = N e^{i\vec{p} \cdot \vec{x} - iEt}, \quad (387)$$

which describes a free particle of energy E and momentum \vec{p} , has $\rho = |N|^2$ and $\vec{j} = |N|^2 \vec{p}/m$.

EXERCISE 10.12 Calculate the eigenfunctions and energy levels for a free particle, enclosed in a box with edges of lengths a , b , and c . [Hint: The presence of the box (because of continuity) requires the wave function to vanish at the edges.]

We now consider the time-independent Schrödinger equation for a central potential

$$V(\vec{r}) = V(|\vec{r}|) = V(r). \quad (388)$$

Since the potential depends only on the distance from the origin, the hamiltonian is spherically symmetric. Therefore, instead of using cartesian coordinates $\vec{x} = \{x, y, z\}$, it is convenient to use spherical coordinates $\vec{x} = \{r, \vartheta, \varphi\}$ defined by

$$\left\{ \begin{array}{l} x = r \sin \vartheta \cos \varphi \\ y = r \sin \vartheta \sin \varphi \\ z = r \cos \vartheta \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{l} r = \sqrt{x^2 + y^2 + z^2} \\ \vartheta = \arctan(z/\sqrt{x^2 + y^2}) \\ \varphi = \arctan(y/x) \end{array} \right\}. \quad (389)$$

First, we express the Laplacian ∇^2 in spherical coordinates. It follows, after some tedious analysis, that

$$\nabla^2 = \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial}{\partial r} \right) + \frac{1}{r^2 \sin \vartheta} \frac{\partial}{\partial \vartheta} \left(\sin \vartheta \frac{\partial}{\partial \vartheta} \right) + \frac{1}{r^2 \sin^2 \vartheta} \frac{\partial^2}{\partial \varphi^2}. \quad (390)$$

To look for solutions, we use again the separation of variable methods, writing $\psi(\vec{x}) = \psi(r, \vartheta, \varphi) = R(r)Y(\vartheta, \varphi)$:

$$-\frac{\hbar^2}{2m} \left[\frac{Y}{r^2} \frac{d}{dr} \left(r^2 \frac{dR}{dr} \right) + \frac{R}{r^2 \sin \vartheta} \frac{\partial}{\partial \vartheta} \left(\sin \vartheta \frac{\partial Y}{\partial \vartheta} \right) + \frac{R}{r^2 \sin^2 \vartheta} \frac{\partial^2 Y}{\partial \varphi^2} \right] + V(r)RY = ERY. \quad (391)$$

We then divide by RY/r^2 and rearrange the terms as

$$-\frac{\hbar^2}{2m} \left[\frac{1}{R} \frac{d}{dr} \left(r^2 \frac{dR}{dr} \right) \right] + r^2(V - E) = \frac{\hbar^2}{2mY} \left[\frac{1}{\sin \vartheta} \frac{\partial}{\partial \vartheta} \left(\sin \vartheta \frac{\partial Y}{\partial \vartheta} \right) + \frac{1}{\sin^2 \vartheta} \frac{\partial^2 Y}{\partial \varphi^2} \right]. \quad (392)$$

Each side is a function of r only and ϑ, φ , so they must be independently equal to a constant \varkappa that we set (for reasons to be seen later) equal to

$$\varkappa = -\frac{\hbar^2}{2m} l(l+1). \quad (393)$$

We obtain two equations:

$$\frac{1}{R} \frac{d}{dr} \left(r^2 \frac{dR}{dr} \right) - \frac{2mr^2}{\hbar^2} (V - E) = l(l+1) \quad (394)$$

and

$$\frac{1}{\sin \vartheta} \frac{\partial}{\partial \vartheta} \left(\sin \vartheta \frac{\partial Y}{\partial \vartheta} \right) + \frac{1}{\sin^2 \vartheta} \frac{\partial^2 Y}{\partial \varphi^2} = -l(l+1)Y. \quad (395)$$

This last equation is the angular equation. Notice that it can be considered an eigenvalue equation for the operator

$$\hat{O} \equiv \frac{1}{\sin \vartheta} \frac{\partial}{\partial \vartheta} \left(\sin \vartheta \frac{\partial}{\partial \vartheta} \right) + \frac{1}{\sin^2 \vartheta} \frac{\partial^2}{\partial \varphi^2}. \quad (396)$$

What is the meaning of this operator?

We take one step back and look at the angular momentum operator. From its classical form $\vec{L} = \vec{r} \times \vec{p}$ we can define the quantum mechanical operator:

$$\hat{\vec{L}} = \hat{\vec{r}} \times \hat{\vec{p}} = -i\hbar \hat{\vec{r}} \times \hat{\vec{\nabla}}. \quad (397)$$

In cartesian coordinates this reads

$$\begin{aligned} \hat{L}_x &= \hat{y}\hat{p}_z - \hat{p}_y\hat{z} = -i\hbar \left(y \frac{\partial}{\partial z} - \frac{\partial}{\partial y} z \right), \\ \hat{L}_y &= \hat{z}\hat{p}_x - \hat{p}_z\hat{x} = -i\hbar \left(z \frac{\partial}{\partial x} - \frac{\partial}{\partial z} x \right), \\ \hat{L}_z &= \hat{x}\hat{p}_y - \hat{p}_x\hat{y} = -i\hbar \left(x \frac{\partial}{\partial y} - \frac{\partial}{\partial x} y \right). \end{aligned} \quad (398)$$

Some very important properties of this vector operator regard its commutator. Consider for example

$$[\hat{L}_x, \hat{L}_y] = [\hat{y}\hat{p}_z - \hat{p}_y\hat{z}, \hat{z}\hat{p}_x - \hat{p}_z\hat{x}] = [\hat{y}\hat{p}_z, \hat{z}\hat{p}_x] - [\hat{p}_y\hat{z}, \hat{z}\hat{p}_x] - [\hat{y}\hat{p}_z, \hat{p}_z\hat{x}] + [\hat{p}_y\hat{z}, \hat{p}_z\hat{x}]. \quad (399)$$

Now remember that $[x_i, x_j] = [p_i, p_j] = 0$ and $[x_i, p_j] = i\hbar\delta_{ij}$. Also $[AB, C] = A[B, C] + [A, C]B$. This simplifies matters a lot

$$[\hat{L}_x, \hat{L}_y] = \hat{y}[\hat{p}_z, \hat{z}]\hat{p}_x - [\hat{p}_y\hat{z}, \hat{z}\hat{p}_x] - [\hat{y}\hat{p}_z, \hat{p}_z\hat{x}] + \hat{p}_y[\hat{z}, \hat{p}_z]\hat{x} = i\hbar(\hat{x}\hat{p}_y - \hat{y}\hat{p}_x) = i/\hat{L}_z. \quad (400)$$

By performing a cyclic permutation of the indexes, we can show that this holds in general:

$$[\hat{L}_i, \hat{L}_j] = i\hbar \varepsilon_{ijk} \hat{L}_k, \quad (401)$$

where

$$\varepsilon_{ijk} = \begin{cases} 1 & \text{if } (i, j, k) \text{ is an even permutation of } (1, 2, 3); \text{ that is } (1, 2, 3), (2, 3, 1), (3, 1, 2), \\ -1 & \text{if } (i, j, k) \text{ is an odd permutation of } (1, 2, 3); \text{ that is } (3, 2, 1), (1, 3, 2), (2, 1, 3), \\ 0 & \text{if any index is repeated; that is if } i = j \text{ or } j = k \text{ or } k = i \end{cases} \quad (402)$$

is the Levi-Civita tensor. (The following relations are useful: $\varepsilon_{ikl}\varepsilon_{imn} = \delta_{km}\delta_{ln} - \delta_{kn}\delta_{lm}$, $\varepsilon_{ikl}\varepsilon_{ikm} = 2\delta_{lm}$, and $\varepsilon_{ikl}\varepsilon_{ikl} = 6$.) Since the different components of the angular momentum do not commute, they do not possess common eigenvalues and there is an uncertainty relation for them. If for example we know with absolute precision the angular momentum along the z direction, we cannot have any knowledge of the components along x and y . From the uncertainty relations,

$$\Delta L_x \Delta L_z \geq \frac{\hbar}{2} |\langle L_y \rangle| \quad \text{and} \quad \Delta L_y \Delta L_z \geq \frac{\hbar}{2} |\langle L_x \rangle|, \quad (403)$$

which derive from (303) and (401), we have that if $\Delta L_z = 0$ (perfect knowledge) then we have a complete uncertainty in L_x and L_y .

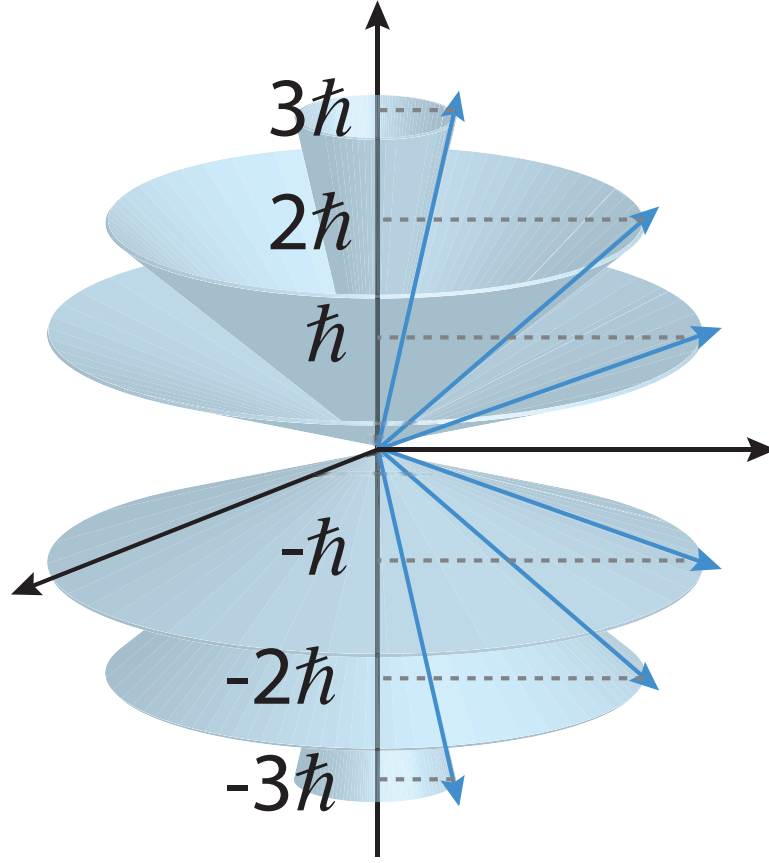


FIG. 43: Graphical representation of the angular momentum, with fixed L_z and L^2 , but complete uncertainty in L_x and L_y .

EXERCISE 10.13 Consider the squared length of the angular momentum vector $\hat{L}^2 = \hat{L}_x^2 + \hat{L}_y^2 + \hat{L}_z^2$. Show that $[\hat{L}^2, \hat{L}_i] = 0$, for $i = x, y, z$.

All in all, we can always know the length of the angular momentum plus one of its components. For example, choosing the z -component, we can represent the angular momentum as a cone, of length $\langle L \rangle$, projection on the z -axis (L_z) and with complete uncertainty of its projection along x and y , see Fig. 43.

We now express the angular momentum using spherical coordinates. This simplifies particularly how the azimuthal angular momentum \hat{L}_z is expressed. It is also convenient to define a new set of orthogonal unit vectors appropriate to the spherical coordinate system (389). These are

$$\begin{aligned}\hat{e}_r &= \frac{\vec{r}}{r} = \cos \varphi \sin \vartheta \hat{e}_x + \sin \varphi \sin \vartheta \hat{e}_y + \cos \vartheta \hat{e}_z \\ \hat{e}_\theta &= \frac{\partial \hat{e}_r}{\partial \theta} = \cos \varphi \cos \vartheta \hat{e}_x + \sin \varphi \cos \vartheta \hat{e}_y - \sin \vartheta \hat{e}_z \\ \hat{e}_\phi &= \frac{1}{\sin \vartheta} \frac{\partial \hat{e}_r}{\partial \varphi} = -\sin \varphi \hat{e}_x + \cos \varphi \hat{e}_y.\end{aligned}\tag{404}$$

We can now rewrite the gradient operator by using the chain rule for partial differentiation as

$$\begin{aligned}\vec{\nabla} &= \hat{e}_x \frac{\partial}{\partial x} + \hat{e}_y \frac{\partial}{\partial y} + \hat{e}_z \frac{\partial}{\partial z} \\ &= \hat{e}_x \left(\frac{\partial r}{\partial x} \frac{\partial}{\partial r} + \frac{\partial \vartheta}{\partial x} \frac{\partial}{\partial \vartheta} + \frac{\partial \varphi}{\partial x} \frac{\partial}{\partial \varphi} \right) + \hat{e}_y \left(\frac{\partial r}{\partial y} \frac{\partial}{\partial r} + \frac{\partial \vartheta}{\partial y} \frac{\partial}{\partial \vartheta} + \frac{\partial \varphi}{\partial y} \frac{\partial}{\partial \varphi} \right) + \hat{e}_z \left(\frac{\partial r}{\partial z} \frac{\partial}{\partial r} + \frac{\partial \vartheta}{\partial z} \frac{\partial}{\partial \vartheta} + \frac{\partial \varphi}{\partial z} \frac{\partial}{\partial \varphi} \right).\end{aligned}\tag{405}$$

The various partial derivatives can be evaluated using (389) to give

$$\begin{aligned}
\frac{\partial r}{\partial x} &= \frac{x}{r} = \cos \varphi \sin \vartheta, & \frac{\partial r}{\partial y} &= \frac{y}{r} = \sin \varphi \sin \vartheta, & \frac{\partial r}{\partial z} &= \frac{z}{r} = \cos \vartheta, \\
\frac{\partial \vartheta}{\partial x} &= \frac{1}{r} \cos \varphi \cos \vartheta, & \frac{\partial \vartheta}{\partial y} &= \frac{1}{r} \sin \varphi \cos \vartheta, & \frac{\partial \vartheta}{\partial z} &= -\frac{1}{r} \sin \vartheta, \\
\frac{\partial \varphi}{\partial x} &= -\frac{\sin \varphi}{r \sin \vartheta}, & \frac{\partial \varphi}{\partial y} &= \frac{\cos \varphi}{r \sin \vartheta}, & \frac{\partial \varphi}{\partial z} &= 0.
\end{aligned} \tag{406}$$

These along with the definitions of the spherical unit vectors can be used to write the gradient as

$$\vec{\nabla} = \hat{e}_r \frac{\partial}{\partial r} + \frac{\hat{e}_\vartheta}{r} \frac{\partial}{\partial \vartheta} + \frac{\hat{e}_\varphi}{r \sin \vartheta} \frac{\partial}{\partial \varphi}. \tag{407}$$

We can now write the angular momentum operator as

$$\begin{aligned}
\hat{L} &= \vec{r} \times \frac{\hbar}{i} \vec{\nabla} = \frac{\hbar}{i} r \hat{e}_r \times \left(\hat{e}_r \frac{\partial}{\partial r} + \frac{\hat{e}_\vartheta}{r} \frac{\partial}{\partial \vartheta} + \frac{\hat{e}_\varphi}{r \sin \vartheta} \frac{\partial}{\partial \varphi} \right) \\
&= \frac{\hbar}{i} r \left(\frac{\hat{e}_\varphi}{r} \frac{\partial}{\partial \vartheta} - \frac{\hat{e}_\vartheta}{r \sin \vartheta} \frac{\partial}{\partial \varphi} \right) = \frac{\hbar}{i} \left(\hat{e}_\varphi \frac{\partial}{\partial \vartheta} - \frac{\hat{e}_\vartheta}{\sin \vartheta} \frac{\partial}{\partial \varphi} \right) \\
&= \frac{\hbar}{i} \left[(-\sin \varphi \hat{e}_x + \cos \varphi \hat{e}_y) \frac{\partial}{\partial \vartheta} - \frac{\cos \varphi \cos \vartheta \hat{e}_x + \sin \varphi \cos \vartheta \hat{e}_y - \sin \vartheta \hat{e}_z}{\sin \vartheta} \frac{\partial}{\partial \varphi} \right] \\
&= \frac{\hbar}{i} \left[\hat{e}_x \left(-\sin \varphi \frac{\partial}{\partial \vartheta} - \frac{\cos \varphi \cos \vartheta}{\sin \vartheta} \frac{\partial}{\partial \varphi} \right) + \hat{e}_y \left(\cos \varphi \frac{\partial}{\partial \vartheta} - \frac{\sin \varphi \cos \vartheta}{\sin \vartheta} \frac{\partial}{\partial \varphi} \right) + \hat{e}_z \frac{\partial}{\partial \varphi} \right].
\end{aligned} \tag{408}$$

From this we can extract the cartesian components of the angular momentum vector in terms of spherical coordinates yielding

$$\begin{aligned}
\hat{L}_x &= i\hbar \left(\sin \varphi \frac{\partial}{\partial \vartheta} + \cot \varphi \cos \vartheta \frac{\partial}{\partial \varphi} \right), \\
\hat{L}_y &= -i\hbar \left(\cos \varphi \frac{\partial}{\partial \vartheta} - \cot \vartheta \sin \varphi \frac{\partial}{\partial \varphi} \right), \\
\hat{L}_z &= -\hbar \frac{\partial}{\partial \varphi}.
\end{aligned} \tag{409}$$

The spherical representation of the components of the angular momentum operator can be used to express the square of the angular momentum operator as

$$\begin{aligned}
\hat{L}^2 &= -\hbar^2 \left[\sin^2 \varphi \frac{\partial^2}{\partial \vartheta^2} + \sin \varphi \cos \varphi \frac{\partial \cot \vartheta}{\partial \vartheta} \frac{\partial}{\partial \vartheta} + \frac{\sin \varphi \cos \varphi \cos \vartheta}{\sin \vartheta} \frac{\partial^2}{\partial \vartheta \partial \varphi} \right. \\
&\quad + \frac{\cos^2 \varphi \cos \vartheta}{\sin \vartheta} \frac{\partial}{\partial \vartheta} + \frac{\sin \varphi \cos \varphi \cos \vartheta}{\sin \vartheta} \frac{\partial^2}{\partial \varphi \partial \vartheta} - \frac{\sin \varphi \cos \varphi \cos^2 \vartheta}{\sin^2 \vartheta} \frac{\partial}{\partial \varphi} \\
&\quad + \frac{\cos^2 \varphi \cos^2 \vartheta}{\sin^2 \vartheta} \frac{\partial^2}{\partial \varphi^2} \\
&\quad + \cos^2 \varphi \frac{\partial^2}{\partial \vartheta^2} - \sin \varphi \cos \varphi \frac{\partial \cot \vartheta}{\partial \vartheta} \frac{\partial}{\partial \vartheta} - \frac{\sin \varphi \cos \varphi \cos \vartheta}{\sin \vartheta} \frac{\partial^2}{\partial \vartheta \partial \varphi} \\
&\quad + \frac{\sin^2 \varphi \cos \vartheta}{\sin \vartheta} \frac{\partial}{\partial \vartheta} - \frac{\sin \varphi \cos \varphi \cos \vartheta}{\sin \vartheta} \frac{\partial^2}{\partial \varphi \partial \vartheta} + \frac{\sin \varphi \cos \varphi \cos^2 \vartheta}{\sin^2 \vartheta} \frac{\partial}{\partial \varphi} \\
&\quad \left. + \frac{\sin^2 \varphi \cos^2 \vartheta}{\sin^2 \vartheta} \frac{\partial^2}{\partial \varphi^2} + \frac{\partial^2}{\partial \varphi^2} \right].
\end{aligned} \tag{410}$$

The colored terms cancel in pairs and the remaining terms can be simplified to give a relation that should be familiar

$$\hat{L}^2 = -\hbar^2 \left[\frac{1}{\sin \vartheta} \frac{\partial}{\partial \vartheta} \left(\sin \vartheta \frac{\partial}{\partial \vartheta} \right) + \frac{1}{\sin^2 \vartheta} \frac{\partial^2}{\partial \varphi^2} \right]. \tag{411}$$

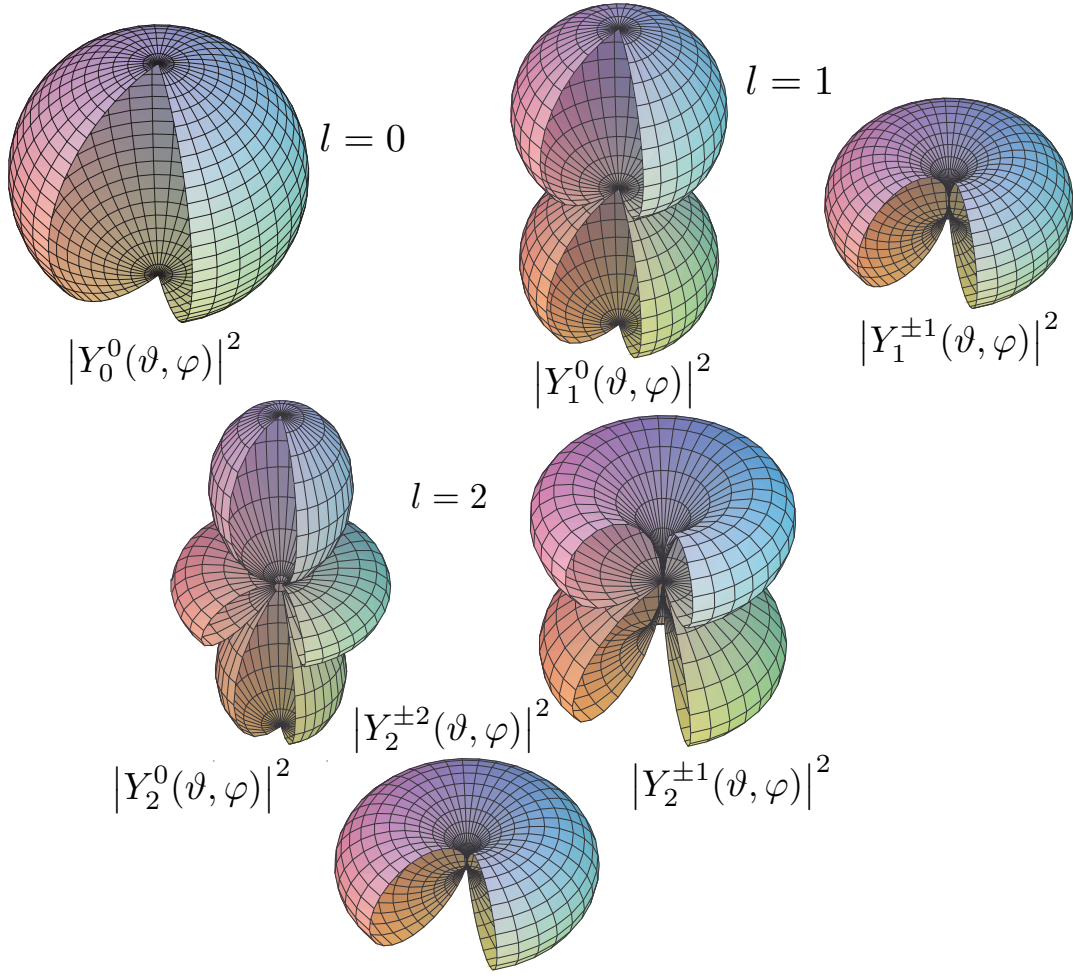


FIG. 44: Representations of $|Y_l^m|^2$ for different sets of quantum numbers. The z axis is the vertical direction. The probability densities have rotational symmetry about the z axis [64].

TABLE I: Associated Legendre polynomials.

$\begin{array}{c c} m \\ \hline l \end{array}$	0	1	2	3
0	$P_0^0 = 1$			
1	$P_1^0 = \cos \vartheta$	$P_1^1 \sin \vartheta$		
2	$P_2^0 = (3 \cos^2 \vartheta - 1)/2$	$P_2^1 = 3 \cos \vartheta \sin \vartheta$	$P_2^2 = 3 \sin^2 \vartheta$	
3	$P_3^0 (5 \cos^3 \vartheta - 3 \cos \vartheta)/2$	$P_3^1 = 3(5 \cos^2 \vartheta - 1)/2 \sin \vartheta$	$P_3^2 = 15 \cos \vartheta \sin^2 \vartheta$	$P_3^3 = 15 \sin^3 \vartheta$

We can then write the eigenvalue equations for these two operators: and

$$\hat{L}^2 Y(\vartheta, \varphi) = \hbar^2 l(l+1) Y(\vartheta, \varphi) \quad \text{and} \quad \hat{L}_z Y(\vartheta, \varphi) = \hbar m Y(\vartheta, \varphi) \quad (412)$$

where we already used the fact that they share common eigenfunctions. Then, by its very nature we can label these eigenfunctions by l and m , i.e. $Y_{l,m}(\vartheta, \varphi)$.

EXERCISE 10.14 Show that the allowed values for l and m_z are integers such that $l = 0, 1, 2, \dots$ and $m_z = l, \dots, l-1, l$. [*Hint:* This result can be inferred from the commutation relationship.]

We now go back to the Schrödinger equation in spherical coordinates and we consider the angular and radial

equation separately to find the energy eigenvalues and eigenfunctions. The angular equation was found to be

$$\frac{1}{\sin \vartheta} \frac{\partial}{\partial \vartheta} \left(\sin \vartheta \frac{\partial Y_l^m(\vartheta, \varphi)}{\partial \vartheta} \right) + \frac{1}{\sin^2 \vartheta} \frac{\partial^2 Y_l^m(\vartheta, \varphi)}{\partial \varphi^2} = -l(l+1) Y_l^m(\vartheta, \varphi). \quad (413)$$

Note that this equation does not depend at all on the potential, thus it will be common to all problems with a central potential.

We can solve the equation by using again separation of variables: $Y(\vartheta, \varphi) = \Theta(\vartheta)\Phi(\varphi)$. By multiplying both sides of the equation by $\sin^2 \vartheta / Y(\vartheta, \varphi)$ we obtain:

$$\frac{1}{\Theta(\vartheta)} \left[\sin \vartheta \frac{d}{d\vartheta} \left(\sin \vartheta \frac{d\Theta}{d\vartheta} \right) \right] + l(l+1) \sin^2 \vartheta = -\frac{1}{\Phi(\varphi)} \frac{d^2 \Phi}{d\varphi^2}. \quad (414)$$

As usual we separate the two equations in the different variables and introduce a constant m^2 :

$$\frac{d^2 \Phi}{d\varphi^2} = -m^2 \Phi(\varphi) \quad (415)$$

and

$$\sin \vartheta \frac{d}{d\vartheta} \left(\sin \vartheta \frac{d\Theta}{d\vartheta} \right) = [m^2 - l(l+1) \sin^2 \vartheta] \Theta(\vartheta). \quad (416)$$

The first equation is easily solved to give $\Phi(\varphi) = e^{im\varphi}$ with $m = 0, \pm 1, \pm 2, \dots$ since we need to impose the periodicity of Φ , such that $\Phi(\varphi + 2\pi) = \Phi(\varphi)$. The solutions to the second equation are associated Legendre polynomials: $\Theta(\vartheta) = AP_l^m(\cos \vartheta)$, the first few of which are in Table I. Note that, as previously found when solving for the eigenvalues of the angular momentum, we have that $m = -l, -l+1, \dots, l$, with $l = 0, 1, \dots$.

The normalized angular eigenfunctions are then spherical harmonic functions, given by the normalized associated Legendre polynomial times the solution to the equation in φ ,

$$Y_l^m(\vartheta, \varphi) = \sqrt{\frac{(2l+1)}{4\pi} \frac{(l-m)!}{(l+m)!}} P_l^m(\cos \vartheta) e^{im\varphi}. \quad (417)$$

As we expect from eigenfunctions, the Spherical Harmonics are orthogonal:

$$\int_{S^2} Y_l^{m*}(\vartheta, \varphi) Y_{l'}^{m'}(\vartheta, \varphi) d\Omega = \delta_{ll'} \delta_{mm'}, \quad (418)$$

where $\int_{S^2} d\Omega = \int_0^\pi \int_0^{2\pi} \sin \vartheta d\vartheta d\varphi$ denotes the integral over the spherical surface. For this phase convention, $Y_{lm}^*(\vartheta, \varphi) = (-1)^m Y_{l,-m}(\vartheta, \varphi)$, and hence for $l = 0, 1, 2$ we obtain

$$\begin{aligned} Y_0^0(\vartheta, \varphi) &= \frac{1}{\sqrt{4\pi}} \\ Y_1^{-1}(\vartheta, \varphi) &= \sqrt{\frac{3}{8\pi}} \sin \vartheta e^{i\varphi} \\ Y_1^0(\vartheta, \varphi) &= \sqrt{\frac{3}{4\pi}} \cos \vartheta \\ Y_1^1(\vartheta, \varphi) &= -\sqrt{\frac{3}{8\pi}} \sin \vartheta e^{i\varphi} \end{aligned}$$

$$\begin{aligned} Y_2^{-2}(\vartheta, \varphi) &= \sqrt{\frac{15}{32\pi}} \sin^2 \vartheta e^{-i2\varphi} \\ Y_2^{-1}(\vartheta, \varphi) &= \sqrt{\frac{15}{8\pi}} \sin \vartheta \cos \vartheta e^{-i\varphi} \\ Y_2^0(\vartheta, \varphi) &= \sqrt{\frac{5}{16\pi}} (-1 + 3 \cos^2 \vartheta) \\ Y_2^1(\vartheta, \varphi) &= -\sqrt{\frac{15}{8\pi}} \sin \vartheta \cos \vartheta e^{i\varphi} \\ Y_2^2(\vartheta, \varphi) &= \sqrt{\frac{15}{32\pi}} \sin^2 \vartheta e^{i2\varphi}; \end{aligned}$$

the surfaces $r = |Y_{lm}(\vartheta, \varphi)|$ for these spherical harmonics are shown in Fig. 44. (For details, see e.g., [64].)

We now turn to the radial equation:

$$\frac{d}{dr} \left(r^2 \frac{dR(r)}{dr} \right) - \frac{2mr^2}{\hbar^2} (V - E) = l(l+1)R(r). \quad (419)$$

To simplify the solution, we introduce a different function $u(r) = rR(r)$. Then the equation reduces to:

$$-\frac{\hbar^2}{2m} \frac{d^2 u}{dr^2} + \left[V + \frac{\hbar}{2m} \frac{l(l+1)}{r^2} \right] u(r) = Eu(r). \quad (420)$$

If we define an effective potential

$$V'(r) = V(r) + \frac{\hbar^2}{2m} \frac{l(l+1)}{r^2}, \quad (421)$$

(420) is very similar to the one-dimensional Schrödinger equation. The second term in this effective potential is called the *centrifugal* term. Solutions can be found for some forms of the potential $V(r)$, by first calculating the equation solutions $u_{n,l}(r)$, then finding $R_{n,l}(r) = u_{n,l}(r)/r$ and finally the wavefunction

$$\psi_{n,l,m}(r, \vartheta, \varphi) = R_{n,l}(r) Y_l^m(\vartheta, \varphi). \quad (422)$$

Note that we need 3 quantum numbers (n, l, m) to define the eigenfunctions of the hamiltonian in three dimensions. For example, we can have a simple spherical well: $V(r) = 0$ for $r < r_0$ and $V(r) = V_0$ otherwise. In the case of $l = 0$, this is the same equation as for the

square well in one dimension. Note, however, that since the boundary conditions need to be such that $R(r)$ is finite for all r , we need to impose that $u(r=0)=0$, hence only the odd solutions are acceptable. For $l > 0$, we can find solutions in terms of Bessel functions [64]. Two other important examples of potential are: the harmonic oscillator potential

$$V(r) = V_0 \frac{r}{r_0^2} - V_0, \quad (423)$$

which is an approximation for any potential close to its minimum, and the Coulomb potential

$$V(r) = -\frac{e^2}{4\pi\epsilon_0} \frac{1}{r}, \quad (424)$$

which describes the atomic potential and in particular the Hydrogen atom.

XI. STERN-GERLACH EXPERIMENT

Within the framework of classical mechanics one can show that an electron in a circular orbit has an angular momentum $\vec{L} = m_e r^2 \vec{\omega}$ and an associated magnetic moment

$$\vec{\mu} = -\frac{e}{2m_e} \vec{L}, \quad (425)$$

where m_e and e are, respectively, the mass and charge of the electron, and r and $\vec{\omega}$ are the radius and angular velocity of the orbital motion. In a magnetic field \vec{B} the atom will be acted on by a torque

$$\vec{\tau} = \vec{\mu} \times \vec{B}, \quad (426)$$

which causes \vec{L} to precess about the direction of \vec{B} with some fixed value of the projection $\mu_z = |\vec{\mu}| \cos \theta$ of its magnetic moment along the direction of the field; see Appendix D. The atom will also have a potential energy $-\vec{\mu} \cdot \vec{B}$, and if the field is inhomogeneous such that at a certain point it is in the z direction and varies strongly with z , then the atom will be acted on by a force

$$F_z = -\vec{\nabla}(-\vec{\mu} \cdot \vec{B}) = \mu_z \frac{\partial B_z}{\partial z}, \quad (427)$$

which may have any of a continuous set of values from $|\vec{\mu}| \partial B_z / \partial z$ to $-|\vec{\mu}| \partial B_z / \partial z$. One would then expect a monoenergetic beam of atoms, initially randomly oriented and passing through an inhomogeneous magnetic field, to be deflected in the $+z$ and $-z$ directions with a distribution of deflection angles that has a maximum value at zero deflection and decreases monotonically in either direction. This is not what is observed experimentally.

In 1921, Stern and Gerlach generated a beam of neutral silver atoms by evaporating silver from an oven [65–68]. The process was performed in a vacuum so that the silver

atoms moved without scattering. The atoms were collimated by slits and sent through a region with a large non-uniform magnetic field. The silver atoms were thus deflected and allowed to strike a cold metallic plate (hereafter refer to as the detection screen). After about 8-10 hours the number of condensed silver atoms was large enough to show a visible trace on the screen. The trace showed two marks implying that the silver atoms had two possible components of μ_z .

From the classical relation (425) we see that the quantum-mechanical operator corresponding to the observable μ_L (the magnetic moment due to the orbital motion) must be

$$\hat{\mu}_L = -\frac{e}{2m_e} \hat{L} = -\frac{e}{2m_e} \hat{r} \times \hat{p}. \quad (428)$$

The operator relation (428) suggests that the magnetic moment $\vec{\mu}_L$ due to the orbital motion is quantized in the same way as the orbital angular momentum. This means that the size and one of the components can have discrete values simultaneously, i.e.

$$|\vec{\mu}_L| = \frac{e}{2m_e} |\vec{L}| = \mu_B \sqrt{l(l+1)}, \quad (429)$$

with $l = 0, 1, 2, \dots$, while for example the z component can take the values

$$(\mu_L)_z = -\frac{e}{2m_e} L_z = -m\mu_B, \quad (430)$$

with $m = 0, \pm 1, \pm 2, \dots, \pm l$. Here the quantity,

$$\mu_B \equiv \frac{e\hbar}{2m_e} \quad (1 \text{ Bohr magneton}) \quad (431)$$

is the natural unit for the magnetic moment for the electron, just as \hbar is the natural unit for angular momenta, $1 \mu_B = 5.788 \times 10^{-5} \text{ eV/T(esla)}$ [69]. We note that the quantization of L_z and $(\mu_L)_z$ corresponds to a so-called “space quantization,” that is, quantized values of the angle between the vector μ_L and the z axis. If the total angular momentum of the atom is “integral” (given by an integral angular-momentum quantum number l), we should then expect to find $2l+1$ discrete deflections, that is, an odd number of pictures of the slit on the detection screen. In plain English, Schrödinger equation predicts an odd number of possible states, also in contradiction with experiment.

The explanation for the even number of images observed in the Stern-Gerlach experiment was developed by Uhlenbeck and Goudsmit [70, 71], and independently by Pauli [72, 73]. It is most easily understood in the hydrogen case: Even with $l = 0$ the electron in the hydrogen atom has a magnetic moment μ_S , which causes the deflection of the orbit of each atom. This magnetic moment is connected with an intrinsic angular momentum of the electron, the so-called intrinsic spin angular momentum S , (spin for short). As all other angular momenta, the spin can be characterized by an angular-momentum

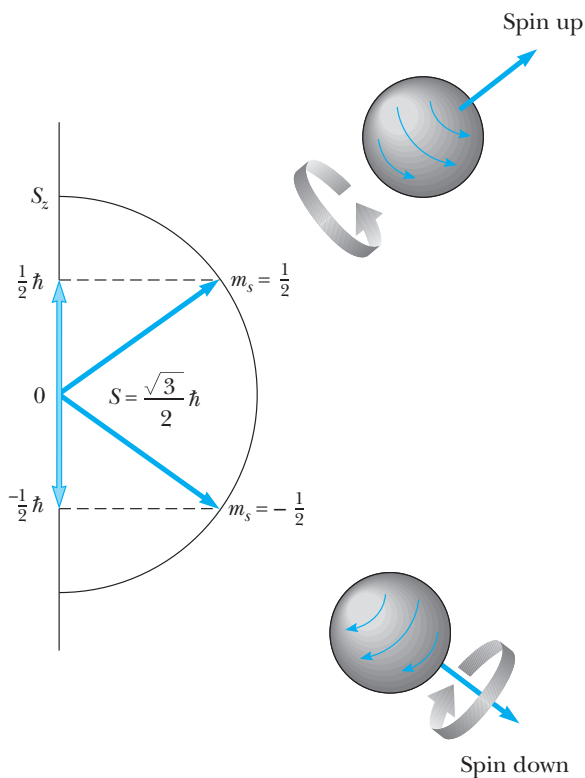


FIG. 45: Two allowed orientations of the spin vector \vec{S} for a spin- $\frac{1}{2}$ particle, such as the electron [74].

quantum number which is usually denoted by s , so that $|\vec{S}| = \hbar\sqrt{s(s+1)}$ and such that the z component can take the values

$$S_z = m_s \hbar, \quad (432)$$

where $m_s = -s, -s+1, -s+2, \dots, +s$. This is analogous to $m = L_z/\hbar$ taking the values $-l, -l+1, -l+2, \dots, +l$ for a given orbital angular-momentum quantum number l . From the general discussion of angular momenta in Sec. XH, it follows that a spin quantum number must in general take one of the values $s = 0, \frac{1}{2}, 1, \frac{3}{2}, 2, \dots$. The number of m_s values is $2s+1$, in analogy with the $2l+1$ values of L_z .

The fact that we observe two discrete deflections in the Stern-Gerlach experiment then leads to the interpretation that in this case $2s+1$ is equal to 2; that is, the spin quantum number s of the electron is equal to $\frac{1}{2}$. This corresponds to $|\vec{S}| = \hbar\sqrt{s(s+1)} = \hbar\sqrt{3/4} = 0.866\hbar$. The two possible values of the magnetic quantum number of the electron spin then are $m_s = \pm\frac{1}{2}$, corresponding to $S_z = \pm\frac{1}{2}\hbar$. These two spin states are commonly denoted by spin up and spin down, see Fig. 45.

Experiment show that the intrinsic magnetic moment connected with the spin is

$$\vec{\mu}_S = -\frac{e}{2m_e} g_e \vec{S}, \quad (433)$$

where the factors in front of \vec{S} is the product of the gyromagnetic ratio we found for the orbital motion, and g_e which is a dimensionless factor known as the gyromagnetic factor of the electron. It is conventional to express g_e in terms of the so called “anomalous magnetic moment”,

$$a_e = (g_e - 2)/2, \quad (434)$$

which is one of most accurately determined quantities experimentally. Recent measurements of the anomalous magnetic moment of the electron reached the fabulous relative precision of 0.7 parts per billion [75, 76],

$$a_e^{\text{exp}} = 11596521807.3(2.8) \times 10^{-13}. \quad (435)$$

Pauli realized that, for intrinsic spin, only a matrix representation is possible [77].¹² The electron (ignoring the spatial dependence of the wave function) can be in two states, called *spin up* and *spin down*, which we denote by

$$\text{spin up} \Leftrightarrow \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{spin down} \Leftrightarrow \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (436)$$

The \hat{S}_z spin operator is defined by

$$\hat{S}_z = \frac{\hbar}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad (437)$$

and acts on the spin up and spin down states by ordinary matrix multiplication. Denoting the spin up and spin down state by $|\uparrow\rangle$ and $|\downarrow\rangle$, it follows that

$$\begin{aligned} \hat{S}_z |\uparrow\rangle &= \frac{\hbar}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{\hbar}{2} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ &= \frac{\hbar}{2} |\uparrow\rangle \end{aligned} \quad (438)$$

and similarly

$$\begin{aligned} \hat{S}_z |\downarrow\rangle &= \frac{\hbar}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = -\frac{\hbar}{2} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ &= -\frac{\hbar}{2} |\downarrow\rangle. \end{aligned} \quad (439)$$

From (438) and (439) we conclude that the spin up state $|\uparrow\rangle$ has eigenvalue $\hbar/2$ under the operator \hat{S}_z , and the spin down state $|\downarrow\rangle$ has eigenvalue $-\hbar/2$. This corresponds precisely to the two spots observed on the detection screen. When the electron is in the up state,

¹² Quantum mechanics was initially developed in the language of matrices by Heisenberg [78] and only later in the language of waves by Schrödinger. Due to the greater familiarity of physicists with waves than with matrices, Schrödinger’s formulation quickly became very popular.

it is deflected upwards by the Stern-Gerlach apparatus, whereas when it is in the down state, it will hit the spot lower on the screen.

According to the rules of quantum mechanics, all linear combinations of $|\uparrow\rangle$ and $|\downarrow\rangle$ should also be physical states. Therefore, any state of the form

$$\alpha \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \beta \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \quad (440)$$

is a viable physical state, for any complex numbers α and β . (Requiring normalization yields $|\alpha|^2 + |\beta|^2 = 1$.) Now, what happens to such a state when it enters the Stern-Gerlach apparatus? We have already seen that the Schrödinger equation (for whatever system of hamiltonian) is linear. Thus, if a state is the sum of two parts, each part will evolve in its own way – as if the other was not there. From this we infer that the $\alpha|\uparrow\rangle$ component will travel upwards, while the $\beta|\downarrow\rangle$ will fly downwards. As a consequence the wave function will *split* in two parts. The total amplitudes of the two parts of the wave function are given by $|\alpha|^2$ and $|\beta|^2$. We know that the measuring device (detecting screen) will then report impact on the upper spot with probability $|\alpha|^2$, and impact on the lower spot with probability $|\beta|^2$. Note that the electron can be: (i) in the up state; (ii) in the down state; (iii) in a superposition. Nevertheless, for any of these possibilities there will only be one place at which detection occurs, so it *always* looks as if the electron was strictly up or down, even if the state entering the apparatus was a mixture [79].

The preceding discussion clarifies the z -component of the spin, but what about the other components, x and y ? Interestingly, Stern-Gerlach apparatuses are like Lego pieces: you can put them together in many different ways, and have lots of fun. Consider drilling a hole in the detecting screen, at the spin up spot, and behind it placing another Stern-Gerlach apparatus. The outcome is not so surprising: the screen after the second apparatus will only detect spin up. Indeed, the hole served as some kind of filter letting through only the $|\uparrow\rangle$ states. Therefore, at a second splitting, there will only be particles going upwards, as confirmed by experiment. Now we can do something more radical. Keep the first magnet in the same position, but twist the second apparatus over an angle of 90° , so that the \vec{B} -field is in the x -direction. In this way it selects states of the \hat{S}_x operator, the x -component of their intrinsic magnetic moment. The outcome is not so surprising, as we again see two spots on the final screen: one with magnetic moment $S_x = \frac{\hbar}{2}$, and one spot for $S_x = -\frac{\hbar}{2}$. As expected, the state $|\uparrow\rangle$ of definite S_z does not have a definite S_x . It is a superposition of an spin up and spin down when viewed in the x -direction. A similar result holds when studying S_y .

Both S_x and S_y are, like S_z , physical quantities which can be measured. They have to correspond to *operators* acting on the spin state. So, just like \hat{S}_z they have to be represented by a two-by-two matrix acting on the states

$\begin{pmatrix} \alpha \\ \beta \end{pmatrix}$. Without loss of generality, we can write the \hat{S}_x and \hat{S}_y operators in terms of complex numbers a_x, \dots, d_y ,

$$\hat{S}_x = \begin{pmatrix} a_x & b_x \\ c_x & d_x \end{pmatrix} \quad \text{and} \quad \hat{S}_y = \begin{pmatrix} a_y & b_y \\ c_y & d_y \end{pmatrix}. \quad (441)$$

How can we guess the form of S_x and S_y ? Because the operators S_x and S_y are observables, they have to be hermitian. Hence the corresponding matrices need to be so too, and (441) becomes

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a^* & c^* \\ b^* & d^* \end{pmatrix}. \quad (442)$$

This is a very restricting property. It is easily seen that (up to linear combinations) there are only 4 hermitian 2-by-2 matrices: the identity $\mathbb{1}$ and the three Pauli matrices σ_1, σ_2 , and σ_3 ,

$$\begin{aligned} \mathbb{1} &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, & \sigma_1 &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \\ \sigma_2 &= \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, & \sigma_3 &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \end{aligned} \quad (443)$$

Note that σ_3 is S_z modulo the factor $\hbar/2$ which has been stripped of. A fond guesser might then conjecture that

$$\hat{S}_x = \frac{\hbar}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad \hat{S}_y = \frac{\hbar}{2} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad (444)$$

which is actually a good guess. To convince you a bit, we will use the above guess to explain the outcome of the “twisted” Stern-Gerlach experiment. On the first place,

$$\hat{S}_x \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{\hbar}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{\hbar}{2} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad (445)$$

so clearly $|\uparrow\rangle$ is *not* a state with definite S_x . However, it can be broken in two parts

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix} + \begin{pmatrix} 1/2 \\ -1/2 \end{pmatrix} \quad (446)$$

And these parts *are* eigenstates of S_x :

$$\begin{aligned} \hat{S}_x \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix} &= \frac{\hbar}{2} \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}, \\ \hat{S}_x \begin{pmatrix} 1/2 \\ -1/2 \end{pmatrix} &= -\frac{\hbar}{2} \begin{pmatrix} 1/2 \\ -1/2 \end{pmatrix}. \end{aligned} \quad (447)$$

This explains the outcome of the “twisted” Stern-Gerlach experiment. The first part selects the state to be in $|\uparrow\rangle$. This is an eigenstate of \hat{S}_z , but not of \hat{S}_x . Hence, when passing through the second \vec{B} -field, the

state gets split again. The part $(\frac{1}{2})$ had positive S_x , hence feels a force $F_x > 0$ and goes to positive x . The part $(-\frac{1}{2})$ has negative S_x , so it is subject to $F_x < 0$ and goes the other way. Similar reasonings hold when considering \hat{S}_z and \hat{S}_y or \hat{S}_x and \hat{S}_y : if an electron is in a definite state with respect to one of the three operators, it will be a superposition of the up and down states along any other direction, i.e. $[\hat{S}_i, \hat{S}_j] = i\hbar \epsilon_{ijk} \hat{S}_k$.

EXERCISE 11.1 (i) Show that $\{\mathbb{1}, \sigma_1, \sigma_2, \sigma_3\}$ are linearly independent. (ii) Prove that $\{\mathbb{1}, \sigma_1, \sigma_2, \sigma_3\}$ form a basis in 2×2 matrix space, by showing that any arbitrary matrix

$$\mathbb{M} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix}$$

can be written on the form $\mathbb{M} = a_0 \mathbb{1} + \vec{a} \cdot \vec{\sigma}$, where $a_0 = \frac{1}{2} \text{Tr}(\mathbb{M})$, $\vec{a} = \frac{1}{2} \text{Tr}(\mathbb{M} \vec{\sigma})$, and $\vec{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ is the Pauli vector.

XII. KLEIN-GORDON EQUATION

Unless otherwise stated hereafter we work with natural (particle physicist's) Heaviside-Lorentz units $\hbar = c = 1$. In natural units the quantities energy, momentum, mass, (length) $^{-1}$, and (time) $^{-1}$ all have the same dimension.

EXERCISE 12.1 In a unit system where $\hbar = c = 1$ show that: (i) $1 \text{ kg} = 5.62 \times 10^{26} \text{ GeV}$; (ii) $1 \text{ GeV}^{-2} = 0.389 \text{ mb}$; (iii) $1 \text{ m} = 5.068 \times 10^{15} \text{ GeV}^{-1}$; (iv) $1 \text{ s} = 1.52 \times 10^{24} \text{ GeV}^{-1}$. (v) The Compton wavelength for an electron is given by $\lambda_c = m_e^{-1}$, calculate the numerical value; (vi) the Bohr radius of a hydrogen atom is given by $r_B = (\alpha m_e)^{-1}$, calculate the numerical value;

(vii) the velocity of an electron in the lowest Bohr orbit is α , calculate the numerical value. Show that: (viii) due to the fact that the electromagnetic interaction is relatively weak, we can use the non-relativistic Schrödinger equation to describe the hydrogen atom; (ix) the energy scale where quantum gravity effects become important is $M_{\text{Pl}} \sim 10^{19} \text{ GeV}$. Estimate the length scale at which this happens. [Hint: $\hbar c = 197.3 \text{ MeV fm}$, $c = 3 \times 10^8 \text{ m/s}$, $\alpha = 1/137$, $m_e = 0.511 \text{ MeV}$, $G = 6.67 \times 10^{-11} \text{ J m/kg}^2$.]

The wave equation (382) violates Lorentz invariance and is not suitable for a particle moving relativistically. It is tempting to repeat the steps of Sec. X but starting from the relativistic energy momentum relation (110) [49]. Making the operator substitution (381), we obtain the hyperbolic wave equation for the quantum mechanical description of a relativistic free particle

$$-\frac{\partial^2 \psi}{\partial t^2} + \nabla^2 \psi = m^2 \psi, \quad (448)$$

which is known as the Klein-Gordon equation [80, 81]. Introducing the covariant form of (381), $p^\mu \rightarrow i\partial^\mu$, with

$$\partial^\mu = \left(\frac{\partial}{\partial t}, -\vec{\nabla} \right) \quad \text{and} \quad \partial_\mu = \left(\frac{\partial}{\partial t}, \vec{\nabla} \right), \quad (449)$$

we can form the invariant (D'Alembertian) operator $\square^2 \equiv \partial_\mu \partial^\mu$ and rewrite (448) as

$$\partial_\mu \partial^\mu \psi + m^2 \psi \equiv (\square^2 + m^2) \psi = 0. \quad (450)$$

Recall $\psi(\vec{x}, t)$ is a scalar, complex-valued wave function. Multiplying (448) by $-i\psi^*$ and the complex conjugate equation by $-i\psi$, and subtracting, leads the continuity equation

$$\partial_t \underbrace{[i(\psi^* \partial_t \psi - \psi \partial_t \psi^*)]}_{\rho} + \vec{\nabla} \cdot \underbrace{[-i(\psi^* \vec{\nabla} \psi - \psi \vec{\nabla} \psi^*)]}_{\vec{j}} = 0. \quad (451)$$

EXERCISE 12.2 Convince yourself that the continuity equation follows from (448).

Considering the motion a free particle of energy E and momentum \vec{p} , described by Klein-Gordon solution,

$$\psi = N e^{i(\vec{p} \cdot \vec{x} - Et)}, \quad (452)$$

from (451) we find $\rho = 2E|N|^2$ and $\vec{j} = 2\vec{p}|N|^2$. We note that the probability density ρ is the timelike component of a 4-vector

$$\rho \propto E = \pm(\vec{p}^2 + m^2)^{1/2}. \quad (453)$$

Thus, in addition to the acceptable $E > 0$ solutions, we have negative energy solutions which have associated a negative probability density. We cannot simply discard the negative energy solutions as we have to work with a complete set of states, and this set inevitably includes the unwanted states.

Pauli and Weisskopf gave a natural interpretation to positive and negative probability densities by inserting the charge e into (451),

$$j^\mu = -ie(\psi^* \partial^\mu \psi - \psi \partial^\mu \psi^*), \quad (454)$$

and interpreting it as the electromagnetic charge-current

density [82]. With this in mind, j^0 represents a charge density, not a probability density, and so the fact that it can be negative is no longer objectionable. In some sense, which we will make clear in a moment, the $E < 0$ solutions may then be regarded as $E > 0$ solutions for particles of opposite charge (antiparticles).

The prescription for handling negative energy configurations was put forward by Stückelberg [83] and by Feynman [84–87]. Expressed most simply, the idea is that a negative energy solution describes a particle which propagates backwards in time or, equivalently, a positive energy *antiparticle* propagating forward in time.

Consider a spin-0 particle of energy E , 3-momentum \vec{p} , and charge e , generally referred to as the “spinless electron.” From (452) and (454), we know that the electromagnetic 4-vector current is

$$j^\mu(e^-) = -2e|N|^2(E, \vec{p}). \quad (455)$$

Now, taking its antiparticle e^+ of the same (E, \vec{p}) , because its charge is $+e$, we obtain

$$\begin{aligned} j^\mu(e^+) &= +2e|N|^2(E, \vec{p}) \\ &= -2e|N|^2(-E, -\vec{p}), \end{aligned} \quad (456)$$

which is exactly the same as the current of the original particle with $-E, -\vec{p}$. Hence, as far as a system is concerned, the emission of an antiparticle with energy E is the same as the absorption of a particle of energy $-E$. In other words, negative-energy particle solutions going backward in time describe positive-energy antiparticle solutions going forward in time. Of course the reason why this identification can be made is simply because

$$e^{-i(-E)(-t)} = e^{-iEt}. \quad (457)$$

A point worth noting at this juncture: No spinless quark or lepton has ever been observed in an experiment. Spinless hadrons exist (*e.g.* the π -meson), but they are complicated composite structures of spin- $\frac{1}{2}$ quarks and spin-1 gluons. The spin-zero leptons, that is, leptons satisfying the Klein-Gordon equation, are completely fictitious objects. We have ignored the complications due to the spin of the electrons, leaving complete developments of Dirac equation [89] to specialized textbooks; see *e.g.* [90].

XIII. THE STANDARD MODEL

The “Standard Model” is our most modern attempt to answer two simple questions that have been perplexing (wo)mankind throughout the epochs: What is the Universe made of? Why is our world the way it is?

A. Quarks, leptons, and gauge bosons

If we look deep inside a rock, we can see that it is made up of only a few types of elementary “point-like”

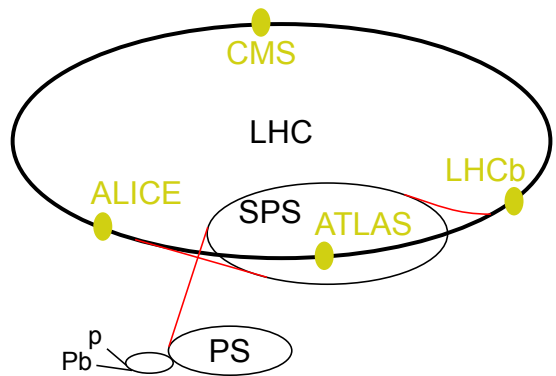


FIG. 46: The four large detectors at the LHC. ATLAS and CMS are designed to cover the full solid angle – a goal that is achieved using a cylindrical configuration with a central barrel and end-caps on both sides that also detect particles traversing the detector under a shallow angle with respect to the beamline (*i.e.* in the large pseudo-rapidity region). LHCb and ALICE are conceptualized for studies at well-defined solid angles. Both consist of a main detector section and an extended forward arm in one direction. Within the chosen solid angle of an experiment – be it the full solid angle or only a fraction of it – no particle should escape detection, except for weakly interacting particles (such as *e.g.* neutrinos) that do not interact on the length scale of a typical detector. Alongside measuring cross sections at central rapidity in pp collisions, ALICE is optimized to study heavy-ion (Pb-Pb nuclei) collisions at a CM energy of 2.76 TeV per nucleon pair. The resulting temperature and energy density are expected to be high enough to produce quark-gluon plasma, a state of matter wherein quarks and gluons are free. Similar conditions are thought to have existed a fraction of the second after the Big Bang before quarks and gluons bound together to form hadrons.

particles. The elementary-particle model accepted today views quarks and leptons as the basic constituents of ordinary matter. By pointlike, we understand that quarks and leptons show no evidence of internal structure at the current limit of our resolution, which is about $r \sim 2 \times 10^{-20}$ m.

The colossal microscope attaining such an incredible resolution is the Large Hadron Collider (or LHC), a machine for collisions of ultrarelativistic protons [91]. A schematic representation of the acceleration process is visible in Fig. 46. At the beginning Hydrogen atoms are injected into the source chamber of the linear accelerator, CERN’s Linac2, where the electrons are stripped off to become packets of protons, which are subsequently accelerated by an electric field. The beam of protons is further accelerated at the booster, the Proton Synchrotron (PS), and the Super Proton Synchrotron (SPS) up to energies of 1.4 GeV, 25 GeV, and 450 GeV, respectively. The packets of protons are then launched into the orbit of the gigantic Large Hadron Collider that has a circumference of 27 km. There are two vacuum pipes containing proton beams travelling in opposite directions. The co-

TABLE II: Relative force strength for protons in a nucleus.

Force	Relative Strength
Strong	1
Electromagnetic	10^{-2}
Weak	10^{-6}
Gravitational	10^{-38}

nunter rotating beams cross over in the four detectors (ATLAS [92], CMS [93], LHCb [94], ALICE [95]) caverns where they can be made to collide. It is the debris of this collisions that are tracked in the detectors. Remarkably, 70% of the energy carried into the collision by the protons emerges perpendicular to the incident beams. At a given transverse energy E_{\perp} , we may roughly estimate the resolution as $r \approx \hbar c/E_{\perp} \approx 2 \times 10^{-19} \text{ TeV m}/E_{\perp}$. For CM collisions of 13 TeV, we obtain a resolution of $2 \times 10^{-20} \text{ m}$ [96, 97]. (Throughout LHC8 refers to the run at $\sqrt{s} = 8 \text{ TeV}$ and LHC14 to the future run at $\sqrt{s} = 14 \text{ TeV}$.)

Now, an understanding of how the world is put together requires a theory of how quarks and leptons interact with one another. Equivalently, it requires a theory of the basic forces of nature. Four such forces have been identified. Two of the forces, gravitation and electromagnetism, have an unlimited range; largely for this reason they are familiar to everyone. They can be felt directly as agencies that pull or push. The remaining forces, which

are called simply the weak force and the strong force, cannot be perceived directly because their influence extends only over a short range, no larger than the radius of an atomic nucleus. The strong force binds together the quarks inside the protons, the neutrons, and various other particles generically called hadrons. Indirectly, it also binds protons and neutrons into atomic nuclei. The weak force is mainly responsible for the decay of certain particles. Its best-known effect is to transmute a down quark into an up quark, which in turn causes a neutron to become a proton plus an electron and a neutrino.

The forces can be characterized on the basis of the following four criteria: the types of particles that experience the force, the relative strength of the force, the previously alluded range over which the force is effective, and the nature of the particles that mediate the force. The electromagnetic force is carried by the photon, the strong force is mediated by gluons, the W and Z bosons transmit the weak force, and the quantum of the gravitational force is called the graviton. A comparison of the (approximate) relative force strengths for two protons inside a nucleus is given in Table II. Though gravity is the most obvious force in daily life, on a nuclear scale it is the weakest of the four forces and its effect at the particle level can nearly always be ignored.

The quarks are fractionally charged spin- $\frac{1}{2}$ strongly interacting objects which are known to form the composites collectively called hadrons [98–100]:

$$\left\{ \begin{array}{ll} q\bar{q} & (\text{quark} + \text{antiquark}) \\ qq\bar{q} & (\text{three quarks}) \end{array} \right. \quad \begin{array}{ll} \text{mesons} & \text{integral spin} \rightarrow \text{Bose-Einstein statistics [101, 102]} \\ \text{baryons} & \text{half-integral spin} \rightarrow \text{Fermi-Dirac statistics [103, 104]} \end{array} \quad (458)$$

There are six different types of quarks, known as flavors, forming three generations; their properties are given in Table III. (Antiquarks have opposite signs of electric charge.)

Quarks are fermions with spin- $\frac{1}{2}$ and therefore should obey the Pauli exclusion principle [105]; see Appendix E. Yet for three particular baryons ($\Delta^{++} = uuu$, $\Delta^{-} = ddd$, and $\Omega^{-} = sss$), all three quarks would have the same quantum numbers, and at least two quarks have their spin in the same direction because there are only two choices, spin up (\uparrow) or spin down (\downarrow). This would seem to violate the exclusion principle!

Not long after the quark theory was proposed, it was suggested that quarks possess another “charge” which enables them to interact strongly with one another. This “charge” is a three-fold degree of freedom which has come to be known as color [106], and so the field theory has taken on the name of quantum chromodynamics, or QCD. Each quark flavor can have three colors usually

designated red, green, and blue. The antiquarks are colored antired, antigreen, and antiblue. Baryons are made up of three quarks, one with each color. Mesons consist of a quark-antiquark pair of a particular color and its anticolor. Both baryons and mesons are thus colorless or white. Because the color is different for each quark, it serves to distinguish them and allows the exclusion principle to hold. Even though quark color was originally an *ad hoc* idea, it soon became the central feature of the theory determining the force binding quarks together in a hadron.

One may wonder what would happen if we try to see a single quark with color by reaching deep inside a hadron. Quarks are so tightly bound to other quarks that extracting one would require a tremendous amount of energy, so much that it would be sufficient to create more quarks. Indeed, such experiments are done at modern particle colliders and all we get is not an isolated quark, but more hadrons (quark-antiquark pairs or triplets). This prop-

erty of quarks, that they are always bound in groups that are colorless, is called confinement. Moreover, the color force has the interesting property that, as two quarks approach each other very closely (or equivalently have high energy), the force between them becomes small. This aspect is referred to as asymptotic freedom [107, 108]. When probed at small distances compared to the size of a hadron (i.e., about $1 \text{ fm} = 10^{-15} \text{ m}$) the “bare” masses of the quarks are those given in Table III. However, the effective quark masses in composite hadrons are significantly larger; namely, 0.3 GeV, 0.3 GeV, 0.5 GeV, 1.5 GeV and 4.9 GeV, for u , d , s , c , and b ; respectively. The lightest flavors are generally stable and are very common in the universe as they are the constituents of protons (uud) and neutrons (ddu). More massive quarks are unstable and rapidly decay; these can only be produced as quark-pairs under high energy conditions, such as in particle accelerators and in cosmic rays.

Leptons are fractionally spin- $\frac{1}{2}$ particles which do not strongly interact. There are six types of leptons, known as flavours, forming three generations (see Table III). Each charged flavor has an associated neutrino.

One important aspect of on-going research is the attempt to find a unified basis for the different forces. For example, the weak and electromagnetic forces are indeed two different manifestations of a single, more fundamental *electroweak* interaction [109–111]. The electroweak theory has had many notable successes, culminating in the discovery of the predicted Higgs boson [112, 113] by the ATLAS [114] and CMS [115] collaborations.

EXERCISE 13.1 The LHC beam is not continuous. Instead, it has a bunch structure. Bunches of particles will collide every 25 ns at $\sqrt{s} = 14 \text{ TeV}$ with a luminosity (number of particles per second per square centimeter) of $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$. The total inelastic proton-proton cross section at 14 TeV is 70 mb. (i) Compute the number of interactions occurring per second as well as every time two bunches collide. These are called minimum bias event and are of almost no interest. (ii) The Higgs particle is produced at the LHC mainly through gluon fusion ($gg \rightarrow H$) which has a cross section of about 40 pb. One of the Higgs discovery channels involves searching for the Higgs decaying to two photons which has a probability (branching ratio) of $\sim 0.2 \times 10^{-2}$. How many bias events are expected to be produced for every Higgs event observed via the two photon channel if one ignores detector effects?

The main properties of the four force carriers are summarized in Table IV. In the next sections we will discuss the properties of the forces in more detail, but the style coverage will remain qualitatively. The discussion will follow closely the heuristic approach of [116–120]. For a rigorous introduction to gauge theories of the strong and electroweak interactions, see e.g. [121–124].

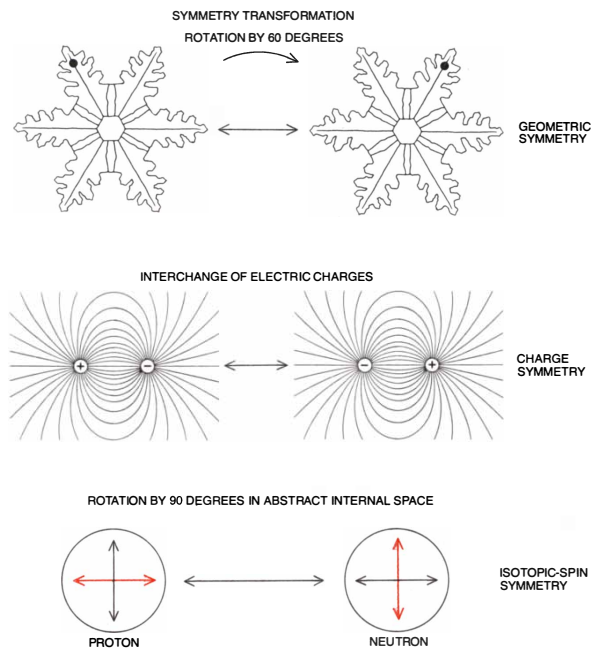


FIG. 47: Discrete and continuous symmetries in nature [118].

B. Gauge symmetries and field theories

We have seen that since the time of Galileo and Newton symmetries and apparent symmetries in the laws of nature have played an important role in the development of physical theories. The most recognizable symmetries are spatial or geometric ones. For instance, in a snowflake the existence of a symmetrical pattern can be detected at first sight. The symmetry can be defined as an invariance in the pattern that is observed when some transformation is enforced to it. For the snowflake, the transformation is a rotation by $\pi/3$, or else by $1/6$ of a circle. In plain English, if the initial position is identified (say, the black dot in the upper left branch of the snowflake shown in Fig. 47) and the snowflake is then rotated by $\pi/3$ (or by any integer multiple of $\pi/3$), no change will be noted. The snowflake is invariant with respect to rotations by $\pi/3$. Applying the same rule, a square is invariant with respect to rotations by $\pi/2$ and a circle is said to have continuous symmetry because rotation by any angle leaves it unchanged. The invariance of the circle is known in the physics vernacular as $U(1)$ symmetry; see Appendix F.

Despite the fact the notion of symmetry had its origin in geometry, it is general enough to encompass invariance with respect to transformations of other kinds. For example, a nongeometric symmetry is the charge symmetry of electromagnetism shown in Fig. 47. Assume a number of electrically charged particles have been arranged in some definite configuration and all the forces acting between pairs of particles have been measured. If the polarity of all the charges is then inverted, the forces remain unchanged.

Another symmetry with nongeometric origin reveals

TABLE III: The three generations of quarks and leptons in the Standard Model.

	Fermion	Short-hand	Generation	Charge	Mass	Spin
Quarks	up	u	I	$+\frac{2}{3}$	$2.3^{+0.7}_{-0.5}$ MeV	$\frac{1}{2}$
	charm	c	II		1.275 ± 0.025 GeV	
	top	t	III		173.21 ± 0.51 GeV	
	down	d	I	$-\frac{1}{3}$	$4.8^{+0.5}_{-0.3}$ MeV	$\frac{1}{2}$
	strange	s	II		95^{+5} MeV	
	bottom	b	III		4.18 ± 0.03 GeV	
Leptons	electron neutrino	ν_e	I	0	< 2 eV 95% CL	$\frac{1}{2}$
	muon neutrino	ν_μ	II		< 0.19 MeV 90% CL	
	tau neutrino	ν_τ	III		< 18.2 MeV 95%CL	
	electron	e	I	-1	0.511 MeV	$\frac{1}{2}$
	muon	μ	II		105.7 MeV	
	tau	τ	III		1.777 GeV	

TABLE IV: The four force carriers in the Standard Model.

Force	Boson	Short-hand	Charge	Mass	Spin
Electromagnetic	photon	γ	0	0	1
Weak	W	W^\pm	± 1	80.385 ± 0.015 GeV	1
Weak	Z	Z^0	0	91.1876 ± 0.0021 GeV	1
Strong	gluon	g	0	0	1

the isotopic spin, a property of neutrons and protons. The essence of this symmetry manifests in the observation that the neutron and the proton are rather similar particles. They differ in mass by only $\approx 10\%$, and aside for their electric charge they are identical in all other properties. Then it looks like all neutrons and protons could be interchanged and the strong interactions would hardly be altered. Should the electromagnetic forces (which depend on electric charge) could somehow be turned off, the isotopic-spin symmetry would be exact; in the real world, however, it is only approximate.

Despite the fact the neutron and proton seem to be distinct particles and it is hard to envisage a state of matter intermediate between them, it turns up that the symmetry with respect to isotopic spin is a continuous symmetry, like the symmetry of the circle rather than like that of a snowflake. We can provide a simplified explanation of why that is so. Contemplate that inside each particle there are imaginary black and red arrows crossing each other representing the proton and neutron component of the particle, respectively. As shown in Fig. 47, if the proton arrow is pointing up (it makes no difference what direction is defined as up) then the particle is a proton, and if the neutron arrow is up then it is a neutron. Intermediate positions correspond to quantum-mechanical superpositions of the two states, and the particle then looks sometimes like a proton and sometimes like a neutron. The symmetry transformation associated with isotopic spin rotates the internal indicators of all protons and neutrons everywhere in spacetime by the same amount and at the same time. If the rotation is

exactly by $\pi/2$, every proton mutates to a neutron and every neutron mutates to a proton. The isotopic spin symmetry, to the extent it is exact, states that no effects of this transformation can be detected.

So far we have discussed symmetries which can be defined as global symmetries. Herein the word global means “occurring everywhere at the same time.” In our characterization of isotopic-spin symmetry this constraint has been made explicit: the internal rotation that transforms protons into neutrons and neutrons into protons is to be carried out everywhere in spacetime at once. Supplementary to global symmetries, which are almost always present in a physical theory, it is desirable to have a “local” symmetry, in which the convention can be decided independently at every point in space and at every instant of time. Even though in general “local” could appear as something of more modest scope than “global”, in fact the requirement of local symmetry places a far more stringent constraint on the structure of a theory. A global symmetry states that some law of physics remains invariant when the same transformation is applied everywhere at the same time. For a local symmetry to be observed the law of physics must retain its validity even when a different transformation takes place at each point in space and time.

The so-called “gauge theories” can be assembled with either a global or a local symmetry (or both), but it is the theories with local symmetry that hold particular interest here. To construct a theory invariant with respect to a local transformation something new must be added: *a force*. Before showing how this comes about, however,

it is necessary to reexamine how forces are described in theories of elementary-particle interactions.

Nowadays, the building blocks of particle theory include not only particles and forces, but also fields. We have already introduced the concept of a field in Sec. III. A field is a quantity defined at every point throughout some region of space and time. For example, the quantity might be temperature and the region might be the surface of the Earth. Then the field consists of temperature values for every point on the Earth's surface.

The temperature is a scalar quantity, because it can be represented by position along a line, or scale. The associated temperature field is a scalar field, in which each point has correspondence with a single number, or magnitude. There are other types of field, such as the vector field, where at each point a vector, or arrow, is drawn. A familiar example of a vector field is the velocity field of a fluid; at each point throughout the volume of the fluid an arrow can be drawn to show the speed and direction of flow. Examples of vector and scalar fields are shown in Fig. 48.

When studying electrically charged objects a field is a handy mechanism for expressing how the force of electromagnetism is conveyed from one point to another. All charged particles are expected to emanate an electromagnetic field; each particle then interacts with the sum of all the fields rather than directly with the other particles.

We have seen that in quantum systems the particles themselves can be represented as fields. For example, the electron can be treated as a packet of waves with some finite extension in space. Conversely, it is often convenient to represent a quantum mechanical field as if it were a particle. The interaction of two particles through their interpenetrating fields can be analyzed as a two particles exchange a third particle, which is dubbed the quantum of the field. As an illustration, if two electrons, each surrounded by an electromagnetic field, approach near each other and bounce apart, we may conjecture they have exchanged a quantum of the electromagnetic field, i.e. a photon.

The exchanged quantum has only a fugacious existence. Shortly after it has been emitted it must be re-absorbed, either by the same particle or by another one, within a finite time. It cannot keep going forever, and it cannot be detected in an experiment. This kind of entities are called virtual particles. The larger their energy, the shorter their existence. In effect a virtual particle borrows or embezzles an amount of energy, but it must repay its debt before the shortage can be noticed. The range of an interaction is linked to the mass of the exchanged quantum. If the quantum field is heavy, more energy must be borrowed in order to support its existence, and the debt must be repaid sooner lest the discrepancy be discovered. The length a virtual particle can travel before it must be re-absorbed is thus reduced and so the associated force has a short range. For the particular case in which the exchanged quantum is massless, the range is infinite.

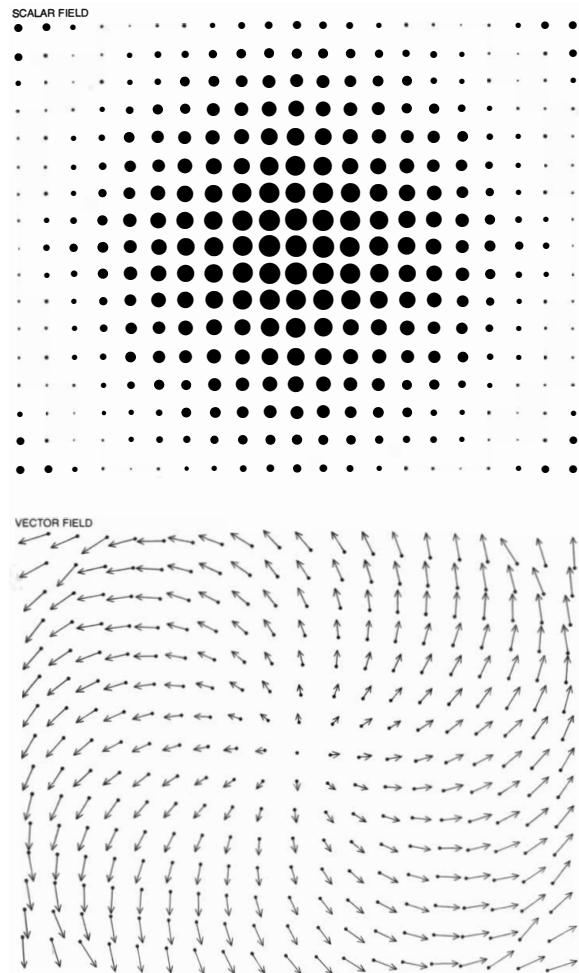


FIG. 48: A scalar field (top) has only a magnitude at each point; in the example shown here the magnitude is given by the area of the dots. A vector field (bottom) has both a magnitude and a direction and can be illustrated by drawing an arrow at each point. The number of components in the fields is reflected in the number of distinct orientations of the particle, which in turn depends on its spin angular momentum. A scalar field has just one component (its value can be given by a single number) and is represented by spin-0 particle with one spin state, or orientation. A vector field in 3-dimensional space has three components (a magnitude and two angles), and it corresponds to a spin-1 particle with three spin states [118].

The number of components in a field coincides with the number of quantum-mechanical states of the field quantum. The number of possible states is in turn associated to the particle's intrinsic spin angular momentum, which can only take discrete values, i.e. the measured magnitude of the spin in fundamental units is always an integer or a half integer. Furthermore, it is not only the magnitude of the spin that is quantized but also its direction or orientation. (As we have seen in Sec. XI, the spin can be defined by a vector parallel to the spin axis, and the projections, or components, of this vector along any direction

in space must have values that are integers or half integers.) The number of possible orientations (a.k.a. spin states) equals twice the magnitude of the spin, plus one. Hence, a particle with a spin- $\frac{1}{2}$, such as the electron, has two spin states. The spin can be oriented parallel to the particle's direction of motion or antiparallel to it. Using the arrow notation introduced in Sec. XI the spin orientation becomes up $|\uparrow\rangle$ or down $|\downarrow\rangle$. A spin-1 particle has three orientations: parallel, antiparallel, and transverse. A spin-0 particle has no spin axis, because all orientations are equivalent; it is said to have just one spin state. A scalar field that has only one component, its magnitude, must be characterized by a field quantum that has also one component: a spin-0 particle. Such a particle is then called a scalar particle. Likewise, a three-component vector field requires a spin-1 field quantum with three spin states: a vector particle. The electromagnetic field is a vector field, and the photon, in conformity with these specifications, has a spin-1. Actually, since the photon is massless and propagates at c it carries a property not shared by particles with finite mass: the transverse spin state does not exist. Even though in some formal interpretation the photon has three spin states, in practice only two of the spin states can be detected.

Electrodynamics describes how light and matter interact. It is the simplest field theory where full agreement between quantum mechanics and special relativity is achieved. We discuss this next.

C. Electromagnetic interaction

We have already noted that electromagnetism is a local field theory of fundamental interactions. The foundation of Maxwell's theory is the proposition that an electric charge is surrounded by an electric field \vec{E} stretching to infinity, and that the movement of an electric charge gives rise to a magnetic field \vec{B} also of infinite extent. Both \vec{E} and \vec{B} are vector quantities, which are defined at each point in space by a magnitude and a direction. In electromagnetism the value of \vec{E} at any given point is determined ultimately by the charge distribution around that point. However, in general it is convenient to define a potential V that is also determined by the charge distribution: the greater the density of charges in a region, the higher its potential. The electric field between two points is then given by the voltage difference between them.

The nature of the electromagnetic symmetry is evident in the following Gedanken-experiment. Assume a system of electric charges is arranged in a laboratory and we measure \vec{E} and its properties. If the charges are stationary there can be no \vec{B} -field. In such experimental footing a global symmetry is readily noted. The symmetry transformation consists in raising the entire laboratory to a high voltage, or in other words to a higher V . If the measurements are then repeated, no change in \vec{E} will be detected. The reason is simple, \vec{E} is determined

only by differences in electric potential ΔV , not by the absolute value of the potential. This property carries a symmetry: \vec{E} is invariant with respect to the addition or subtraction of an arbitrary overall potential. Note that this is a global symmetry, because the result of the experiment remains constant only if V is changed everywhere at the same time. If V were raised in one region and not in another, any experiment that crossed the boundary would be affected by the potential difference.

A complete theory of electromagnetic fields must contain not only static arrays of charges but also moving charges. To consolidate this, the global symmetry of the theory must be converted into a local symmetry. If the electric field were the only one acting between charged particles, it would not have a local symmetry. Indeed if the charges are in motion not only \vec{E} but also \vec{B} are at play. It is the effects of \vec{B} that restore the local symmetry. In the same way \vec{E} depends ultimately on the distribution of charges but can conveniently be derived from V , so the \vec{B} -field is generated by the motion of the charges but is more easily described as resulting from a magnetic potential \vec{A} . It is in this system of potential fields that local transformations can be carried out leaving all the original electric and magnetic fields unaltered. The system of dual, interconnected fields has an exact local symmetry even though \vec{E} alone does not. Any local change in V can be combined with a compensating change in \vec{A} in such a way that \vec{E} and \vec{B} are invariant.

Maxwell's theory of electromagnetism is a classical field theory, but a related symmetry can be demonstrated in the quantum theory of electromagnetic interactions. We have seen that in quantum mechanics massive particles must be characterized by a wave or a field. In particular, this convention can be adopted for electrons. In their quantum description a change in V yields a change in the phase of the electron wave. The electron is spin- $\frac{1}{2}$ particle and therefore it has two spin states (parallel and antiparallel). This implies that the associated field should have two components. Each component is characterized by a complex number. The electron field is a moving packet of waves, which are oscillations in the amplitudes of the real and the imaginary components of the field. At this stage, it is worthwhile to point out that this field is not the electric field emanating from the electron but instead is a massive field. It would exist even if the electron had no electric charge. What the field defines is the probability of finding the electron in a specified spin state at a given point in space and at a given instant in time. The probability is given by the sum of the squares of the real and the imaginary parts of the field.

We have seen that in the absence of electromagnetic fields the frequency of the oscillations in the electron field is proportional to the energy of the electron $E = \hbar\omega$, and the wavelength of the oscillations is proportional to the momentum $\lambda = h/p$. In order to completely define the oscillations an additional quantity must be determined: the phase. Recall that the phase measures the displace-

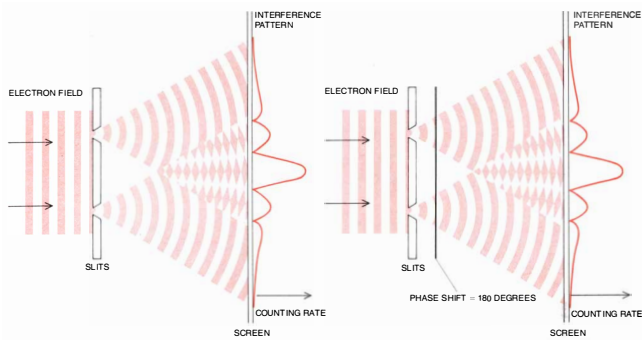


FIG. 49: Global gauge symmetry of electromagnetism [118].

ment of the wave from some arbitrary reference point and is generally expressed by the angle ϕ . If at some point the real part of the oscillation, say, has its maximum positive amplitude, then at that point $\phi = 0$. On the hand, where the real part falls to zero $\phi = \pi/2$, and where it reaches its negative maximum $\phi = \pi$. In general the imaginary part of the amplitude is out of phase by $\pi/2$ with the real part, so that whenever one part has a maximal value the other part is zero.

It is evident that the only way to determine ϕ is to disentangle the contributions of the real and the imaginary parts of the amplitude, which is a *mission impossible*. The sum of the squares of the real and the imaginary parts can be measured, but there is no way of resolving at any given point or at any moment how much of the total derives from the real part and how much from the imaginary part. As a matter of fact, an exact symmetry of the theory demands that the two contributions are indistinguishable. Phase differences of the field at two points or at two moments (of course) can be measured, but never the absolute phase. The conclusion that the phase of an electron wave is inaccessible to measurement has a corollary: *the phase cannot have an effect on the outcome of any conceivable experiment*. If it did so, such an experiment could be used to determine the phase. Hence the electron field exhibits a symmetry with respect to arbitrary changes of phase. In summary any phase angle ϕ can be added to or subtracted from the electron field and the results of all possible experiments will remain invariant.

We have dropped enough clues and hints that a really attentive reader might guess this symmetry principle can be best illustrated by considering a double-slit experiment of electrons. Consider again a beam of massive particles (electrons in this case) passing through two narrow slits in a screen, as shown in Fig. 49. The number of particles reaching a second screen located at a distance $D \gg d$ is counted. As expected, the distribution of electrons across the surface of the second screen forms an interference pattern of alternating peaks and valleys.

The quantum-mechanical interpretation of this experiment is that the electron wave splits into two segments on striking the first screen and the two waves then in-

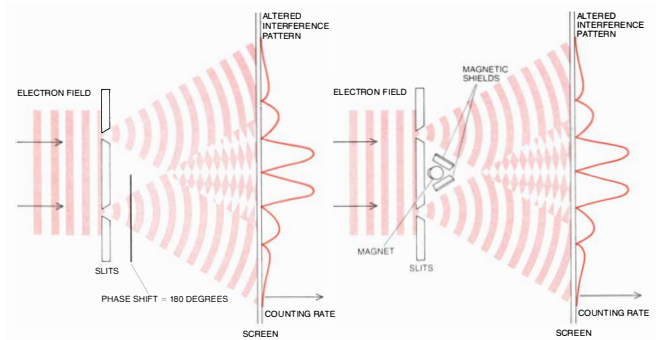


FIG. 50: Local gauge symmetry of electromagnetism [118].

terfere with each other. Again, if the waves are in phase the interference is constructive and many electrons are counted at the second screen. However, if the waves are out of phase a destructive interference reduces the count. Undoubtedly, it is only the difference in phase $\Delta\phi$ that regulates the pattern formed. In other words, if we shift both waves phases by the same amount, the phase difference at each point would be unaffected and the same pattern of constructive and destructive interference would be observed. This particular type of symmetries, in which the phase of a quantum field can be adjusted at will, are the so-called “gauge symmetries.” Despite the fact the absolute value of the phase is irrelevant to the outcome of experiments, in constructing a theory of electrons it is still necessary to specify such a phase. The choice of a particular value is known as a gauge-fixing condition.

The preceding electron symmetry is a global symmetry. Note that the phase of the field must be shifted in the same way everywhere in space at the same time. It is easily seen that a theory of electron fields alone, with no other forms of matter or radiation, is not invariant with respect to a corresponding local gauge transformation. Consider again the double-slit electron experiment. An initial experiment is carried out as before and the interference pattern is recorded. Then the experiment is repeated, but one of the slits is fitted with the electron-optical equivalent of a half-wave plate, a device that shifts the phase of a wave by π . When the waves emanating from the two slits now interfere, the phase difference between them will be altered by π . As a result wherever the interference was constructive in the first experiment it will now be destructive, and vice versa. The positions of all the peaks and depressions on the observed interference pattern will be interchanged; see Fig. 50.

Pretend we want to make the theory consistent with a local gauge symmetry, assuming e.g. another field could be added that would compensate for the changes in electron phase. The new field would of course have to accomplish more than restore the defects in this one experiment. It would have to preserve the invariance of all observable quantities when the phase of the electron field is altered in any way from point to point in space and from time to time. Strictly speaking, the phase shift

must be allowed to vary as an arbitrary function of position and time.

At first glance it may look improbable to find a field that can accommodate these specifications. However, on a second look we can figure out that the required field is a vector, corresponding to a field quantum with a spin of one unit. We can also immediately realize that the field must have infinite range, as there is no limit to the distance over which the phases of the electron fields might have to be reconciled. The need for infinite range entails that the field quantum must be massless. These are the characteristics of a familiar field: the electromagnetic field, whose quantum is the photon. How does the electromagnetic field protect the gauge invariance of the electron field? Recall that the effect of the electromagnetic field is to transmit forces between charged particles. These forces can modify the state of motion of the particles. Of particular interest here the forces can change the phase. When an electron absorbs or emits a photon, the phase of the electron field is shifted. We have already seen that \vec{E} itself exhibits an exact local symmetry; by describing the two fields together the symmetry can be extended to both of them.

The connection between the two fields relies in the interaction of \vec{E} with the charge of the electron. Because of this interaction the propagation of an electron wave through the \vec{E} field can be described properly only if V is specified. Likewise, to describe an electron in a magnetic field \vec{A} must be specified. Once these two potentials are assigned definite values the phase of the electron wave becomes fixed everywhere. However, the local symmetry of electromagnetism allows V to take any arbitrary value, which can be chosen independently at every point in space and at every instant of time. For this reason the phase of the electron matter field can also take on any value at any point, but the phase will always be consistent with the convention adopted for V and \vec{A} . This local structure can be reproduced in the double-slit experiment because the effects of an arbitrary shift in the phase of the electron wave can be mimicked by applying an electromagnetic field. As an illustration, the change in the observed interference pattern caused by interposing a half-wave plate in front of one slit could be generated instead by placing the slits between the poles of a magnet; see Fig. 50. By observing the resulting pattern it would be impossible to discern which procedure had been followed. Since the gauge-fixing condition for V and A can be chosen locally, so can the phase of the electron field. The theory that results from combining electron massive fields with electromagnetic fields is known as quantum electrodynamics (QED).

We have seen that quantum mechanics predicts probabilities of events. These probabilities must not be negative, and all the probabilities taken together must add up to one. In addition, energies must be assigned positive values but should not be infinite. It is not immediately apparent whether QED could satisfy all of these conditions. One challenge repeatedly shows up in any

attempt to calculate the probability that one particle will scatter off another. This is the case even in the simplest electromagnetic interaction, such as the interaction between two electrons. The likeliest sequence of events in Bhabha scattering is that one electron emits a single virtual photon and the other electron absorbs it. Many more complicated exchanges are also possible, however; indeed, their number is infinite. For example, the electrons could interact by exchanging two photons, or three, and so on. The total probability of the interaction is determined by the sum of the contributions of all the events. This sum can be most easily carried out using Feynman diagrams, i.e. by drawing diagrams of the events in one spatial dimension and one time dimension, which are shorthand representations of a well-defined mathematical procedure for tabulating all of these contributions [84–87]. A notably troublesome class of diagrams are those that include “loops,” such as the loop in spacetime that is formed when a virtual photon is emitted and later reabsorbed by the same electron. As was shown above, the maximum energy of a virtual particle is limited only by the time needed for it to reach its destination. When a virtual photon is emitted and reabsorbed by the same particle, the distance covered and the time required can be reduced to zero, and so the maximum energy can be infinite. For this reason some diagrams with loops make an infinite contribution to the strength of the interaction. These infinities would spoil even the description of an isolated electron: because the electron can emit and reabsorb virtual particles it has infinite mass and infinite charge. The cure for this plague of infinities is the procedure called renormalization. Roughly speaking, it works by finding one negative infinity for each positive infinity, so that in the sum of all the possible contributions the infinities cancel and a finite residue could be obtained. The finite residue is the theory’s prediction. It is uniquely determined by the requirement that all interaction probabilities come out finite and positive. The rationale of this procedure can be summarized as follows. When a measurement is made on an electron, what is actually measured is not the mass or the charge of the pointlike particle with which the theory begins but the properties of the electron together with its enveloping cloud of virtual particles. Only the net mass and charge, the measurable quantities, are required to be finite at all stages of the calculation. The properties of the pointlike object, which are called the “bare” mass and the “bare” charge, are not well defined.

The symmetry properties of QED are unquestionably appealing [125–131]. Moreover, QED has yielded results that are in agreement with experiment to an accuracy of about one part in a billion [132], which makes the theory the most accurate physical theory ever devised. It is the model for theories of the other fundamental forces and the standard by which such theories are judged.

EXERCISE 13.2 Consider that the fields \vec{E} and \vec{B}

are the components of a second rank tensor,

$$F^{\mu\nu} = \begin{pmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & -B_z & B_y \\ E_y & B_z & 0 & -B_x \\ E_z & -B_y & B_x & 0 \end{pmatrix}, \quad (459)$$

and that $j^\mu = (\rho, \vec{j})$ is a four-vector. Maxwell's equations of classical electrodynamics are, in vacuo,

$$\vec{\nabla} \times \vec{E} + \frac{\partial \vec{B}}{\partial t} = 0, \quad (460)$$

$$\vec{\nabla} \cdot \vec{E} = \rho, \quad (461)$$

$$\vec{\nabla} \times \vec{B} - \frac{\partial \vec{E}}{\partial t} = \vec{j}, \quad (462)$$

$$\vec{\nabla} \cdot \vec{B} = 0, \quad (463)$$

where we are using Heaviside-Lorentz rationalized units. (i) Show that in terms of the antisymmetric field strength tensor $F^{\mu\nu}$ Maxwell's equations can be written in the compact forms

$$\begin{aligned} \partial_\mu F^{\mu\nu} &= j^\nu \\ \partial^\alpha F^{\beta\gamma} + \partial^\beta F^{\gamma\alpha} + \partial^\gamma F^{\alpha\beta} &= 0, \end{aligned} \quad (464)$$

while the current conservation

$$\partial_\nu j^\nu = \frac{\partial \rho}{\partial t} + \vec{\nabla} \cdot \vec{j} = 0,$$

follows as a natural compatibility. (ii) Show that if the electric and magnetic fields can be written in terms of a four-vector potential $A^\mu = (V, \vec{A})$ as

$$\vec{E} = -\frac{\partial \vec{A}}{\partial t} - \vec{\nabla} V, \quad \vec{B} = \vec{\nabla} \times \vec{A}, \quad (465)$$

then the homogeneous Maxwell's equations (460) and (463) are trivially satisfied, while the inhomogeneous ones, (461) and (462), follow from the covariant equation

$$\square^2 A^\mu - \partial^\mu (\partial_\nu A^\nu) = j^\mu. \quad (466)$$

The tensor $F^{\mu\nu}$ can be written now as $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$. [Hint: note that $\vec{\nabla} \times (\vec{\nabla} \times \vec{A}) = -\nabla^2 \vec{A} + \vec{\nabla}(\vec{\nabla} \cdot \vec{A})$]. (iii) Verify that Maxwell's equations remain invariant under arbitrary Lorentz boosts.

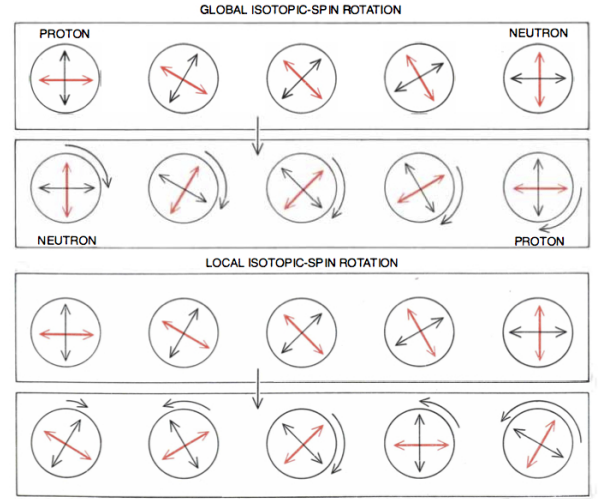


FIG. 51: Isotopic-spin symmetry [118].

D. Yang-Mills bonanza

In early 1954, in an attempt to explain the strong interaction, Yang and Mills generalized the symmetry concepts associated with QED [133]. The symmetry at play in the Yang-Mills theory is isotopic-spin symmetry, the guideline stating that the strong interactions of matter remain invariant (or nearly so) when the identities of neutrons and protons are interchanged. In the global symmetry any rotation of the internal arrows that pinpoint the isotopic-spin state must be made simultaneously everywhere. Postulating a local symmetry allows the orientation of the arrows to fluctuate independently from place to place and from time to time. In essence, rotations of the arrows can depend on any arbitrary function of position and time. This freedom to adopt different conventions for the identity of a nuclear particle in different places constitutes a local gauge symmetry. As in other examples where a global symmetry is converted into a local one, the invariance can be maintained only if something more is added to the theory. Because the Yang-Mills theory is more complicated than earlier gauge theories it turns out that a large number of degrees of freedom must be added. By and on itself, when isotopic-spin rotations are made arbitrarily from point to point in spacetime, the laws of physics remain invariant only if six new fields are introduced. They are all vector fields which have infinite range.

The Yang-Mills fields are constructed using electromagnetism as a prototype, and actually two of them can be identified with the ordinary electric and magnetic fields. In plain English, they describe the field of the photon. The remaining Yang-Mills fields can also be taken in pairs and interpreted as electric and magnetic fields, but the “photons” they describe differ in a central respect from the well known properties of the photon: they are still massless spin-1 particles, but they carry an

electric charge. One “photon” is positive and the other one negative. The wicked effects of “charged photons” become most apparent when a local symmetry transformation is applied more than once to the same particle. We have seen that in QED the symmetry operation is a local change in the phase of the electron field, each such phase shift being accompanied by an interaction with the electromagnetic field. It is simple to picture an electron undergoing two phase shifts in succession, say by absorbing a photon and later emitting one. Intuition guess that if the sequence of the phase shifts were reversed, so that first a photon was emitted and later one was absorbed, the end result would be the same. This is indeed the case. An unlimited series of phase shifts can be made, and the final result will be simply the algebraic sum of all the shifts no matter what their sequence. The symmetry of electromagnetism can be described by phase rotations in a circle, i.e. a $U(1)$ symmetry.

In the Yang-Mills theory, where the symmetry operation is a local rotation of the isotopic-spin arrow, the result of multiple transformations is indeed different. Assume a hadron is subjected to a gauge transformation \mathcal{A} , followed shortly by a second transformation \mathcal{B} ; at the end of the sequence the isotopic spin arrow is found in the orientation that corresponds to a proton. Now assume the same transformation were applied to the same hadron but in reverse sequence: \mathcal{B} followed by \mathcal{A} . Generically, the expected final state will not necessarily be the same, e.g. the particle may be a neutron instead of a proton. The net effect of the two transformations depends explicitly on the sequence in which they are applied. Because of this contrast QED is called an Abelian theory and the Yang-Mills theory is called a non-Abelian one. The terms are mooch from group theory. Abelian groups are made up of transformations that, when they are applied one after another, have the commutative property; non-Abelian groups are not commutative. A day-to-day example is the group of rotations. All possible rotations of a 2-dimensional object are commutative, and so the group of such rotations is Abelian. As an illustration, rotations by $\pi/3$ and $-\pi/2$ degrees yield a net rotation of $-\pi/6$ degrees no matter which is applied first. For a three-dimensional object free to rotate about three axes the commutative law does not hold, and the group of three-dimensional rotations is non-Abelian; an illustration is exhibited in Fig. 52.

The Yang-Mills theory has proved to be of paramount importance in the development of the Standard Model, but as it was originally formulated it does not describe the real world. A first objection to the theory is that isotopic-spin symmetry is exact, with the outcome that protons and neutrons are indistinguishable. This prediction is of course excluded by experiment. Even more troubling is the prediction of electrically charged “photons.” The photon is necessarily massless because it must have an infinite range. The existence of any electrically charged particle lighter than the electron would modify the world beyond recognition. Obviously, no such parti-

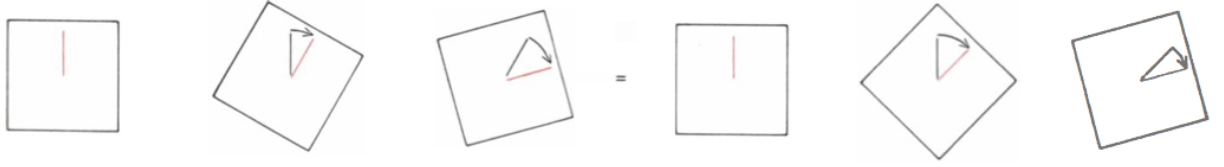
cle has been observed. One possible pathway to contest these problems is to artificially endow the charged field quanta with a mass greater than zero. Note that by imposing a mass on the quanta of the charged fields we can confine them to a finite range. If the mass is large enough, the range can be made as small as is desired. If the long-range effects are removed then the existence of the charged fields can be easily reconciled with experimental observations. Furthermore, by selection the neutral Yang-Mills field as the only real long-range we can automatically distinguish protons from neutrons. Since the long-range field is simply the electromagnetic field, the proton and the neutron can be distinguished by their differing interactions with it, or expressly by their differing electric charges. After introducing this adjustment the local symmetry of the Yang-Mills theory would no longer be exact but approximate, since rotation of the isotopic-spin arrow would now have observable consequences. That is not a radical objection: approximate symmetries are usually commonplace in nature, e.g. the bilateral symmetry of the human body is only approximate. What is more, at distance scales much smaller than the range of the massive components of the Yang-Mills field, the local symmetry becomes better and better. Therefore, in principle the microscopic structure of the theory could remain locally symmetric, but not its predictions of macroscopic, observable events. The Yang-Mills theory begun as a model of the strong interactions, but as we discuss below, the real interest in it centers on applications to the weak interactions.

E. Parity violation

Parity says something about the symmetry of a system. The parity transformation inverts an object completely (back to front, side to side, and top to bottom) by the simultaneous flip in the sign of all three spatial coordinates (x, y, z) to $(-x, -y, -z)$, i.e. by a point reflection. As an illustration, in Fig. 53, we show a parity transformation, which is equivalent to a reflection (in the $x - y$ plane) together with a rotation through 180° (around the z axis). Note that during the transformation a right-handed system changes to a left-handed system. As a consequence, a point reflection can also be thought of as a test for chirality of a physical phenomenon, in that a parity inversion transforms a phenomenon into its mirror image.

In classical physics, space inversion is merely a geometrical transformation, a rule to map each point in space to its image by inversion through a chosen origin. The time is left unaffected. A particle trajectory, for example, would be mapped onto another possible trajectory. Wigner showed that in quantum mechanics, parity is more than a transformation, it is a physical observable whose value can be experimentally measured [134]. In order to gain an understanding of parity in quantum systems, let us examine the symmetry properties of the

ABELIAN TRANSFORMATION



NON-ABELIAN TRANSFORMATION

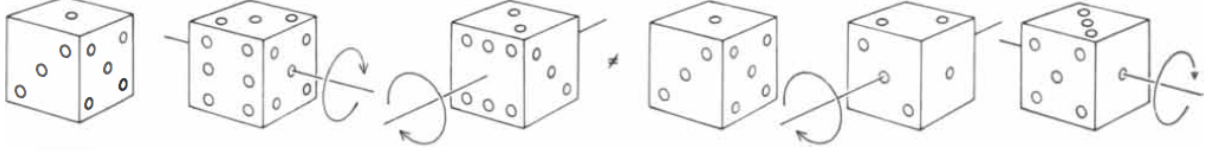


FIG. 52: Effects of repeated Abelian and non-Abelian transformations [118].

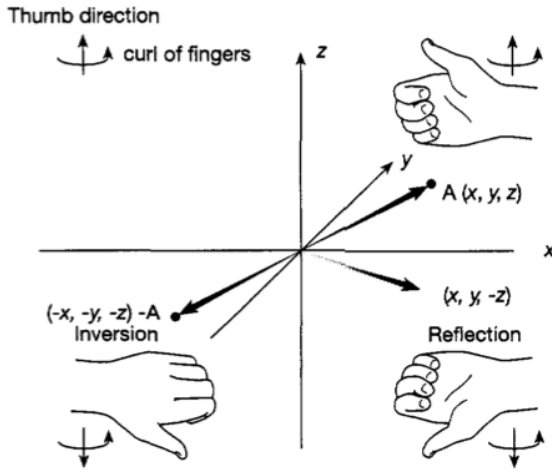


FIG. 53: Parity transformation.

wavefunction $\psi(\vec{x}, t)$ under space inversion. We define a new operator, \hat{P} which transforms $\psi(\vec{x}, t)$ into $\psi(-\vec{x}, t)$, i.e.

$$\hat{P} \psi(\vec{x}, t) = \psi(-\vec{x}, t). \quad (467)$$

If the system is symmetric ($\psi(\vec{x}, t) = \psi(-\vec{x}, t)$) or antisymmetric ($\psi(\vec{x}, t) = -\psi(-\vec{x}, t)$) then it obeys a simple eigenvalue problem when operated on by

$$\hat{P} \psi(\vec{x}, t) = P \psi(\vec{x}, t), \quad (468)$$

where $P \in \{-1, 1\}$ is the parity of the system. The possible results of measurements are then ± 1 , and the corresponding quantum states are said to possess even or odd parity, respectively.

Parity has three important properties which we should state before we proceed: (i) Parity is a multiplicative

quantum number. This means that the quantity that is conserved in parity conserving interactions is not the sum of the parities of the constituent particles $\sum_j P_j$, but is instead the product of these constituent parities $\prod_j P_j$. This is actually true for any operator \hat{O} which, when applied twice, does nothing: $\hat{O}^2 = \mathbb{1}$. The parity operator, which performs a mirror reflection $\vec{x} \rightarrow -\vec{x}$, if applied twice, does nothing to the original system, and therefore fits this rule. (ii) Besides the parity connected to its spatial state, a particle a has an intrinsic parity P_a . The total parity of the particle is the product of the intrinsic and spatial parities. If the overall wavefunction of a particle (or system of particles) contains spherical harmonics then we must take this into account to get the total parity of the particle (or system of particles). For a wavefunction containing spherical harmonics, it is easily seen that the spatial parity is given by:

$$\hat{P} \psi(r, \theta, \phi) = (-1)^l \psi(r, \theta, \phi). \quad (469)$$

The total parity of a state consisting of particles a and b is then $(-1)^L P_a P_b$ where L is their relative orbital momentum, and P_a and P_b are the intrinsic parity of the two particles (iii) Vector quantities change sign under a parity transformation, e.g. $\vec{p} \rightarrow -\vec{p}$. However, axial vectors (a.k.a. pseudovectors) like the angular momentum, does not switch sign: $\vec{L} = \vec{r} \times \vec{p} \rightarrow \vec{L} = -\vec{r} \times -\vec{p}$. In this direction, it is interesting to see what happens if we take the scalar product of \vec{S} (which is an axial vector) and the unit vector in the direction of the momentum \hat{p} (which is a vector), i.e. $\lambda = \vec{S} \cdot \hat{p}$. Usually scalars do not change sign under reflection but, surprisingly, we can clearly see that, since \hat{p} changes sign but \vec{S} does not, $\lambda \rightarrow -\lambda$. This is not a usual scalar, it is actually a pseudoscalar. This particular pseudoscalar, the projection of the spin onto the direction of a particle's momentum is called the helicity of the particle. For example, for a spin- $\frac{1}{2}$ particle,

the eigenvalues of the helicity operator are

$$\lambda = \begin{cases} +\frac{1}{2} & \text{positive helicity, } \Rightarrow \\ -\frac{1}{2} & \text{negative helicity, } \Leftarrow \end{cases} \quad (470)$$

Left handed particles are defined as having a negative helicity while right handed particles are defined as having a positive helicity.

EXERCISE 13.3 Convince yourself that wavefunctions containing spherical harmonics satisfy (469), i.e., $\hat{P} [R(r) Y_m^l(\theta, \phi)] = (-1)^l R(r) Y_m^l(\theta, \phi)$.

The parity of the neutron, the proton, and the electron is conventionally taken to be equal to +1. Let us assume that parity is conserved in the strong interaction. No experimental indication is reported so far to sustain the contrary. We can determine the intrinsic parity of the pion by studying pion capture by a deuteron, $\pi^- d \rightarrow nn$. The pion is known to have spin-0, the deuteron spin-1, and the neutron spin- $\frac{1}{2}$. The internal parity of the deuteron is +1. The pion is captured by the deuteron from the ground state, implying $l = 0$ in the initial state. So the total angular momentum quantum number of the initial state is $j = 1$. The parity of the initial state is $(-1)^0 P_\pi P_d = (-1)^0 P_\pi P_d = P_\pi$. The parity of the final state is $P_n P_n (-1)^l = (-1)^l$. Because neutrons are identical fermions, the only allowed state has spin oriented in opposite directions. To conserve angular $j = 1$ which requires $l = 1$, and so $P_\pi = -1$.

It is the role of parity in quantum mechanics that was shown by Wigner to be the explanation for Laporte's selection rule in atomic spectroscopy: atomic states undergoing photon absorption or emission always end up, after the transition, in a final state of opposite parity [135]. We can use the properties defined above to show that this rule is simply a statement of the law of conservation of parity. The intrinsic parities of both the proton and the electron are +1; actually, by convention all fermions have even parity. However, the electron state also has a spatial parity of $(-1)^l$, where l is the orbital angular momentum. Thus, the total parity of the hydrogen atom is $P_{H_i} = (+1)(+1)(-1)^l = (-1)^l$. Using Laporte's rule, we know that the final state of an atom which has undergone photon absorption or emission will have a parity opposite to its original state. Thus, the wavefunction of the hydrogen atom will have a parity of $P_{H_f} = (-1)P_{H_i} = (-1)^{l+1}$ after the transition. Including the intrinsic parity of the photon, which is defined as -1, we see that the total parity of the system in its final state is $P_{\text{tot}_f} = P_{H_f} P_\gamma = (-1)^{l+2} = (-1)^l = P_{H_i}$. Thus we see that Laporte's rule (the atom's wavefunction must change parity after photon emission or absorption) is nothing more than a statement of the conservation of parity in electromagnetic interactions.

In the early 1950's there was an "odd" experimental observation: two particles with identical mass, spin, charge, lifetime, etcetera decayed (weakly) into states of

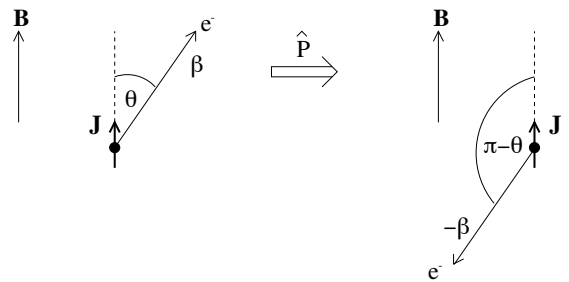


FIG. 54: The Wu experiment [138].

opposite parity:

$$\begin{aligned} \theta^+ &\rightarrow \pi^+ \pi^0 & (P_\theta = +1) \\ \tau^+ &\rightarrow \begin{cases} \pi^+ \pi^0 \pi^0 \\ \pi^+ \pi^+ \pi^- \end{cases} & (P_\tau = -1). \end{aligned} \quad (471)$$

There are two radical hypotheses that can explain the θ - τ puzzle: (i) there are two particles with identical properties except for parity; (ii) parity is not conserved in the weak interaction. Lee and Yang soon realized that if parity is violated by the weak force then reflection symmetry would also be broken, implying that the average value of some pseudoscalars would not be zero in weak interactions [136].

EXERCISE 13.4 Assuming parity conservation is an exact symmetry of nature and that the quarks and antiquarks of the mesons do not have any relative orbital angular momentum, determine the parities of τ^+ and θ^+ [137]. [Hint: The mesons all have well-defined parity values which turn out to be -1, i.e. they are parity eigenstates $\hat{P}|\pi^+\rangle = -|\pi^+\rangle$; along with being spin 0.]

The classical experiment which actually established that parity is not a good symmetry of nature operates with the archetype of weak processes; namely, the β -decay of $^{60}\text{Co} \rightarrow ^{60}\text{Ni} e^- \bar{\nu}_e$ [136, 138]. This is a weak interaction as it involves a neutrino, which has no electromagnetic or strong charge. The cobalt nuclei were cooled to temperatures around 0.01 K and placed in a magnetic field. At such low temperatures, the magnetic moments of the nuclei cause the nuclear spins to align themselves with the magnetic field direction, so the nuclei are all polarised along the field direction. The experiment then simply consisted of measuring the angular distribution of the emitted electrons. The pseudoscalar to be measured in this case is the projection of the electron's momentum on the spin \vec{J} of the cobalt nucleus. Consider a system where \vec{J} is lined up on the $+z$ axis with a mirror at the $x-y$ plane as shown in Fig. 54. In the reflection of the system \vec{J} still points in the same direction as it had before since it is a pseudovector. On the other hand, the momentum of an electron \vec{p}_e which had been at an angle of θ relative to the $+z$ axis is at an angle $\pi - \theta$ in the reflection, because momentum is a vector. Accordingly,

the correlation $\vec{J} \cdot \hat{p}_e$ is manifestly parity violating. If the process obeys reflection symmetry then electron will be just as likely to be emitted at an angle θ as they would be at $\pi - \theta$. Does this make sense with regard to the magnetic field? Yes; since $\vec{B} = \vec{\nabla} \times \vec{A}$, then \vec{B} is itself an axial vector like \vec{J} . Hence, the nuclear spin is still aligned along the magnetic field, as this alignment is due to the electromagnetic force which conserves parity. The angular distribution of the emitted electrons is found to be $1 - \beta \cos \theta$, where $\beta = v/c$ is the electron velocity, meaning more electrons were emitted in a direction opposite the magnetic field than along it (see Appendix G). In the parity inverted experiment, the electron comes off at $\pi - \theta$ so if parity is a symmetry, then the rate for electrons at θ must equal the rate at $\pi - \theta$. However, a distribution of $1 - \beta \cos \theta$ clearly does not have the same rate, since $1 - \beta \cos(\pi - \theta) = 1 + \beta \cos \theta \neq 1 - \beta \cos \theta$. Hence, this simple observation is all that is required to conclude that parity is not a symmetry of the weak interactions. One thing to note is the velocity dependence. For the electrons emitted with $v \rightarrow 0$, the distribution becomes isotropic, which clearly then is parity symmetric. In contrast, for electrons with $v \rightarrow c$, the maximum parity violation occurs.

Another symmetry that we might expect nature to respect is the charged conjugation C, i.e. swapping the sign of all charges. For example, if we suddenly swapped the charge of every proton and electron in the universe, we would expect nothing to change since the force between them would be the same. Electromagnetism, gravity and the strong interaction all obey the C-symmetry. However, it is easily seen that weak interactions violate the C-symmetry.

F. Electroweak interaction

Experiments in the late 1950's established that (charged-current) weak interactions are left-handed, and motivated the construction of a manifestly parity-violating theory of the weak interactions. The electroweak theory takes three crucial clues from experiment: (i) the existence of left-handed weak-isospin doublets,

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix} \quad \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix} \quad \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix} \quad (472)$$

and

$$\begin{pmatrix} u \\ d \end{pmatrix} \quad \begin{pmatrix} c \\ s \end{pmatrix} \quad \begin{pmatrix} t \\ b \end{pmatrix}; \quad (473)$$

(ii) the universal strength of the weak interactions; (iii) the idealization that neutrinos are massless. To incorporate electromagnetism into a theory of the weak interactions, we add to the $SU(2)_L$ isospin family symmetry (suggested by the first two experimental clues)

TABLE V: Weak-isospin and hypercharge quantum numbers.

Lepton	T	T^3	Q	Y	Quark	T	T^3	Q	Y
ν_e	$\frac{1}{2}$	$\frac{1}{2}$	0	$-\frac{1}{2}$	u_L	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{1}{6}$
e_L^-	$\frac{1}{2}$	$-\frac{1}{2}$	-1	$-\frac{1}{2}$	d_L	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{3}$	$\frac{1}{6}$
					u_R	0	0	$\frac{2}{3}$	$\frac{2}{3}$
e_R^-	0	0	-1	-1	d_R	0	0	$-\frac{1}{3}$	$-\frac{1}{3}$

a $U(1)_Y$ weak-hypercharge phase symmetry (see Appendix F). The electroweak theory then implies two sets of gauge fields a weak isovector, \vec{W} , and a weak isoscalar B [109–111]. From the fermionic point of view, the electroweak gauge symmetry $SU(2)_L \otimes U(1)_Y$ is chiral: the left-handed and the right-handed fermions form different types of multiplets. Specifically, all the left-handed quarks and leptons form $SU(2)$ doublets, whereas the right-handed quarks and leptons are $SU(2)$ singlets. As a consequence, the weak interactions do not respect the parity or the charge-conjugation symmetries. Note that the particle-antiparticle conjugation (456) does not alter the chirality of particles. Therefore, a left-handed neutrino would be taken by charge conjugation into a left-handed antineutrino, which does not interact in the Standard Model. This property is what is meant by the “maximal violation” of C-symmetry in the weak interaction. We define the electric charge quantum number as $Q = T^3 + Y$, where the weak-isospin and hypercharge quantum numbers of leptons and quarks are given in Table V [121].

A central question of the Standard Model is why the electroweak forces are asymmetrical: electromagnetism is long-ranged, whereas the weak nuclear force is short-ranged (less than 10^{-17} m). According to Heisenberg's uncertainty principle, this limited range implies that the force particles must have a mass approaching \sim GeV energies. The answer is that both these forces are actually symmetrical, but their symmetry is hidden, or “broken.” A simple analogy is provided by the familiar phenomenon of ferromagnetism. The equations governing the electrons and iron nuclei in a bar of iron obey rotational symmetry, so that the free energy of the bar is the same whether one end is made the north pole by magnetization or the south. Strictly speaking, the free energy of a ferromagnet is related to its magnetization M by

$$G = \alpha M^2 + \beta M^4. \quad (474)$$

At high temperatures the curve of energy versus magnetization has a simple U shape ($\alpha > 0$ and $\beta > 0$) that has the same rotational symmetry as the underlying equations, see Fig. 55. The equilibrium state, the state of lowest energy at the bottom of the U , is also a state of zero magnetization, which shares this symmetry. On the other hand, when the temperature is lowered, the lowest point on the U -shaped curve humps upward so that the curve resembles a W with rounded corners ($\alpha < 0$ and

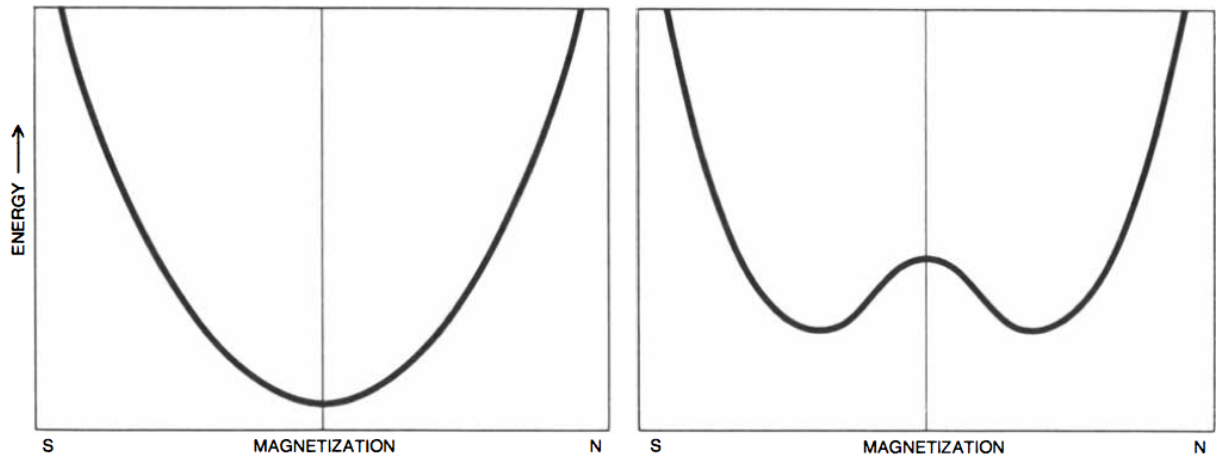


FIG. 55: Free energy as a function of the magnetization for a bar magnet at high temperature (left) or at low temperature (right). The magnet naturally seeks a state of minimum free energy. At high temperature this is a state of zero magnetization, a state that exhibits perfect symmetry between north and south. At low temperature the equilibrium state shifts to one of nonzero magnetization, which can be either north or south, even though the free energy curve is still perfectly symmetrical between north and south. In this case we say that the symmetry is spontaneously broken [117].

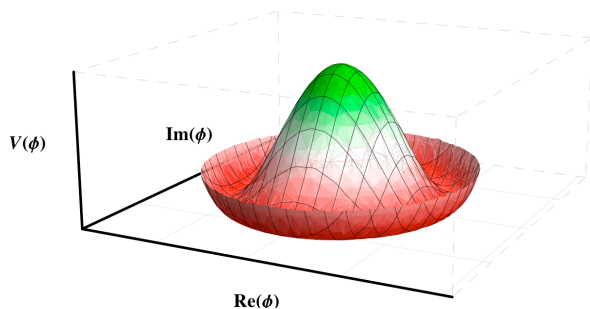


FIG. 56: Mexican hat potential, $V(\phi) = \mu^2(\phi^*\phi) + |\lambda|(\phi^*\phi)^2$, of complex scalar field ϕ [139].

$\beta > 0$), see Fig. 55. The curve still has the same rotational symmetry as the underlying equations, but now the equilibrium state has a definite nonzero magnetization, which can be either north or south but which in either case no longer exhibits the rotational symmetry of the equations. We say in such cases that the symmetry is spontaneously broken. The spontaneous symmetry breaking then describes systems where the equations of motion obey certain symmetries, but the lowest energy solutions do not exhibit that symmetry.

To approach our goal of generating a mass for the gauge bosons we first note that the Mexican hat potential of Fig. 56 has a circle of minima. The minima of this system are thus degenerate; there are multiple states with the same vacuum energy. The different orientations in the complex plane define different states. The orientation of these states is comparable to the direction of alignment of the spins in the ferromagnet. Note that the potential is symmetric under rotations in ϕ space, but the ground states are asymmetric. Applying the $U(1)$ trans-

formation to any of the vacuum states will rotate it to a different orientation that describes a different physical state.

In the electroweak theory the Higgs field brings about the symmetry breaking [112, 113]. The process is analogous to that of the ferromagnet, but it requires more dimensions than we can comfortably draw. The idea is that there are originally four Higgses which are all parts of a single “complex Higgs doublet”

$$\phi \equiv \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}, \quad (475)$$

with weak hypercharge $Y_\phi = +1/2$. The scalar potential is given by

$$V(\phi) = \mu^2(\phi^\dagger\phi) + |\lambda|(\phi^\dagger\phi)^2. \quad (476)$$

The electroweak symmetry is spontaneously broken if the parameter $\mu^2 < 0$. In the unified theory, where electroweak symmetry is unbroken, these four Higgses can be rotated into one another and the physics does not change. However, when the system rolls into the bottom of the Mexican hat one of the Higgses obtains a vacuum expectation value (VEV)

$$\langle\phi\rangle_0 = \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix}, \quad (477)$$

while the others do not, where $v = \sqrt{-\mu^2/|\lambda|}$. Performing a “rotation” then moves the VEV from one Higgs to the others and the symmetry is broken as the four Higgses are no longer being treated equally; see Appendix H. The electroweak symmetry makes all the electroweak force particles massless. The broken symmetry gives masses

to the gauge bosons, thereby restricting their range. The W_1 and W_2 combine into the W^+ and W^- by eating the charged Higgses. A similar story goes through for the W_3 and B , which combine and eat the neutral Higgs to form the massive Z boson. Meanwhile, the photon is the leftover combination of the W_3 and B . There are no more Higgses to eat, so the photon remains massless. The apparent extra degree of freedom is actually spurious, because it corresponds only to the freedom to make a gauge transformation. Such a transformation leads to the massive scalar H , usually referred to as the Higgs boson.

Fermion masses arise from Yukawa interactions, which couple the right-handed fermion singlets to the left-handed fermion doublets and the Higgs field [140]. In the process of spontaneous symmetry breaking these interactions lead to charged fermion masses, but leave the neutrinos massless.

The electroweak symmetry is more abstract than the symmetry of the ferromagnet. It means the freedom to decide which leptons are electrons and which are neutrinos or how to label up and down quarks. In the symmetrical case, the lepton-naming convention is set independently at each point in space. What one person calls an electron, another might call some mixture of electron and neutrino, and it would make no difference to their predictions. In the broken symmetry, the convention is fixed everywhere. What one person calls an electron, all do.

G. Strong interaction

The development of a successful gauge theory of the strong interaction, which is unique to hadrons, cannot not be undertaken until an inherent property about the hadrons is understood: they are not elementary particles. A model of hadrons as composite objects was proposed independently by Gell-Mann and Ne'eman [98–100]. In this model hadrons are composed of point-like particles named quarks. A hadron can be made up of quarks according to either of archetypes in (458). Binding together three quarks leads to a baryon, a class of hadrons that includes the neutron and the proton. Combining one quark and one antiquark makes a meson, a class typified by the pions. Every known hadron can be accounted for as one of these allowed composites of quarks. The keystone of any theory of the strong interactions is to explain the peculiar rules for building hadrons out of quarks. The structure of a meson is not so hard to account for: since the meson is made out of a quark and an antiquark, it is merely necessary to assume that the quarks carry some property analogous to electric charge. The binding of a quark and an antiquark would then be explained on the principle that opposite charges attract, just as they do in electromagnetism. However, the structure of the baryons is a far profound enigma. To describe how three quarks can produce a bound state we must assume that three like charges attract.

The theory that describes the strong force, QCD, lays down exactly these interactions. The analogue of electric charge is a property called color. The term color refers to the fact that the rules for producing hadrons can be expressed succinctly by requiring all allowed combinations of quarks to be white, or colorless. The quarks are assigned the primary colors red, green and blue, whereas the antiquarks have the complementary anticolors cyan, magenta and yellow. Each of the quark flavors comes in all three colors, so that the introduction of the color charge triples the number of distinguishable quarks.

There are two prescriptions of producing the white color with the available quark pigments: by mixing all three primary colors or by mixing one primary color with its complementary anticolor. The baryons are made according to the first prescription: the three quarks in a baryon are required to have different colors. In a meson a color is always accompanied by its complementary anticolor.

QCD is a non-Abelian gauge theory. The gauge symmetry is an invariance with respect to local transformations of quark color. It is easy to visualize a global color symmetry. The quark colors, like the isotopic-spin states of hadrons, could be indicated by the orientation of an arrow in some imaginary internal space. Successive rotations of $\frac{1}{3}$ of a circle would change a quark from red to green to blue and back to red again. Thus, in a baryon there would be three arrows with one arrow set to each of the three colors. By definition a global symmetry transformation must affect all three arrows in the same way and at the same time. For example, as shown in Fig. 57 all three arrows might rotate clockwise a third of a turn. As a result of such a transformation all three quarks would change color, but all observable properties of the hadron would remain as before. In particular there would still be one quark of each color, and so the baryon would remain colorless.

QCD requires that the color invariance be retained even when the symmetry transformation is a local one. In the absence of forces or interactions the invariance is of course lost. Hence a local transformation can change the color of one quark but leave the other quarks unaltered, which would give the hadron a net color. As in QED, we restore the invariance with respect to local symmetry operations by introducing new fields. In QCD the fields needed are analogous to the electromagnetic field but are much more complicated; they have eight times as many components as the electromagnetic field has. It is these fields that govern the strong force.

The quanta of the color fields are called gluons (as they glue the quarks together). There are eight of them, they are all massless, and have a spin-1; i.e. they are massless vector bosons like the photon. In addition, the gluons are electrically neutral like the photon, but they are not color-neutral. Each gluon carries one color and one anticolor. There are nine possible combinations of a color and an anticolor, but one of them is equivalent to white and is excluded, leaving eight definite gluon fields.

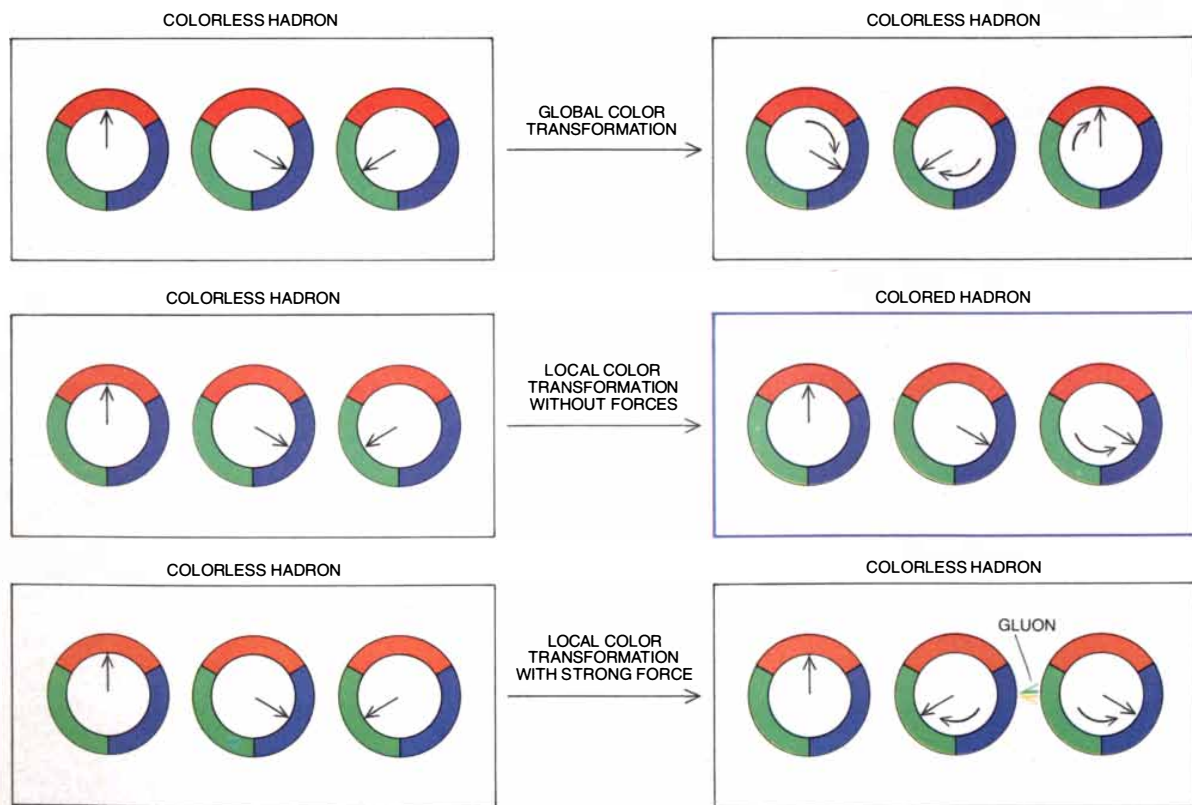


FIG. 57: Color symmetry [118].

The gluons preserve local color symmetry as follows: a quark is allowed to change its color, and it can do so independently of all other quarks, but every color transformation must be accompanied by the emission of a gluon, just as an electron can shift its phase only by emitting a photon. The gluon, propagating at the speed of light, is subsequently absorbed by another quark, which will have its color shifted in exactly the way needed to compensate for the original change. An example is shown in Fig. 57: a green quark changes its color to blue and in the process emits a gluon that bears the colors green and antiblue. The gluon is then absorbed by a blue quark, and in the ensuing reaction the blue of the quark and the antiblue of the gluon annihilate each other, leaving the second quark with a net color of green. As a consequence, in the final state as in the initial state there is one green quark and one blue quark. Because of the continual arbitration of the gluons there can be no net change in the color of a hadron, even though the quark colors vary freely from point to point, see Fig. 58. All hadrons remain white, and the strong force is nothing more than the system of interactions needed to maintain that condition.

Notwithstanding the complex structure of the gluon fields, QED and QCD are remarkably similar in form. Most notably, we have seen that the gluon and the photon are identical in their spin and in their lack of mass and electric charge. It is hence intriguing that quarks

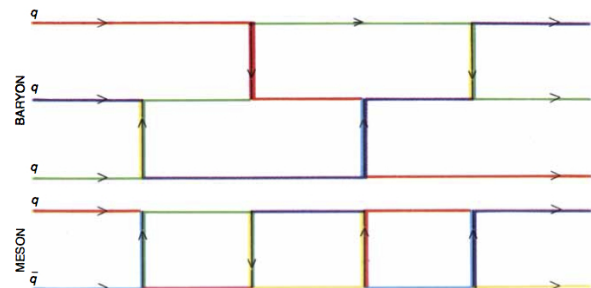


FIG. 58: Gluon exchange [118].

interactions are very different from those of electrons. Both electrons and quarks form bound states: electrons form atoms and quarks form hadrons. Electrons, however, are also observed as independent particles; a small quantity of energy suffices to isolate an electron by ionizing an atom. An isolated quark has never been detected. It seems to be impossible to ionize a hadron, no matter how much energy is supplied. The quarks are evidently bound so tightly that they cannot be pried apart; paradoxically, however, probes of the internal structure of hadrons show quarks moving freely, as if they were not bound at all. Gluons too have not been seen directly in experiments.

The resolution to this conundrum was developed not

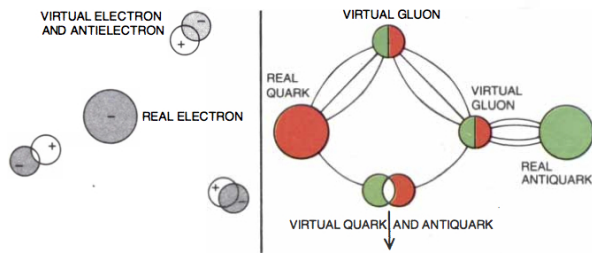


FIG. 59: Polarization of the vacuum [118].

by modifying the color fields but by examining their properties in greater detail. We have seen that an isolated electron is surrounded by a cloud of virtual e^+e^- pairs, which it constantly emits and reabsorbs; see Fig. 59. It is the charged virtual particles in this cloud that under ordinary conditions camouflage the “infinite” negative bare charge of the electron. In the vicinity of the bare charge the e^+e^- pairs become slightly polarized: the virtual positrons, under the attractive influence of the bare charge, stay closer to it on the average than the virtual electrons, which are repelled. As a result the bare charge is partially neutralized; what we observe at long range is the difference between the bare charge and the screening charge of the virtual positrons. Only when a probe approaches to within less than about 10^{-10} cm do the unscreened effects of the bare charge become significant.

It is feasible to guess the same process would operate among color charges, and indeed it does. A red quark is surrounded by pairs of quarks and antiquarks, and the antired charges in the cloud are attracted to the central quark and tend to screen its charge. However, in QCD there is a competing effect which is not present in QED. Whereas the photon carries no electric charge and therefore has no direct influence on the screening of electrons, gluons do bear an $SU(3)_C$ color charge. This distinction expresses the fact that QED is a $U(1)$ Abelian theory and QCD is an $SU(3)$ non-Abelian one (see Appendix F). Virtual gluon pairs also encompass a cloud around a colored quark (see Fig. 59), but it turns out that the gluons tend to enhance the color charge rather than attenuate it. This can be understood visualizing that the red component of a gluon is actually attracted to a red quark and hence could add its charge to the total effective charge. If there are no more than 16 flavors of quark (and as of today only 6 have been observed), the “antiscreening” by gluons is the dominant component. As a result of this “antiscreening” the effective color charge of a quark grows larger at long range than it is close by. A distant quark reacts to the combined fields of the central quark and the reinforcing gluon charges; at close range, once the gluon cloud has been penetrated, only the smaller bare charge is effective. The quarks inside the hadron thus move as if they were connected by rubber bands: at very close range, where the bands are feeble, the quarks move

almost independently, but at a large distances, where the bands are stretched tense, the quarks are tightly bound [107, 108]. Therefore, quarks and gluons could be permanently confined in hadrons. By its very nature the color symmetry is an exact symmetry and the colors of particles are completely indistinguishable. The theory is a pure gauge theory of the kind first proposed by Yang and Mills. The gauge fields are inherently long-range and fundamentally are similar to the photon field. However, the constraints on those fields from quantum mechanics are so strong that the observed interactions are quite unlike those of QED and even lead to the imprisonment of an entire class of particles.

EXERCISE 13.5 A proton or a neutron can sometimes “violate” conservation of energy by emitting and then reabsorbing a π meson, which has a mass of 135 MeV. This is possible as long as the π meson is reabsorbed within a short enough time consistent with the uncertainty principle. (i) Consider $p \rightarrow p\pi$. By what amount ΔE is energy conservation violated? (Ignore any kinetic energies.) (ii) For how long a time can the π meson exist? (iii) Assuming the π meson to travel at very nearly the speed of light, how far from the proton can it go? (This procedure, gives us an estimate of the range of the nuclear force, because protons and neutrons are held together in the nucleus by exchanging π mesons.)

XIV. MULTIDISCIPLINE APPROACH TO THE UV COMPLETION OF THE STANDARD MODEL

The conspicuously well-known accomplishments of the $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ Standard Model (SM) of strong and electroweak forces can be considered as the apotheosis of the gauge symmetry principle to describe particle interactions. Most spectacularly, the recent discovery [114, 115] of a new boson with scalar quantum numbers and couplings compatible with those of the SM Higgs has possibly plugged the final remaining experimental hole in the SM, cementing the theory further.

However, the saga of the SM is still exhilarating because it leaves all questions of consequence unanswered. The most evident of unanswered questions is the huge disparity between the strength of gravity and of the SM forces. Even if one abandons this hierarchy problem, which does not conflict with any experimental measurement, the SM has many other (perhaps more basic) shortcomings. Roughly speaking, the SM is incapable of explaining some well established observational results. Among the most notable of these are neutrino masses and the presence of a large non-baryonic dark matter (DM) component of the energy density in the universe. In addition, luminosity distance measurements of Type Ia supernovae strongly imply the presence of some unknown form of energy density, related to otherwise empty space, which appears to dominate the recent gravitational dynamics of the universe and yields a stage of cosmic accel-

eration. We still have no solid clues as to the nature of such dark energy (or perhaps more accurately dark pressure). In this section we give a brief description of the most characteristic features of these fundamental physical phenomena that the SM does not adequately explain.

A. Neutrino oscillations

At present, convincing experimental evidence exists for (time dependent) oscillatory transitions $\nu_\alpha \rightleftharpoons \nu_\beta$ between the different neutrino flavors. The simplest and most direct interpretation of the atmospheric data is that of muon neutrino oscillations [141–144]. The evidence of atmospheric ν_μ disappearing is now at $> 15\sigma$, most likely converting to ν_τ . The angular distribution of contained events shows that for neutrino energy $E_\nu \sim 1$ GeV, the deficit comes mainly from $L_{\text{atm}} \sim 10^2 - 10^4$ km. These results have been confirmed by the KEK-to-Kamioka (K2K) experiment which observes the disappearance of accelerator ν_μ 's at a distance of 250 km and finds a distortion of their energy spectrum with a CL of $2.5 - 4\sigma$ [145–147]. Data collected by the Sudbury Neutrino Observatory (SNO) in conjunction with data from Super-Kamiokande (SK) show that solar ν_e 's convert to ν_μ or ν_τ with CL of more than 7σ [148–152]. The KamLAND Collaboration has measured the flux of $\bar{\nu}_e$ from distant reactors and find that $\bar{\nu}_e$'s disappear over distances of about 180 km [153]. All these data suggest that the neutrino eigenstates that travel through space are not the flavor states that we measured through the weak force, but rather mass eigenstates.¹³ This is the first compelling experimental evidence for new physics beyond the SM.

The superposition of neutrino mass eigenstates produced in association with the charged lepton of flavor α is the state we refer to as the neutrino of flavor α . The flavor eigenstates and the mass eigenstates are related by a unitary transformation U (mass-to-flavor mixing matrix, fundamental to particle physics)

$$\begin{aligned} |\nu_\alpha\rangle &= \sum_i U_{\alpha i} |\nu_i\rangle \Leftrightarrow |\nu_i\rangle = \sum_\alpha (U^\dagger)_{i\alpha} |\alpha\rangle \\ &= \sum_\alpha U_{\alpha i}^* |\nu_\alpha\rangle, \end{aligned} \quad (478)$$

with $U^\dagger U = 1$, that is to say $\sum_i U_{\alpha i} U_{\beta i}^* = \delta_{\alpha\beta}$ and $\sum_\alpha U_{\alpha i} U_{\alpha j}^* = \delta_{ij}$ [156–158]. For antineutrinos we have to replace $U_{\alpha i}$ by $U_{\alpha i}^*$, i.e.

$$|\bar{\nu}_\alpha\rangle = \sum_i U_{\alpha i}^* |\bar{\nu}_i\rangle. \quad (479)$$

A summary of the oscillation length for 3 neutrino species

¹³ Contrariwise, charged leptons are states of definite mass and hence cannot undergo oscillations [154, 155].

is exhibited in Fig. 60.¹⁴

For many years, the sparse data on the angle θ_{13} allowed consistency with zero. However, in spring of 2012, the angle was definitively measured to be nonzero (but still small on the scale of $\theta_{23} \sim 45^\circ$ and $\theta_{12} \sim 35^\circ$), $\theta_{13} \approx 9^\circ$ [161–163]. At present, the low statistics statistics limits our capacity to disentangle neutrino flavors with sufficient precision to be sensitive to small θ_{13} , granting us the privilege of certain amount of laziness in subsequent explanation. To simplify the discussion, we will adopt maximal mixing for atmospheric $\nu_\mu \rightleftharpoons \nu_\tau$ neutrinos (i.e. $\theta_{23} \sim 45^\circ$) along with a negligible $|U_{e3}|^2 = \sin^2(\theta_{13})$. The latter approximation allows us to ignore CP violation and assume real matrix elements. (The small effects of nonzero θ_{13} have been investigated in [164].)

With our simplifying assumptions in mind, one can define a mass basis as follows,

$$|\nu_1\rangle = \sin\theta_\odot |\nu^\star\rangle + \cos\theta_\odot |\nu_e\rangle, \quad (480)$$

$$|\nu_2\rangle = \cos\theta_\odot |\nu^\star\rangle - \sin\theta_\odot |\nu_e\rangle, \quad (481)$$

and

$$|\nu_3\rangle = \frac{1}{\sqrt{2}}(|\nu_\mu\rangle + |\nu_\tau\rangle), \quad (482)$$

where $\theta_\odot \sim 34^\circ$ is the solar mixing angle and

$$|\nu^\star\rangle = \frac{1}{\sqrt{2}}(|\nu_\mu\rangle - |\nu_\tau\rangle) \quad (483)$$

is the eigenstate orthogonal to $|\nu_3\rangle$. Inversion of the neutrino mass-to-flavor mixing matrix leads to

$$|\nu_e\rangle = \cos\theta_\odot |\nu_1\rangle - \sin\theta_\odot |\nu_2\rangle \quad (484)$$

and

$$|\nu^\star\rangle = \sin\theta_\odot |\nu_1\rangle + \cos\theta_\odot |\nu_2\rangle. \quad (485)$$

Finally, by adding (482) and (483) one obtains the ν_μ flavor eigenstate,

$$|\nu_\mu\rangle = \frac{1}{\sqrt{2}} [|\nu_3\rangle + \sin\theta_\odot |\nu_1\rangle + \cos\theta_\odot |\nu_2\rangle], \quad (486)$$

and by subtracting these same equations the ν_τ eigenstate.

The evolution in time of the ν_i component of a neutrino initially born as ν_α in the rest frame of that component is described by Schrödinger's equation,

$$|\nu_i(\tau_i)\rangle = e^{-im_i\tau_i} |\nu_i(0)\rangle, \quad (487)$$

¹⁴ A sterile neutrino is a neutral lepton with no ordinary weak interactions except those induced by mixing [160]. They are present in most extensions of the SM and in principle can have any mass. Throughout we only consider the three weakly-interacting neutrino species.

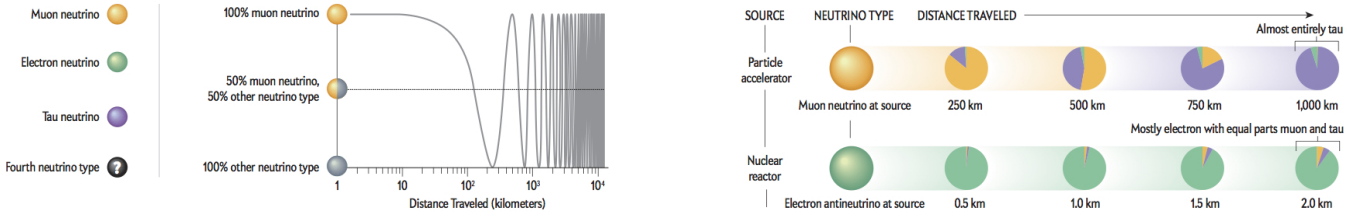


FIG. 60: The graph on the left illustrates the probability that a muon neutrino, having traversed a given distance, will switch flavors. In practice, oscillation distances depend on the neutrino's energy. The illustration on the right shows idealized source-to-detector oscillation patterns for experiments at particle accelerators and nuclear reactors [159].

where m_i is the mass of ν_i and τ_i is the proper time. In the lab frame, the Lorentz invariant phase factor may be written as $e^{-i(E_i t - p_i L)}$, where t , L , E_i , and p_i , are respectively, the time, the position, the energy, and the momentum of ν_i in the lab frame. Since the neutrino is extremely relativistic $t \approx L$ and

$$E_i = \sqrt{p^2 + m_i^2} = p\sqrt{1 + (m_i/p)^2} \approx p + m_i^2/2p, \quad (488)$$

where we have used that $(1 + x^2)^{1/2} = 1 + x^2/2 + \dots$. Hence, from (478) it follows that the state vector of a neutrino born as ν_α after propagation of distance L becomes

$$|\nu_\alpha(L)\rangle \approx \sum_i U_{\alpha i} e^{-i(m_i^2/2E)L} |\nu_i\rangle, \quad (489)$$

where $E \approx p$ is the average energy of the various mass eigenstate components of the neutrino. Using the unitarity of U to invert (478), from (489) one finds that

$$|\nu_\alpha(L)\rangle \approx \sum_\beta \left[\sum_i U_{\alpha i} e^{-i(m_i^2/2E)L} U_{\beta i} \right] |\nu_\beta\rangle. \quad (490)$$

In other words, the propagating mass eigenstates acquire relative phases giving rise to flavor oscillations. Thus, after traveling a distance L an initial state ν_α becomes a superposition of all flavors, with probability of transition to flavor β , $|\langle\nu_\beta|\nu_\alpha(L)\rangle|^2$, given by [165]

$$P(\nu_\alpha \rightarrow \nu_\beta) = \delta_{\alpha\beta} - 4 \sum_{i>j} U_{\alpha i} U_{\beta i} U_{\alpha j} U_{\beta j} \sin^2 \Delta_{ij},$$

where $\Delta_{ij} \sim \delta m_{ij}^2 L/2E$, and $\delta m_{ij}^2 = m_i^2 - m_j^2$.

For $\Delta_{ij} \gg 1$, the phases will be erased by uncertainties in L and E . Consequently, averaging over $\sin^2 \Delta_{ij}$ one finds

$$P(\nu_\alpha \rightarrow \nu_\beta) = \delta_{\alpha\beta} - 2 \sum_{i>j} U_{\alpha i} U_{\beta i} U_{\alpha j} U_{\beta j}. \quad (491)$$

Now, using $2 \sum_{1>j} = \sum_{i,j} - \sum_{i=j}$, (491) can be re-

written as

$$\begin{aligned} P(\nu_\alpha \rightarrow \nu_\beta) &= \delta_{\alpha\beta} - \sum_{i,j} U_{\alpha i} U_{\beta i} U_{\alpha j} U_{\beta j} \\ &+ \sum_i U_{\alpha i} U_{\beta i} U_{\alpha i} U_{\beta i} \\ &= \delta_{\alpha\beta} - \left(\sum_i U_{\alpha i} U_{\beta i} \right)^2 + \sum_i U_{\alpha i}^2 U_{\beta i}^2. \end{aligned} \quad (492)$$

Since $\delta_{\alpha\beta} = \delta_{\alpha\beta}^2$, the first and second terms in (493) cancel each other, yielding

$$P(\nu_\alpha \rightarrow \nu_\beta) = \sum_i U_{\alpha i}^2 U_{\beta i}^2. \quad (493)$$

EXERCISE 14.1 It is helpful to envision the cosmic ray engines as machines where protons are accelerated and (possibly) permanently confined by the magnetic fields of the acceleration region. The production of neutrons and charged pions and subsequent decay produces both neutrinos and cosmic rays: the former via $\pi^+ \rightarrow \mu^+ \nu_\mu \rightarrow e^+ \nu_e \nu_\mu \bar{\nu}_\mu$ (and the conjugate process), the latter via neutron diffusion from the region of the confined protons. Consequently, the expectation for the relative fluxes of each neutrino flavor at production in the cosmic sources, $(\alpha_e : \alpha_\mu : \alpha_\tau)_S$, is nearly $(1 : 2 : 0)_S$. After neutrino oscillations decohere over the astronomical propagation distances the flavor conversion is properly described by the mean oscillation probability. Show that the flux of pionic cosmic neutrinos should arrive at Earth with democratic flavor ratios, $(\alpha_e : \alpha_\mu : \alpha)_\oplus \approx (1 : 1 : 1)_\oplus$. (ii) Show that the flux of antineutrinos, which originates via neutron β -decay, arrives at Earth with $\sim (3 : 1 : 1)_\oplus$ [166].

The most economic way to get massive neutrinos would be to introduce the right-handed neutrino states (having no gauge interactions, these sterile states would be essentially undetectable) and obtain a Dirac mass term through a Yukawa coupling [165].

EXERCISE 14.2 On 23 February 1987, astronomers were startled by the observation of a new supernova in the Large Magellanic Cloud, a satellite galaxy of our Milky Way. However, the first observation of

TABLE VI: Observation times and inferred neutrino energies in the Kamioka experiment. The uncertainties on observation times are small enough to be neglected.

Event No.	Time t_{obs} [s]	Energy E [MeV]
1	0 (def)	21.3 ± 2.9
2	0.107	14.8 ± 3.2
3	0.303	8.9 ± 2.0
4	0.324	10.6 ± 2.7
5	0.507	14.4 ± 2.9

this supernova occurred several hours earlier by two neutrino-detection experiments [167, 168]. The fact that the neutrinos all arrived within a few seconds of each other after traveling for more than 100,000 lightyears allows us to put tight constraints on the mass of the neutrino [169]. (i) Suppose light takes a time t_0 to reach us from the location of the supernova. How long would it take a neutrino of energy E and mass $m \ll E$ to reach the Earth (work to lowest non-trivial order in m)? For future reference, the light travel time is approximately $t_0 = 5.3 \times 10^{12}$ s. (ii) The observation times and neutrino energies for the first 5 neutrinos observed by the Kamioka detector in Japan are given in Table VI. The first neutrino is defined to have arrived at time $t_{\text{obs}} = 0$. For any given neutrino, t_{obs} is the sum of its emission time (compared to neutrino #1) and its travel time (again, subtracting neutrino #1's travel time). Assume that all the neutrinos have the same mass and plot their emission times vs. the common value of $m^2 c^4$ (in eV^2). From this plot, argue that, if all these neutrinos were emitted within 4 s of each other (a conservative upper limit), the maximum neutrino mass is no more than $13 \text{ eV}/c^2$. [Hint: For simplicity, you can drop the error bars.] (iii) Consider only events 1 and 3. If neutrino 1 was emitted no more than 1 second later than event 3, what is the maximum neutrino mass? In this part, consider the error bars.

B. Dark matter

The evidence that DM is required to make sense of our Universe has been building for some time. In 1933 Zwicky found that the velocity dispersion of galaxies in the Coma cluster of galaxies was far too large to be supported by the luminous matter [170]. In the 1970's, Rubin and collaborators measured the rotational velocities of stars in spiral galaxies and also found evidence for non-luminous matter [171–173]. Spiral galaxies are flat rotating systems. The stars and gas in the disk are moving in nearly circular orbits, with the gravitational field of the galaxy providing the inward acceleration required for the circular motion. Observed rotation curves usually exhibit a

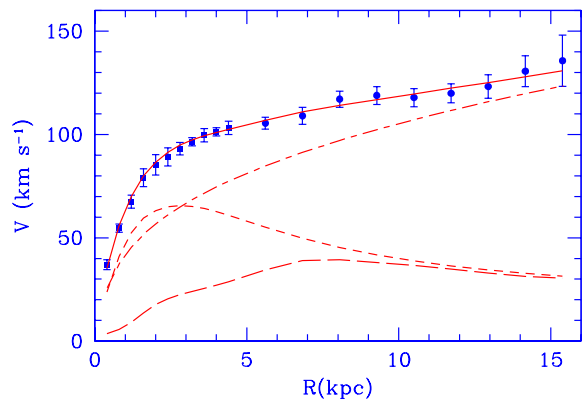


FIG. 61: M33 rotation curve (points) compared with the best fit model (continuous line). Also shown the halo contribution (dashed- dotted line), the stellar disk (short dashed line) and the gas contribution (long dashed line) [174].



FIG. 62: The Bullet Cluster.

characteristic flat behavior at large distances, i.e. out towards, and even far beyond, the edge of the visible disks (see a typical example in Fig. 61 [174]). To a fair approximation, assuming Newtonian gravity, the rotational velocity $v(r)$ at radius r is related to the total mass $M(r)$ within radius r by the equation

$$v^2(r) = \frac{GM(r)}{r}, \quad (494)$$

where G is the gravitational constant. Here,

$$M(r) = 4\pi \int \rho(r) r^2 dr, \quad (495)$$

where $\rho(r)$ is the mass density profile, which should be falling $\propto 1/\sqrt{r}$ beyond the optical disc. The fact that $v(r)$ is approximately constant implies the existence of a DM halo, with $M(r) \propto r$ and $\rho \propto 1/r^2$.

The most compelling evidence for DM is that observed at the Bullet Cluster [175]. In Fig. 62 we show a composite image of the Bullet Cluster (1E 0657-558) that shows the X-ray light detected by Chandra in purple, (an image

from Magellan and the Hubble space telescope of) the optical light in white and orange, and the DM map (drawn up using data on gravitational lensing from Magellan and European Space Observatory telescopes at Paranal) in blue. Galaxy clusters contain not only the galaxies ($\sim 2\%$ of the mass), but also intergalactic plasma ($\sim 10\%$ of the mass), and (assuming the null hypothesis) DM ($\sim 88\%$ of the mass). Over time, the gravitational attraction of all these parts naturally push all the parts to be spatially coincident. If two galaxy clusters were to collide/merge, we will observe each part of the cluster to behave differently. Galaxies will behave as collisionless particles but the plasma will experience ram pressure. Throughout the collision of two clusters, the galaxies will then become separated from the plasma. This is seen clearly in the Bullet Cluster, which is undergoing a high-velocity (around 4500 km/s) merger, evident from the spatial distribution of the hot, X-ray emitting gas. The galaxies of both concentrations are spatially separated from the (purple) X-ray emitting plasma. The DM clump (blue), revealed by the weak-lensing map, is coincident with the collisionless galaxies, but lies ahead of the collisional gas. As the two clusters cross, the intergalactic plasma in each cluster interacts with the plasma in the other cluster and slows down. However, the dark matter in each cluster does not interact at all, passing right through without disruption. This difference in interaction causes the DM to sail ahead of the hot plasma, separating each cluster into two components: DM (and collisionless galaxies) in the lead and the hot interstellar plasma lagging behind.

The non-gravitational interactions of DM may be with any of the known particles or, for a complex hidden dynamical DM sector [176], with other currently unknown particles. A complete research program in DM therefore requires a diverse set of experiments that together probe all possible types of couplings [177]. At a qualitative level, the complementarity may be illustrated by the following observations that follow from basic features of each approach. *(i)* With the information at hand one can ask the question: “How often will a DM particle at Earth’s location in the halo interact with a particular nucleus, and what is the expected distribution of recoil energies imparted to the nucleus?” Direct detection experiments typically operate in deep underground laboratories to reduce the background. The majority of present experiments use one of two detector technologies: cryogenic detectors, operating at temperatures below 100 mK, detect the heat produced when a particle hits an atom in a crystal absorber such as germanium. Noble liquid detectors detect the flash of scintillation light produced by a particle collision in liquid xenon or argon. *(ii)* DM may be detected indirectly when DM pair-annihilates somewhere, producing something, which is detected somehow. Searches for neutrinos are unique among indirect searches in that they are, given certain assumptions, probes of scattering cross sections, not annihilation cross sections, and so compete directly with the direct detection searches. The idea behind neu-

trino searches is the following: when DM particles pass through the Sun or the Earth, they may scatter and be slowed below escape velocity. Once captured, they then settle to the center, where their densities and annihilation rates are greatly enhanced. Although most of their annihilation products are immediately absorbed, neutrinos are not. Some of the resulting neutrinos then travel to the surface of the Earth, where they may convert to charged leptons through $\nu q \rightarrow l^+ q'$, and the charged leptons may be detected. Alternatively, photons from DM annihilation at the Galactic center or in other galaxies can be seen by γ -ray telescopes. *(iii)* Particle colliders, such as the LHC, could produce DM particles that escape the detector and be discovered as an excess of events with missing energy or momentum. *(iv)* The particle properties of DM are constrained through its impact on astrophysical observables. DM distributions and substructure in galaxies are unique probes of the “warmth” of DM and hidden DM properties, such as its self-interaction strength, and they measure the effects of DM properties on structure formation in the Universe. Examples include the self-interaction of DM particles affecting central DM densities in galaxies (inferred from rotation velocity or velocity dispersion measures), the mass of the DM particle affecting DM substructure in galaxies (inferred from strong lensing data), and the annihilation of DM in the early Universe affecting CMB fluctuations.

C. Dark energy

The expansion history of the cosmos can be determined using as a “standard candle” any distinguishable class of astronomical objects of known intrinsic brightness that can be identified over a wide distance range. As the light from such beacons travels to Earth through an expanding universe, the cosmic expansion stretches not only the distances between galaxy clusters, but also the very wavelengths of the photons en route. By the time the light reaches us, the spectral wavelength λ has thus been redshifted by precisely the same incremental factor $z \equiv \delta\lambda/\lambda$ by which the cosmos has been stretched in the time interval since the light left its source. That time interval is the speed of light times the object’s distance from Earth, which can be determined by comparing its apparent brightness to a nearby standard of the same class of astrophysical objects. The recorded redshift and brightness of each such object thus provide a measurement of the total integrated expansion of the universe since the time the light was emitted. A collection of such measurements, over a sufficient range of distances, would yield an entire historical record of the universe’s expansion.

In Fig. 63 we show the observed magnitude versus redshift for well-measured distant and (in the inset) nearby type Ia supernovae. The faintness (or distance) of the high-redshift supernovae in Fig. 63 comes as a dramatic surprise. In the simplest cosmological models, the expansion history of the cosmos is determined entirely by

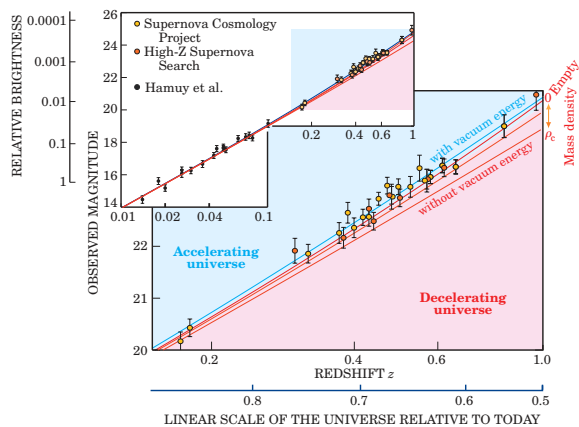


FIG. 63: Observed magnitude versus redshift is plotted for well-measured distant [178, 179] and (in the inset) nearby [180, 181] type Ia supernovae. For clarity, measurements at the same redshift are combined. At redshifts beyond $z = 0.1$ (distances greater than about 109 light-years), the cosmological predictions (indicated by the curves) begin to diverge, depending on the assumed cosmic densities of mass and vacuum energy. The red curves represent models with zero vacuum energy and mass densities ranging from the critical density ρ_c down to zero (an empty cosmos). The best fit (blue line) assumes a mass density of about $\rho_c/3$ plus a vacuum energy density twice that large, implying an accelerating cosmic expansion [182].

its mass density. The greater the density, the more the expansion is slowed by gravity. Thus, in the past, a high-mass-density universe would have been expanding much faster than it does today. So one should not have to look far back in time to especially distant (faint) supernovae to find a given integrated expansion (redshift). Conversely, in a low-mass-density universe one would have to look farther back. But there is a limit to how low the mean mass density could be. After all, we are here, and the stars and galaxies are here. All that mass surely puts a lower limit on how far—that is, to what level of faintness we must look to find a given redshift. However, the high-redshift supernovae in Fig. 63 are fainter than would be expected even for an empty cosmos.

If these data are correct, the obvious implication is that the simplest model of cosmology must be too simple. The next to simplest model includes an expansionary term in the equation of motion driven by the cosmological constant Λ , which competes against gravitational collapse. The best fit to the 1998 supernova data shown in Figs. 63 and 64 implies that, in the present epoch, the vacuum energy density ρ_Λ is larger than the energy density attributable to mass ρ_m . Therefore, the cosmic expansion is now accelerating. If the universe has no large-scale curvature, as the recent measurements of the CMB strongly indicate, we can say quantitatively that about 70% of the total energy density is vacuum energy and 30% is mass.

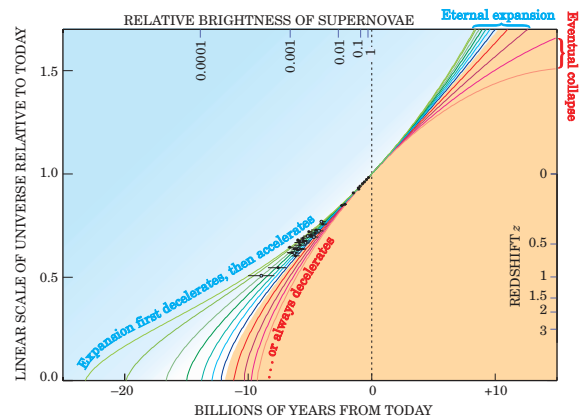


FIG. 64: The history of cosmic expansion, as measured by the high-redshift supernovae (the black data points), assuming flat cosmic geometry. The scale factor R of the universe is taken to be 1 at present, so it equals $1/(1+z)$. The curves in the blue shaded region represent cosmological models in which the accelerating effect of vacuum energy eventually overcomes the decelerating effect of the mass density. These curves assume vacuum energy densities ranging from $0.95 \rho_c$ (top curve) down to $0.4 \rho_c$. In the yellow shaded region, the curves represent models in which the cosmic expansion is always decelerating due to high mass density. They assume mass densities ranging (left to right) from $0.8 \rho_c$ up to $1.4 \rho_c$. In fact, for the last two curves, the expansion eventually halts and reverses into a cosmic collapse [182].

In units of the critical density

$$\rho_c = \frac{3H}{8\pi G}, \quad (496)$$

one usually writes this result as $\Omega_\Lambda \equiv \rho_\Lambda/\rho_c \approx 0.7$ and $\Omega_m \equiv \rho_m/\rho_c \approx 0.3$ [183]. Only about 4.6% of the mass are atoms. Hence, more than 95% of the energy density in the universe is in a form that has never been directly detected in the laboratory. The actual density of atoms is equivalent to roughly 1 proton per 4 m^3 .

The SM has no natural place for a vacuum energy density of the modest magnitude required by the astrophysical data. The simplest estimates would predict a vacuum energy 10^{120} times greater. So enormous Λ would have engendered an acceleration so rapid that stars and galaxies could never have formed. Therefore it has long been assumed that there must be some underlying symmetry that precisely cancels the vacuum energy. Now, however, the supernova data appear to require that such a cancellation would have to leave a remainder of about one part in 10^{120} . That degree of fine tuning is most unappealing. The cosmological constant model requires yet another fine tuning. In the cosmic expansion, mass density becomes ever more dilute. Since the end of inflation, it has fallen by very many orders of magnitude. However, the vacuum energy density ρ_Λ , a property of empty space itself, stays constant. It seems a remarkable and implausible coincidence that the mass density, just in the present epoch, is within a factor of

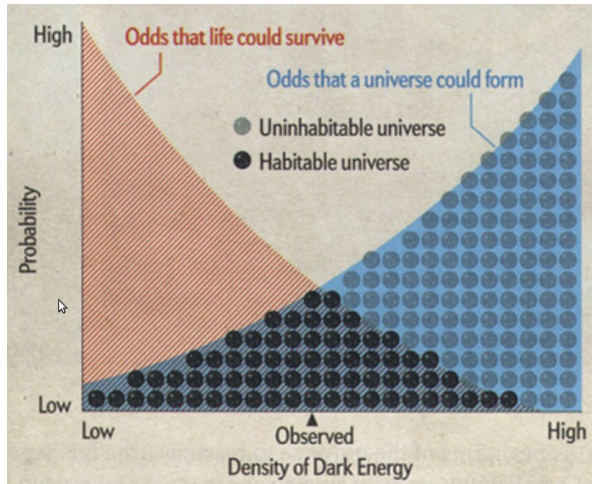


FIG. 65: The higher the value of Λ the most probable it is for the universe to form, but the more hostile the universe would be to life [189].

2 of the vacuum energy density [184]. Given these two fine-tuning coincidences, it seems likely that the SM is missing some fundamental physics.

EXERCISE 14.3. (i) Write Gauss's law for a galaxy of mass m inside a homogeneous and isotropic universe of mass density ρ , and show that

$$\frac{d^2 R/dt^2}{R} = -\frac{4\pi}{3}G\rho, \quad (497)$$

where R is the distance of the galaxy to the center of the distribution. (ii) Obtain an expression for dR/dt from the fact that the total energy of the galaxy vanishes. P.S. You have now derived Friedman's equations for a flat universe in the Newtonian limit.

D. Across the Multiverse

The extraordinary level of fine tuning required in generating a cosmological constant consistent with observations motivated a radically different approach in deriving the laws of nature. The key insight of this new viewpoint is no longer symmetries and invariants, but rather statistics. A remarkable fact about our universe is that the physical constants has just the right values allow for complex structures, including living things. The so-called multiverse provides an explanation for this apparent coincidence: If all possible values occur in a large enough collection of universes, then viable ones for life will surely be found somewhere [185–188]. This reasoning has been applied, in particular, to explaining the density of dark energy that is speeding up the expansion of the universe today. Indeed, it is the only scientifically based option we have right now. Inflation is the extremely rapid exponential expansion of the early universe by a factor

of at least 10^{78} in volume, driven by a negative-pressure vacuum energy density. Following the inflationary period, the universe continued to expand, but at a slower rate. As a direct consequence of this expansion, all of the observable universe originated in a small causally connected region. Inflation answers the classic conundrum of the Big Bang cosmology: why does the universe appear flat, homogeneous, and isotropic in accordance with the cosmological principle when one would expect, on the basis of the physics of the Big Bang, a highly curved, heterogeneous universe? In theories of eternal inflation, the inflationary phase of the multiverse's expansion lasts forever in at least some regions of the multiverse. Because these regions expand exponentially rapidly, most of the volume of the multiverse at any given time is inflating. All models of eternal inflation produce an infinite multiverse. The process of eternal inflation, endows each universe with a random density of dark energy. Relatively few universes have zero or a low dark energy density; most have higher values (blue area in Fig. 65). But too much dark energy tears apart the complex structures needed to sustain life (red area in Fig. 65). So most habitable universes should have a middling density of dark energy (peak of overlap region).

XV. WHERE DO WE STAND AND WHERE ARE WE GOING

The LHC physics program has just started. The ATLAS and CMS experiments have reported the discovery of a new boson with a mass around 126 GeV consistent with the SM Higgs particle (within current experimental and theoretical uncertainties). This marks just the beginning of the physics exploration of the LHC and it opens a new chapter in the study of the mechanism of electroweak symmetry breaking. The physics accessible at the high-luminosity phase of the LHC14 extends well beyond that of the earlier LHC8 program. Concurrent with LHC14 measurements will be many interesting missions to measure cosmological parameters, as well as dedicated experiments designed to detect cosmic rays, neutrinos, and DM particles. This new arsenal of data will certainly provide an ideal testing ground for the many interesting theories that extend the SM (see, e.g. [139, 190, 191]), and, at the same time, a unique opportunity to view similar physics from various different points of view. Alternatively, it is possible that we may be guessing the future while holding too small a deck of cards and LHC14 will open a new world that we did not anticipate

Acknowledgments

L.A.A. is supported by U.S. National Science Foundation (NSF) CAREER Award PHY1053663 and by the National Aeronautics and Space Administration (NASA) Grant No. NNX13AH52G; he thanks the Center for Cos-

mology and Particle Physics at New York University for its hospitality. Any opinions, findings, and conclusions or recommendations expressed in this material are those

of the author and do not necessarily reflect the views of the NSF or NASA.

-
- [1] L. Euler, *Leonhardi Euleri Opera Omnia*, (Edited by the Euler Commission of the Swiss Academy of Science, 1911) ISBN 978-3764314750.
 - [2] J. L. Lagrange, *Mécanique Analytique* (Courier 1811; reissued by Cambridge University Press, 2009) ISBN 978-1-108-00174-8.
 - [3] M. Faraday, Phil. Trans. R. Soc. Lond. **128**, 125 (1838).
 - [4] H. R. Hertz, Annalen Phys. **267**, 983 (1887).
 - [5] J. J. Thomson, Philos. Mag. **44**, 293 (1897).
 - [6] P. Lenard, Annalen Phys. **2**, 359 (1900).
 - [7] P. Lenard, Annalen Phys. **313**, 149 (1902).
 - [8] M. Planck, Verh. d. deutsch. phys. Ges. **2**, 202 (1900).
 - [9] M. Planck, Verh. d. deutsch. phys. Ges. **2**, 237 (1900).
 - [10] M. Planck, Annalen Phys. **4**, 553 (1901).
 - [11] G. Kirchhoff, Philos. Mag. **20**, 1 (1860).
 - [12] J. Stefan, Wiener Ber. **79**, 391 (1879).
 - [13] L. Boltzmann, Annalen Phys. **22**, 291 (1884).
 - [14] W. Wien, Annalen Phys. **52**, 132 (1894).
 - [15] Lord Rayleigh (J. W. Strutt), Philos. Mag. **49**, 539 (1900).
 - [16] Lord Rayleigh (J. W. Strutt), Nature **72**, 54 (1905).
 - [17] J. H. Jeans, Proc. R. Soc. A **76**, 546 (1905).
 - [18] G. B. Rybicki and A. P. Lightman, *Radiative Processes in Astrophysics*, (John Wiley & Sons, Massachusetts, 1979) ISBN 978-0-471-82759-7.
 - [19] J. C. Maxwell, Philos. Mag. **19**, 19 (1860).
 - [20] J. C. Maxwell, Philos. Mag. **20**, 21 (1860).
 - [21] L. Boltzmann, Wiener Ber. **76**, 373 (1877).
 - [22] A. H. Carter, *Classical and Statistical Thermodynamics* (Addison-Wesley, Boston, 2001) ISBN 0-13-779208-5.
 - [23] G. F. Smoot, astro-ph/9705101.
 - [24] J. C. Mather *et al.*, Astrophys. J. **420**, 439 (1994).
 - [25] A. Einstein, Annalen Phys. **17**, 132 (1905).
 - [26] R. Millikan, "A Direct Determination of h ," Phys. Rev. **4**, 73 (1914).
 - [27] R. Millikan, Phys. Rev. **7** 355 (1916).
 - [28] A. H. Compton, Phys. Rev. **21**, 483 (1923).
 - [29] J. J. Balmer, Annalen Phys. **261**, 80 (1885).
 - [30] J. R. Rydberg, Kongl. Sven. vetensk.-akad. hand. **23**, no. 11 (1890).
 - [31] J. R. Rydberg, Philos. Mag. **29**, 331 (1890).
 - [32] W. Ritz, Phys. Z. **9**, 521 (1908).
 - [33] T. Lyman, Memoirs of the American Academy of Arts and Sciences (New Series) **13**, 125 (1906).
 - [34] T. Lyman, Nature **93**, 241 (1914).
 - [35] F. Paschen Annalen Phys. **332**, 537 (1908).
 - [36] P. A. Tipler and R. A. Llewellyn, *Modern Physics*, (W. H. Freeman and Company, New York, 2008) ISBN 0-7167-7550-6.
 - [37] J. J. Thomson, Philos. Mag. **7**, 237 (1904).
 - [38] E. Rutherford, Philos. Mag. **21**, 669 (1911).
 - [39] N. Bohr, Philos. Mag. **26**, 1, 476, 857 (1913).
 - [40] N. Bohr, Nature **92**, 231 (1913).
 - [41] C. A. Coulomb, Histoire de l'Académie Royale des Sciences (Imprimerie Royale), 569, 578 (1788).
 - [42] H. D. Young and R. A. Freedman *University Physics* (Addison-Wesley, Boston, 2012) ISBN 978-0-321-69686-1.
 - [43] W. Wilson, Philos. Mag. **29**, 795 (1915).
 - [44] A. Sommerfeld, Ann. Phys. **51**, 1 (1916).
 - [45] A. Einstein, Phys. Z. **18**, 121 (1917).
 - [46] J. Larmor, Philos. Mag. **44**, 503 (1897).
 - [47] L. de Broglie, Comptes Rendus **177**, 507, 548, 630 (1923).
 - [48] L. de Broglie, Found. Phys. **1**, 5 (1970).
 - [49] L. A. Anchordoqui, arXiv:1509.08868 [physics.pop-ph].
 - [50] C. Jönsson, Z. Phys. **161**, 454 (1961).
 - [51] K. Krane, *Modern Physics*, (John Wiley & Sons, Massachusetts, 2012) ISBN 978-1-118-06114-5.
 - [52] R. Gähler and A. Zeilinger, Am. J. Phys. **59**, 316 (1991).
 - [53] W. Heisenberg, Z. Phys. **43**, 172 (1927).
 - [54] R. Feynman, *The Character of Physical Law*, (MIT Press, Cambridge, MA 1982) ISBN 978-0-262-06016-5.
 - [55] E. Schrödinger, Annalen Phys. **79**, 361, 409, 734 (1926).
 - [56] E. Schrödinger, Annalen Phys. **80**, 437 (1926).
 - [57] E. Schrödinger, Annalen Phys. **81**, 109 (1926).
 - [58] E. Schrödinger, Naturwiss. **14**, 664 (1926).
 - [59] E. Schrödinger, Phys. Rev. **28**, 1049 (1926).
 - [60] M. Born, Z. Phys. **38**, 803 (1926).
 - [61] H. P. Robertson, Phys. Rev. **34**, 163 (1929).
 - [62] E. Schrödinger, Bulg. J. Phys. **26**, 193 (1999) [Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.) **19**, 296 (1930)] [quant-ph/9903100].
 - [63] L. Mandelstam and I. G. Tamm, J. Phys. USSR **9**, 249 (1945).
 - [64] L. A. Anchordoqui and T. C. Paul, *Mathematical models of physics problems* (Nova, New York, 2013) ISBN 978-1-62618-600-2.
 - [65] O. Stern, Z. Phys. **7**, 249 (1921).
 - [66] W. Gerlach and O. Stern, Z. Phys. **8**, 110 (1922).
 - [67] W. Gerlach and O. Stern, Z. Phys. **9**, 349 (1922).
 - [68] W. Gerlach and O. Stern, Z. Phys. **9**, 353 (1922).
 - [69] W. Gerlach and O. Stern, Ann. Phys. **74**, 673 (1924).
 - [70] G. E. Uhlenbeck and S. A. Goudsmit, Naturwiss. **13**, 953 (1925).
 - [71] G. E. Uhlenbeck and S. Goudsmit, Nature **117**, 264 (1926);
 - [72] W. Pauli, Z. Phys. **31**, 373 (1925).
 - [73] W. Pauli, Z. Physik **43**, 601 (1927).
 - [74] R. A. Serway, C. J. Moses, C. A. Moyer, *Modern Physics*, (Brooks/Cole, California, 2005) ISBN 0-534-49339-4.
 - [75] B. C. Odom, D. Hanneke, B. D'Urso and G. Gabrielse, Phys. Rev. Lett. **97**, 030801 (2006) [Erratum-ibid. **99**, 039902 (2007)].
 - [76] D. Hanneke, S. Fogwell and G. Gabrielse, Phys. Rev. Lett. **100**, 120801 (2008) [arXiv:0801.1134 [physics.atom-ph]].
 - [77] W. Pauli, Z. Phys. **36**, 336 (1926).
 - [78] W. Heisenberg, Z. Phys. **33**, 879 (1925).
 - [79] B. Gaasbeek, arXiv:1007.4184 [quant-ph].
 - [80] O. Klein, Z. Phys. **37**, 895 (1926) [Surveys High Energ.

- Phys. **5**, 241 (1986)].
- [81] W. Gordon Z. Phys. **40**, 117 (1926).
 - [82] W. Pauli and V. Weisskopf, Helv. Phys. Acta **7**, 709 (1934).
 - [83] E. C. G. Stückelberg, Helv. Phys. Acta. **14**, 322 (1941); **14**, 558 (1941); **15**, 23 (1942).
 - [84] R. P. Feynman, Phys. Rev. **74**, 939 (1948). doi:10.1103/PhysRev.74.939
 - [85] R. P. Feynman, Phys. Rev. **74**, 1430 (1948). doi:10.1103/PhysRev.74.1430
 - [86] R. P. Feynman, Phys. Rev. **76**, 749 (1949). doi:10.1103/PhysRev.76.749
 - [87] R. P. Feynman, Phys. Rev. **76**, 769 (1949). doi:10.1103/PhysRev.76.769
 - [88] R. P. Feynman, Phys. Rev. **74**, 939 (1948); **76**, 749 (1949).
 - [89] P. A. M. Dirac, Proc. Roy. Soc. Lond. A **117**, 610 (1928).
 - [90] J. D. Bjorken and S. D. Drell, *Relativistic quantum mechanics*, (McGraw-Hill, New York, 1965) ISBN 978-0-07-005493-6.
 - [91] L. Evans and P. Bryant, JINST **3**, S08001 (2008). doi:10.1088/1748-0221/3/08/S08001
 - [92] G. Aad *et al.* [ATLAS Collaboration], JINST **3**, S08003 (2008). doi:10.1088/1748-0221/3/08/S08003
 - [93] S. Chatrchyan *et al.* [CMS Collaboration], JINST **3**, S08004 (2008). doi:10.1088/1748-0221/3/08/S08004
 - [94] A. A. Alves, Jr. *et al.* [LHCb Collaboration], JINST **3**, S08005 (2008). doi:10.1088/1748-0221/3/08/S08005
 - [95] K. Aamodt *et al.* [ALICE Collaboration], JINST **3**, S08002 (2008). doi:10.1088/1748-0221/3/08/S08002
 - [96] G. Aad *et al.* [ATLAS Collaboration], arXiv:1512.01530 [hep-ex].
 - [97] V. Khachatryan *et al.* [CMS Collaboration], arXiv:1512.01224 [hep-ex].
 - [98] M. Gell-Mann, CTSL-20, TID-12608.
 - [99] Y. Ne'eman, Nucl. Phys. **26**, 222 (1961).
 - [100] M. Gell-Mann, Phys. Lett. **8**, 214 (1964).
 - [101] S. N. Bose, Z. Phys. **26**, 178 (1924).
 - [102] A. Einstein, [Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.) **22**, 261 (1924); **1**, 3 (1925); **3**, 18 (1925)].
 - [103] E. Fermi, Rend. Lincei **3**, 145 (1926); Z. Phys. **36**, 902 (1926).
 - [104] P. A. M. Dirac, Proc. R. Soc. Lond. Ser. A **112**, 661 (1926).
 - [105] W. Pauli, Z. Phys. **31**:765 (1925).
 - [106] H. Fritzsch, M. Gell-Mann and H. Leutwyler, Phys. Lett. B **47**, 365 (1973).
 - [107] D. J. Gross and F. Wilczek, Phys. Rev. Lett. **30**, 1343 (1973).
 - [108] H. D. Politzer, Phys. Rev. Lett. **30**, 1346 (1973).
 - [109] S. L. Glashow, Nucl. Phys. **22**, 579 (1961).
 - [110] S. Weinberg, Phys. Rev. Lett. **19**, 1264 (1967).
 - [111] A. Salam, Conf. Proc. C **680519**, 367 (1968).
 - [112] P. W. Higgs, Phys. Rev. Lett. **13**, 508 (1964).
 - [113] F. Englert and R. Brout, Phys. Rev. Lett. **13**, 321 (1964).
 - [114] G. Aad *et al.* [ATLAS Collaboration], Phys. Lett. B **710**, 49 (2012) [arXiv:1202.1408 [hep-ex]].
 - [115] S. Chatrchyan *et al.* [CMS Collaboration], Phys. Lett. B **710**, 26 (2012) [arXiv:1202.1488 [hep-ex]].
 - [116] G. 't Hooft, *In search of the ultimate building blocks*, (Cambridge University Press, UK, 1996) ISBN 978-0-521-57883-7.
 - [117] S. Weinberg, Sci. Am. **231N1**, 50 (1974).
 - [118] G. 't Hooft, Sci. Am. **242N6**, 90 (1980).
 - [119] C. Quigg, Sci. Am. **252N4**, 64 (1985).
 - [120] M. J. G. Veltman, Sci. Am. **255**, no. 5, 88 (1986).
 - [121] F. Halzen and A. D. Martin, *Quarks and leptons: an introductory course in modern particle physics*, (John Wiley & Sons, New York, 1984) ISBN 0-471-88741-2
 - [122] V. D. Barger and R. J. N. Phillips, *Collider physics*, Front. Phys. **71**, 1 (1991) ISBN 0-201-14945-1
 - [123] C. Quigg, *Gauge theories of the strong, weak and electromagnetic interactions*, Front. Phys. **56**, 1 (1983) ISBN 978-0805360202
 - [124] L. Anchordoqui and F. Halzen, arXiv:0906.1271.
 - [125] J. S. Schwinger, Phys. Rev. **74**, 1439 (1948).
 - [126] J. S. Schwinger, Phys. Rev. **75**, 651 (1948). doi:10.1103/PhysRev.75.651
 - [127] S. Tomonaga, Prog. Theor. Phys. **1**, 27 (1946).
 - [128] R. P. Feynman, Rev. Mod. Phys. **20**, 367 (1948).
 - [129] F. J. Dyson, Phys. Rev. **75**, 486 (1949).
 - [130] F. J. Dyson, Phys. Rev. **75**, 1736 (1949).
 - [131] R. P. Feynman, Phys. Rev. **80**, 440 (1950). doi:10.1103/PhysRev.80.440
 - [132] J. S. Schwinger, Phys. Rev. **73**, 416 (1948).
 - [133] C. N. Yang and R. L. Mills, Phys. Rev. **96**, 191 (1954).
 - [134] E. P. Wigner, Nachr. Ges. Wiss. Gott. Math. Phys. Kl., 375 (1927).
 - [135] O. Laporte, Z. Phys. **23**, 135 (1924).
 - [136] T. D. Lee and C. N. Yang, Phys. Rev. **104**, 254 (1956).
 - [137] A. Lesov, arXiv:0911.0058 [physics.hist-ph].
 - [138] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes and R. P. Hudson, Phys. Rev. **105**, 1413 (1957).
 - [139] F. Quevedo, S. Krippendorf and O. Schlotterer, arXiv:1011.1491 [hep-th].
 - [140] H. Yukawa, Proc. Phys. Math. Soc. Jap. **17**, 48 (1935) [Prog. Theor. Phys. Suppl. **1**, 1]. doi:10.1143/PTPS.1.1
 - [141] Y. Fukuda *et al.* [Super-Kamiokande Collaboration], Phys. Lett. B **433**, 9 (1998) doi:10.1016/S0370-2693(98)00476-6 [hep-ex/9803006].
 - [142] Y. Fukuda *et al.* [Super-Kamiokande Collaboration], Phys. Rev. Lett. **81**, 1562 (1998) doi:10.1103/PhysRevLett.81.1562 [hep-ex/9807003].
 - [143] S. Fukuda *et al.* [Super-Kamiokande Collaboration], Phys. Rev. Lett. **85**, 3999 (2000) doi:10.1103/PhysRevLett.85.3999 [hep-ex/0009001].
 - [144] Y. Ashie *et al.* [Super-Kamiokande Collaboration], Phys. Rev. D **71**, 112005 (2005) doi:10.1103/PhysRevD.71.112005 [hep-ex/0501064].
 - [145] S. H. Ahn *et al.* [K2K Collaboration], Phys. Lett. B **511**, 178 (2001) doi:10.1016/S0370-2693(01)00647-5 [hep-ex/0103001].
 - [146] M. H. Ahn *et al.* [K2K Collaboration], Phys. Rev. Lett. **90**, 041801 (2003) doi:10.1103/PhysRevLett.90.041801 [hep-ex/0212007].
 - [147] M. H. Ahn *et al.* [K2K Collaboration], Phys. Rev. Lett. **93**, 051801 (2004) doi:10.1103/PhysRevLett.93.051801 [hep-ex/0402017].
 - [148] Y. Fukuda *et al.* [Super-Kamiokande Collaboration], Phys. Rev. Lett. **81**, 1158 (1998) [Phys. Rev. Lett. **81**, 4279 (1998)] doi:10.1103/PhysRevLett.81.1158 [hep-ex/9805021].
 - [149] Y. Fukuda *et al.* [Super-Kamiokande Collaboration], Phys. Rev. Lett. **82**, 1810 (1999) doi:10.1103/PhysRevLett.82.1810 [hep-ex/9812009].
 - [150] S. Fukuda *et al.* [Super-Kamiokande Collab-

- oration], Phys. Rev. Lett. **86**, 5656 (2001) doi:10.1103/PhysRevLett.86.5656 [hep-ex/0103033].
- [151] S. Fukuda *et al.* [Super-Kamiokande Collaboration], Phys. Lett. B **539**, 179 (2002) doi:10.1016/S0370-2693(02)02090-7 [hep-ex/0205075].
- [152] S. N. Ahmed *et al.* [SNO Collaboration], Phys. Rev. Lett. **92**, 181301 (2004) doi:10.1103/PhysRevLett.92.181301 [nucl-ex/0309004].
- [153] T. Araki *et al.* [KamLAND Collaboration], Phys. Rev. Lett. **94**, 081801 (2005) doi:10.1103/PhysRevLett.94.081801 [hep-ex/0406035].
- [154] S. Pakvasa, Lett. Nuovo Cim. **31**, 497 (1981). doi:10.1007/BF02778099
- [155] E. K. Akhmedov, JHEP **0709**, 116 (2007) doi:10.1088/1126-6708/2007/09/116 [arXiv:0706.1216 [hep-ph]].
- [156] B. Pontecorvo, Sov. Phys. JETP **6**, 429 (1957) [Zh. Eksp. Teor. Fiz. **33**, 549 (1957)].
- [157] B. Pontecorvo, Sov. Phys. JETP **26**, 984 (1968) [Zh. Eksp. Teor. Fiz. **53**, 1717 (1967)].
- [158] Z. Maki, M. Nakagawa and S. Sakata, Prog. Theor. Phys. **28**, 870 (1962).
- [159] M. Hirsch, H. Päs and W. Porod, Sci. Am. **308**, no. 4, 23 (2013). doi:10.1038/scientificamerican0413-40
- [160] K. N. Abazajian *et al.*, arXiv:1204.5379 [hep-ph].
- [161] F. P. An *et al.* [DAYA-BAY Collaboration], Phys. Rev. Lett. **108**, 171803 (2012) [arXiv:1203.1669 [hep-ex]].
- [162] J. K. Ahn *et al.* [RENO Collaboration], Phys. Rev. Lett. **108**, 191802 (2012) [arXiv:1204.0626 [hep-ex]].
- [163] F. P. An *et al.* [Daya Bay Collaboration], Chin. Phys. C **37**, 011001 (2013) [arXiv:1210.6327 [hep-ex]].
- [164] L. Fu, C. M. Ho and T. J. Weiler, Phys. Lett. B **718**, 558 (2012) [arXiv:1209.5382 [hep-ph]].
- [165] M. C. Gonzalez-Garcia and Y. Nir, Rev. Mod. Phys. **75**, 345 (2003) doi:10.1103/RevModPhys.75.345 [hep-ph/0202058].
- [166] L. A. Anchordoqui, H. Goldberg, F. Halzen and T. J. Weiler, Phys. Lett. B **593**, 42 (2004) doi:10.1016/j.physletb.2004.04.054 [astro-ph/0311002].
- [167] K. Hirata *et al.* [Kamiokande-II Collaboration], Phys. Rev. Lett. **58**, 1490 (1987). doi:10.1103/PhysRevLett.58.1490
- [168] R. M. Bionta *et al.*, Phys. Rev. Lett. **58**, 1494 (1987). doi:10.1103/PhysRevLett.58.1494
- [169] W. D. Arnett and J. L. Rosner, Phys. Rev. Lett. **58**, 1906 (1987). doi:10.1103/PhysRevLett.58.1906
- [170] F. Zwicky, Helv. Phys. Acta **6**, 110 (1933).
- [171] V. C. Rubin and W. K. Ford, Jr., Astrophys. J. **159**, 379 (1970). doi:10.1086/150317
- [172] V. C. Rubin, N. Thonnard and W. K. Ford, Jr., Astrophys. J. **238**, 471 (1980). doi:10.1086/158003
- [173] V. C. Rubin, D. Burstein, W. K. Ford, Jr. and N. Thonnard, Astrophys. J. **289**, 81 (1985). doi:10.1086/162866
- [174] E. Corbelli and P. Salucci, Mon. Not. Roy. Astron. Soc. **311**, 441 (2000) doi:10.1046/j.1365-8711.2000.03075.x [astro-ph/9909252].
- [175] D. Clowe, M. Bradac, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones and D. Zaritsky, Astrophys. J. **648**, L109 (2006) doi:10.1086/508162 [astro-ph/0608407].
- [176] K. R. Dienes and B. Thomas, Phys. Rev. D **85**, 083523 (2012) doi:10.1103/PhysRevD.85.083523 [arXiv:1106.4546 [hep-ph]].
- [177] J. L. Feng, Ann. Rev. Astron. Astrophys. **48**, 495 (2010) doi:10.1146/annurev-astro-082708-101659 [arXiv:1003.0904 [astro-ph.CO]].
- [178] A. G. Riess *et al.* [Supernova Search Team Collaboration], Astron. J. **116**, 1009 (1998) doi:10.1086/300499 [astro-ph/9805201].
- [179] S. Perlmutter *et al.* [Supernova Cosmology Project Collaboration], Astrophys. J. **517**, 565 (1999) doi:10.1086/307221 [astro-ph/9812133].
- [180] M. Hamuy *et al.*, Astron. J. **106**, 2392 (1993).
- [181] M. Hamuy, M. M. Phillips, J. Maza, N. B. Suntzeff, R. A. Schommer and R. Aviles, Astron. J. **109**, 1 (1995). doi:10.1086/117251
- [182] S. Perlmutter, Phys. Today, April 2003.
- [183] S. Weinberg, *Cosmology*, (Oxford University Press, UK, 2008) ISBN 9780198526827
- [184] N. A. Bahcall, J. P. Ostriker, S. Perlmutter and P. J. Steinhardt, Science **284**, 1481 (1999) doi:10.1126/science.284.5419.1481 [astro-ph/9906463].
- [185] S. Weinberg, Phys. Rev. Lett. **59**, 2607 (1987). doi:10.1103/PhysRevLett.59.2607
- [186] S. Weinberg, Rev. Mod. Phys. **61**, 1 (1989). doi:10.1103/RevModPhys.61.1
- [187] L. Susskind, [hep-th/0302219].
- [188] S. Kachru, R. Kallosh, A. D. Linde and S. P. Trivedi, Phys. Rev. D **68**, 046005 (2003) doi:10.1103/PhysRevD.68.046005 [hep-th/0301240].
- [189] G. F. R. Ellis Sci. Am. **23**, 76 (2014).
- [190] J. D. Wells, arXiv:0909.4541 [hep-ph].
- [191] C. Quigg, hep-ph/0404228.
- [192] S. Weinberg, *Lectures on quantum mechanics* (Cambridge University Press, UK, 2013) ISBN 978-1-107-02872-2
- [193] L. P. Gilbert, M. Belloni, M. A. Doncheski, and R. W. Robinett, Eur. J. Phys. **26**, 815 (2005) [quant-ph/0512156].
- [194] J. D. Jackson, *Classical Electrodynamics*, (John Wiley & Sons, New York, 1962). ISBN 9780471309321.
- [195] M. S. Longair, *High-energy astrophysics* (Cambridge University Press, UK, 2011) ISBN 978-0521756181

Answers and Comments on the Exercises

9.1 $\int_0^\infty P(E) dE = \int_0^\infty e^{-E/(kT)} / (kT) dE = -(kT) e^{-E/(kT)} / (kT) \Big|_0^\infty = 1$. To calculate $\int_0^\infty E P(E) dE$ we adopt the following change of variables, $u = E$ and $dv = e^{-E/(kT)} dE$, yielding $du = dE$ and $v = -kT e^{-E/(kT)}$. Then, $\int_0^\infty E e^{-E/(kT)} dE = uv - \int v du = E(-kT) e^{-E/(kT)} \Big|_0^\infty - \int_0^\infty (-kT) e^{-E/(kT)} dE = kT \int_0^\infty e^{-E/(kT)} dE = (kT)^2$. Thus

$\langle E \rangle = \int_0^\infty E P(E) dE = (kT)^2/(kT) = kT$. We now take $E_n = nh\nu = nhc/\lambda$ to compute (230),

$$\langle E \rangle = \frac{\sum_{n=0}^\infty \frac{nhc}{\lambda kT} e^{-nhc/(\lambda kT)}}{\frac{1}{kT} \sum_{n=0}^\infty e^{-nhc/(\lambda kT)}} = kT \frac{\sum_{n=0}^\infty n\alpha e^{-n\alpha}}{\sum_{n=0}^\infty e^{-n\alpha}}, \quad (498)$$

where $\alpha = hc/(\lambda kT)$. First we note that

$$-\alpha \frac{d}{d\alpha} \ln \left[\sum_{n=0}^\infty e^{-n\alpha} \right] = \frac{\sum_{n=0}^\infty n\alpha e^{-n\alpha}}{\sum_{n=0}^\infty e^{-n\alpha}}, \quad (499)$$

and so substituting (499) into (498) we have

$$\langle E \rangle = kT \left[-\alpha \frac{d}{d\alpha} \ln \left(\sum_{n=0}^\infty e^{-n\alpha} \right) \right] = -\frac{hc}{\lambda} \left[\frac{d}{d\alpha} \ln \left(\sum_{n=0}^\infty e^{-n\alpha} \right) \right] = -\frac{hc}{\lambda} \left\{ \frac{d}{d\alpha} \ln [(1 - e^{-\alpha})^{-1}] \right\}, \quad (500)$$

where in the last equality we have used the sum of a geometric series $\sum_{n=0}^\infty e^{-n\alpha} = (1 - e^{-\alpha})^{-1}$. Now, we calculate

$$\frac{d}{d\alpha} \ln(1 - e^{-\alpha})^{-1} = (-1) \frac{(1 - e^{-\alpha})^{-2} e^{-\alpha}}{(1 - e^{-\alpha})^{-1}} = \frac{-e^{-\alpha}}{1 - e^{-\alpha}} \left(\frac{e^\alpha}{e^\alpha} \right) = -\frac{1}{(e^\alpha - 1)}. \quad (501)$$

Substituting (501) into (500) we obtain (230).

9.2 In the frequency domain (211) can be rewritten as

$$F(T) = \pi \int_0^\infty \frac{2h\nu^3}{c^2} \frac{1}{e^{h\nu/(kT)} - 1} d\nu. \quad (502)$$

This is an integral over frequency alone. Change the variable to $x = h\nu/(kT)$ to obtain an integral of the form $\int [x^3(e^x - 1)^{-1}] dx$. Solutions of this type of integral include the Riemann zeta function, which is defined in Appendix B. It is straightforward to show that the result of the integration can be written as $F(T) = \sigma T^4$, with $\sigma = 2\pi^5 k^4/(15c^2 h^3) = 5.67 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$.

9.3 The first derivative of (231) in the frequency domain reads

$$\frac{\partial}{\partial \nu} B_\nu(\nu, T) = \frac{2h\nu^2}{c^2 [e^{h\nu/(kT)} - 1]^2} \left\{ 3 [e^{h\nu/(kT)} - 1] - \frac{h\nu}{kT} e^{h\nu/(kT)} \right\}. \quad (503)$$

This derivative is only equal to zero when the numerator is equal to zero. The corresponding denominator is larger than zero for $0 < \nu < \infty$. Demanding the numerator to vanish we obtain $3(e^{x_\nu} - 1) = x_\nu e^{x_\nu}$, with $x_\nu = h\nu/(kT)$. This transcendental function can only be solved numerically. We find $x_\nu \simeq 2.82$, and in a further step $\nu_{\text{ext}}/T = x_\nu k/h$, where ν_{ext} is the frequency at which the extreme (either a minimum or a maximum) of the Planck function in the frequency domain occurs. It can simply be proved that for this extreme the second derivative fulfills the condition $\partial^2 B_\nu / \partial \nu^2 < 0$ so that the extreme corresponds to a maximum, ν_{max} . The form of Wien's displacement law in terms of maximum spectral emittance per unit wavelength is derived using similar methods, but starting with the form of Planck's law expressed in the wavelength domain. The effective result is to substitute 3 for 5 in the equation for the peak frequency, i.e. $5(e^{x_\lambda} - 1) = x_\lambda e^{x_\lambda}$, where $x_\lambda = hc/(\lambda kT)$. This solves with $x_\lambda = 4.96$, yielding $\lambda_{\text{max}} T = hc/(x_\nu k) \simeq 2.90 \times 10^{-3} \text{ m K}$.

9.4 (i) We can obtain a first estimate of the surface temperature of the Sun from the sensitivity of the human eye to light in the range 400 – 700 nm. Assuming that the evolution worked well, i.e. that the human eye uses optimal the light from the Sun, and that the atmosphere is for all frequencies in the visible range similarly transparent, we identify the maximum in Wien's law with the center of the frequency range visible for the human eye. Thus we set $\lambda_{\text{max}, \odot} \approx 550 \text{ nm}$, and obtain $T_\odot \approx 5270 \text{ K}$ for the surface temperature of the Sun. (ii) The bolometric luminosity L of a star is given by the product of its surface $A = 4\pi R^2$ and the radiation emitted per area σT^4 , i.e., $L = 4\pi R^2 \sigma T^4$. The radiant flux is defined by $F = L/A$, so that we recover the well known inverse-square law for the energy flux at the distance $r > R$ outside of the star, $F = L/(4\pi r^2)$. The validity of the inverse-square law $F(r) \propto r^{-2}$ relies on the assumptions that no radiation is absorbed and that relativistic effects can be neglected. The later condition requires, in particular, that the relative velocity of observer and source is small compared to the velocity of light.

The energy flux received from the Sun at the distance of the Earth, $r_{\text{SE}} = 1 \text{ AU}$, is equal to $F = 1365 \text{ W/m}^2$. The solar luminosity follows then as $L_{\odot} = 4\pi d^2 F = 4 \times 10^{33} \text{ erg s}^{-1}$, and serves as a convenient unit in stellar astrophysics. The Stefan-Boltzmann law can then be used to define, with $R_{\odot} \approx 7 \times 10^{10} \text{ cm}$, the effective temperature of the Sun, $T_{\odot} \approx 5780 \text{ K}$. (iii) The average temperature on the surface of Neptune is $T_{\text{N}} = L_{\odot}/(4\pi r_{\text{N}}^2 \sigma) = 73 \text{ K}$.

9.5 The total energy density in the blackbody radiation is

$$u = \int_0^{\infty} \frac{8\pi hc}{\lambda^5} d\lambda \frac{1}{e^{hc/\lambda kT} - 1}. \quad (504)$$

Change the variable $x = hc/\lambda kT$ to obtain an integral of the form $\int [x^3(e^x - 1)^{-1}]dx$, which can be solved using the Riemann zeta function, see Appendix B; we obtain

$$u = \frac{8\pi^5 (kT)^4}{15(hc)^3} = 7.56464 \times 10^{-15} (T/\text{K})^4 \text{ erg/cm}^3. \quad (505)$$

(Recall that $1 \text{ J} \equiv 10^7 \text{ erg} = 6.24 \times 10^{18} \text{ eV}$.) We can easily interpret the Planck distribution in terms of quanta of light or photons. Each photon has an energy $E_{\gamma} = hc/\lambda$. Hence the number dn_{γ} of photons per unit volume in blackbody radiation in a narrow range of wavelengths from λ to $\lambda + d\lambda$ is

$$dn_{\gamma} = \frac{du}{hc/\lambda} = \frac{8\pi}{\lambda^4} d\lambda \frac{1}{e^{hc/\lambda kT} - 1}. \quad (506)$$

Then the total number of photons per unit volume is

$$n_{\gamma} = \int_0^{\infty} dn_{\gamma} = 8\pi \left(\frac{kT}{hc} \right)^3 \int_0^{\infty} \frac{x^2 dx}{e^x - 1}, \quad (507)$$

where $x = hc/(\lambda kT)$. The integral cannot be expressed in terms of elementary functions, but $\int [x^2(e^x - 1)^{-1}]dx = \Gamma(3)\zeta(3) \approx 2.4$, see Appendix B. Therefore, the number photon density is

$$n_{\gamma} = 60.42198 \left(\frac{kT}{hc} \right)^3 = 20.28 \left(\frac{T}{\text{K}} \right)^3 \text{ photons cm}^{-3} \approx 400 \text{ photons cm}^{-3},$$

and the average photon energy is

$$\langle E_{\gamma} \rangle = u/n_{\gamma} = 3.73 \times 10^{-16} (T/\text{K}) \text{ erg}. \quad (508)$$

For a temperature of 2.726 K, the number density of CMB photons is $\approx 410 \text{ cm}^{-3}$ and the typical photon energy is $\approx 6 \times 10^{-4} \text{ eV}$ in agreement with the values adopted in exercise 8.9. (iii) Now, let's consider what happens to blackbody radiation in an expanding universe. Suppose the size of the universe changes by a factor f , for example, if it doubles in size, then $f = 2$. As predicted by the Doppler effect, the wavelengths will change in proportion to the size of the universe to a new value $\lambda' = f\lambda$. After the expansion, the energy density du' in the new wavelength range λ' to $\lambda' + d\lambda'$ is less than the original energy density du in the old wavelength range λ to $\lambda + d\lambda$, for two different reasons: (1) since the volume of the universe has increased by a factor of f^3 , as long as no photons have been created or destroyed, the number of photons per unit volume has decreased by a factor of $1/f^3$; (2) the energy of each photon is inversely proportional to its wavelength, and therefore is decreased by a factor of $1/f$. It follows that the energy density is decreased by an overall factor $1/f^3 \times 1/f = 1/f^4$:

$$du' = \frac{1}{f^4} du = \frac{8\pi hc}{\lambda^5 f^4} d\lambda \frac{1}{e^{hc/\lambda kT} - 1}. \quad (509)$$

If we rewrite the previous equation in terms of the new wavelengths λ' , it becomes

$$du' = \frac{8\pi hc}{\lambda'^5} d\lambda' \frac{1}{e^{hf/\lambda' kT} - 1}, \quad (510)$$

which is exactly the same as the old formula for du in terms of λ and $d\lambda$, except that T has been replaced by a new temperature $T' = T/f$. Therefore, we conclude that freely expanding blackbody radiation remains described by the Planck formula, but with a temperature that drops in inverse proportion to the scale of expansion.

9.6 (i) According to classical theory, the energy in a light wave is spread out uniformly and continuously over the wavefront. Assuming that all absorption of light occurs in the top atomic layer of the metal, that each atom absorbs an equal amount of energy proportional to its cross sectional area, A , and that each atom somehow funnels this energy into one of its electrons, we find that each electron absorbs an energy K in time t given by $K = \epsilon I A t$ where ϵ is a fraction accounting for less than 100% light absorption. Because the most energetic electrons are held in the metal by a surface energy barrier φ , these electrons will be emitted with K_{\max} once they have absorbed enough energy to overcome the barrier. We can express this as $K_{\max} = \epsilon I A t - \varphi$. Thus, classical theory predicts that for a fixed absorption period, t , at low light intensities when $\epsilon I A t < \varphi$, no electrons must be emitted. At higher intensities, when $\epsilon I A t > \varphi$ electrons should be emitted with higher kinetic energies the higher the light intensity. Therefore, classical predictions contradict experiment at both very low and very high light intensities. (ii) According to classical theory, the intensity of a light wave is proportional to the square of the amplitude of the electric field, $|\vec{E}_0|^2$, and it is this electric field amplitude that increases with increasing intensity and imparts an increasing acceleration and kinetic energy to an electron. Substituting I with a quantity proportional to $|\vec{E}_0|^2$ in part (i) shows that K_{\max} should not depend at all on the frequency of the classical light wave, again contradicting the experimental results. (iii) To estimate the time lag between the start of illumination and the emission of electrons, we assume that an electron must accumulate just enough light energy to overcome the work function. Setting $K_{\max} = 0$ gives $0 = \epsilon I A t - \varphi$ or $t = \varphi/(\epsilon I A)$. Taking $\epsilon = 1$ and the cross sectional area of the atom $A = \pi r^2$, where $r = 1.0 \times 10^{-8}$ cm is a typical atomic radius, we have $t = 1.2 \times 10^7$ s ≈ 130 days. Thus we see that the classical calculation of the time lag for photoemission does not agree with the experimental result, disagreeing by a factor of 10^{16} !

9.7 At the threshold wavelength the photoelectrons have just enough energy to overcome the work function, so $K_{\max} = 0$. Hence we have $\varphi = hc/\lambda_0 = 3.808$ eV. When 259.8 nm light is used, $eV_0 = hc/\lambda - \varphi = 0.964$ eV, so $V_0 = 0.964$ V.

9.8 (i) Consider an incident photon of frequency ν_0 which is scattered by a stationary electron to give a photon of frequency ν at an angle θ with respect to the original photon. Conservation of energy gives

$$E_0 + mc^2 = E + E', \quad (511)$$

while conservation of 3-momentum gives $\vec{p}_0 = \vec{p} + \vec{p}'$ or $\vec{p}' = \vec{p}_0 - \vec{p}$. Squaring this $\vec{p}'^2 = \vec{p}_0^2 + \vec{p}^2 - 2\vec{p}_0 \cdot \vec{p}$ and using $E_0 = |\vec{p}_0|c$, $E = |\vec{p}|c$, and $E'^2 = (\vec{p}'c)^2 + m^2c^4$, we get $E'^2 - m^2c^4 = E_0^2 + E^2 - 2E_0E \cos \theta$. Now, from (511) the left-hand side of the previous relation equals $(E_0 - E)^2 + 2mc^2(E_0 - E)$, thus yielding

$$2mc^2(E_0 - E) = 2E_0E - 2E_0E \cos \theta. \quad (512)$$

Using $E_0 = hc/\lambda_0$ and $E = hc/\lambda$ we obtain the well known formula for the change in wavelength as a function of angle

$$\lambda - \lambda_0 = \frac{h}{mc}(1 - \cos \theta) \quad (513)$$

The quantity $h/(mc)$ is known as the Compton wavelength (in this case, of the electron). (ii) Since λ_c is very small, high energy radiation (X-rays) is needed to observe the effect. If we choose a wavelength of 7×10^{-9} cm for the X-rays we estimate for a maximal scattering angle an effect of $\Delta\lambda/\lambda_0 = 2\lambda_c/\lambda_{\text{Xray}} \approx 0.07$.

9.9 The dominant energy loss is from electric dipole radiation, which obeys the Larmor formula (257). For an electron of charge $-e$ and mass m_e in an orbit of radius r about a fixed nucleus of charge $+e$, the radial component of the nonrelativistic force law, $\vec{F} = m_e \vec{a}$, tells us that $e^2/r^2 = m_e a_r \approx m_e v_\theta^2/r$, in the adiabatic approximation that the orbit remains nearly circular at all times. In the same approximation, $a_\theta \ll a_r$, i.e., $a \approx a_r$, and hence,

$$\frac{dE}{dt} = -\frac{2e^6}{3r^4 m_e^2 c^3} = -\frac{2}{3} \frac{r_0^3}{r^4} m_e c^3, \quad (514)$$

where $r_0 = e^2/(m_e c^2) = 2.8 \times 10^{-15}$ m is the classical electron radius. The total nonrelativistic energy (kinetic plus potential) is

$$E = -\frac{e^2}{r} + \frac{1}{2} m_e v^2 = -\frac{e^2}{2r} = -\frac{r_0}{r} m_e c^2. \quad (515)$$

Equating the time derivative of (515) to (514), we have

$$\frac{dE}{dt} = \frac{r_0}{2r^2} \dot{r} m_e c^2 = -\frac{2}{3} \frac{r_0^3}{r^4} m_e c^3, \quad \text{or equivalently} \quad r^2 \dot{r} = \frac{1}{3} \frac{dr^3}{dt} = -\frac{4}{3} r_0^2 c, \quad \text{yielding} \quad r^3 = a_0^3 - 4r_0^2 c t. \quad (516)$$

The time to fall to the origin is then

$$t_{\text{fall}} = \frac{a_0^3}{4r_0^2 c} = 1.6 \times 10^{-11} \text{ s}. \quad (517)$$

This is of the order of magnitude of the lifetime of an excited hydrogen atom, whose ground state, however, appears to have infinite lifetime.

9.10 (i) The average kinetic energy of the neutrons is $K = 3kT/2 = 0.0379 \text{ eV}$. (ii) The neutrons are non-relativistic so the momentum is given by $p = \sqrt{2mK} = \sqrt{2mc^2 K}/c = 8.44 \times 10^3 \text{ eV}/c$, yielding $\lambda = hc/(pc) = 0.147 \text{ nm}$.

10.1 Perhaps the most direct solution is to begin with the continuity equation (284), substitute in the definitions of ρ and j , and then prove the equality. First, calculate the partial derivatives:

$$\frac{\partial j}{\partial x} = -\frac{i\hbar}{2m} \left(\frac{\partial \psi^*}{\partial x} \frac{\partial \psi}{\partial x} + \psi^* \frac{\partial^2 \psi}{\partial x^2} - \frac{\partial^2 \psi^*}{\partial x^2} \psi - \frac{\partial \psi^*}{\partial x} \frac{\partial \psi}{\partial x} \right) = -\frac{i\hbar}{2m} \left(\psi^* \frac{\partial^2 \psi}{\partial x^2} - \frac{\partial^2 \psi^*}{\partial x^2} \psi \right) \quad (518)$$

and

$$\frac{\partial \rho}{\partial t} = \frac{\partial \psi^*}{\partial t} \psi + \psi^* \frac{\partial \psi}{\partial t}. \quad (519)$$

The connection between the time and space derivatives is given by rearranging the Schrödinger equation and its complex conjugate

$$\frac{\partial \psi}{\partial t} = \frac{i\hbar}{2m} \frac{\partial^2 \psi}{\partial x^2} - \frac{i}{\hbar} V \psi \quad \text{and} \quad \frac{\partial \psi^*}{\partial t} = -\frac{i\hbar}{2m} \frac{\partial^2 \psi^*}{\partial x^2} + \frac{i}{\hbar} V \psi^*. \quad (520)$$

Substituting (520) into (519)

$$\frac{\partial \rho}{\partial t} = -\frac{i\hbar}{2m} \frac{\partial^2 \psi^*}{\partial x^2} \psi + \frac{i}{\hbar} V \psi^* \psi + \frac{i\hbar}{2m} \psi^* \frac{\partial^2 \psi}{\partial x^2} - \frac{i}{\hbar} V \psi^* \psi = -\frac{\partial j}{\partial x}. \quad (521)$$

The continuity equation is equivalent to conservation of probability. One way to see this is to integrate the continuity equation over x , with the added restriction that ψ and $\partial\psi/\partial x$ go to zero as $x \rightarrow \pm\infty$,

$$0 = \int_{-\infty}^{+\infty} dx \left(\frac{\partial \rho}{\partial t} + \frac{\partial j}{\partial x} \right) = \frac{\partial}{\partial t} \int_{-\infty}^{+\infty} dx |\psi|^2 + \int_{-\infty}^{+\infty} dx \frac{\partial j}{\partial x} = \frac{\partial}{\partial t} \int_{-\infty}^{+\infty} dx |\psi|^2 + j(x) \Big|_{-\infty}^{+\infty} = \frac{\partial}{\partial t} \int_{-\infty}^{+\infty} dx |\psi|^2. \quad (522)$$

The last integral is the total probability (otherwise known as the normalization), and is shown to be constant with respect to time.

10.2 (i) From the identity (296) it follows that:

$$\langle (A^\dagger)^\dagger \phi | \psi \rangle = \langle \phi | A^\dagger \psi \rangle = \langle A^\dagger \psi | \phi \rangle^* = \langle \psi | \hat{A} \phi \rangle^* = (\langle A \phi | \psi \rangle^*)^* = \langle \hat{A} \phi | \psi \rangle \Rightarrow (\hat{A}^\dagger)^\dagger = \hat{A} \quad (523)$$

and

$$\langle \phi | \hat{A} \hat{B} \psi \rangle = \langle \hat{A}^\dagger \phi | B \psi \rangle = \langle \hat{B}^\dagger \hat{A}^\dagger \phi | \psi \rangle \Rightarrow (AB)^\dagger = B^\dagger A^\dagger. \quad (524)$$

10.3 (i) The operator \hat{x} is hermitian because

$$\int_{-\infty}^{+\infty} (\hat{x}\phi)^* \psi dx = \int_{-\infty}^{+\infty} (x\phi(x))^* \psi(x) dx = \int_{-\infty}^{+\infty} \phi^* x \psi dx = \int_{-\infty}^{+\infty} \psi^* \hat{x} \phi dx. \quad (525)$$

(ii) The operator \hat{p} is hermitian because

$$\int (\hat{p}\phi)^* \psi dx = \int_{-\infty}^{+\infty} \left(-i\hbar \frac{\partial \phi}{\partial x} \right)^* \psi dx = i\hbar \int_{-\infty}^{+\infty} \left(\frac{\partial \phi}{\partial x} \right)^* \psi dx \quad (526)$$

and after integration by parts, recognizing that the wave function tends to zero as $x \rightarrow \text{infy}$, the right-hand side of (526) becomes

$$-i\hbar \int_{-\infty}^{+\infty} \phi^* \frac{\partial \psi}{\partial x} dx = \int_{-\infty}^{+\infty} \phi^* \hat{p} \psi dx. \quad (527)$$

10.4 We may assert without proof that the expectation value of a physical observable is real, i.e. $\langle \psi | \hat{A} \psi \rangle = \langle \psi | \hat{A} \psi \rangle^*$. Now,

$$\langle \psi | \hat{A} \psi \rangle^* = \left[\int_{-\infty}^{+\infty} \psi^*(x) \hat{A} \psi(x) dx \right]^* = \int_{-\infty}^{+\infty} \psi(x) [\hat{A} \psi(x)]^* dx = \int_{-\infty}^{+\infty} [\hat{A} \psi(x)]^* \psi dx = \langle \hat{A} \psi | \psi \rangle, \quad (528)$$

so from (297) it follows that physical observables are represented by hermitian operators.

10.5 (ii) $[\hat{A}, \hat{B}]/2 + \{\hat{A}, \hat{B}\}/2 = (\hat{A}\hat{B} - \hat{B}\hat{A})/2 + (\hat{A}\hat{B} + \hat{B}\hat{A})/2 = \hat{A}\hat{B}$. (iii) $\{\hat{A}, \hat{B}\}^\dagger = (\hat{A}\hat{B})^\dagger + (\hat{B}\hat{A})^\dagger = \hat{B}^\dagger \hat{A}^\dagger + \hat{A}^\dagger \hat{B}^\dagger = \{\hat{A}, \hat{B}\}$, so the anticommutator is hermitian. $[\hat{A}, \hat{B}]^\dagger = (\hat{A}\hat{B})^\dagger - (\hat{B}\hat{A})^\dagger = \hat{B}^\dagger \hat{A}^\dagger - \hat{A}^\dagger \hat{B}^\dagger = -(\hat{A}\hat{B} - \hat{B}\hat{A}) = -[\hat{A}, \hat{B}]$, so the commutator is anti-Hermitian. (iv) An anti-hermitian operator is equal to the negative of its hermitian conjugate, that is $\hat{A}^\dagger = -\hat{A}$. In inner products this means $\langle \phi | \hat{A} \psi \rangle = \langle \hat{A}^\dagger \phi | \psi \rangle = -\langle \hat{A} \phi | \psi \rangle$. The expectation value of an anti-hermitian operator is: $\langle \psi | \hat{A} \psi \rangle = \langle \hat{A}^\dagger \psi | \psi \rangle = -\langle \hat{A} \psi | \psi \rangle = -\langle A \rangle^*$. But $\langle \psi | \hat{A} \psi \rangle = \langle A \rangle$, so $\langle A \rangle = -\langle A \rangle^*$, which means the expectation value must be pure imaginary.

10.6 Define new Hermitian operators $\hat{A}' = \hat{A} - \langle \hat{A} \rangle$ and $\hat{B}' = \hat{B} - \langle \hat{B} \rangle$. Then, using the Schwarz's inequality we obtain $\langle \hat{A}'^2 \rangle \langle \hat{B}'^2 \rangle \geq |\langle \hat{A}' \hat{B}' \rangle|^2$, or $\Delta A \Delta B \geq |\langle \hat{A}' \hat{B}' \rangle| = |\langle [\hat{A}', \hat{B}'] \rangle / 2 + \langle \{\hat{A}', \hat{B}'\} \rangle / 2| \geq |\langle [\hat{A}', \hat{B}'] \rangle| / 2$. Since the expectation value of the commutator is imaginary and the anticommutator is real, each makes a positive contribution to the absolute value, and the anticommutator can be dropped without changing the inequality in the last step. So, $\Delta A \Delta B \geq |\langle [\hat{A}', \hat{B}'] \rangle| / 2 = |\langle [\hat{A}, \hat{B}] - [\langle \hat{A} \rangle, \langle \hat{B} \rangle] - [\langle \hat{A} \rangle, \hat{B}] + [\hat{A}, \langle \hat{B} \rangle]| / 2 = |\langle [\hat{A}, \hat{B}] \rangle| / 2$. Note that $\langle \hat{A} \rangle$ and $\langle \hat{B} \rangle$ are just numbers, so they commute with the operators and the commutators involving them are 0.

10.7 For nonrelativistic quantum mechanics, it is not so surprising that time and space are treated differently, with position being an operator and *not time*. After all, this is also what happens in Newtonian mechanics: time is absolute, and part of the background, and all other observables are functions of time. This paradigm underlies the formulation of the fundamental problem of Newtonian physics: to determine how a system evolves in time. Time cannot be an observable because an observable is a function of what we consider the system's "state", but the state is considered a function of time in the first place (so time is the independent variable). In deriving the time-energy uncertainty principle one should be careful in defining the meaning of the standard deviation Δt . It is well known that the total energy of an isolated quantum mechanical system in distinction to a classical one, does not, in general, have a definite constant value. Instead of this the probability to obtain in a measurement any specified value of the energy of the system remains constant in time. The energy can only be determined exactly in the special case of a stationary state. But in this case, as easily seen, all dynamical variables or, more exactly, their distribution functions, remain constant in time. In other words, the *definiteness* of the total energy of the system entails the *constancy* with respect to the time of all dynamical variables. It can be concluded that there must exist a general connection between the dispersion of the total energy of the system and the time variation of coordinates, momenta, etcetera. The uncertainty relation with which are concern gives a quantitative formulation of this connection. Let A and B denote any two quantities and at the same time the corresponding Hermitian operators. From (303) we have

$$\Delta A \Delta B \geq \frac{1}{2} \langle AB - BA \rangle \quad (529)$$

where ΔA and ΔB are the standards of the quantities A and B and $\langle \cdot \rangle$ denotes as usual the quantum mechanical average. In addition, it is easily seen that

$$\frac{\hbar}{2} \frac{\partial \langle B \rangle}{\partial t} = i \langle (HB - BH) \rangle \quad (530)$$

where H is the Hamiltonian of the system not depending explicitly on time [192]. Putting in (529) $A \equiv H$ we obtain, with the help of (530) the desired uncertainty relation for energy, in the form of the following inequality:

$$\Delta H \Delta B \geq \frac{\hbar}{2} \left| \frac{\partial \langle B \rangle}{\partial t} \right|. \quad (531)$$

This relation gives, thus, the connection between the standard ΔH of the total energy of an isolated system, the standard ΔB of some other dynamical quantity and the rate of change of the average value of this quantity. The relation (531) can be put in a different form. The absolute value of an integral cannot exceed the integral of the absolute value of the integrand. Thus, integrating (531) from t to $t + \delta t$ and taking into account that ΔH is constant one gets

$$\Delta H \delta t \geq \frac{\hbar}{2} \frac{|\langle B_{t+\delta t} \rangle - \langle B_t \rangle|}{\Delta B}, \quad (532)$$

where the denominator of the right-hand side denotes the average value of the standard ΔB during the time δt . Sometimes (especially in the case of a continuous spectrum of eigenvalues) it is convenient to refer the variations of the average value of a dynamical quantity to its standard. In such a case it is convenient to introduce a special notation Δt for the shortest time, during which the average value of a certain quantity is changed by an amount equal to the standard of this quantity. Δt can be called the standard time. With the help of this notation we can rewrite (532) in the form of an uncertainty relation

$$\Delta H \Delta t \equiv \Delta E \Delta t \geq \hbar/2. \quad (533)$$

10.8 We want to project $|\phi\rangle$ onto each basis vector (this gives the expansion coefficients) and then sum the coefficients times the basis vectors: $|\phi\rangle = (\sum_n \psi_n \langle \psi_n |) |\phi\rangle$.

10.9 For $E < V_0$, the solutions for zones I and II are

$$\psi_I(x) = A \sin(kx) + B \cos(kx) \quad \text{and} \quad \psi_{II}(x) = C e^{-\kappa x} + D e^{\kappa x}, \quad (534)$$

where $k = \sqrt{2mE}/\hbar$ and $\kappa = \sqrt{2m(V_0 - E)}/\hbar$. Imposing the boundary conditions we have

$$\psi_I(x=0) = 0 \Rightarrow B = 0, \quad (535)$$

$$\psi_{II}(x=2L) = 0 \Rightarrow C e^{-2\kappa L} + D e^{2\kappa L} = 0 \Rightarrow D = -C e^{-4\kappa L}. \quad (536)$$

Imposing continuity of the wave function at $x = L$ we obtain

$$\psi_I(x=L) = \psi_{II}(x=L) \Rightarrow A \sin kL = C e^{-\kappa L} + D e^{\kappa L} \quad (537)$$

and

$$\psi'_I(x=L) = \psi'_{II}(x=L) \Rightarrow A k \cos kL = \kappa (-C e^{-\kappa L} + D e^{\kappa L}). \quad (538)$$

Substituting the expression for D in (537) and (538) we have

$$A \sin(kL) = C(e^{-\kappa L} - e^{-3\kappa L}) \quad \text{and} \quad A k \cos kL = -C \kappa (e^{-\kappa L} + e^{-3\kappa L}). \quad (539)$$

Taking the ratio of these two expressions we get

$$\kappa \tan(kL) = -k \frac{1 - e^{-2\kappa L}}{1 + e^{-2\kappa L}}, \quad \text{or equivalently} \quad \kappa \tan(kL) = -k \tanh(2\kappa L). \quad (540)$$

(iii) For $E > V_0$, the solutions for zones I and II are

$$\psi_I(x) = A \sin(k_1 x) \quad \text{and} \quad \psi_{II}(x) = B \sin(k_2 x). \quad (541)$$

Imposing the continuity condition on $\psi(x)$ and $\psi'(x)$ at $x = L$ we obtain

$$k_2 \tan(k_1 L) = -k_1 \tan(k_2 L). \quad (542)$$

In Fig. 66 we show the first four eigenstates for $E < 0$ and $E > 0$.

10.10 The wavefunction $\psi(x)$ for a particle with energy E in a potential $V(x)$ satisfies the Schrödinger equation (279). Inside the well ($0 \leq x \leq L$), the particle is free. The wave function is $\psi_I(x) = A \sin(kx)$, where $k = \sqrt{2mE}/\hbar$. Outside the well ($L < x < \infty$), the potential has constant value $V > E$. The wave function is $\psi_{II}(x) = B e^{-\kappa x}$, where

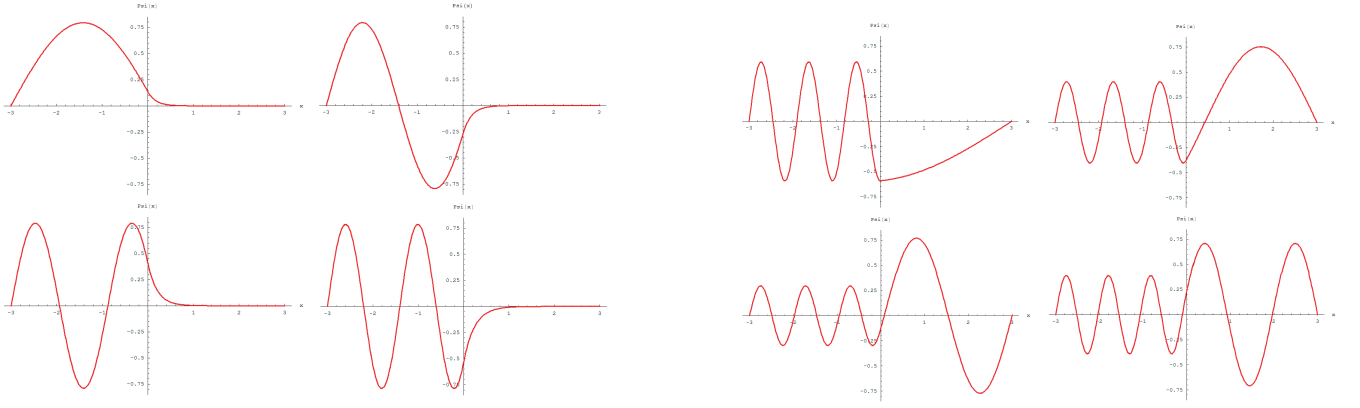


FIG. 66: The first four energy eigenstates for the asymmetric infinite square well with $E < V_0$ (left) and $E > V_0$ (right). In all of the images we have taken $\hbar = 2m = 1$, $V_0 = 33$, and $L = 3$ [193].

$\kappa = \sqrt{2m(V_0 - E)}/\hbar$. $\psi(x)$ and its derivative are continuous at $x = L$, then $A \sin kL = B e^{-\kappa L}$ and $A k \cos kL = -B \kappa e^{-\kappa L}$, from which

$$k \cot kL = -\kappa. \quad (543)$$

Now, since $\cot^2 \theta + 1 = \csc^2 \theta$, (543) can be rewritten as $k^2(\csc^2 \theta - 1) = -\kappa^2$, where $\theta = kL$. After some algebra the transcendental equation can be rewritten as

$$\theta \csc \theta = \pm a, \quad (544)$$

where $a = \sqrt{2mV_0L^2}/\hbar$. Note that (544) are equations for the allowed values of k . The equation with the positive sign yields values of θ in the second quadrant. The equation with the negative sign yields values of θ in the fourth quadrant. Since $\sin \theta \leq \theta \forall \theta$, it follows from (544) that there are no bound states if $2mV_0L^2/\hbar^2 \leq 1$.

10.11 (i) In region I, where $x < 0$, Schrödinger equation is given by

$$\frac{\partial^2 \psi_I}{\partial x^2} + \frac{2mE}{\hbar^2} \psi_I = 0 \Rightarrow \psi_I = A e^{ik_1 x} + B e^{-ik_1 x}, \quad (545)$$

while in region II, where $x > 0$, we have

$$\frac{\partial^2 \psi_{II}}{\partial x^2} + \frac{2m(E + V_0)}{\hbar^2} \psi_{II} = 0 \Rightarrow \psi_{II} = C e^{ik_2 x}. \quad (546)$$

Demanding continuity of ψ and ψ' at $x = 0$ we obtain

$$\psi_I(x=0) = \psi_{II}(x=0) \Rightarrow A + B = C \quad (547)$$

and

$$\psi'_I(x=0) = \psi'_{II}(x=0) \Rightarrow ik_1(A - B) = ik_2 C \quad (548)$$

Substituting (547) into (548) we get

$$ik_1(A - B) = ik_2(A + B) \quad \text{or equivalently} \quad A(k_1 - k_2) = B(k_1 + k_2). \quad (549)$$

This leads to $B/A = (k_1 - k_2)/(k_1 + k_2)$. The reflection coefficient is then

$$R = \left| \frac{B}{A} \right|^2 = \left(\frac{k_1 - k_2}{k_1 + k_2} \right)^2 = \left(\frac{\sqrt{E} - \sqrt{E + V_0}}{\sqrt{E} + \sqrt{E + V_0}} \right)^2, \quad (550)$$

whereas the transmissivity is given by

$$T = 1 - R = \frac{4k_1 k_2}{(k_1 + k_2)^2} = \frac{4\sqrt{E(E + V_0)}}{(\sqrt{E + V_0} + \sqrt{E})^2}. \quad (551)$$

(ii) Taking $V_0 = 12$ MeV and $E = 4$ MeV we obtain $T = 8/9$.

10.12 The time-independent Schrödinger equation of a particle of mass m , which is constrained to remain within a finite region of space (“a box”) is given by

$$-\frac{\hbar^2}{2m} \nabla^2 \psi = E\psi. \quad (552)$$

Let $k^2 = 2mE/\hbar^2$, and note that it is real. This equation can be solved with the help of the separation of variables technique. Start out by trying a solution of the following form $\psi(x, y, z) = X(x)Y(y)Z(z)$. Substitution of this solution into the time-independent Schrödinger equation yields: $YZX'' + XZY'' + XYZ'' = -k^2XYZ$. Then divide both sides of the equation by ψ , obtain $X''/X + Y''/Y + Z''/Z = -k^2$, and note that each of the three terms on the right-hand side is independent of the others, because x, y, z are independent variables. In order for their sum to be equal to a constant, $-k^2$, each of those terms must be independently equal to a constant, such that the sum of all three constants is equal to $-k^2$. Denote those three constants by $-k_x^2, -k_y^2, -k_z^2$, respectively, such that Schrödinger equation now translates into three ordinary differential equations: $X'' = -k_x^2X, Y'' = -k_y^2Y, Z'' = -k_z^2Z$. The solutions to these equations are: $X(x) = A \sin(k_x x) + B \cos(k_x x), Y(y) = C \sin(k_y y) + D \cos(k_y y), Z(z) = F \sin(k_z z) + G \cos(k_z z)$, where A, B, C, D, F , and G are (complex) undetermined parameters. Since the infinitely high walls do not allow the particle to leave the box, the wave function is zero at all times for $(x, y, z) < (0, 0, 0)$ and $(x, y, z) > (a, b, c)$, and hence $\psi(0, 0, 0) = \psi(a, b, c) = 0$, because the wave function needs to be continuous. Imposing $\psi(0, 0, 0) = 0$ implies $B = D = G = 0$, whereas applying the second boundary condition $\psi(a, b, c) = 0$ yields $k_x a = n\pi, k_y b = m\pi$, and $k_z c = l\pi$, with $n, m, l \in \mathbb{Z}$. The particle is equally likely to be found everywhere, $\int_0^a \int_0^b \int_0^c |\psi(x, y, z)|^2 dx dy dz = 1$, and so $N = ACF$ can be determined from the requirement that the wave function is normalized, i.e.

$$|N|^2 \int_0^a \int_0^b \int_0^c \sin^2(n\pi x/a) \sin^2(m\pi y/b) \sin^2(l\pi z/c) dx dy dz = \frac{1}{8} |N|^2 abc \Rightarrow |N| = \sqrt{\frac{8}{abc}}. \quad (553)$$

All in all, the stationary states of a particle in a 3-dimensional box are given by

$$\psi_{nlm}(x, y, z) = \sqrt{\frac{8}{abc}} \sin(n\pi x/a) \sin(m\pi y/b) \sin(l\pi z/c), \quad (554)$$

and the corresponding energy levels are

$$E_{n,m,l} = \frac{\hbar^2 \pi^2}{2m} \left(\frac{n^2}{a^2} + \frac{m^2}{b^2} + \frac{l^2}{c^2} \right). \quad (555)$$

10.13 Consider for instance the commutator $[\hat{L}^2, \hat{L}_z]$:

$$\begin{aligned} [\hat{L}^2, \hat{L}_z] &= [\hat{L}_x^2 + \hat{L}_y^2 + \hat{L}_z^2, \hat{L}_z] \quad \text{from the definition of } \hat{L}^2 \\ &= [\hat{L}_x^2, \hat{L}_z] + [\hat{L}_y^2, \hat{L}_z] + [\hat{L}_z^2, \hat{L}_z] \\ &= [\hat{L}_x^2, \hat{L}_z] + [\hat{L}_y^2, \hat{L}_z] \quad \text{since } \hat{L}_z \text{ commutes with itself} \\ &= \hat{L}_x \hat{L}_x \hat{L}_z - \hat{L}_z \hat{L}_x \hat{L}_x + \hat{L}_y \hat{L}_y \hat{L}_z - \hat{L}_z \hat{L}_y \hat{L}_y. \end{aligned} \quad (556)$$

We can use the commutation relation $[\hat{L}_z, \hat{L}_x] = i\hbar \hat{L}_y$ to rewrite the first term on the right-hand side as $\hat{L}_x \hat{L}_x \hat{L}_z = \hat{L}_x \hat{L}_z \hat{L}_x - i\hbar \hat{L}_x \hat{L}_y$, and the second term as $\hat{L}_z \hat{L}_x \hat{L}_x = \hat{L}_x \hat{L}_z \hat{L}_x + i\hbar \hat{L}_y \hat{L}_z$. In a similar way, we can use $[\hat{L}_y, \hat{L}_z] = i\hbar \hat{L}_x$ to rewrite the third term as $\hat{L}_y \hat{L}_y \hat{L}_z = \hat{L}_y \hat{L}_z \hat{L}_y + i\hbar \hat{L}_y \hat{L}_x$, and the fourth term $\hat{L}_z \hat{L}_y \hat{L}_y = \hat{L}_y \hat{L}_z \hat{L}_y - i\hbar \hat{L}_x \hat{L}_y$. Thus, on substituting in we find that

$$[\hat{L}^2, \hat{L}_z] = -i\hbar \hat{L}_x \hat{L}_y - i\hbar \hat{L}_y \hat{L}_x + i\hbar \hat{L}_y \hat{L}_x + i\hbar \hat{L}_x \hat{L}_y = 0. \quad (557)$$

By performing a cyclic permutation of the indexes, we can show that this holds in general, i.e. $[\hat{L}^2, \hat{L}_i] = 0$, for $i = x, y, z$.

10.14 Assume that the eigenvalues of \hat{L}^2 and \hat{L}_z are unknown and denote them λ and μ . We introduce two new operators, the raising and lowering operators $\hat{L}_+ = \hat{L}_x + i\hat{L}_y$ and $\hat{L}_- = \hat{L}_x - i\hat{L}_y$. The commutator with L_z is

$[\hat{L}_z, \hat{L}_\pm] = \pm \hbar \hat{L}_\pm$ (while they of course commute with L^2). Now consider the function $f_\pm = \hat{L}_\pm f$, where f is an eigenfunction of \hat{L}^2 and \hat{L}_z :

$$\hat{L}^2 f_\pm = \hat{L}_\pm \hat{L}^2 f = \hat{L}_\pm \lambda f = \lambda f_\pm \quad \text{and} \quad \hat{L}_z f_\pm = [\hat{L}_z, \hat{L}_\pm] f + \hat{L}_\pm \hat{L}_z f = \pm \hbar \hat{L}_\pm f + \hat{L}_\pm \mu f = (\mu \pm \hbar) f_\pm. \quad (558)$$

Then $f_\pm = \hat{L}_\pm f$ is also an eigenfunction of \hat{L}^2 and \hat{L}_z . Moreover, we can keep finding eigenfunctions of \hat{L}_z with higher and higher eigenvalues $\mu' = \mu + \hbar + \hbar + \dots$, by applying the \hat{L}_+ operator (or lower and lower with \hat{L}_-), while the \hat{L}^2 eigenvalue is fixed. Of course there is a limit, since we want $\mu' \leq \lambda$. Then there is a maximum eigenfunction such that $\hat{L}_+ f_M = 0$ and we set the corresponding eigenvalue to $\hbar l_M$. Now note that we can write \hat{L}^2 instead of using $\hat{L}_{x,y}$ by using \hat{L}_\pm :

$$\hat{L}^2 = \hat{L}_- \hat{L}_+ + \hat{L}_z^2 + \hbar \hat{L}_z. \quad (559)$$

Using this relationship on f_M we find:

$$\hat{L}^2 f_M = \lambda f_M \Rightarrow (\hat{L}_- \hat{L}_+ + \hat{L}_z^2 + \hbar \hat{L}_z) f_M = [0 + \hbar^2 l_M^2 + \hbar(\hbar l_M)] f_M \Rightarrow \lambda = \hbar^2 l_M(l_M + 1). \quad (560)$$

In the same way, there is also a minimum eigenvalue l_m and eigenfunction such that $\hat{L}_- f_m = 0$ and we can find $\lambda = \hbar^2 l_m(l_m - 1)$. Since λ is always the same, we also have $l_m(l_m - 1) = l_M(l_M + 1)$, with solution $l_m = -l_M$ (the other solution would have $l_m > l_M$). Finally, we have found that the eigenvalues of L_z are between $+\hbar l$ and $-\hbar l$ with integer increases, so that $l = -l + N$ giving $l = N/2$: that is, l is either an integer or a half-integer. We thus set $\lambda = \hbar^2 l(l + 1)$ and $\mu = \hbar m$, with $m = -l, -l + 1, \dots, l$.

11.1 (i) First we show that the set $\{\mathbf{1}, \sigma_1, \sigma_2, \sigma_3\}$ is linearly independent. Suppose

$$\alpha \mathbf{1} + \beta \sigma_1 + \zeta \sigma_2 + \xi \sigma_3 = \begin{pmatrix} \alpha + \xi & \beta - i\zeta \\ \beta + i\zeta & \alpha - \xi \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}. \quad (561)$$

Then $\alpha = -\xi$ and $\alpha = \xi \Leftrightarrow \alpha = \xi = 0$. Similarly, $\beta = -i\zeta$ and $\beta = i\zeta$, which implies $\beta = \zeta = 0$. Now we show that the vectors $\{\mathbf{1}, \sigma_1, \sigma_2, \sigma_3\}$ span the 2×2 matrix space. Let

$$\begin{aligned} \mathbb{M} &= \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \\ &= \frac{1}{2}(m_{11} + m_{22}) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \frac{1}{2}(m_{11} - m_{22}) \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} + \frac{1}{2}(m_{12} + m_{21}) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + \frac{i}{2}(m_{12} - m_{21}) \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \\ &= \frac{1}{2}(m_{11} + m_{22}) \mathbf{1} + \frac{1}{2}(m_{12} + m_{21}) \sigma_1 + \frac{i}{2}(m_{12} - m_{21}) \sigma_2 + \frac{1}{2}(m_{11} - m_{22}) \sigma_3. \end{aligned} \quad (562)$$

Note that

$$\frac{1}{2} \text{Tr} [\mathbb{M}] = \frac{1}{2}(m_{11} + m_{22}) \quad (563)$$

and so the first term in (562) can be written as $\frac{1}{2} \text{Tr} [\mathbb{M}] \mathbf{1}$. Now,

$$\begin{aligned} \frac{1}{2} \text{Tr} [\mathbb{M} \sigma_1] &= \frac{1}{2} \text{Tr} \begin{pmatrix} m_{12} & m_{11} \\ m_{22} & m_{21} \end{pmatrix} = \frac{1}{2}(m_{12} + m_{21}) \\ \frac{1}{2} \text{Tr} [\mathbb{M} \sigma_2] &= \frac{1}{2} \text{Tr} \begin{pmatrix} i m_{12} & -i m_{11} \\ i m_{22} & -i m_{21} \end{pmatrix} = \frac{1}{2}(m_{12} - m_{21}) \\ \frac{1}{2} \text{Tr} [\mathbb{M} \sigma_3] &= \frac{1}{2} \text{Tr} \begin{pmatrix} m_{11} & -m_{12} \\ m_{21} & -m_{22} \end{pmatrix} = \frac{1}{2}(m_{11} - m_{22}). \end{aligned}$$

We define, $\mathbb{M} \vec{\sigma} = (\mathbb{M} \sigma_1, \mathbb{M} \sigma_2, \mathbb{M} \sigma_3)$ so that the last three terms in (562) can be written as $\frac{1}{2} \text{Tr} [\mathbb{M} \vec{\sigma}] \cdot \vec{\sigma}$. Therefore, any 2×2 matrix can be written as $\mathbb{M} = a_0 \mathbf{1} + \vec{a} \cdot \vec{\sigma}$, where $a_0 = \frac{1}{2} \text{Tr} [\mathbb{M}]$ and $\vec{a} = \frac{1}{2} \text{Tr} [\mathbb{M} \vec{\sigma}]$.

12.1 (i) Consider the equation $E = mc^2$. If we use SI units and solve for E with $m = 1$ kg, we find

$$E = (1 \text{ kg}) \cdot (3 \times 10^8 \text{ m/s})^2 = 9 \times 10^{16} \text{ J}. \quad (564)$$

Converting from J to GeV using $1 \text{ eV} = 1.6 \times 10^{-19} \text{ J}$ and $10^9 \text{ eV} = 1 \text{ GeV}$, we find

$$E = 9 \times 10^{16} \text{ J} \left(\frac{1 \text{ eV}}{1.6 \times 10^{-19} \text{ J}} \right) \left(\frac{1 \text{ GeV}}{10^9 \text{ eV}} \right) = \frac{9 \times 10^{16}}{(1.6 \times 10^{-19}) \cdot (10^9)} \text{ GeV} = 5.6 \times 10^{26} \text{ GeV}. \quad (565)$$

Plugging in this value of E into our original equation $E = mc^2$ and setting $c = 1$, we can now solve for 1 kg:

$$(5.6 \times 10^{26} \text{ GeV}) = (1 \text{ kg})(1)^2 = 5.6 \times 10^{26} \text{ GeV}. \quad (566)$$

(ii) We are given that $\hbar c = 197.3 \text{ MeVfm}$. We can solve this equation for MeV^{-1} , set $\hbar = 1$ and $c = 1$, convert to GeV using $1 \text{ GeV} = 10^3 \text{ MeV}$, and square both sides to get:

$$\text{MeV}^{-1} = \frac{197.3 \text{ fm}}{\hbar c} \Rightarrow \frac{1}{\text{MeV}} \left(\frac{10^3 \text{ MeV}}{1 \text{ GeV}} \right) = \frac{197.3 \text{ fm}}{1} \quad (567)$$

$$1 \text{ GeV}^{-1} = 0.1973 \text{ fm} \Rightarrow 1 \text{ GeV}^{-2} = 0.0389 \text{ fm}^2 \quad (568)$$

Converting from fm^2 to mb (note: b stands for barn, which is a measure of area) using $1 \text{ m} = 10^{15} \text{ fm}$, $1 \text{ m}^2 = 10^{28} \text{ b}$, and $1 \text{ b} = 10^3 \text{ mb}$, we find

$$1 \text{ GeV}^{-2} = 0.0389 \text{ fm}^2 \left(\frac{1 \text{ m}}{10^{15} \text{ fm}} \right)^2 \left(\frac{10^{28} \text{ b}}{1 \text{ m}^2} \right) \left(\frac{10^3 \text{ mb}}{1 \text{ b}} \right) = 0.389 \text{ mb}. \quad (569)$$

(iii) We are given that $\hbar c = 197.3 \text{ MeVfm}$. We can solve this equation for fm, set $\hbar = 1$ and $c = 1$, convert from fm to m using $1 \text{ m} = 10^{15} \text{ fm}$, and convert from MeV to GeV using $1 \text{ GeV} = 10^3 \text{ MeV}$ to get:

$$\begin{aligned} \text{fm} &= \frac{\hbar c}{197.3 \text{ MeV}} \\ \text{fm} \left(\frac{1 \text{ m}}{10^{15} \text{ fm}} \right) &= \frac{1}{197.3 \text{ MeV}} \left(\frac{10^3 \text{ MeV}}{1 \text{ GeV}} \right) \\ 1 \text{ m} &= 5.068 \times 10^{15} \text{ GeV}^{-1}. \end{aligned} \quad (570)$$

(iv) We know that $c = 3 \times 10^8 \text{ m/s}$ in SI units. Therefore, we can solve for 1 s by setting $c = 1$ and plugging in that $1 \text{ m} = 5.068 \times 10^{15} \text{ GeV}^{-1}$:

$$1 \text{ s} = \frac{3 \times 10^8 \text{ m}}{c} = (3 \times 10^8) \cdot (5.068 \times 10^{15} \text{ GeV}^{-1}) = 1.5 \times 10^{24} \text{ GeV}^{-1}. \quad (571)$$

(v) Converting from MeV to GeV using that $1 \text{ GeV} = 10^3 \text{ MeV}$, we can find the Compton wavelength to be

$$\lambda_c = \frac{1}{0.511 \text{ MeV}} \left(\frac{10^3 \text{ MeV}}{1 \text{ GeV}} \right) = 1.957 \times 10^3 \text{ GeV}^{-1}. \quad (572)$$

Using that $1 \text{ m} = 5.068 \times 10^{15} \text{ GeV}^{-1}$ and $1 \text{ m} = 10^{15} \text{ fm}$, we find that

$$\lambda_c = 1.957 \times 10^3 \text{ GeV}^{-1} \left(\frac{1 \text{ m}}{5.068 \times 10^{15} \text{ GeV}^{-1}} \right) \left(\frac{10^{15} \text{ fm}}{1 \text{ m}} \right) = 386 \text{ fm}. \quad (573)$$

(vi) As you can see, $r_B = \lambda_c/\alpha$, where $\alpha = 1/137$. Therefore,

$$r_B = \frac{386 \text{ fm}}{1/137} = 5.29 \times 10^4 \text{ fm}. \quad (574)$$

(vii) In a unit system where $\hbar = 1$ and $c = 1$, $v = \alpha = 1/137$. To convert this into SI units where $c \neq 1$, we must restore the \hbar and c variables. Here, recognize that the 1 in the numerator must be c in order to get the correct units for velocity. Therefore,

$$v = \frac{1}{137} = \frac{c}{137} = \frac{3 \times 10^8 \text{ m/s}}{137} = 2.2 \times 10^6 \text{ m/s}. \quad (575)$$

(viii) Since we just saw that

$$v = \left(\frac{1}{137} \right) c = 0.007c \Rightarrow v \ll c, \quad (576)$$

the non-relativistic Schrödinger equation can be used to describe the hydrogen atom. (ix) To estimate the length scale, let's find the associated Compton wavelength for M_{P1} using that $1 \text{ m} = 5.068 \times 10^{15} \text{ GeV}^{-1}$ and $1 \text{ m} = 10^{15} \text{ fm}$:

$$\lambda_{P1} \sim \frac{1}{10^{19}} \text{ GeV}^{-1} \left(\frac{1 \text{ m}}{5.068 \times 10^{15} \text{ GeV}^{-1}} \right) \left(\frac{10^{15} \text{ fm}}{1 \text{ m}} \right) \sim 1.97 \times 10^{-20} \text{ fm} \sim 10^{-20} \text{ fm} \quad (577)$$

12.2 Multiplying the Klein-Gordon equation by $-i\psi^*$ and the complex conjugate equation by $-i\psi$, and subtracting it follows that

$$-i\psi^* \partial_\mu \partial^\mu \psi - i\psi^* m^2 \psi + i\psi \partial_\mu \partial^\mu \psi^* + i\psi m^2 \psi^* = -i\psi^* \partial_\mu \partial^\mu \psi + i\psi \partial_\mu \partial^\mu \psi^* = 0. \quad (578)$$

Note that the last equality holds because

$$\partial_\mu (\psi^* \partial^\mu \psi) = \partial_\mu \psi^* \partial^\mu \psi + \psi^* \partial_\mu \partial^\mu \psi. \quad (579)$$

13.1 (i) The number of minimum bias events per second can be computed as follows:

$$\frac{\Delta N}{\Delta t} = L \times \sigma \quad (580)$$

where L is the luminosity and σ the cross section; this yields

$$\frac{\Delta N}{\Delta t} = 70 \times 10^{34} \text{ mb cm}^{-2} \text{ s}^{-1} = 7 \times 10^8 \text{ s}^{-1}. \quad (581)$$

Thus, we get 700 million interactions per second. Since the beams collide every 25 ns, the number of events produced per bunch crossing will be $\Delta N = 17.5$. (ii) To estimate the Higgs production rate we proceed in a similar way, but since we are only looking for Higgs decaying into two photons we need to multiply the Higgs production cross section by the corresponding branching fraction,

$$\frac{\Delta N}{\Delta t} = \sigma(pp \rightarrow H) \times \text{BR}(H \rightarrow \gamma\gamma) \times L = 8 \times 10^{-4} \text{ s}^{-1}. \quad (582)$$

The ratio of the two rates is

$$R = \frac{7 \times 10^8}{8 \times 10^{-4}} \sim 10^{12}. \quad (583)$$

This implies that LHC gets 1000 billion minimum bias events for every single Higgs event we observe decaying into two photons. Clearly a very sophisticated selection procedure must be employed to be able to select the Higgs events from the background.

13.2 (i) Let's show that inhomogeneous Maxwell's equations take the compact form

$$\partial_\mu F^{\mu\nu} = j^\nu. \quad (584)$$

This can be done just by separating temporal and spatial components and looking at the matrix form (459) of the tensor $F^{\mu\nu}$. Choosing $\nu = 0$, we find

$$j^0 = \rho = \partial_\mu F^{\mu 0} = \partial_0 F^{00} + \partial_1 F^{10} + \partial_2 F^{20} + \partial_3 F^{30}. \quad (585)$$

Substituting the matrix elements of $F^{\mu\nu}$ into (585) and identifying the partial derivatives, this becomes

$$\rho = \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} = \vec{\nabla} \cdot \vec{E} \quad (586)$$

which is Gauss's law (461). Next, consider the spatial components of j^ν . For $\nu = 1$, we have

$$\begin{aligned} j^1 &= \partial_0 F^{01} + \partial_1 F^{11} + \partial_2 F^{21} + \partial_3 F^{31} \\ j_x &= -\frac{\partial E_x}{\partial t} + \frac{\partial B_z}{\partial y} - \frac{\partial B_y}{\partial z} = -\frac{\partial E_x}{\partial t} + (\vec{\nabla} \times \vec{B})_x, \end{aligned} \quad (587)$$

and analogous relations can be obtained for $\nu = 2$ and $\nu = 3$. Thus we get

$$\vec{j} = -\partial_t \vec{E} + \vec{\nabla} \times \vec{B}, \quad (588)$$

i.e. (462).

The homogeneous equations (460) and (463) follow from the relation

$$\partial^\alpha F^{\beta\gamma} + \partial^\beta F^{\gamma\alpha} + \partial^\gamma F^{\alpha\beta} = 0. \quad (589)$$

Once again we consider here temporal and spatial components separately. Owing to permutation symmetry it is enough to take into account four cases: $(\alpha, \beta, \gamma) = (0, 0, 0)$, $(0, 0, i)$, $(0, i, j)$ and (i, j, k) . Moreover, taking into account that $F^{\mu\nu}$ is antisymmetric, it is easily seen that in the first two cases the left hand side of (589) trivially vanishes. For $(\alpha, \beta, \gamma) = (0, i, j)$ we have

$$\partial_t F^{ij} + \partial^i F^{j0} + \partial^j F^{0i} = 0.$$

Here if $i = 1, j = 2$, by looking at (459) we get

$$\frac{\partial F^{12}}{\partial t} - \frac{\partial F^{20}}{\partial x} - \frac{\partial F^{01}}{\partial y} = -\frac{\partial B_z}{\partial t} - \frac{\partial E_y}{\partial x} + \frac{\partial E_x}{\partial y} = -\left(\frac{\partial \vec{B}}{\partial t} + \vec{\nabla} \times \vec{E}\right)_z = 0.$$

Analogous relations can be obtained for $i = 2, j = 3$ and $i = 3, j = 1$, while the cases $i = j$ yield trivial identities. Thus we end up with (460). Finally let us consider the case $(\alpha, \beta, \gamma) = (i, j, k)$. From the permutation symmetry, and noting that the sum in the left hand side of (589) is zero if any two indices are equal, we just need to evaluate the case $i = 1, j = 2, k = 3$. One has

$$\partial^1 F^{23} + \partial^2 F^{31} + \partial^3 F^{12} = -\frac{\partial(-B_x)}{\partial x} - \frac{\partial(-B_y)}{\partial y} - \frac{\partial(-B_z)}{\partial z} = \vec{\nabla} \cdot \vec{B} = 0,$$

i.e. (463).

Now current conservation follows immediately from the fact that $F^{\mu\nu}$ is antisymmetric: one has

$$\partial_\nu \partial_\mu F^{\mu\nu} = \partial_\nu(j^\nu) = 0. \quad (590)$$

(ii) Let's verify that we can retrieve the homogeneous Maxwell's equations from (465). By substituting (465) into the left hand side of (460) we get

$$\begin{aligned} \vec{\nabla} \times \vec{E} + \partial_t \vec{B} &= \vec{\nabla} \times \left(-\partial_t \vec{A} - \vec{\nabla} V\right) + \partial_t \left(\vec{\nabla} \times \vec{A}\right) \\ &= \vec{\nabla} \times \left(-\partial_t \vec{A}\right) - \vec{\nabla} \times \left(\vec{\nabla} V\right) + \partial_t \left(\vec{\nabla} \times \vec{A}\right) \\ &= -\partial_t \left(\vec{\nabla} \times \vec{A}\right) - \vec{\nabla} \times \left(\vec{\nabla} V\right) + \partial_t \left(\vec{\nabla} \times \vec{A}\right) \\ &= -\vec{\nabla} \times \left(\vec{\nabla} V\right) \\ &= 0, \end{aligned} \quad (591)$$

i.e. (460) is satisfied. Now (463) can be verified just by substituting $\vec{B} = \vec{\nabla} \times \vec{A}$, since the divergence of the curl is always zero:

$$\begin{aligned} \vec{\nabla} \cdot \vec{B} &= \vec{\nabla} \cdot \left(\vec{\nabla} \times \vec{A}\right) \\ &= 0. \end{aligned} \quad (592)$$

To obtain Eqs. (461) and (462), we first note that the d'Alembertian is $\square^2 = (\partial_t^2 - \nabla^2)$. Thus we can rewrite (466) as

$$(\partial_t^2 - \nabla^2) A^\mu - \partial^\mu (\partial_\nu A^\nu) = j^\mu. \quad (593)$$

Let's look at the $\mu = 0$ component of the left hand side of (593). Recall, $A^0 = V$ and $\partial_\nu A^\nu = \partial_t V + \vec{\nabla} \cdot \vec{A}$. We get

$$\begin{aligned} (\partial_t^2 - \nabla^2) V - \partial_t (\partial_\nu A^\nu) &= \partial_t^2 V - \nabla^2 V - \partial_t \left(\partial_t V + \vec{\nabla} \cdot \vec{A} \right) \\ &= \partial_t^2 V - \nabla^2 V - \partial_t^2 V - \partial_t \left(\vec{\nabla} \cdot \vec{A} \right) \\ &= -\partial_t \left(\vec{\nabla} \cdot \vec{A} \right) - \nabla^2 V \\ &= \vec{\nabla} \cdot \left(-\partial_t \vec{A} \right) + \vec{\nabla} \cdot \left(-\vec{\nabla} V \right) \\ &= \vec{\nabla} \cdot \left(-\partial_t \vec{A} - \vec{\nabla} V \right) \\ &= \vec{\nabla} \cdot \vec{E}, \end{aligned}$$

where we have used (465) in the last line. Noting that the $\mu = 0$ component of the right hand side of (593) is ρ , it is seen that we have retrieved (461). Finally, let's look at the spatial components of (593), $\mu = 1, 2, 3$. Once again using $\partial_\nu A^\nu = \partial_t V + \vec{\nabla} \cdot \vec{A}$, and recalling the $\mu = 1, 2, 3$ components of A^μ are the components of the vector potential \vec{A} , we obtain

$$\begin{aligned} (\partial_t^2 - \nabla^2) \vec{A} + \vec{\nabla} (\partial_\nu A^\nu) &= \partial_t^2 \vec{A} - \nabla^2 \vec{A} + \vec{\nabla} \left(\partial_t V + \vec{\nabla} \cdot \vec{A} \right) \\ &= -\nabla^2 \vec{A} + \vec{\nabla} \left(\vec{\nabla} \cdot \vec{A} \right) + \partial_t^2 \vec{A} + \vec{\nabla} (\partial_t V). \end{aligned}$$

Using the triple cross product identity given in the hint, $\vec{\nabla} \times (\vec{\nabla} \times \vec{A}) = -\nabla^2 \vec{A} + \vec{\nabla} (\vec{\nabla} \cdot \vec{A})$, and rearranging the derivatives,

$$\begin{aligned} (\partial_t^2 - \nabla^2) \vec{A} + \vec{\nabla} (\partial_\nu A^\nu) &= \vec{\nabla} \times (\vec{\nabla} \times \vec{A}) - \partial_t \left(-\partial_t \vec{A} \right) - \partial_t \left(-\vec{\nabla} V \right) \\ &= \vec{\nabla} \times (\vec{\nabla} \times \vec{A}) - \partial_t \left(-\partial_t \vec{A} - \vec{\nabla} V \right) \\ &= \vec{\nabla} \times \vec{B} - \partial_t \vec{E}, \end{aligned} \quad (594)$$

where in the last line we have used (465). We know that the $\mu = 1, 2, 3$ components of the right hand side of (593) are the components of \vec{j} , thus we get (462). Therefore, we have shown that Maxwell's equations can be retrieved from (465) and (466).

Notice that (466) is equivalent to the compact expression (464), written in terms of the four-vector potential. Indeed, we have

$$\partial_\mu F^{\mu\nu} = \partial_\mu (\partial^\mu A^\nu - \partial^\nu A^\mu) = \square^2 A^\nu - \partial^\nu (\partial_\mu A^\mu). \quad (595)$$

Moreover, using the covariant notation it is easy to see that (589) —and thus inhomogeneous Maxwell's equations— follow from $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$. Indeed, one has

$$\partial^\alpha F^{\beta\gamma} + \partial^\beta F^{\gamma\alpha} + \partial^\gamma F^{\alpha\beta} = \partial^\alpha \partial^\beta A^\gamma - \partial^\alpha \partial^\gamma A^\beta + \partial^\beta \partial^\gamma A^\alpha - \partial^\beta \partial^\alpha A^\gamma + \partial^\gamma \partial^\alpha A^\beta - \partial^\gamma \partial^\beta A^\alpha = 0. \quad (596)$$

(iii) Given that Maxwell's Equations can be written in the covariant form discussed above, they must remain invariant under arbitrary Lorentz boosts. To show this explicitly, let us start with Maxwell's equations in an unprimed frame,

$$\partial_\alpha F^{\alpha\beta} = j^\beta \quad (597)$$

$$\partial_\alpha F_{\beta\gamma} + \partial_\beta F_{\gamma\alpha} + \partial_\gamma F_{\alpha\beta} = 0, \quad (598)$$

apply an arbitrary Lorentz boost, and show that they have the same form in the boosted, primed frame.

First, let us boost (597):

$$\Lambda^\gamma_\beta \partial_\alpha F^{\alpha\beta} = \Lambda^\gamma_\beta j^\beta. \quad (599)$$

Using the properties of the delta function δ^α_β (remember, it's just the identity matrix), we can rewrite this as

$$\Lambda^\gamma_\beta \partial_\alpha \delta^\alpha_\tau F^{\tau\beta} = \Lambda^\gamma_\beta j^\beta. \quad (600)$$

Now from (76) in [49] (norm is preserved under Lorentz transformations) we see that $\delta^\alpha_\beta = \Lambda_\gamma^\alpha \Lambda^\gamma_\beta$. Thus we have

$$\Lambda^\gamma_\beta \partial_\alpha \Lambda_\delta^\alpha \Lambda^\delta_\tau F^{\tau\beta} = \Lambda^\gamma_\beta j^\beta. \quad (601)$$

By rearranging, we get

$$(\Lambda_\delta^\alpha \partial_\alpha)(\Lambda^\delta_\tau \Lambda^\gamma_\beta F^{\tau\beta}) = (\Lambda^\gamma_\beta j^\beta) \quad (602)$$

$$\partial'_\delta F'^{\delta\gamma} = j'^\gamma, \quad (603)$$

which means that (597) remains invariant under arbitrary Lorentz boosts.

Finally, the invariance of (598) is rather trivial. An arbitrary boost leads to

$$\begin{aligned} \Lambda_\delta^\alpha \Lambda_\rho^\beta \Lambda_\sigma^\gamma (\partial_\alpha F_{\beta\gamma} + \partial_\beta F_{\gamma\alpha} + \partial_\gamma F_{\alpha\beta}) &= 0 \\ (\Lambda_\delta^\alpha \partial_\alpha)(\Lambda_\rho^\beta \Lambda_\sigma^\gamma F_{\beta\gamma}) + (\Lambda_\rho^\beta \partial_\beta)(\Lambda_\sigma^\gamma \Lambda_\delta^\alpha F_{\gamma\alpha}) + (\Lambda_\sigma^\gamma \partial_\gamma)(\Lambda_\delta^\alpha \Lambda_\rho^\beta F_{\alpha\beta}) &= 0 \\ \partial'_\delta F'_{\rho\sigma} + \partial'_\rho F'_{\sigma\delta} + \partial'_\sigma F'_{\delta\rho} &= 0. \end{aligned} \quad (604)$$

Therefore, we have shown that Maxwell's equations remain invariant under arbitrary Lorentz boosts.

13.3 For any system bound by a central potential, $V(r)$, the wave function can be decomposed into radial and angular parts, with the angular parts described by spherical harmonics: $\psi(r, \vartheta, \varphi) = R(r)Y_l^m(\vartheta, \varphi)$. The spherical harmonics are given by (417). The parity operation on spherical coordinates changes r to $-r$, ϑ to $\pi - \vartheta$, and φ to $\varphi + \pi$. Thus, $e^{im\varphi}$ goes to $e^{im\varphi + im\pi} = e^{im\pi} e^{im\varphi} = (-1)^m e^{im\varphi}$ and $P_l^m(\cos \vartheta)$ goes to $P_l^m(\cos(\pi - \vartheta)) = (-1)^{l+m} P_l^m(\cos \vartheta)$. Assembling this yields $PY_l^m = (-1)^l Y_l^m$.

13.4 The following decays had been observed: $\tau^+ \rightarrow \pi^- \pi^+ \pi^+$ and $\theta^+ \rightarrow \pi^+ \pi^0$. Let us find the parities of the τ^+ and the θ^+ from these decays. First off, we know that the pions have an intrinsic parity of -1 . Now, we have to consider the spatial contribution to parity in these decays. The τ , the θ , and the pions are all spin-0 particles. Thus, the total initial angular momentum in both of the above decays is $J = L + S = 0$, where $L = 0$ is the external angular momentum and $S = 0$ is the intrinsic spin. In the decay of the θ , the orbital angular momentum of the two pions must be equal to zero in order to conserve the total angular momentum. In this case the spatial contribution to parity is $(-1)^l = (-1)^0 = +1$. The total parity of the final state is then $P = P_{\pi^+} P_{\pi^0} P_{\text{spatial}} = (-1)(-1)(+1) = +1$. Thus, by parity conservation, $P(\theta) = +1$. In the decay of the τ , on the other hand, the situation is not so simple. The total orbital angular momentum has two components: The first is given by the angular momentum between the two π^+ . The second, by the angular momentum of the remaining π^- about the center of mass of the two π^+ . This sum must be equal to zero in order to conserve total angular momentum. This implies that these two components of the total orbital angular momentum must have the same magnitude. In this case the spatial component of parity is given by the product of the parities given by the two components discussed above $P_{\text{spatial}} = (-1)^l (-1)^l = +1$. So, the total parity of the final state of the τ decay, and thus of the τ itself, is $P_\tau = P_{\pi^+} P_{\pi^+} P_{\pi^-} P_{\text{spatial}} = (-1)^3 (+1) = -1$. So the two particles have different parities.

13.5 (i) $\Delta E \sim m_\pi = 135 \text{ MeV}$; (ii) $\Delta t \sim \hbar/(2\Delta E) = 2.44 \times 10^{-24} \text{ s}$; (iii) $\Delta x = c\Delta t = 0.73 \text{ fm}$.

14.1 From (493) it follows that the probabilities for flavor oscillation are

$$P(\nu_\mu \rightarrow \nu_\mu) = P(\nu_\mu \rightarrow \nu_\tau) = \frac{1}{4} (\cos^4 \theta_\odot + \sin^4 \theta_\odot + 1) \simeq 0.4, \quad (605)$$

$$P(\nu_\mu \rightarrow \nu_e) = P(\nu_e \rightarrow \nu_\mu) = P(\nu_e \rightarrow \nu_\tau) = \sin^2 \theta_\odot \cos^2 \theta_\odot \simeq 0.2, \quad (606)$$

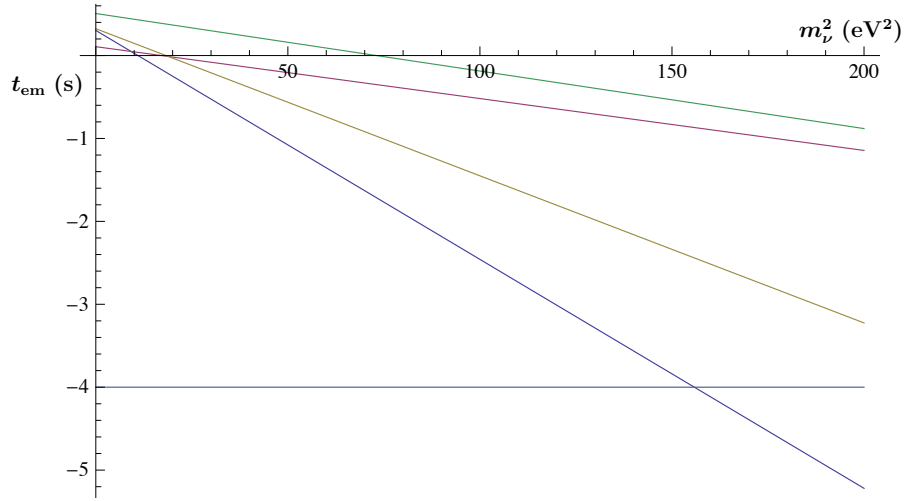


FIG. 67: Emission time as a function of the square neutrino mass. The 4 s emission window is indicated by the horizontal line.

and

$$P(\nu_e \rightarrow \nu_e) = \cos^4 \theta_\odot + \sin^4 \theta_\odot \simeq 0.6. \quad (607)$$

(i) For $(1 : 2 : 0)_S$, the earth expectation for ν_μ and ν_τ is $2 \times 0.4 + 1 \times 0.2 = 1$ and the ν_e expectation is $2 \times 0.2 + 1 \times 0.6 = 1$, that is $(1 : 1 : 1)_\oplus$. (ii) For $n \rightarrow pe^- \bar{\nu}_e$, we have $(1 : 0 : 0)_S$, yielding $(3 : 1 : 1)_\oplus$.

14.2 (i) The travel time is determined by $v = dx/dt$. Given that SN1987A was a fixed distance d away, $t_0 = d/c$. On the other hand, a massive neutrino satisfies the relation $E = \gamma m_\nu c^2$, where $\gamma = (1 - v^2/c^2)^{-1/2}$ is the Lorentz factor. Equating these two relations we see that

$$\frac{v}{c} = \sqrt{1 - \frac{m_\nu^2 c^4}{E^2}} \approx 1 - \frac{1}{2} \frac{m_\nu^2 c^4}{E^2} + \dots. \quad (608)$$

Hence, the neutrino travel time is

$$\Delta t = t_{\text{em}} - t_{\text{obs}} = \frac{d}{v} - \frac{ct_0}{v} \approx t_0 \left(1 + \frac{1}{2} \frac{m_\nu^2 c^4}{E^2} + \dots \right). \quad (609)$$

(ii) The emission time of each neutrino relative to the first one is then

$$t_{\text{em}} = t_{\text{obs}} - \frac{t_0 m_\nu^2 c^4}{2 E^2} + \frac{t_0 m_\nu^2 c^4}{2 E_1^2}, \quad (610)$$

where E_1 is the energy of the first neutrino to arrive. As shown in Fig. 67, neutrino 3 would need to have been emitted more than 4 s earlier than neutrino 1 if $m_\nu^2 > 156 \text{ eV}^2$. This translates into a limit on the neutrino mass to be no more than about 12.5 eV, as desired. (iii) The larger the neutrino mass, the larger the difference in the speeds of the two neutrinos, which makes it more difficult for the third neutrino to have been emitted only 1 s before neutrino 1. However, the difference in speeds is less if the energy of neutrino 1, E_1 , is at the bottom of its range and the energy of neutrino 3, E_3 , is at the top of its range. Therefore, we take $E_1 = 18.4 \text{ MeV}$ and $E_3 = 10.9 \text{ MeV}$. The largest possible mass is the one that saturates the inequality

$$-1 \leq t_{\text{obs}} - \frac{t_0 m_\nu^2 c^4}{2} \left(\frac{1}{E_3^2} - \frac{1}{E_1^2} \right) \quad (611)$$

where t_{obs} corresponds to neutrino 3. The corresponding limit is $m_\nu < 9.5 \text{ eV}$.

14.3 (i) Gauss's law states that the force due to gravity is only a function of the gravitational field inside the sphere of radius R . The gravitational field outside the sphere of radius R will result in a net zero force on the galaxy. Using

Newton's second law, $\vec{F} = m\vec{a}$, and noting that the net gravitational force will be in the radial direction, which means $a = d^2R/dt^2 \equiv \ddot{R}$, we can say

$$m\ddot{R} = -\frac{GmM}{R^2} = -\frac{Gm}{R^2} \left(\frac{4}{3}\pi R^3 \rho \right), \quad (612)$$

thus,

$$\ddot{R} = -\frac{4\pi}{3}G\rho R \quad (613)$$

and so

$$\frac{\ddot{R}}{R} = -\frac{4\pi}{3}G\rho, \quad (614)$$

which is the desired result. (ii) The total energy of the galaxy is given by the sum of its kinetic and potential energy,

$$E_{\text{tot}} = \frac{1}{2}m\dot{R}^2 - \frac{GmM}{R}. \quad (615)$$

Using that the total energy is zero and $M = \frac{4}{3}\pi R^3 \rho$,

$$\frac{1}{2}m\dot{R}^2 - \frac{Gm}{R} \left(\frac{4}{3}\pi R^3 \rho \right) = 0 \quad (616)$$

which implies

$$\frac{1}{2}\dot{R}^2 = \frac{4\pi}{3}G\rho R^2 \quad (617)$$

yielding

$$\left(\frac{\dot{R}}{R} \right)^2 = \frac{8\pi}{3}G\rho. \quad (618)$$

This is Friedman's equation for a flat universe in the Newtonian limit. For comparison, the full general relativistic Friedman's equation is [183]

$$\left(\frac{\dot{R}}{R} \right)^2 = \frac{8\pi}{3}G\rho - \frac{kc^2}{R^2}, \quad (619)$$

where k is the curvature constant, which is zero for a flat universe, and R has the meaning of the scale factor in this equation. The critical density is defined as the density of a spatially flat universe, i.e. $k = 0$. Identifying $H = \dot{R}/R$ as the Hubble parameter, from (618) we obtain (496). From (614) it follows that in the absence of vacuum energy the cosmic expansion is always decelerating.

TABLE VII: Riemann zeta function.

z	$\zeta(z)$
1	diverges
2	$1.644934 \dots \equiv \pi^2/6$
3	$1.202057 \dots$
4	$1.082323 \dots \equiv \pi^4/90$

Appendix B: Riemann zeta function

The Riemann zeta function $\zeta(z)$ is defined as

$$\zeta(z) \equiv \sum_{n=1}^{\infty} \frac{1}{n^z}. \quad (\text{B1})$$

It is easily seen that, for $\text{Re}(z) > 1$,

$$\zeta(z) = \frac{1}{\Gamma(z)} \int_0^{\infty} \frac{x^{z-1}}{e^x - 1} dx, \quad (\text{B2})$$

where the gamma function is related to the factorial function as $\Gamma(z) \equiv (z-1)!$. This enables one to evaluate integrals of the type (B2). The relevant values of $\zeta(z)$ are tabulated in Table VII.

Appendix C: Radiation from an accelerated charge

The path that leads from Maxwell's equations to the solution relevant in the case of the radiation field generated by an accelerated non-relativistic charge is somewhat arduous. It can be followed in your preferred electrodynamics book (see .e.g. [194]), and will not be given here. The essential insights can, however, be understood from an argument of Thomson as presented in [195].

Consider a charge q stationary at the origin O of some inertial frame of reference S at time $t = 0$. The charge then suffers a small acceleration to velocity Δv in the short time interval Δt . After a time t , we can distinguish between the field configuration inside and outside a sphere of radius $r = ct$ centred on the origin of S . Outside this sphere, the field lines do not yet know that the charge has moved away from the origin and so the field lines are radial, centred on O . Inside this sphere, the field lines are radial about the origin of the frame of reference centred on the moving charge. Between these two regions, there is a thin shell (perturbed zone) of thickness $c\Delta t$ in which the electric field lines are connected in a non radial way. In Fig. 68 we draw the electric field lines that result from this arrangement at a time t .

There must be a component of the electric field in the \hat{i}_θ direction. This “pulse” of accelerated charged electromagnetic field is propagated away from the charge at the speed of light and is the energy loss of the accelerated charged particle. The increment in velocity Δv is very small, $\Delta v \ll c$, and therefore it can be assumed that the field lines are radial at $t = 0$ and also at time t in the

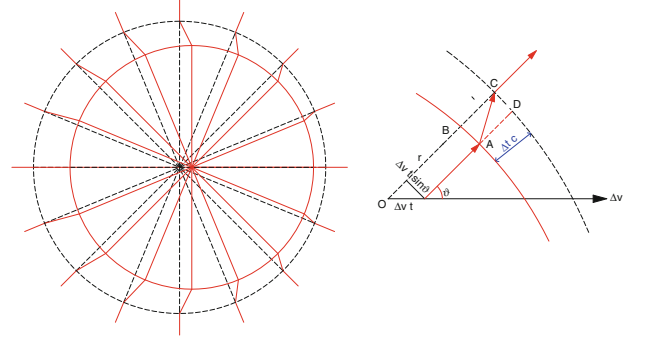


FIG. 68: Schematic view of the electric field lines at time t due to a charged particle accelerated to a velocity $\Delta v \ll c$ in a time interval Δt .

frame of reference S . Consider a small cone of electric field lines at angle θ with respect to the acceleration vector of the charge at $t = 0$ and at some later time t when the charge is moving at a constant velocity Δv . We join up electric field lines through the thin shell of thickness $c\Delta t$ as shown in Fig. 68. The strength of the E_θ component of the field is given by number of field lines per unit area in the \hat{i}_θ direction. From the geometry of the diagram, you can read that the ratio of the tangential to the radial field line components in the perturbed zone is

$$\frac{E_\theta}{E_r} = \frac{\Delta v t \sin \theta}{c\Delta t}. \quad (\text{C1})$$

Coulomb's law for the radial component E_r a distance r from a charge q is (in Gaussian cgs units)

$$E_r = \frac{q}{r^2}, \quad (\text{C2})$$

where $r = ct$. We can therefore deduce the tangential field component and find

$$\begin{aligned} E_\theta &= \frac{q(\Delta v/\Delta t) \sin \theta}{c^2 r} \\ &= \frac{q\dot{v} \sin \theta}{rc^2}. \end{aligned} \quad (\text{C3})$$

Note that according to Coulomb's law the radial component of the field decreases as r^{-2} , but the field in the pulse decreases only as r^{-1} . This is a characteristics of the radiation field in the far zone. The only electrical field component that is relevant for radiation is that which is perpendicular to the direction of propagation, i.e. E_θ . It is the one we consider further here.

We may now calculate the energy flux carried by this disturbance. The energy flux transported by electromagnetic fields is given by the Poynting vector

$$\vec{S} = c \vec{E} \times \vec{B}. \quad (\text{C4})$$

The magnetic field is equal and perpendicular to the electric field in electromagnetic radiation:

$$\vec{B} = \hat{n} \times \vec{E}. \quad (\text{C5})$$

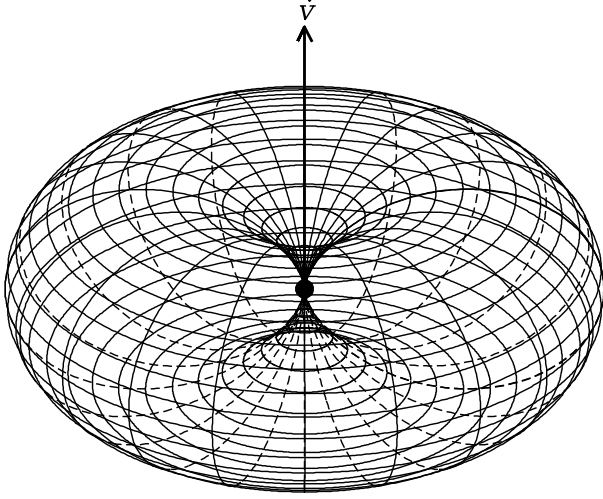


FIG. 69: The power pattern of Larmor radiation from a charged particle shown for an acceleration vector \vec{v} tilted 60° from the line of sight. The power received in any direction is proportional to the component of \vec{v} perpendicular to the line of sight.

The rate loss of energy (total power radiated) through the solid angle $d\Omega$ at distance r from the charge is therefore

$$\begin{aligned} P &= \frac{dE}{dt} d\Omega = |\vec{S}| r^2 d\Omega \\ &= \frac{q^2 \dot{v}^2}{4\pi c^3} \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi} \frac{\sin^2 \theta}{r^2} r^2 \sin \theta d\theta d\phi \\ &= \frac{q^2 \dot{v}^2}{2c^3} \int_{\theta=0}^{\pi} \sin^3 \theta d\theta. \end{aligned} \quad (C6)$$

Evaluating this integral gives $\int_{\theta=0}^{\pi} \sin^3 \theta d\theta = 4/3$ so the total power emitted is

$$P = \frac{2}{3} \frac{q^2 \dot{v}^2}{c^3}. \quad (C7)$$

This is the so-called Larmor formula. It states that any charged particle radiates when accelerated and that the total radiated power is proportional to the square of the acceleration. The charge radiates with a dipolar power pattern that looks like a doughnut whose axis is parallel to \vec{v} , see Fig. 69.

Appendix D: Magnetic dipole moment

As you very well know, a *magnetic dipole* is created by a flow of electric charge around an infinitesimal loop. The *magnetic dipole moment* $\vec{\mu}$ is a measure of the strength of the magnetic dipole, or equivalently to its ability to align with a given external magnetic field \vec{B} . For a planar loop encircling an area A , the magnetic moment is

$$\vec{\mu} = iA\hat{n}, \quad (D1)$$

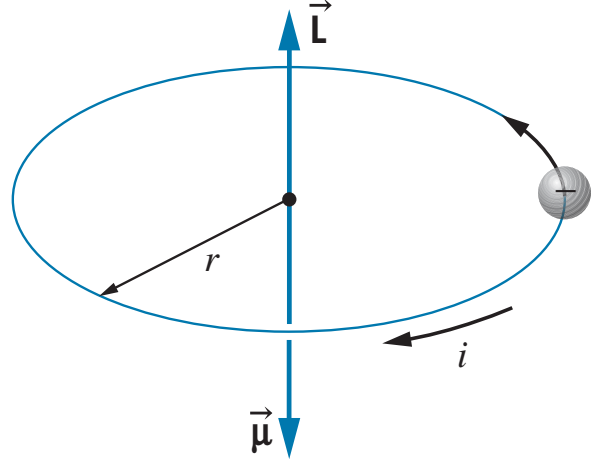


FIG. 70: A circulating negative charge is represented as a current loop. Because the charge is negative, \vec{L} and $\vec{\mu}$ point in opposite directions.

where i is the current and \hat{n} is the unit vector perpendicular to the loop plane. The energy of the unperturbed system depends on the angle θ between the magnetic moment and the external magnetic field,

$$U = -\vec{\mu} \cdot \vec{B} = -|\vec{\mu}|B_0 \cos \theta, \quad (D2)$$

where $\vec{B} = B_0 \hat{k}$. The torque on the magnetic moment is given by (426).

In Fig. 70 we show a particle with mass m and charge $q(<0)$ which is kept moving in a classical circular orbit by a central field $V(r)$. This constitutes a current loop of radius r , velocity v and revolution frequency $\nu = v/(2\pi r)$, which leads to a current $i = q\nu$ and a magnetic moment

$$|\vec{\mu}| = iA = q \frac{v}{2\pi r} \pi r^2 = \frac{1}{2} q r v = \frac{1}{2} |\vec{r} \times \vec{v}|. \quad (D3)$$

Hence the magnetic moment due to the motion of the charge is proportional to the orbital angular momentum

$$\vec{\mu}_L = \frac{q}{2m} \vec{L}. \quad (D4)$$

Note that for this kind of classical motion the angular momentum \vec{L} and the magnetic moment $\vec{\mu}_L$ are not constants of motion when \vec{B} differs from zero. Indeed, using Newton's second law and (426) we have that

$$\frac{d\vec{L}}{dt} = \vec{\tau} = \vec{\mu}_L \times \vec{B} = -\frac{q}{2m} \vec{B} \times \vec{L}. \quad (D5)$$

Now, using (D4) we obtain

$$\frac{d\vec{\mu}_L}{dt} = \frac{q}{2m} \frac{d\vec{L}}{dt} = -\left(\frac{q}{2m}\right)^2 \vec{B} \times \vec{L} = -\frac{q}{2m} \vec{B} \times \vec{\mu}_L. \quad (D6)$$

Note that the magnitude of the magnetic moment does not change since

$$\begin{aligned}\frac{d|\vec{L}|^2}{dt} &= \frac{d(\vec{L} \cdot \vec{L})}{dt} = 2\vec{L} \cdot \frac{d\vec{L}}{dt} = 2\vec{L} \cdot (\vec{\mu}_L \times \vec{B}) \\ &= 2\vec{L} \cdot \left(-\frac{q}{2m}\vec{B} \times \vec{L}\right) = 0.\end{aligned}\quad (\text{D7})$$

Similarly the angle between $\vec{\mu}_L$ and \vec{B} (i.e., the projection of $\vec{\mu}_L$ along the z axis) does not change, since

$$\begin{aligned}\frac{d(\vec{B} \cdot \vec{L})}{dt} &= 0 + \vec{B} \cdot \frac{d\vec{L}}{dt} = \vec{B} \cdot (\vec{\mu}_L \times \vec{B}) \\ &= -\frac{q}{2m}\vec{B} \cdot (\vec{L} \times \vec{B}) = 0.\end{aligned}\quad (\text{D8})$$

The equations of motion for the magnetic moment follow from (D6)

$$\begin{aligned}\frac{d\mu_{L,x}}{dt} &= -\omega_0 \mu_{L,y}, \\ \frac{d\mu_{L,y}}{dt} &= \omega_0 \mu_{L,x}, \\ \frac{d\mu_{L,z}}{dt} &= 0,\end{aligned}\quad (\text{D9})$$

where $\omega_0 = -q B_0/(2m)$ is the Larmor frequency. For the initial condition $\vec{\mu}_L(0) = \mu_{L,x}(0)\hat{i} + \mu_{L,z}(0)\hat{k}$, i.e. $\theta \neq 0$, the solutions are

$$\begin{aligned}\mu_{L,x}(t) &= \mu_{L,x}(0) \cos(\omega_0 t) \\ \mu_{L,y}(t) &= \mu_{L,x}(0) \sin(\omega_0 t).\end{aligned}\quad (\text{D10})$$

That is, the magnetic moment precesses around the z -axis at the Larmor frequency.

Appendix E: Pauli exclusion principle

Consider the simplest case of a two-particle system. The wave function is $\psi(\vec{r}_1, \vec{r}_2)$. If we assume that there is no interaction between the two particles, we will be able to describe the states using separation of variables:

$$\psi(\vec{r}_1, \vec{r}_2) = \psi_a(\vec{r}_1)\psi_b(\vec{r}_2) \quad (\text{E1})$$

where a and b label two different single-particle states. Implicit in this expression is the assumption that we can distinguish the two particles by some mean and link particle one to the position 1 and the state a . However, if we consider two identical particles (such as two electrons, two photons, two neutrons) there is no physical mean to distinguish them. Even if we try to measure them in order to keep track of which one is which, we know that in the process we destroy the state (by the wavefunction collapse) so not even this is a possibility.

In quantum mechanics identical particles are fundamentally indistinguishable. Then the expression above does not correctly describe the state anymore. In order

to faithfully describe a state in which we cannot know if particle a or b is at r_1 or r_2 , we can take a linear combination of these two possibilities:

$$\psi(\vec{r}_1, \vec{r}_2) = A_1 \psi_a(\vec{r}_1)\psi_b(\vec{r}_2) + A_2 \psi_b(\vec{r}_1)\psi_a(\vec{r}_2). \quad (\text{E2})$$

Now, since the two possibilities have the same probability, we have $|A_1| = |A_2| = 1/\sqrt{2}$. Then there are two possible combinations:

$$\psi(\vec{r}_1, \vec{r}_2) = \frac{1}{\sqrt{2}}[\psi_a(\vec{r}_1)\psi_b(\vec{r}_2) \pm \psi_b(\vec{r}_1)\psi_a(\vec{r}_2)]. \quad (\text{E3})$$

These two combinations describe two types of particle. The combination with the plus sign describes bosons, particles that are invariant under exchange of a particle pair. The combination with the minus sign describes fermions. All particles with integer spin are bosons, whereas all particles with half-integer spin are fermions.

We can define an operator \hat{P} that interchanges the two particles:

$$\hat{P}[\psi(\vec{r}_1, \vec{r}_2)] = \psi(\vec{r}_2, \vec{r}_1) \quad (\text{E4})$$

Since of course $\hat{P}[\hat{P}[\psi(\vec{r}_1, \vec{r}_2)]] = \psi(\vec{r}_1, \vec{r}_2)$, we have that $\hat{P}^2 = 1$. Then the eigenvalues of \hat{P} must be ± 1 . If two particles are identical, then the hamiltonian is invariant with respect to their exchange and $[H, \hat{P}] = 0$. Then we can find energy eigenfunctions that are common eigenfunctions of the exchange operator, or $\psi(\vec{r}_1, \vec{r}_2) = \pm \psi(\vec{r}_2, \vec{r}_1)$. Then if the system is initially in such a state, it will be always be in a state with the same exchange symmetry. For the considerations above, however, we have seen that the wavefunction is not only allowed, but it must be in a state with a definite symmetry:

$$\psi(\vec{r}_1, \vec{r}_2) = \begin{cases} \psi(\vec{r}_2, \vec{r}_1) & \text{bosons} \\ -\psi(\vec{r}_2, \vec{r}_1) & \text{fermions} \end{cases}. \quad (\text{E5})$$

From the form of the allowed wave function for fermions, it follows that two fermions cannot occupy the same state. Assume that $\psi_a(\vec{r}) = \psi_b(\vec{r})$, then we always have that

$$\psi_f(\vec{r}_1, \vec{r}_2) = \frac{1}{\sqrt{2}}[\psi_a(\vec{r}_1)\psi_b(\vec{r}_2) - \psi_b(\vec{r}_1)\psi_a(\vec{r}_2)] = 0. \quad (\text{E6})$$

This is the well-known Pauli exclusion principle. Notice that of course it applies to any fermions. For example, it applies to electrons, and this is the reason why electrons do not pile up in the lowest energy level of the atomic structure, but form a shell model.

Appendix F: A little group theory

A group is a set of elements S plus a composition rule \otimes , such that: (i) Combining two elements under the rule gives another element of the group, i.e. if $E, E' \in S$ then $E \otimes E' = E''$, with $E'' \in S$. (ii) The composition rule is

associative $E \otimes (E' \otimes E'') = (E \otimes E') \otimes E''$. (iii) There is an identity element $\mathbf{1}$, such that $E \otimes \mathbf{1} = \mathbf{1} \otimes E = E$ (iv) Every element has a unique inverse E^{-1} such that $E \otimes E^{-1} = E^{-1} \otimes E = \mathbf{1}$.

If the elements of a group are differentiable with respect to their parameters, the group is a Lie group. For a Lie group, any element can be written in the form

$$E(\theta_1, \theta_2, \dots, \theta_n) = \exp \left(\sum_{i=1}^n i\theta_i F_i \right). \quad (\text{F1})$$

The quantities F_i are the generators of the group. The quantities θ_i are the parameters of the group: a set of i real numbers that are needed to specify a particular element of the group. Note that the number of generators and parameters are the same. There is one generator for each parameter.

The group $U(1)$ is the set of all one dimensional, complex unitary matrices. The group has one generator $F = 1$, and one parameter, θ . It simply produces a complex phase change

$$E(\theta) = e^{-i\theta F} = e^{-i\theta}. \quad (\text{F2})$$

$U(1)$ is a Lie group,

$$\frac{dE}{d\theta} = iE. \quad (\text{F3})$$

Since the generator F , commutes with itself, the group elements also commute

$$E(\theta_1) \otimes E(\theta_2) = e^{-i\theta_1} e^{-i\theta_2} = e^{-i\theta_2} e^{-i\theta_1}. \quad (\text{F4})$$

Such groups are called Abelian groups.

The group $U(2)$ is the set of all two dimensional, complex unitary matrices. A complex $n \times n$ matrix has $2n^2$ real parameters. The unitary condition constraint removes n^2 of these. The group $U(2)$ then has four generators and four parameters

$$E(\theta_0, \theta_1, \theta_2, \theta_3) = e^{-i\theta_j F_j} \quad (\text{F5})$$

where $j = 0, 1, 2, 3$. The generators are: $F_i = \sigma_i/2$, where σ_0 denotes the identity matrix. The operators represented by the elements of $U(2)$ act on two dimensional complex vectors. The operations generated by $F_0 = \sigma_0/2$ simply change the complex phase of both components of the vector by the same amount. In general we are not so interested in these operations.

The group $SU(2)$ is the set of all two dimensional, complex unitary matrices with unit determinant. The unit determinant constraint removes one more parameter. The group $SU(2)$ then has three generators and three parameters

$$E(\theta_1, \theta_2, \theta_3) = e^{-i\theta_j F_j}, \quad (\text{F6})$$

where $j = 1, 2, 3$. The generators of $SU(2)$ are a set of three linearly independent, traceless 2×2 hermitian

matrices: $F_j = \tau_j/2$, where $\tau_j = \sigma_j$ with $j = 1, 2, 3$. Since the generators do not commute with one another, $[\sigma_i, \sigma_j] = 2i\epsilon_{ijk}\sigma_k$, this is a non-Abelian group. The Pauli matrices obey the following anticommutation relations $\{\sigma_i, \sigma_j\} = 2\delta_{ij}\mathbf{1}$.

The group $SU(3)$ is the set of all three dimensional, complex unitary matrices with unit determinant. This set has $2(3)^2 - (3)^2 - 1 = 8$ parameters and generators

$$E(\theta_1, \theta_2, \dots, \theta_8) = e^{-i\theta_j F_j}, \quad (\text{F7})$$

where $j = 1, 2, \dots, 8$. The generators of $SU(3)$ are a set of eight linearly independent, traceless 3×3 hermitian matrices. Since there are eight generators, the $SU(3)$ elements represent rotations of complex three component vectors in an eight dimensional space. The generators are

$$\begin{aligned} \lambda_1 &= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} & \lambda_2 &= \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \\ \lambda_3 &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} & \lambda_4 &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \\ \lambda_5 &= \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix} & \lambda_6 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \\ \lambda_7 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} & \lambda_8 &= \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \end{aligned} \quad (\text{F8})$$

The structure of $SU(3)$ is: $[\lambda_a, \lambda_b] = 2if_{abc}\lambda_c$, where $f_{123} = 1$, $f_{458} = f_{678} = \sqrt{3}/2$, $f_{147} = f_{516} = f_{246} = f_{257} = f_{345} = f_{637} = 1/2$; f_{abc} is totally antisymmetric.

Appendix G: Wu experiment

In the Wu experiment, ^{60}Co beta decays into $^{60}\text{Ni}^*$ (an excited form of Ni), which de-excites via gamma radiation to the ground state of Ni [136, 138]. From the angular distribution of the gamma rays, one can obtain the polarization of the original excited state of nickel. The result, roughly speaking, is that if the original Co (spin $J = 5$) is fully polarized ($J_z = 5$), then Ni^* has $J = 4, J_z = 4$. Again roughly speaking, we can regard the transition as if a particle with $J = 1, J_z = 1$ were to decay into a back-to-back pair $e^- \bar{\nu}_e$, with the electron emitted at an angle θ with respect to the original spin of the decaying spin-1 “particle.”

Consider an effective spin-1 particle decaying into an e and a $\bar{\nu}_e$. We have a Hamiltonian that interacts with a spin-1 particle at rest and as a consequence it decays back-to-back. If the original state of the decaying spin -1 particle is $|\Lambda\rangle$, then the *colinear* amplitude for decay to an angle θ for an electron with helicity λ_e , antineutrino

helicity $\lambda_{\bar{\nu}_e}$ is

$$\begin{aligned} \mathcal{M}_{\lambda_e \lambda_{\bar{\nu}_e}}^\Lambda(\theta) &= \langle \lambda_e \lambda_{\bar{\nu}_e} \theta | H | \Lambda \rangle = \sum_{\Lambda'} \langle \lambda_e \lambda_{\bar{\nu}_e} 0 | H | \Lambda' \rangle \langle \Lambda' | \Lambda \rangle \\ &= \langle \lambda_e \lambda_{\bar{\nu}_e} 0 | H | \lambda_e - \lambda_{\bar{\nu}_e} \rangle d_{\lambda_e - \lambda_{\bar{\nu}_e}, \Lambda}^1(\theta), \end{aligned} \quad (G1)$$

where conservation of angular momentum along the z' axis (the direction of the electron) implies $\Lambda' = \lambda_e - \lambda_{\bar{\nu}_e}$. For given Λ , say $\Lambda = 1$, there are four matrix elements:

$$\mathcal{M}_{\frac{1}{2} \frac{1}{2}}^1(\theta) = a d_{0 \ 1}^1(\theta) \quad (G2)$$

$$\mathcal{M}_{-\frac{1}{2} -\frac{1}{2}}^1(\theta) = \bar{a} d_{0 \ 1}^1(\theta) \quad (G3)$$

$$\mathcal{M}_{\frac{1}{2} -\frac{1}{2}}^1(\theta) = b d_{1 \ 1}^1(\theta) \quad (G4)$$

$$\mathcal{M}_{-\frac{1}{2} \frac{1}{2}}^1(\theta) = \bar{b} d_{-1 \ 1}^1(\theta), \quad (G5)$$

where for example, the collinear matrix element $b = \langle \frac{1}{2} - \frac{1}{2} 0 | H | 1 \rangle$, $\bar{b} = \langle -\frac{1}{2} + \frac{1}{2} 0 | H | -1 \rangle$, etcetera. Squaring and adding helicity amplitudes, we find the decay rate for electrons into angle θ with respect to the initial polarization direction

$$\begin{aligned} \frac{d\Gamma}{d\Omega} &\propto \frac{1}{2} (|a|^2 + |\bar{a}|^2) \sin^2 \theta \\ &+ \frac{1}{4} (|b|^2 + |\bar{b}|^2) (1 + \cos^2 \theta) \\ &+ \frac{1}{2} (|b|^2 - |\bar{b}|^2) \cos \theta. \end{aligned} \quad (G6)$$

The decay will be up-down symmetric iff $|b| = |\bar{b}|$. From the defining equations above, we can see that \bar{b} is the mirror image of b i.e., it is b with all the spins around the z' axis reversed. Thus, if an asymmetry is observed, parity is violated. Moreover, the asymmetry is predicted to go as $\cos \theta$. For weak interaction, we have maximal parity violation as $b = 0$. If $|b| = |\bar{b}|$ then there is no dependence with $\cos \theta$ and parity is conserved.

Appendix H: Electroweak symmetry breaking

Herein we demonstrate that the vacuum (477) breaks the gauge symmetry. The vacuum state $\langle \phi \rangle_0$ is invariant under a symmetry operation $\exp(i\theta_j F_j)$ corresponding

to the generator F_j provided that $\exp(i\theta_j F_j) \langle \phi \rangle_0 = \langle \phi \rangle_0$, that is, if $F_j \langle \phi \rangle_0 = 0$ (see Appendix F). We easily compute that

$$\begin{aligned} \frac{1}{2} \tau_1 \langle \phi \rangle_0 &= \frac{1}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix} \\ &= \frac{1}{2} \begin{pmatrix} v/\sqrt{2} \\ 0 \end{pmatrix} \neq 0 \quad \text{broken!} \\ \frac{1}{2} \tau_2 \langle \phi \rangle_0 &= \frac{1}{2} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix} \\ &= \frac{1}{2} \begin{pmatrix} -iv/\sqrt{2} \\ 0 \end{pmatrix} \neq 0 \quad \text{broken!} \\ \frac{1}{2} \tau_3 \langle \phi \rangle_0 &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix} \\ &= \frac{1}{2} \begin{pmatrix} 0 \\ -v/\sqrt{2} \end{pmatrix} \neq 0 \quad \text{broken!} \\ Y \langle \phi \rangle_0 &= Y_\phi \langle \phi \rangle_0 = +\frac{1}{2} \langle \phi \rangle_0 \\ &= \frac{1}{2} \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix} \neq 0 \quad \text{broken!} \end{aligned} \quad (H1)$$

However, if we examine the effect of the electric charge operator Q on the (electrically neutral) vacuum state, we find that

$$\begin{aligned} Q \langle \phi \rangle_0 &= \left(\frac{1}{2} \tau_3 + Y \right) \langle \phi \rangle_0 = \begin{pmatrix} Y_\phi + \frac{1}{2} & 0 \\ 0 & Y_\phi - \frac{1}{2} \end{pmatrix} \langle \phi \rangle_0 \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{unbroken!} \end{aligned} \quad (H2)$$

The original four generators are all broken, but the electric charge is not. It appears that we have accomplished our goal of breaking $SU(2)_L \otimes U(1)_Y \rightarrow U(1)_{\text{em}}$. We expect the photon to remain massless, and expect the gauge bosons that correspond to the generators τ_1 , τ_2 , and $\kappa \equiv (\frac{1}{2} \tau_3 - Y)$ to acquire masses.