

CHAPTER 6 CONCEPT CHECK ANSWERS

- 1. (a)** What is a one-to-one function? How can you tell if a function is one-to-one by looking at its graph?

A function f is one-to-one if it never takes on the same value twice; that is, $f(x_1) \neq f(x_2)$ whenever $x_1 \neq x_2$. When looking at a graph, use the Horizontal Line Test: a function is one-to-one if and only if no horizontal line intersects its graph more than once.

- (b)** If f is a one-to-one function, how is its inverse function f^{-1} defined? How do you obtain the graph of f^{-1} from the graph of f ?

If f is a one-to-one function with domain A and range B , then its inverse function f^{-1} has domain B and range A and is defined by

$$f^{-1}(y) = x \iff f(x) = y$$

for any y in B .

The graph of f^{-1} is obtained by reflecting the graph of f about the line $y = x$.

- (c)** If f is a one-to-one function and $f'(f^{-1}(a)) \neq 0$, write a formula for $(f^{-1})'(a)$.

$$(f^{-1})'(a) = \frac{1}{f'(f^{-1}(a))}$$

- 2. (a)** What are the domain and range of the natural exponential function $f(x) = e^x$?

Domain: \mathbb{R} Range: $(0, \infty)$

- (b)** What are the domain and range of the natural logarithmic function $g(x) = \ln x$?

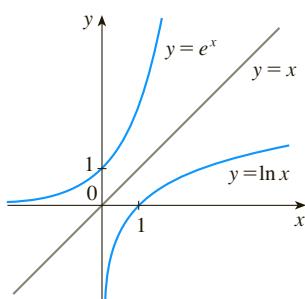
Domain: $(0, \infty)$ Range: \mathbb{R}

- (c)** How are the functions $f(x) = e^x$ and $g(x) = \ln x$ related?

They are inverses of each other.

- (d)** How are the graphs of these functions related? Sketch these graphs by hand, using the same axes.

The graphs are reflections of one another about the line $y = x$.



- (e)** If b is a positive number, $b \neq 1$, write an equation that expresses $\log_b x$ in terms of $\ln x$.

$$\log_b x = \frac{\ln x}{\ln b}$$

- 3. (a)** How is the inverse sine function $f(x) = \sin^{-1} x$ defined? What are its domain and range?

The inverse sine function $f(x) = \sin^{-1} x$ is defined as

$$\sin^{-1} x = y \iff \sin y = x$$

and $-\pi/2 \leq y \leq \pi/2$

Its domain is $-1 \leq x \leq 1$ and its range is $-\pi/2 \leq y \leq \pi/2$.

- (b)** How is the inverse cosine function $f(x) = \cos^{-1} x$ defined? What are its domain and range?

The inverse cosine function $f(x) = \cos^{-1} x$ is defined as

$$\cos^{-1} x = y \iff \cos y = x$$

and $0 \leq y \leq \pi$

Its domain is $-1 \leq x \leq 1$ and its range is $0 \leq y \leq \pi$.

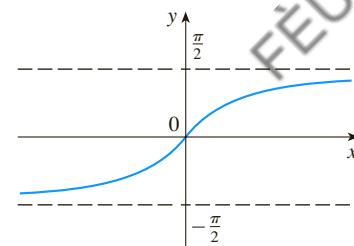
- (c)** How is the inverse tangent function $f(x) = \tan^{-1} x$ defined? What are its domain and range? Sketch its graph.

The inverse tangent function $f(x) = \tan^{-1} x$ is defined as

$$\tan^{-1} x = y \iff \tan y = x$$

and $-\pi/2 < y < \pi/2$

Its domain is \mathbb{R} and its range is $-\pi/2 < y < \pi/2$.



- 4.** Write the definitions of the hyperbolic functions $\sinh x$, $\cosh x$, and $\tanh x$.

$$\sinh x = \frac{e^x - e^{-x}}{2} \quad \cosh x = \frac{e^x + e^{-x}}{2}$$

$$\tanh x = \frac{\sinh x}{\cosh x} = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

- 5.** State the derivative of each function.

(a) $y = e^x$: $y' = e^x$

(b) $y = b^x$: $y' = b^x \ln b$

(c) $y = \ln x$: $y' = 1/x$

(d) $y = \log_b x$: $y' = 1/(x \ln b)$

(e) $y = \sin^{-1} x$: $y' = 1/\sqrt{1-x^2}$

(f) $y = \cos^{-1} x$: $y' = -1/\sqrt{1-x^2}$

(continued)

CHAPTER 6 CONCEPT CHECK ANSWERS (continued)

(g) $y = \tan^{-1}x$: $y' = 1/(1 + x^2)$

(h) $y = \sinh x$: $y' = \cosh x$

(i) $y = \cosh x$: $y' = \sinh x$

(j) $y = \tanh x$: $y' = \operatorname{sech}^2 x$

(k) $y = \sinh^{-1}x$: $y' = 1/\sqrt{1 + x^2}$

(l) $y = \cosh^{-1}x$: $y' = 1/\sqrt{x^2 - 1}$

(m) $y = \tanh^{-1}x$: $y' = 1/(1 - x^2)$

6. (a) How is the number e defined?

e is the number such that $\lim_{h \rightarrow 0} \frac{e^h - 1}{h} = 1$.

- (b) Express e as a limit.

$$e = \lim_{x \rightarrow 0} (1 + x)^{1/x}$$

- (c) Why is the natural exponential function $y = e^x$ used more often in calculus than the other exponential functions $y = b^x$?

The differentiation formula for $y = b^x$ [$dy/dx = b^x \ln b$] is simplest when $b = e$ because $\ln e = 1$.

- (d) Why is the natural logarithmic function $y = \ln x$ used more often in calculus than the other logarithmic functions $y = \log_b x$?

The differentiation formula for $y = \log_b x$ [$dy/dx = 1/(x \ln b)$] is simplest when $b = e$ because $\ln e = 1$.

7. (a) Write a differential equation that expresses the law of natural growth.

If $y(t)$ is the value of a quantity y at time t , then

$$\frac{dy}{dt} = ky \quad \text{where } k > 0 \text{ is a constant}$$

- (b) Under what circumstances is this an appropriate model for population growth?

The equation in part (a) is an appropriate model when there is enough room and nutrition to support growth.

- (c) What are the solutions of this equation?

If $y(0) = y_0$, then the solutions are $y(t) = y_0 e^{kt}$.

8. (a) What does l'Hospital's Rule say?

l'Hospital's Rule says that if the limit of a quotient of functions is an indeterminate form of type $0/0$ or ∞/∞ , then the limit is equal to the limit of the quotient of their derivatives:

$$\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = \lim_{x \rightarrow a} \frac{f'(x)}{g'(x)}$$

- (b) How can you use l'Hospital's Rule if you have a product $f(x)g(x)$, where $f(x) \rightarrow 0$ and $g(x) \rightarrow \infty$ as $x \rightarrow a$?

Write fg as $\frac{f}{1/g}$ or $\frac{g}{1/f}$ so that the limit becomes an indeterminate form of type $0/0$ or ∞/∞ .

- (c) How can you use l'Hospital's Rule if you have a difference $f(x) - g(x)$, where $f(x) \rightarrow \infty$ and $g(x) \rightarrow \infty$ as $x \rightarrow a$?

Convert the difference into a quotient by using a common denominator, rationalizing, factoring, or by some other method.

- (d) How can you use l'Hospital's Rule if you have a power $[f(x)]^{g(x)}$ where $f(x) \rightarrow 0$ and $g(x) \rightarrow 0$ as $x \rightarrow a$?

By taking the natural logarithm of both sides of $y = [f(x)]^{g(x)}$, we can convert the power to a product by writing $\ln y = g(x) \ln f(x)$. Alternatively, we can write the function as an exponential: $[f(x)]^{g(x)} = e^{g(x) \ln f(x)}$.

9. State whether each of the following limit forms is indeterminate. Where possible, state the limit.

- (a) $\frac{0}{0}$: indeterminate. L'Hospital's Rule can be applied to this form. Note that every derivative is a limit of this form.

- (b) $\frac{\infty}{\infty}$: indeterminate. L'Hospital's Rule can be applied to this form.

- (c) $\frac{0}{\infty}$: not indeterminate. A limit of this form has value 0.

- (d) $\frac{\infty}{0}$: not indeterminate. A limit of this form could equal ∞ , $-\infty$, or may not exist (but it cannot equal a finite value).

- (e) $\infty + \infty$: not indeterminate. A limit of this form is equal to ∞ .

- (f) $\infty - \infty$: indeterminate

- (g) $\infty \cdot \infty$: not indeterminate. A limit of this form is equal to ∞ .

- (h) $\infty \cdot 0$: indeterminate

- (i) 0^0 : indeterminate

- (j) 0^∞ : not indeterminate. A limit of this form has value 0.

- (k) ∞^0 : indeterminate

- (l) 1^∞ : indeterminate

CHAPTER 7 CONCEPT CHECK ANSWERS

- 1.** State the rule for integration by parts. In practice, how do you use it?

To integrate $\int f(x)g'(x) dx$, let $u = f(x)$ and $v = g(x)$. Then $\int u dv = uv - \int v du$.

In practice, try to choose $u = f(x)$ to be a function that becomes simpler when differentiated (or at least not more complicated) at long as $dv = g'(x) dx$ can be readily integrated to give v .

- 2.** How do you evaluate $\int \sin^m x \cos^n x dx$ if m is odd? What if n is odd? What if m and n are both even?

If m is odd, use $\sin^2 x = 1 - \cos^2 x$ to write all sine factors except one in terms of cosine. Then substitute $u = \cos x$.

If n is odd, use $\cos^2 x = 1 - \sin^2 x$ to write all cosine factors except one in terms of sine. Then substitute $u = \sin x$.

If m and n are even, use the half-angle identities

$$\sin^2 x = \frac{1}{2}(1 - \cos 2x) \quad \cos^2 x = \frac{1}{2}(1 + \cos 2x)$$

- 3.** If the expression $\sqrt{a^2 - x^2}$ occurs in an integral, what substitution might you try? What if $\sqrt{a^2 + x^2}$ occurs? What if $\sqrt{x^2 - a^2}$ occurs?

If $\sqrt{a^2 - x^2}$ occurs, try $x = a \sin \theta$; if $\sqrt{a^2 + x^2}$ occurs, try $x = a \tan \theta$, and if $\sqrt{x^2 - a^2}$ occurs, try $x = a \sec \theta$.

- 4.** What is the form of the partial fraction decomposition of a rational function $P(x)/Q(x)$ if the degree of P is less than the degree of Q and $Q(x)$ has only distinct linear factors? What if a linear factor is repeated? What if $Q(x)$ has an irreducible quadratic factor (not repeated)? What if the quadratic factor is repeated?

For distinct linear factors,

$$\frac{P(x)}{Q(x)} = \frac{A_1}{a_1x + b_1} + \frac{A_2}{a_2x + b_2} + \cdots + \frac{A_k}{a_kx + b_k}$$

If the linear factor $a_1x + b_1$ is repeated r times, then we must include all the terms

$$\frac{B_1}{a_1x + b_1} + \frac{B_2}{(a_1x + b_1)^2} + \cdots + \frac{B_r}{(a_1x + b_1)^r}$$

If $Q(x)$ has an irreducible quadratic factor (not repeated), then we include a term of the form

$$\frac{Ax + B}{ax^2 + bx + c}$$

If the irreducible quadratic factor is repeated r times, then we include all the terms

$$\frac{A_1x + B_1}{ax^2 + bx + c} + \frac{A_2x + B_2}{(ax^2 + bx + c)^2} + \cdots + \frac{A_rx + B_r}{(ax^2 + bx + c)^r}$$

- 5.** State the rules for approximating the definite integral

$\int_a^b f(x) dx$ with the Midpoint Rule, the Trapezoidal Rule, and Simpson's Rule. Which would you expect to give the best estimate? How do you approximate the error for each rule?

Let $a \leq x \leq b$, $I = \int_a^b f(x) dx$, and $\Delta x = (b - a)/n$.

Midpoint Rule:

$$I \approx M_n = \Delta x [f(\bar{x}_1) + f(\bar{x}_2) + \cdots + f(\bar{x}_n)]$$

where \bar{x}_i is the midpoint of $[x_{i-1}, x_i]$.

Trapezoidal Rule:

$$I \approx T_n = \frac{\Delta x}{2} [f(x_0) + 2f(x_1) + 2f(x_2) + \cdots + 2f(x_{n-1}) + f(x_n)]$$

where $x_i = a + i \Delta x$.

Simpson's Rule:

$$I \approx S_n = \frac{\Delta x}{3} [f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + \cdots + 2f(x_{n-2}) + 4f(x_{n-1}) + f(x_n)]$$

where n is even.

We would expect the best estimate to be given by Simpson's Rule.

Suppose $|f''(x)| \leq K$ and $|f^{(4)}(x)| \leq L$ for $a \leq x \leq b$. The errors in the Midpoint, Trapezoidal, and Simpson's Rules are given by, respectively,

$$|E_M| \leq \frac{K(b-a)^3}{24n^2} \quad |E_T| \leq \frac{K(b-a)^3}{12n^2}$$

$$|E_S| \leq \frac{L(b-a)^5}{180n^4}$$

- 6.** Define the following improper integrals.

(a) $\int_a^\infty f(x) dx = \lim_{t \rightarrow \infty} \int_a^t f(x) dx$

(b) $\int_{-\infty}^b f(x) dx = \lim_{t \rightarrow -\infty} \int_t^b f(x) dx$

(c) $\int_{-\infty}^\infty f(x) dx = \int_{-\infty}^a f(x) dx + \int_a^\infty f(x) dx$, where a is any real number (assuming that both integrals are convergent)

- 7.** Define the improper integral $\int_a^b f(x) dx$ for each of the following cases.

- (a) f has an infinite discontinuity at a .

If f is continuous on $(a, b]$, then

$$\int_a^b f(x) dx = \lim_{t \rightarrow a^+} \int_t^b f(x) dx$$

if this limit exists (as a finite number).

- (b) f has an infinite discontinuity at b .

If f is continuous on $[a, b)$, then

$$\int_a^b f(x) dx = \lim_{t \rightarrow b^-} \int_a^t f(x) dx$$

if this limit exists (as a finite number).

- (c) f has an infinite discontinuity at c , where $a < c < b$.

If both $\int_a^c f(x) dx$ and $\int_c^b f(x) dx$ are convergent, then

$$\int_a^b f(x) dx = \int_a^c f(x) dx + \int_c^b f(x) dx$$

- 8.** State the Comparison Theorem for improper integrals.

Suppose that f and g are continuous functions with $f(x) \geq g(x) \geq 0$ for $x \geq a$.

- (a) If $\int_a^\infty f(x) dx$ is convergent, then $\int_a^\infty g(x) dx$ is convergent.
 (b) If $\int_a^\infty g(x) dx$ is divergent, then $\int_a^\infty f(x) dx$ is divergent.

CHAPTER 8 CONCEPT CHECK ANSWERS

1. (a) How is the length of a curve defined?

We can approximate a curve C by a polygon with vertices P_i along C . The length L of C is defined to be the limit of the lengths of these inscribed polygons:

$$L = \lim_{n \rightarrow \infty} \sum_{i=1}^n |P_{i-1}P_i|$$

(b) Write an expression for the length of a smooth curve given by $y = f(x)$, $a \leq x \leq b$.

$$L = \int_a^b \sqrt{1 + [f'(x)]^2} dx$$

(c) What if x is given as a function of y ?

If $x = g(y)$, $c \leq y \leq d$, then $L = \int_c^d \sqrt{1 + [g'(y)]^2} dy$.

2. (a) Write an expression for the surface area of the surface obtained by rotating the curve $y = f(x)$, $a \leq x \leq b$, about the x -axis.

$$S = \int_a^b 2\pi f(x) \sqrt{1 + [f'(x)]^2} dx$$

(b) What if x is given as a function of y ?

If $x = g(y)$, $c \leq y \leq d$, then $S = \int_c^d 2\pi y \sqrt{1 + [g'(y)]^2} dy$.

(c) What if the curve is rotated about the y -axis?

$$S = \int_a^b 2\pi x \sqrt{1 + [f'(x)]^2} dx$$

or $S = \int_c^d 2\pi g(y) \sqrt{1 + [g'(y)]^2} dy$

3. Describe how we can find the hydrostatic force against a vertical wall submersed in a fluid.

We divide the wall into horizontal strips of equal height Δx and approximate each by a rectangle with horizontal length $f(x_i)$ at depth x_i . If δ is the weight density of the fluid, then the hydrostatic force is

$$F = \lim_{n \rightarrow \infty} \sum_{i=1}^n \delta x_i f(x_i) \Delta x = \int_a^b \delta x f(x) dx$$

4. (a) What is the physical significance of the center of mass of a thin plate?

The center of mass is the point at which the plate balances horizontally.

(b) If the plate lies between $y = f(x)$ and $y = 0$, where $a \leq x \leq b$, write expressions for the coordinates of the center of mass.

$$\bar{x} = \frac{1}{A} \int_a^b x f(x) dx \quad \text{and} \quad \bar{y} = \frac{1}{A} \int_a^b \frac{1}{2} [f(x)]^2 dx$$

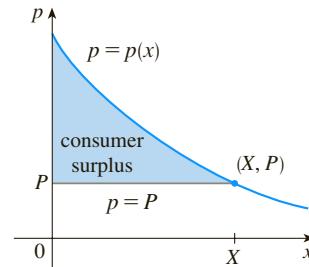
where $A = \int_a^b f(x) dx$.

5. What does the Theorem of Pappus say?

If a plane region \mathcal{R} that lies entirely on one side of a line ℓ in its plane is rotated about ℓ , then the volume of the resulting solid is the product of the area of \mathcal{R} and the distance traveled by the centroid of \mathcal{R} .

6. Given a demand function $p(x)$, explain what is meant by the consumer surplus when the amount of a commodity currently available is X and the current selling price is P . Illustrate with a sketch.

The consumer surplus represents the amount of money saved by consumers in purchasing the commodity at price P [when they were willing to purchase it at price $p(x)$], corresponding to an amount demanded of X .



7. (a) What is the cardiac output of the heart?

It is the volume of blood pumped by the heart per unit time, that is, the rate of flow into the aorta.

(b) Explain how the cardiac output can be measured by the dye dilution method.

An amount A of dye is injected into part of the heart and its concentration $c(t)$ leaving the heart is measured over a time interval $[0, T]$ until the dye has cleared. The cardiac output is given by $A/\int_0^T c(t) dt$.

8. What is a probability density function? What properties does such a function have?

Given a random variable X , its probability density function f is a function such that $\int_a^b f(x) dx$ gives the probability that X lies between a and b . The function f has the properties that $f(x) \geq 0$ for all x , and $\int_{-\infty}^{\infty} f(x) dx = 1$.

9. Suppose $f(x)$ is the probability density function for the weight of a female college student, where x is measured in pounds.

(a) What is the meaning of the integral $\int_0^{130} f(x) dx$?

It represents the probability that a randomly chosen female college student weighs less than 130 pounds.

(b) Write an expression for the mean of this density function.

$$\mu = \int_{-\infty}^{\infty} x f(x) dx = \int_0^{\infty} x f(x) dx$$

[since $f(x) = 0$ for $x < 0$]

(c) How can we find the median of this density function?

The median of f is the number m such that

$$\int_m^{\infty} f(x) dx = \frac{1}{2}$$

10. What is a normal distribution? What is the significance of the standard deviation?

A normal distribution corresponds to a random variable X that has a probability density function with a bell-shaped graph and equation given by

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$$

where μ is the mean and the positive constant σ is the standard deviation. σ measures how spread out the values of X are.

CHAPTER 9 CONCEPT CHECK ANSWERS

1. (a) What is a differential equation?

It is an equation that contains an unknown function and one or more of its derivatives.

(b) What is the order of a differential equation?

It is the order of the highest derivative that occurs in the equation.

(c) What is an initial condition?

It is a condition of the form $y(t_0) = y_0$.

2. What can you say about the solutions of the equation $y' = x^2 + y^2$ just by looking at the differential equation?

The equation tells us that the slope of a solution curve at any point (x, y) is $x^2 + y^2$. Note that $x^2 + y^2$ is always positive except at the origin, where $y' = x^2 + y^2 = 0$. Thus there is a horizontal tangent at $(0, 0)$ but nowhere else and the solution curves are increasing everywhere.

3. What is a direction field for the differential equation $y' = F(x, y)$?

A direction field (or slope field) for the differential equation $y' = F(x, y)$ is a two-dimensional graph consisting of short line segments with slope $F(x, y)$ at point (x, y) .

4. Explain how Euler's method works.

Euler's method says to start at the point given by the initial value and proceed in the direction indicated by the direction field. Stop after a short time, look at the slope at the new location, and proceed in that direction. Keep stopping and changing direction according to the direction field until the approximation is complete.

5. What is a separable differential equation? How do you solve it?

It is a differential equation in which the expression for dy/dx can be factored as a function of x times a function of y , that is, $dy/dx = g(x)f(y)$. We can solve the equation by rewriting it as $[1/f(y)]dy = g(x)dx$, integrating both sides, and solving for y .

6. What is a first-order linear differential equation? How do you solve it?

A first-order linear differential equation is a differential equation that can be put in the form

$$\frac{dy}{dx} + P(x)y = Q(x)$$

where P and Q are continuous functions on a given interval. To solve such an equation, we multiply both sides by the integrating factor $I(x) = e^{\int P(x) dx}$ to put it in the form

$(I(x)y)' = I(x)Q(x)$. We then integrate both sides and solve for y .

7. (a) Write a differential equation that expresses the law of natural growth. What does it say in terms of relative growth rate?

If $P(t)$ is the value of a quantity y at time t and if the rate of change of P with respect to t is proportional to its size

$$P(t) \text{ at any time, then } \frac{dP}{dt} = kP.$$

In this case the relative growth rate, $\frac{1}{P} \frac{dP}{dt}$, is constant.

(b) Under what circumstances is this an appropriate model for population growth?

It is an appropriate model under ideal conditions: unlimited environment, adequate nutrition, absence of predators and disease.

(c) What are the solutions of this equation?

If $P(0) = P_0$, the initial value, then the solutions are $P(t) = P_0e^{kt}$.

8. (a) Write the logistic differential equation.

The logistic differential equation is

$$\frac{dP}{dt} = kP\left(1 - \frac{P}{M}\right)$$

where M is the carrying capacity.

(b) Under what circumstances is this an appropriate model for population growth?

It is an appropriate model for population growth if the population grows at a rate proportional to the size of the population in the beginning, but eventually levels off and approaches its carrying capacity because of limited resources.

9. (a) Write Lotka-Volterra equations to model populations of food-fish (F) and sharks (S).

$$\frac{dF}{dt} = kF - aFS \quad \text{and} \quad \frac{dS}{dt} = -rS + bFS$$

(b) What do these equations say about each population in the absence of the other?

In the absence of sharks, an ample food supply would support exponential growth of the fish population, that is, $dF/dt = kF$, where k is a positive constant. In the absence of fish, we assume that the shark population would decline at a rate proportional to itself, that is $dS/dt = -rS$, where r is a positive constant.

CHAPTER 10 CONCEPT CHECK ANSWERS

1. (a) What is a parametric curve?

A parametric curve is a set of points of the form $(x, y) = (f(t), g(t))$, where f and g are functions of a variable t , the parameter.

(b) How do you sketch a parametric curve?

Sketching a parametric curve, like sketching the graph of a function, is difficult to do in general. We can plot points on the curve by finding $f(t)$ and $g(t)$ for various values of t , either by hand or with a calculator or computer. Sometimes, when f and g are given by formulas, we can eliminate t from the equations $x = f(t)$ and $y = g(t)$ to get a Cartesian equation relating x and y . It may be easier to graph that equation than to work with the original formulas for x and y in terms of t .

2. (a) How do you find the slope of a tangent to a parametric curve?

You can find dy/dx as a function of t by calculating

$$\frac{dy}{dx} = \frac{dy/dt}{dx/dt} \quad \text{if } dx/dt \neq 0$$

(b) How do you find the area under a parametric curve?

If the curve is traced out once by the parametric equations $x = f(t)$, $y = g(t)$, $\alpha \leq t \leq \beta$, then the area is

$$A = \int_a^b y \, dx = \int_\alpha^\beta g(t) f'(t) \, dt$$

[or $\int_\beta^\alpha g(t) f'(t) \, dt$ if the leftmost point is $(f(\beta), g(\beta))$ rather than $(f(\alpha), g(\alpha))$].

3. Write an expression for each of the following:

(a) The length of a parametric curve

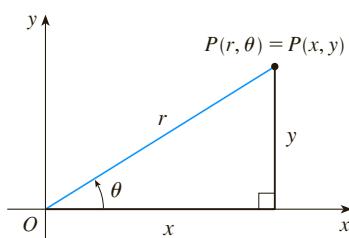
If the curve is traced out once by the parametric equations $x = f(t)$, $y = g(t)$, $\alpha \leq t \leq \beta$, then the length is

$$\begin{aligned} L &= \int_\alpha^\beta \sqrt{(dx/dt)^2 + (dy/dt)^2} \, dt \\ &= \int_\alpha^\beta \sqrt{[f'(t)]^2 + [g'(t)]^2} \, dt \end{aligned}$$

(b) The area of the surface obtained by rotating a parametric curve about the x -axis

$$\begin{aligned} S &= \int_\alpha^\beta 2\pi y \sqrt{(dx/dt)^2 + (dy/dt)^2} \, dt \\ &= \int_\alpha^\beta 2\pi g(t) \sqrt{[f'(t)]^2 + [g'(t)]^2} \, dt \end{aligned}$$

4. (a) Use a diagram to explain the meaning of the polar coordinates (r, θ) of a point.



(b) Write equations that express the Cartesian coordinates (x, y) of a point in terms of the polar coordinates.

$$x = r \cos \theta \quad y = r \sin \theta$$

(c) What equations would you use to find the polar coordinates of a point if you knew the Cartesian coordinates?

To find a polar representation (r, θ) with $r \geq 0$ and $0 \leq \theta < 2\pi$, first calculate $r = \sqrt{x^2 + y^2}$. Then θ is specified by $\tan \theta = y/x$. Be sure to choose θ so that (r, θ) lies in the correct quadrant.

5. (a) How do you find the slope of a tangent line to a polar curve?

$$\begin{aligned} \frac{dy}{dx} &= \frac{\frac{dy}{d\theta}}{\frac{dx}{d\theta}} = \frac{\frac{d}{d\theta}(y)}{\frac{d}{d\theta}(x)} = \frac{\frac{d}{d\theta}(r \sin \theta)}{\frac{d}{d\theta}(r \cos \theta)} \\ &= \frac{\left(\frac{dr}{d\theta}\right) \sin \theta + r \cos \theta}{\left(\frac{dr}{d\theta}\right) \cos \theta - r \sin \theta} \quad \text{where } r = f(\theta) \end{aligned}$$

(b) How do you find the area of a region bounded by a polar curve?

$$A = \int_a^b \frac{1}{2} r^2 \, d\theta = \int_a^b \frac{1}{2} [f(\theta)]^2 \, d\theta$$

(c) How do you find the length of a polar curve?

$$\begin{aligned} L &= \int_a^b \sqrt{(dx/d\theta)^2 + (dy/d\theta)^2} \, d\theta \\ &= \int_a^b \sqrt{r^2 + (dr/d\theta)^2} \, d\theta \\ &= \int_a^b \sqrt{[f(\theta)]^2 + [f'(\theta)]^2} \, d\theta \end{aligned}$$

6. (a) Give a geometric definition of a parabola.

A parabola is a set of points in a plane whose distances from a fixed point F (the focus) and a fixed line l (the directrix) are equal.

(b) Write an equation of a parabola with focus $(0, p)$ and directrix $y = -p$. What if the focus is $(p, 0)$ and the directrix is $x = -p$?

In the first case an equation is $x^2 = 4py$ and in the second case, $y^2 = 4px$.

7. (a) Give a definition of an ellipse in terms of foci.

An ellipse is a set of points in a plane the sum of whose distances from two fixed points (the foci) is a constant.

(b) Write an equation for the ellipse with foci $(\pm c, 0)$ and vertices $(\pm a, 0)$.

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

where $a \geq b > 0$ and $c^2 = a^2 - b^2$.

(continued)

CHAPTER 10 CONCEPT CHECK ANSWERS (continued)

8. (a) Give a definition of a hyperbola in terms of foci.

A hyperbola is a set of points in a plane the difference of whose distances from two fixed points (the foci) is a constant. This difference should be interpreted as the larger distance minus the smaller distance.

- (b) Write an equation for the hyperbola with foci $(\pm c, 0)$ and vertices $(\pm a, 0)$.

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$$

where $c^2 = a^2 + b^2$.

- (c) Write equations for the asymptotes of the hyperbola in part (b).

$$y = \pm \frac{b}{a} x$$

9. (a) What is the eccentricity of a conic section?

If a conic section has focus F and corresponding directrix l , then the eccentricity e is the fixed ratio $|PF|/|Pl|$ for points P of the conic section.

- (b) What can you say about the eccentricity if the conic section is an ellipse? A hyperbola? A parabola?

$e < 1$ for an ellipse; $e > 1$ for a hyperbola; $e = 1$ for a parabola

- (c) Write a polar equation for a conic section with eccentricity e and directrix $x = d$. What if the directrix is $x = -d$? $y = d$? $y = -d$?

$$\text{directrix } x = d: r = \frac{ed}{1 + e \cos \theta}$$

$$x = -d: r = \frac{ed}{1 - e \cos \theta}$$

$$y = d: r = \frac{ed}{1 + e \sin \theta}$$

$$y = -d: r = \frac{ed}{1 - e \sin \theta}$$

CHAPTER 11 CONCEPT CHECK ANSWERS

1. (a) What is a convergent sequence?

A convergent sequence $\{a_n\}$ is an ordered list of numbers where $\lim_{n \rightarrow \infty} a_n$ exists.

(b) What is a convergent series?

A series $\sum a_n$ is the *sum* of a sequence of numbers. It is convergent if the partial sums $s_n = \sum_{i=1}^n a_i$ approach a finite value, that is, $\lim_{n \rightarrow \infty} s_n$ exists as a real number.

(c) What does $\lim_{n \rightarrow \infty} a_n = 3$ mean?

The terms of the sequence $\{a_n\}$ approach 3 as n becomes large.

(d) What does $\sum_{n=1}^{\infty} a_n = 3$ mean?

By adding sufficiently many terms of the series, we can make the partial sums as close to 3 as we like.

2. (a) What is a bounded sequence?

A sequence $\{a_n\}$ is bounded if there are numbers m and M such that $m \leq a_n \leq M$ for all $n \geq 1$.

(b) What is a monotonic sequence?

A sequence is monotonic if it is either increasing or decreasing for all $n \geq 1$.

(c) What can you say about a bounded monotonic sequence?

Every bounded, monotonic sequence is convergent.

3. (a) What is a geometric series? Under what circumstances is it convergent? What is its sum?

A geometric series is of the form

$$\sum_{n=1}^{\infty} ar^{n-1} = a + ar + ar^2 + \dots$$

It is convergent if $|r| < 1$ and its sum is $\frac{a}{1-r}$.

(b) What is a p -series? Under what circumstances is it convergent?

A p -series is of the form $\sum_{n=1}^{\infty} \frac{1}{n^p}$. It is convergent if $p > 1$.

4. Suppose $\sum a_n = 3$ and s_n is the n th partial sum of the series. What is $\lim_{n \rightarrow \infty} a_n$? What is $\lim_{n \rightarrow \infty} s_n$?

If $\sum a_n = 3$, then $\lim_{n \rightarrow \infty} a_n = 0$ and $\lim_{n \rightarrow \infty} s_n = 3$.

5. State the following.

(a) The Test for Divergence

If $\lim_{n \rightarrow \infty} a_n$ does not exist or if $\lim_{n \rightarrow \infty} a_n \neq 0$, then the series $\sum_{n=1}^{\infty} a_n$ is divergent.

(b) The Integral Test

Suppose f is a continuous, positive, decreasing function on $[1, \infty)$ and let $a_n = f(n)$.

- If $\int_1^{\infty} f(x) dx$ is convergent, then $\sum_{n=1}^{\infty} a_n$ is convergent.
- If $\int_1^{\infty} f(x) dx$ is divergent, then $\sum_{n=1}^{\infty} a_n$ is divergent.

(c) The Comparison Test

Suppose that $\sum a_n$ and $\sum b_n$ are series with positive terms.

- If $\sum b_n$ is convergent and $a_n \leq b_n$ for all n , then $\sum a_n$ is also convergent.
- If $\sum b_n$ is divergent and $a_n \geq b_n$ for all n , then $\sum a_n$ is also divergent.

(d) The Limit Comparison Test

Suppose that $\sum a_n$ and $\sum b_n$ are series with positive terms.

If $\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = c$, where c is a finite number and $c > 0$, then either both series converge or both diverge.

(e) The Alternating Series Test

If the alternating series

$$\sum_{n=1}^{\infty} (-1)^{n-1} b_n = b_1 - b_2 + b_3 - b_4 + b_5 - b_6 + \dots$$

where $b_n > 0$ satisfies (i) $b_{n+1} \leq b_n$ for all n and

(ii) $\lim_{n \rightarrow \infty} b_n = 0$, then the series is convergent.

(f) The Ratio Test

- If $\lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| = L < 1$, then the series $\sum_{n=1}^{\infty} a_n$ is absolutely convergent (and therefore convergent).

- If $\lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| = L > 1$ or $\lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| = \infty$, then the series $\sum_{n=1}^{\infty} a_n$ is divergent.

- If $\lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| = 1$, the Ratio Test is inconclusive.

(g) The Root Test

- If $\lim_{n \rightarrow \infty} \sqrt[n]{|a_n|} = L < 1$, then the series $\sum_{n=1}^{\infty} a_n$ is absolutely convergent (and therefore convergent).

- If $\lim_{n \rightarrow \infty} \sqrt[n]{|a_n|} = L > 1$ or $\lim_{n \rightarrow \infty} \sqrt[n]{|a_n|} = \infty$, then the series $\sum_{n=1}^{\infty} a_n$ is divergent.

- If $\lim_{n \rightarrow \infty} \sqrt[n]{|a_n|} = 1$, the Root Test is inconclusive.

6. (a) What is an absolutely convergent series?

A series $\sum a_n$ is called absolutely convergent if the series of absolute values $\sum |a_n|$ is convergent.

(b) What can you say about such a series?

If a series $\sum a_n$ is absolutely convergent, then it is convergent.

(c) What is a conditionally convergent series?

A series $\sum a_n$ is called conditionally convergent if it is convergent but not absolutely convergent.

(continued)

CHAPTER 11 CONCEPT CHECK ANSWERS (continued)

- 7.** (a) If a series is convergent by the Integral Test, how do you estimate its sum?

The sum s can be estimated by the inequality

$$s_n + \int_{n+1}^{\infty} f(x) dx \leq s \leq s_n + \int_n^{\infty} f(x) dx$$

where s_n is the n th partial sum.

- (b) If a series is convergent by the Comparison Test, how do you estimate its sum?

We first estimate the remainder for the comparison series. This gives an upper bound for the remainder of the original series (as in Example 11.4.5).

- (c) If a series is convergent by the Alternating Series Test, how do you estimate its sum?

We can use a partial sum s_n of an alternating series as an approximation to the total sum. The size of the error is guaranteed to be no more than $|a_{n+1}|$, the absolute value of the first neglected term.

- 8.** (a) Write the general form of a power series.

A power series centered at a is

$$\sum_{n=0}^{\infty} c_n(x - a)^n$$

- (b) What is the radius of convergence of a power series?

Given the power series $\sum_{n=0}^{\infty} c_n(x - a)^n$, the radius of convergence is:

- (i) 0 if the series converges only when $x = a$,
- (ii) ∞ if the series converges for all x , or
- (iii) a positive number R such that the series converges if $|x - a| < R$ and diverges if $|x - a| > R$.

- (c) What is the interval of convergence of a power series?

The interval of convergence of a power series is the interval that consists of all values of x for which the series converges. Corresponding to the cases in part (b), the interval of convergence is (i) the single point $\{a\}$, (ii) $(-\infty, \infty)$, or (iii) an interval with endpoints $a - R$ and $a + R$ that can contain neither, either, or both of the endpoints.

- 9.** Suppose $f(x)$ is the sum of a power series with radius of convergence R .

- (a) How do you differentiate f ? What is the radius of convergence of the series for f' ?

If $f(x) = \sum_{n=0}^{\infty} c_n(x - a)^n$, then $f'(x) = \sum_{n=1}^{\infty} nc_n(x - a)^{n-1}$ with radius of convergence R .

- (b) How do you integrate f ? What is the radius of convergence of the series for $\int f(x) dx$?

$\int f(x) dx = C + \sum_{n=0}^{\infty} c_n \frac{(x - a)^{n+1}}{n + 1}$ with radius of convergence R .

- 10.** (a) Write an expression for the n th-degree Taylor polynomial of f centered at a .

$$T_n(x) = \sum_{i=0}^n \frac{f^{(i)}(a)}{i!} (x - a)^i$$

- (b) Write an expression for the Taylor series of f centered at a .

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!} (x - a)^n$$

- (c) Write an expression for the Maclaurin series of f .

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n \quad [a = 0 \text{ in part (b)}]$$

- (d) How do you show that $f(x)$ is equal to the sum of its Taylor series?

If $f(x) = T_n(x) + R_n(x)$, where $T_n(x)$ is the n th-degree Taylor polynomial of f and $R_n(x)$ is the remainder of the Taylor series, then we must show that

$$\lim_{n \rightarrow \infty} R_n(x) = 0$$

- (e) State Taylor's Inequality.

If $|f^{(n+1)}(x)| \leq M$ for $|x - a| \leq d$, then the remainder $R_n(x)$ of the Taylor series satisfies the inequality

$$|R_n(x)| \leq \frac{M}{(n + 1)!} |x - a|^{n+1} \quad \text{for } |x - a| \leq d$$

- 11.** Write the Maclaurin series and the interval of convergence for each of the following functions.

(a) $\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n, \quad R = 1$

(b) $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}, \quad R = \infty$

(c) $\sin x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}, \quad R = \infty$

(d) $\cos x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!}, \quad R = \infty$

(e) $\tan^{-1} x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{2n+1}, \quad R = 1$

(f) $\ln(1+x) = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{x^n}{n}, \quad R = 1$

- 12.** Write the binomial series expansion of $(1 + x)^k$. What is the radius of convergence of this series?

If k is any real number and $|x| < 1$, then

$$(1 + x)^k = \sum_{n=0}^{\infty} \binom{k}{n} x^n = 1 + kx + \frac{k(k-1)}{2!} x^2 + \frac{k(k-1)(k-2)}{3!} x^3 + \dots$$

The radius of convergence for the binomial series is 1.

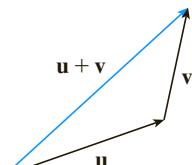
CHAPTER 12 CONCEPT CHECK ANSWERS

1. What is the difference between a vector and a scalar?

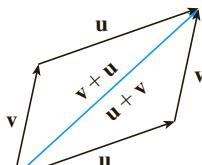
A scalar is a real number, whereas a vector is a quantity that has both a real-valued magnitude and a direction.

2. How do you add two vectors geometrically? How do you add them algebraically?

To add two vectors geometrically, we can use either the Triangle Law or the Parallelogram Law:



Triangle Law



Parallelogram Law

Algebraically, we add the corresponding components of the vectors.

3. If a is a vector and c is a scalar, how is ca related to a geometrically? How do you find ca algebraically?

For $c > 0$, ca is a vector with the same direction as a and length $|c|$ times the length of a . If $c < 0$, ca points in the direction opposite to a and has length $|c|$ times the length of a . Algebraically, to find ca we multiply each component of a by c .

4. How do you find the vector from one point to another?

The vector from point $A(x_1, y_1, z_1)$ to point $B(x_2, y_2, z_2)$ is given by

$$\langle x_2 - x_1, y_2 - y_1, z_2 - z_1 \rangle$$

5. How do you find the dot product $\mathbf{a} \cdot \mathbf{b}$ of two vectors if you know their lengths and the angle between them? What if you know their components?

If θ is the angle between the vectors \mathbf{a} and \mathbf{b} , then

$$\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}| |\mathbf{b}| \cos \theta$$

If $\mathbf{a} = \langle a_1, a_2, a_3 \rangle$ and $\mathbf{b} = \langle b_1, b_2, b_3 \rangle$, then

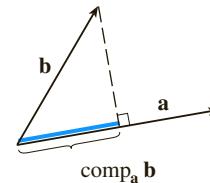
$$\mathbf{a} \cdot \mathbf{b} = a_1 b_1 + a_2 b_2 + a_3 b_3$$

6. How are dot products useful?

The dot product can be used to find the angle between two vectors. In particular, it can be used to determine whether two vectors are orthogonal. We can also use the dot product to find the scalar projection of one vector onto another. Additionally, if a constant force moves an object, the work done is the dot product of the force and displacement vectors.

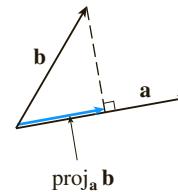
7. Write expressions for the scalar and vector projections of \mathbf{b} onto \mathbf{a} . Illustrate with diagrams.

Scalar projection of \mathbf{b} onto \mathbf{a} : $\text{comp}_{\mathbf{a}} \mathbf{b} = \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}|}$



Vector projection of \mathbf{b} onto \mathbf{a} :

$$\text{proj}_{\mathbf{a}} \mathbf{b} = \left(\frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}|} \right) \frac{\mathbf{a}}{|\mathbf{a}|} = \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}|^2} \mathbf{a}$$



8. How do you find the cross product $\mathbf{a} \times \mathbf{b}$ of two vectors if you know their lengths and the angle between them? What if you know their components?

If θ is the angle between \mathbf{a} and \mathbf{b} ($0 \leq \theta \leq \pi$), then $\mathbf{a} \times \mathbf{b}$ is the vector with length $|\mathbf{a} \times \mathbf{b}| = |\mathbf{a}| |\mathbf{b}| \sin \theta$ and direction orthogonal to both \mathbf{a} and \mathbf{b} , as given by the right-hand rule. If

$$\mathbf{a} = \langle a_1, a_2, a_3 \rangle \quad \text{and} \quad \mathbf{b} = \langle b_1, b_2, b_3 \rangle$$

then

$$\mathbf{a} \times \mathbf{b} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}$$

$$= \langle a_2 b_3 - a_3 b_2, a_3 b_1 - a_1 b_3, a_1 b_2 - a_2 b_1 \rangle$$

9. How are cross products useful?

The cross product can be used to create a vector orthogonal to two given vectors and it can be used to compute the area of a parallelogram determined by two vectors. Two nonzero vectors are parallel if and only if their cross product is $\mathbf{0}$. In addition, if a force acts on a rigid body, then the torque vector is the cross product of the position and force vectors.

(continued)

CHAPTER 12 CONCEPT CHECK ANSWERS

(continued)

10. (a) How do you find the area of the parallelogram determined by \mathbf{a} and \mathbf{b} ?

The area of the parallelogram determined by \mathbf{a} and \mathbf{b} is the length of the cross product: $|\mathbf{a} \times \mathbf{b}|$.

- (b) How do you find the volume of the parallelepiped determined by \mathbf{a} , \mathbf{b} , and \mathbf{c} ?

The volume of the parallelepiped determined by \mathbf{a} , \mathbf{b} , and \mathbf{c} is the magnitude of their scalar triple product: $|\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})|$.

11. How do you find a vector perpendicular to a plane?

If an equation of the plane is known, it can be written in the form $ax + by + cz + d = 0$. A normal vector, which is perpendicular to the plane, is $\langle a, b, c \rangle$ (or any nonzero scalar multiple of $\langle a, b, c \rangle$). If an equation is not known, we can use points on the plane to find two nonparallel vectors that lie in the plane. The cross product of these vectors is a vector perpendicular to the plane.

12. How do you find the angle between two intersecting planes?

The angle between two intersecting planes is defined as the acute angle θ between their normal vectors. If \mathbf{n}_1 and \mathbf{n}_2 are the normal vectors, then

$$\cos \theta = \frac{\mathbf{n}_1 \cdot \mathbf{n}_2}{|\mathbf{n}_1| |\mathbf{n}_2|}$$

13. Write a vector equation, parametric equations, and symmetric equations for a line.

A vector equation for a line that is parallel to a vector \mathbf{v} and that passes through a point with position vector \mathbf{r}_0 is $\mathbf{r} = \mathbf{r}_0 + t\mathbf{v}$. Parametric equations for a line through the point (x_0, y_0, z_0) and parallel to the vector $\langle a, b, c \rangle$ are

$$x = x_0 + at \quad y = y_0 + bt \quad z = z_0 + ct$$

while symmetric equations are

$$\frac{x - x_0}{a} = \frac{y - y_0}{b} = \frac{z - z_0}{c}$$

14. Write a vector equation and a scalar equation for a plane.

A vector equation of a plane that passes through a point with position vector \mathbf{r}_0 and that has normal vector \mathbf{n} (meaning \mathbf{n} is orthogonal to the plane) is $\mathbf{n} \cdot (\mathbf{r} - \mathbf{r}_0) = 0$ or, equivalently, $\mathbf{n} \cdot \mathbf{r} = \mathbf{n} \cdot \mathbf{r}_0$.

A scalar equation of a plane through a point (x_0, y_0, z_0) with normal vector $\mathbf{n} = \langle a, b, c \rangle$ is

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0$$

15. (a) How do you tell if two vectors are parallel?

Two (nonzero) vectors are parallel if and only if one is a scalar multiple of the other. In addition, two nonzero vectors are parallel if and only if their cross product is $\mathbf{0}$.

- (b) How do you tell if two vectors are perpendicular?

Two vectors are perpendicular if and only if their dot product is 0.

- (c) How do you tell if two planes are parallel?

Two planes are parallel if and only if their normal vectors are parallel.

16. (a) Describe a method for determining whether three points P , Q , and R lie on the same line.

Determine the vectors $\vec{PQ} = \mathbf{a}$ and $\vec{PR} = \mathbf{b}$. If there is a scalar t such that $\mathbf{a} = t\mathbf{b}$, then the vectors are parallel and the points must all lie on the same line.

Alternatively, if $\vec{PQ} \times \vec{PR} = \mathbf{0}$, then \vec{PQ} and \vec{PR} are parallel, so P , Q , and R are collinear.

An algebraic method is to determine an equation of the line joining two of the points, and then check whether or not the third point satisfies this equation.

- (b) Describe a method for determining whether four points P , Q , R , and S lie in the same plane.

Find the vectors $\vec{PQ} = \mathbf{a}$, $\vec{PR} = \mathbf{b}$, $\vec{PS} = \mathbf{c}$. Then $\mathbf{a} \times \mathbf{b}$ is normal to the plane formed by P , Q , and R , and so S lies on this plane if $\mathbf{a} \times \mathbf{b}$ and \mathbf{c} are orthogonal, that is, if $(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c} = 0$.

Alternatively, we can check if the volume of the parallelepiped determined by \mathbf{a} , \mathbf{b} , and \mathbf{c} is 0 (see Example 12.4.5).

An algebraic method is to find an equation of the plane determined by three of the points, and then check whether or not the fourth point satisfies this equation.

17. (a) How do you find the distance from a point to a line?

Let P be a point not on the line L that passes through the points Q and R and let $\mathbf{a} = \vec{QR}$, $\mathbf{b} = \vec{QP}$. The distance from the point P to the line L is

$$d = \frac{|\mathbf{a} \times \mathbf{b}|}{|\mathbf{a}|}$$

- (b) How do you find the distance from a point to a plane?

Let $P_0(x_0, y_0, z_0)$ be any point in the plane $ax + by + cz + d = 0$ and let $P_1(x_1, y_1, z_1)$ be a point not in the plane. If $\mathbf{b} = \vec{P_0P_1} = \langle x_1 - x_0, y_1 - y_0, z_1 - z_0 \rangle$, then the distance D from P_1 to the plane is equal to the absolute value of the scalar projection of \mathbf{b} onto the plane's normal vector $\mathbf{n} = \langle a, b, c \rangle$:

$$D = |\text{comp}_{\mathbf{n}} \mathbf{b}| = \frac{|\mathbf{n} \cdot \mathbf{b}|}{|\mathbf{n}|} = \frac{|ax_1 + by_1 + cz_1 + d|}{\sqrt{a^2 + b^2 + c^2}}$$

- (c) How do you find the distance between two lines?

Two skew lines L_1 and L_2 can be viewed as lying on two parallel planes, each with normal vector $\mathbf{n} = \mathbf{v}_1 \times \mathbf{v}_2$, where \mathbf{v}_1 and \mathbf{v}_2 are the direction vectors of L_1 and L_2 . After choosing one point on L_1 and determining the equation of the plane containing L_2 , we can proceed as in part (b). (See Example 12.5.10.)

(continued)

CHAPTER 12 CONCEPT CHECK ANSWERS (continued)

18. What are the traces of a surface? How do you find them?

The traces of a surface are the curves of intersection of the surface with planes parallel to the coordinate planes. We can find the trace in the plane $x = k$ (parallel to the yz -plane) by setting $x = k$ and determining the curve represented by the resulting equation. Traces in the planes $y = k$ (parallel to the xz -plane) and $z = k$ (parallel to the xy -plane) are found similarly.

19. Write equations in standard form of the six types of quadric surfaces.

Equations for the quadric surfaces symmetric with respect to the z -axis are as follows.

$$\text{Ellipsoid: } \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$$

$$\text{Cone: } \frac{z^2}{c^2} = \frac{x^2}{a^2} + \frac{y^2}{b^2}$$

Elliptic paraboloid:

$$\frac{z}{c} = \frac{x^2}{a^2} + \frac{y^2}{b^2}$$

Hyperboloid of one sheet:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1$$

Hyperboloid of two sheets:

$$-\frac{x^2}{a^2} - \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$$

Hyperbolic paraboloid:

$$\frac{z}{c} = \frac{x^2}{a^2} - \frac{y^2}{b^2}$$

CHAPTER 13 CONCEPT CHECK ANSWERS

- 1.** What is a vector function? How do you find its derivative and its integral?

A vector function is a function whose domain is a set of real numbers and whose range is a set of vectors. To find the derivative or integral, we can differentiate or integrate each component function of the vector function.

- 2.** What is the connection between vector functions and space curves?

A continuous vector function \mathbf{r} defines a space curve that is traced out by the tip of the moving position vector $\mathbf{r}(t)$.

- 3.** How do you find the tangent vector to a smooth curve at a point? How do you find the tangent line? The unit tangent vector?

The tangent vector to a smooth curve at a point P with position vector $\mathbf{r}(t)$ is the vector $\mathbf{r}'(t)$. The tangent line at P is the line through P parallel to the tangent vector $\mathbf{r}'(t)$. The unit tangent vector is $\mathbf{T}(t) = \frac{\mathbf{r}'(t)}{|\mathbf{r}'(t)|}$.

- 4.** If \mathbf{u} and \mathbf{v} are differentiable vector functions, c is a scalar, and f is a real-valued function, write the rules for differentiating the following vector functions.

(a) $\mathbf{u}(t) + \mathbf{v}(t)$

$$\frac{d}{dt} [\mathbf{u}(t) + \mathbf{v}(t)] = \mathbf{u}'(t) + \mathbf{v}'(t)$$

(b) $c\mathbf{u}(t)$

$$\frac{d}{dt} [c\mathbf{u}(t)] = c\mathbf{u}'(t)$$

(c) $f(t)\mathbf{u}(t)$

$$\frac{d}{dt} [f(t)\mathbf{u}(t)] = f'(t)\mathbf{u}(t) + f(t)\mathbf{u}'(t)$$

(d) $\mathbf{u}(t) \cdot \mathbf{v}(t)$

$$\frac{d}{dt} [\mathbf{u}(t) \cdot \mathbf{v}(t)] = \mathbf{u}'(t) \cdot \mathbf{v}(t) + \mathbf{u}(t) \cdot \mathbf{v}'(t)$$

(e) $\mathbf{u}(t) \times \mathbf{v}(t)$

$$\frac{d}{dt} [\mathbf{u}(t) \times \mathbf{v}(t)] = \mathbf{u}'(t) \times \mathbf{v}(t) + \mathbf{u}(t) \times \mathbf{v}'(t)$$

(f) $\mathbf{u}(f(t))$

$$\frac{d}{dt} [\mathbf{u}(f(t))] = f'(t)\mathbf{u}'(f(t))$$

- 5.** How do you find the length of a space curve given by a vector function $\mathbf{r}(t)$?

If $\mathbf{r}(t) = \langle f(t), g(t), h(t) \rangle$, $a \leq t \leq b$, and the curve is traversed exactly once as t increases from a to b , then the length is

$$L = \int_a^b |\mathbf{r}'(t)| dt = \int_a^b \sqrt{[f'(t)]^2 + [g'(t)]^2 + [h'(t)]^2} dt$$

- 6. (a)** What is the definition of curvature?

The curvature of a curve is $\kappa = \left| \frac{d\mathbf{T}}{ds} \right|$ where \mathbf{T} is the unit tangent vector.

- (b)** Write a formula for curvature in terms of $\mathbf{r}'(t)$ and $\mathbf{T}'(t)$.

$$\kappa(t) = \frac{|\mathbf{T}'(t)|}{|\mathbf{r}'(t)|}$$

- (c)** Write a formula for curvature in terms of $\mathbf{r}'(t)$ and $\mathbf{r}''(t)$.

$$\kappa(t) = \frac{|\mathbf{r}'(t) \times \mathbf{r}''(t)|}{|\mathbf{r}'(t)|^3}$$

- (d)** Write a formula for the curvature of a plane curve with equation $y = f(x)$.

$$\kappa(x) = \frac{|f''(x)|}{[1 + (f'(x))^2]^{3/2}}$$

- 7. (a)** Write formulas for the unit normal and binormal vectors of a smooth space curve $\mathbf{r}(t)$.

Unit normal vector: $\mathbf{N}(t) = \frac{\mathbf{T}'(t)}{|\mathbf{T}'(t)|}$

Binormal vector: $\mathbf{B}(t) = \mathbf{T}(t) \times \mathbf{N}(t)$

- (b)** What is the normal plane of a curve at a point? What is the osculating plane? What is the osculating circle?

The normal plane of a curve at a point P is the plane determined by the normal and binormal vectors \mathbf{N} and \mathbf{B} at P . The tangent vector \mathbf{T} is orthogonal to the normal plane.

The osculating plane at P is the plane determined by the vectors \mathbf{T} and \mathbf{N} . It is the plane that comes closest to containing the part of the curve near P .

The osculating circle at P is the circle that lies in the osculating plane of C at P , has the same tangent as C at P , lies on the concave side of C (toward which \mathbf{N} points), and has radius $\rho = 1/\kappa$ (the reciprocal of the curvature). It is the circle that best describes how C behaves near P ; it shares the same tangent, normal, and curvature at P .

(continued)

CHAPTER 13 CONCEPT CHECK ANSWERS (continued)

8. (a) How do you find the velocity, speed, and acceleration of a particle that moves along a space curve?

If $\mathbf{r}(t)$ is the position vector of the particle on the space curve, the velocity vector is $\mathbf{v}(t) = \mathbf{r}'(t)$, the speed is given by $|\mathbf{v}(t)|$, and the acceleration is $\mathbf{a}(t) = \mathbf{v}'(t) = \mathbf{r}''(t)$.

- (b) Write the acceleration in terms of its tangential and normal components.

$\mathbf{a} = a_T \mathbf{T} + a_N \mathbf{N}$, where $a_T = v'$ and $a_N = \kappa v^2$ ($v = |\mathbf{v}|$ is speed and κ is the curvature).

9. State Kepler's Laws.

- A planet revolves around the sun in an elliptical orbit with the sun at one focus.
- The line joining the sun to a planet sweeps out equal areas in equal times.
- The square of the period of revolution of a planet is proportional to the cube of the length of the major axis of its orbit.

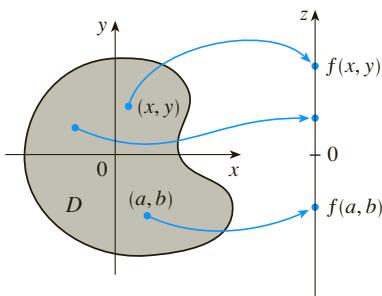
CHAPTER 14 CONCEPT CHECK ANSWERS

1. (a) What is a function of two variables?

A function f of two variables is a rule that assigns to each ordered pair (x, y) of real numbers in its domain a unique real number denoted by $f(x, y)$.

(b) Describe three methods for visualizing a function of two variables.

One way to visualize a function of two variables is by graphing it, resulting in the surface $z = f(x, y)$. Another method is a contour map, consisting of level curves $f(x, y) = k$ (k a constant), which are horizontal traces of the graph of the function projected onto the xy -plane. Also, we can use an arrow diagram such as the one below.



2. What is a function of three variables? How can you visualize such a function?

A function f of three variables is a rule that assigns to each ordered triple (x, y, z) in its domain a unique real number $f(x, y, z)$. We can visualize a function of three variables by examining its level surfaces $f(x, y, z) = k$, where k is a constant.

3. What does

$$\lim_{(x, y) \rightarrow (a, b)} f(x, y) = L$$

mean? How can you show that such a limit does not exist?

$\lim_{(x, y) \rightarrow (a, b)} f(x, y) = L$ means that the values of $f(x, y)$ approach the number L as the point (x, y) approaches the point (a, b) along any path that is within the domain of f . We can show that a limit at a point does not exist by finding two different paths approaching the point along which $f(x, y)$ has different limits.

4. (a) What does it mean to say that f is continuous at (a, b) ?

A function f of two variables is continuous at (a, b) if

$$\lim_{(x, y) \rightarrow (a, b)} f(x, y) = f(a, b)$$

(b) If f is continuous on \mathbb{R}^2 , what can you say about its graph?

If f is continuous on \mathbb{R}^2 , its graph will appear as a surface without holes or breaks.

5. (a) Write expressions for the partial derivatives $f_x(a, b)$ and $f_y(a, b)$ as limits.

$$f_x(a, b) = \lim_{h \rightarrow 0} \frac{f(a + h, b) - f(a, b)}{h}$$

$$f_y(a, b) = \lim_{h \rightarrow 0} \frac{f(a, b + h) - f(a, b)}{h}$$

(b) How do you interpret $f_x(a, b)$ and $f_y(a, b)$ geometrically? How do you interpret them as rates of change?

If $f(a, b) = c$, then the point $P(a, b, c)$ lies on the surface S given by $z = f(x, y)$. We can interpret $f_x(a, b)$ as the slope of the tangent line at P to the curve of intersection of the vertical plane $y = b$ and S . In other words, if we restrict ourselves to the path along S through P that is parallel to the xz -plane, then $f_x(a, b)$ is the slope at P looking in the positive x -direction. Similarly, $f_y(a, b)$ is the slope of the tangent line at P to the curve of intersection of the vertical plane $x = a$ and S .

If $z = f(x, y)$, then $f_x(x, y)$ can be interpreted as the rate of change of z with respect to x when y is fixed. Thus $f_x(a, b)$ is the rate of change of z (with respect to x) when y is fixed at b and x is allowed to vary from a . Similarly, $f_y(a, b)$ is the rate of change of z (with respect to y) when x is fixed at a and y is allowed to vary from b .

(c) If $f(x, y)$ is given by a formula, how do you calculate f_x and f_y ?

To find f_x , regard y as a constant and differentiate $f(x, y)$ with respect to x . To find f_y , regard x as a constant and differentiate $f(x, y)$ with respect to y .

6. What does Clairaut's Theorem say?

If f is a function of two variables that is defined on a disk D containing the point (a, b) and the functions f_{xy} and f_{yx} are both continuous on D , then Clairaut's Theorem states that $f_{xy}(a, b) = f_{yx}(a, b)$.

7. How do you find a tangent plane to each of the following types of surfaces?

(a) A graph of a function of two variables, $z = f(x, y)$

If f has continuous partial derivatives, an equation of the tangent plane to the surface $z = f(x, y)$ at the point (x_0, y_0, z_0) is

$$z - z_0 = f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0)$$

(b) A level surface of a function of three variables, $F(x, y, z) = k$

The tangent plane to the level surface $F(x, y, z) = k$ at $P(x_0, y_0, z_0)$ is the plane that passes through P and has normal vector $\nabla F(x_0, y_0, z_0)$:

$$F_x(x_0, y_0, z_0)(x - x_0) + F_y(x_0, y_0, z_0)(y - y_0)$$

$$+ F_z(x_0, y_0, z_0)(z - z_0) = 0$$

(continued)

CHAPTER 14 CONCEPT CHECK ANSWERS (continued)

8. Define the linearization of f at (a, b) . What is the corresponding linear approximation? What is the geometric interpretation of the linear approximation?

The linearization of f at (a, b) is the linear function whose graph is the tangent plane to the surface $z = f(x, y)$ at the point $(a, b, f(a, b))$:

$$L(x, y) = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b)$$

The linear approximation of f at (a, b) is

$$f(x, y) \approx f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b)$$

Geometrically, the linear approximation says that function values $f(x, y)$ can be approximated by values $L(x, y)$ from the tangent plane to the graph of f at $(a, b, f(a, b))$ when (x, y) is near (a, b) .

9. (a) What does it mean to say that f is differentiable at (a, b) ?

If $z = f(x, y)$, then f is differentiable at (a, b) if Δz can be expressed in the form

$$\Delta z = f_x(a, b) \Delta x + f_y(a, b) \Delta y + \varepsilon_1 \Delta x + \varepsilon_2 \Delta y$$

where ε_1 and $\varepsilon_2 \rightarrow 0$ as $(\Delta x, \Delta y) \rightarrow (0, 0)$. In other words, a differentiable function is one for which the linear approximation as stated above is a good approximation when (x, y) is near (a, b) .

- (b) How do you usually verify that f is differentiable?

If the partial derivatives f_x and f_y exist near (a, b) and are continuous at (a, b) , then f is differentiable at (a, b) .

10. If $z = f(x, y)$, what are the differentials dx , dy , and dz ?

The differentials dx and dy are independent variables that can be given any values. If f is differentiable, the differential dz is then defined by

$$dz = f_x(x, y) dx + f_y(x, y) dy$$

11. State the Chain Rule for the case where $z = f(x, y)$ and x and y are functions of one variable. What if x and y are functions of two variables?

Suppose that $z = f(x, y)$ is a differentiable function of x and y , where $x = g(t)$ and $y = h(t)$ are both differentiable functions of t . Then z is a differentiable function of t and

$$\frac{dz}{dt} = \frac{\partial f}{\partial x} \frac{dx}{dt} + \frac{\partial f}{\partial y} \frac{dy}{dt}$$

If $z = f(x, y)$ is a differentiable function of x and y , where $x = g(s, t)$ and $y = h(s, t)$ are differentiable functions of s and t , then

$$\frac{\partial z}{\partial s} = \frac{\partial z}{\partial x} \frac{\partial x}{\partial s} + \frac{\partial z}{\partial y} \frac{\partial y}{\partial s} \quad \frac{\partial z}{\partial t} = \frac{\partial z}{\partial x} \frac{\partial x}{\partial t} + \frac{\partial z}{\partial y} \frac{\partial y}{\partial t}$$

12. If z is defined implicitly as a function of x and y by an equation of the form $F(x, y, z) = 0$, how do you find $\frac{\partial z}{\partial x}$ and $\frac{\partial z}{\partial y}$?

If F is differentiable and $\frac{\partial F}{\partial z} \neq 0$, then

$$\frac{\frac{\partial z}{\partial x}}{\frac{\partial F}{\partial z}} = -\frac{\frac{\partial F}{\partial x}}{\frac{\partial F}{\partial z}} \quad \frac{\frac{\partial z}{\partial y}}{\frac{\partial F}{\partial z}} = -\frac{\frac{\partial F}{\partial y}}{\frac{\partial F}{\partial z}}$$

13. (a) Write an expression as a limit for the directional derivative of f at (x_0, y_0) in the direction of a unit vector $\mathbf{u} = \langle a, b \rangle$. How do you interpret it as a rate? How do you interpret it geometrically?

The directional derivative of f at (x_0, y_0) in the direction of \mathbf{u} is

$$D_{\mathbf{u}}f(x_0, y_0) = \lim_{h \rightarrow 0} \frac{f(x_0 + ha, y_0 + hb) - f(x_0, y_0)}{h}$$

if this limit exists.

We can interpret it as the rate of change of f (with respect to distance) at (x_0, y_0) in the direction of \mathbf{u} .

Geometrically, if P is the point $(x_0, y_0, f(x_0, y_0))$ on the graph of f and C is the curve of intersection of the graph of f with the vertical plane that passes through P in the direction of \mathbf{u} , then $D_{\mathbf{u}}f(x_0, y_0)$ is the slope of the tangent line to C at P .

- (b) If f is differentiable, write an expression for $D_{\mathbf{u}}f(x_0, y_0)$ in terms of f_x and f_y .

$$D_{\mathbf{u}}f(x_0, y_0) = f_x(x_0, y_0) a + f_y(x_0, y_0) b$$

14. (a) Define the gradient vector ∇f for a function f of two or three variables.

If f is a function of two variables, then

$$\nabla f(x, y) = \langle f_x(x, y), f_y(x, y) \rangle = \frac{\partial f}{\partial x} \mathbf{i} + \frac{\partial f}{\partial y} \mathbf{j}$$

For a function f of three variables,

$$\nabla f(x, y, z) = \langle f_x(x, y, z), f_y(x, y, z), f_z(x, y, z) \rangle$$

$$= \frac{\partial f}{\partial x} \mathbf{i} + \frac{\partial f}{\partial y} \mathbf{j} + \frac{\partial f}{\partial z} \mathbf{k}$$

- (b) Express $D_{\mathbf{u}}f$ in terms of ∇f .

$$D_{\mathbf{u}}f(x, y) = \nabla f(x, y) \cdot \mathbf{u}$$

$$\text{or } D_{\mathbf{u}}f(x, y, z) = \nabla f(x, y, z) \cdot \mathbf{u}$$

(continued)

CHAPTER 14 CONCEPT CHECK ANSWERS (continued)

(c) Explain the geometric significance of the gradient.

The gradient vector of f gives the direction of maximum rate of increase of f . On the graph of $z = f(x, y)$, ∇f points in the direction of steepest ascent. Also, the gradient vector is perpendicular to the level curves or level surfaces of a function.

15. What do the following statements mean?

(a) f has a local maximum at (a, b) .

f has a local maximum at (a, b) if $f(x, y) \leq f(a, b)$ when (x, y) is near (a, b) .

(b) f has an absolute maximum at (a, b) .

f has an absolute maximum at (a, b) if $f(x, y) \leq f(a, b)$ for all points (x, y) in the domain of f .

(c) f has a local minimum at (a, b) .

f has a local minimum at (a, b) if $f(x, y) \geq f(a, b)$ when (x, y) is near (a, b) .

(d) f has an absolute minimum at (a, b) .

f has an absolute minimum at (a, b) if $f(x, y) \geq f(a, b)$ for all points (x, y) in the domain of f .

(e) f has a saddle point at (a, b) .

f has a saddle point at (a, b) if $f(a, b)$ is a local maximum in one direction but a local minimum in another.

16. (a) If f has a local maximum at (a, b) , what can you say about its partial derivatives at (a, b) ?

If f has a local maximum at (a, b) and the first-order partial derivatives of f exist there, then $f_x(a, b) = 0$ and $f_y(a, b) = 0$.

(b) What is a critical point of f ?

A critical point of f is a point (a, b) such that $f_x(a, b) = 0$ and $f_y(a, b) = 0$ or one of these partial derivatives does not exist.

17. State the Second Derivatives Test.

Suppose the second partial derivatives of f are continuous on a disk with center (a, b) , and suppose that $f_{xx}(a, b) = 0$ and $f_{yy}(a, b) = 0$ [that is, (a, b) is a critical point of f]. Let

$$D = D(a, b) = f_{xx}(a, b)f_{yy}(a, b) - [f_{xy}(a, b)]^2$$

- If $D > 0$ and $f_{xx}(a, b) > 0$, then $f(a, b)$ is a local minimum.
- If $D > 0$ and $f_{xx}(a, b) < 0$, then $f(a, b)$ is a local maximum.
- If $D < 0$, then $f(a, b)$ is not a local maximum or minimum. The point (a, b) is a saddle point of f .

18. (a) What is a closed set in \mathbb{R}^2 ? What is a bounded set?

A closed set in \mathbb{R}^2 is one that contains all its boundary points. If one or more points on the boundary curve are omitted, the set is not closed.

A bounded set is one that is contained within some disk. In other words, it is finite in extent.

(b) State the Extreme Value Theorem for functions of two variables.

If f is continuous on a closed, bounded set D in \mathbb{R}^2 , then f attains an absolute maximum value $f(x_1, y_1)$ and an absolute minimum value $f(x_2, y_2)$ at some points (x_1, y_1) and (x_2, y_2) in D .

(c) How do you find the values that the Extreme Value Theorem guarantees?

- Find the values of f at the critical points of f in D .
- Find the extreme values of f on the boundary of D .
- The largest of the values from the above steps is the absolute maximum value; the smallest of these values is the absolute minimum value.

19. Explain how the method of Lagrange multipliers works in finding the extreme values of $f(x, y, z)$ subject to the constraint $g(x, y, z) = k$. What if there is a second constraint $h(x, y, z) = c$?

To find the maximum and minimum values of $f(x, y, z)$ subject to the constraint $g(x, y, z) = k$ [assuming that these extreme values exist and $\nabla g \neq \mathbf{0}$ on the surface $g(x, y, z) = k$], we first find all values of x, y, z , and λ where $\nabla f(x, y, z) = \lambda \nabla g(x, y, z)$ and $g(x, y, z) = k$. (Thus we are finding the points from the constraint where the gradient vectors ∇f and ∇g are parallel.) Evaluate f at all the resulting points (x, y, z) ; the largest of these values is the maximum value of f , and the smallest is the minimum value of f .

If there is a second constraint $h(x, y, z) = c$, then we find all values of x, y, z, λ , and μ such that

$$\nabla f(x, y, z) = \lambda \nabla g(x, y, z) + \mu \nabla h(x, y, z)$$

Again we find the extreme values of f by evaluating f at the resulting points (x, y, z) .

CHAPTER 15 CONCEPT CHECK ANSWERS

1. Suppose f is a continuous function defined on a rectangle $R = [a, b] \times [c, d]$.

- (a) Write an expression for a double Riemann sum of f . If $f(x, y) \geq 0$, what does the sum represent?

A double Riemann sum of f is

$$\sum_{i=1}^m \sum_{j=1}^n f(x_{ij}^*, y_{ij}^*) \Delta A$$

where ΔA is the area of each subrectangle and (x_{ij}^*, y_{ij}^*) is a sample point in each subrectangle. If $f(x, y) \geq 0$, this sum represents an approximation to the volume of the solid that lies above the rectangle R and below the graph of f .

- (b) Write the definition of $\iint_R f(x, y) dA$ as a limit.

$$\iint_R f(x, y) dA = \lim_{m, n \rightarrow \infty} \sum_{i=1}^m \sum_{j=1}^n f(x_{ij}^*, y_{ij}^*) \Delta A$$

- (c) What is the geometric interpretation of $\iint_R f(x, y) dA$ if $f(x, y) \geq 0$? What if f takes on both positive and negative values?

If $f(x, y) \geq 0$, $\iint_R f(x, y) dA$ represents the volume of the solid that lies above the rectangle R and below the surface $z = f(x, y)$. If f takes on both positive and negative values, then $\iint_R f(x, y) dA$ is $V_1 - V_2$, where V_1 is the volume above R and below the surface $z = f(x, y)$, and V_2 is the volume below R and above the surface.

- (d) How do you evaluate $\iint_R f(x, y) dA$?

We usually evaluate $\iint_R f(x, y) dA$ as an iterated integral according to Fubini's Theorem:

$$\iint_R f(x, y) dA = \int_a^b \int_c^d f(x, y) dy dx = \int_c^d \int_a^b f(x, y) dx dy$$

- (e) What does the Midpoint Rule for double integrals say?

The Midpoint Rule for double integrals says that we approximate the double integral $\iint_R f(x, y) dA$ by the double Riemann sum $\sum_{i=1}^m \sum_{j=1}^n f(\bar{x}_i, \bar{y}_j) \Delta A$, where the sample points (\bar{x}_i, \bar{y}_j) are the centers of the subrectangles.

- (f) Write an expression for the average value of f .

$$f_{ave} = \frac{1}{A(R)} \iint_R f(x, y) dA$$

where $A(R)$ is the area of R .

2. (a) How do you define $\iint_D f(x, y) dA$ if D is a bounded region that is not a rectangle?

Since D is bounded, it can be enclosed in a rectangular region R . We define a new function F with domain R by

$$F(x, y) = \begin{cases} f(x, y) & \text{if } (x, y) \text{ is in } D \\ 0 & \text{if } (x, y) \text{ is in } R \text{ but not in } D \end{cases}$$

Then we define

$$\iint_D f(x, y) dA = \iint_R F(x, y) dA$$

- (b) What is a type I region? How do you evaluate $\iint_D f(x, y) dA$ if D is a type I region?

A region D is of type I if it lies between the graphs of two continuous functions of x , that is,

$$D = \{(x, y) \mid a \leq x \leq b, g_1(x) \leq y \leq g_2(x)\}$$

where g_1 and g_2 are continuous on $[a, b]$. Then

$$\iint_D f(x, y) dA = \int_a^b \int_{g_1(x)}^{g_2(x)} f(x, y) dy dx$$

- (c) What is a type II region? How do you evaluate $\iint_D f(x, y) dA$ if D is a type II region?

A region D is of type II if it lies between the graphs of two continuous functions of y , that is,

$$D = \{(x, y) \mid c \leq y \leq d, h_1(y) \leq x \leq h_2(y)\}$$

where h_1 and h_2 are continuous on $[c, d]$. Then

$$\iint_D f(x, y) dA = \int_c^d \int_{h_1(y)}^{h_2(y)} f(x, y) dx dy$$

- (d) What properties do double integrals have?

$$\begin{aligned} \iint_D [f(x, y) + g(x, y)] dA &= \iint_D f(x, y) dA + \iint_D g(x, y) dA \end{aligned}$$

$$\iint_D cf(x, y) dA = c \iint_D f(x, y) dA$$

where c is a constant

- If $f(x, y) \geq g(x, y)$ for all (x, y) in D , then

$$\iint_D f(x, y) dA \geq \iint_D g(x, y) dA$$

- If $D = D_1 \cup D_2$, where D_1 and D_2 don't overlap except perhaps on their boundaries, then

$$\iint_D f(x, y) dA = \iint_{D_1} f(x, y) dA + \iint_{D_2} f(x, y) dA$$

$$\iint_D 1 dA = A(D), \text{ the area of } D.$$

- If $m \leq f(x, y) \leq M$ for all (x, y) in D , then

$$mA(D) \leq \iint_D f(x, y) dA \leq MA(D)$$

(continued)

CHAPTER 15 CONCEPT CHECK ANSWERS (continued)

- 3.** How do you change from rectangular coordinates to polar coordinates in a double integral? Why would you want to make the change?

We may want to change from rectangular to polar coordinates in a double integral if the region D of integration is more easily described in polar coordinates:

$$D = \{(r, \theta) \mid \alpha \leq \theta \leq \beta, h_1(\theta) \leq r \leq h_2(\theta)\}$$

To evaluate $\iint_D f(x, y) dA$, we replace x by $r \cos \theta$, y by $r \sin \theta$, and dA by $r dr d\theta$ (and use appropriate limits of integration):

$$\iint_D f(x, y) dA = \int_{\alpha}^{\beta} \int_{h_1(\theta)}^{h_2(\theta)} f(r \cos \theta, r \sin \theta) r dr d\theta$$

- 4.** If a lamina occupies a plane region D and has density function $\rho(x, y)$, write expressions for each of the following in terms of double integrals.

(a) The mass: $m = \iint_D \rho(x, y) dA$

(b) The moments about the axes:

$$M_x = \iint_D y \rho(x, y) dA \quad M_y = \iint_D x \rho(x, y) dA$$

(c) The center of mass:

$$(\bar{x}, \bar{y}), \quad \text{where } \bar{x} = \frac{M_y}{m} \quad \text{and} \quad \bar{y} = \frac{M_x}{m}$$

(d) The moments of inertia about the axes and the origin:

$$I_x = \iint_D y^2 \rho(x, y) dA$$

$$I_y = \iint_D x^2 \rho(x, y) dA$$

$$I_0 = \iint_D (x^2 + y^2) \rho(x, y) dA$$

- 5.** Let f be a joint density function of a pair of continuous random variables X and Y .

- (a) Write a double integral for the probability that X lies between a and b and Y lies between c and d .

$$P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f(x, y) dy dx$$

- (b) What properties does f possess?

$$f(x, y) \geq 0 \quad \iint_{\mathbb{R}^2} f(x, y) dA = 1$$

- (c) What are the expected values of X and Y ?

$$\text{The expected value of } X \text{ is } \mu_1 = \iint_{\mathbb{R}^2} xf(x, y) dA$$

$$\text{The expected value of } Y \text{ is } \mu_2 = \iint_{\mathbb{R}^2} yf(x, y) dA$$

- 6.** Write an expression for the area of a surface with equation $z = f(x, y)$, $(x, y) \in D$.

$$A(S) = \iint_D \sqrt{[f_x(x, y)]^2 + [f_y(x, y)]^2 + 1} dA$$

(assuming that f_x and f_y are continuous).

- 7. (a)** Write the definition of the triple integral of f over a rectangular box B .

$$\iiint_B f(x, y, z) dV = \lim_{l, m, n \rightarrow \infty} \sum_{i=1}^l \sum_{j=1}^m \sum_{k=1}^n f(x_{ijk}^*, y_{ijk}^*, z_{ijk}^*) \Delta V$$

where ΔV is the volume of each sub-box and $(x_{ijk}^*, y_{ijk}^*, z_{ijk}^*)$ is a sample point in each sub-box.

- (b)** How do you evaluate $\iiint_B f(x, y, z) dV$?

We usually evaluate $\iiint_B f(x, y, z) dV$ as an iterated integral according to Fubini's Theorem for Triple Integrals: If f is continuous on $B = [a, b] \times [c, d] \times [r, s]$, then

$$\iiint_B f(x, y, z) dV = \int_r^s \int_c^d \int_a^b f(x, y, z) dx dy dz$$

Note that there are five other orders of integration that we can use.

- (c)** How do you define $\iiint_E f(x, y, z) dV$ if E is a bounded solid region that is not a box?

Since E is bounded, it can be enclosed in a box B as described in part (b). We define a new function F with domain B by

$$F(x, y, z) = \begin{cases} f(x, y, z) & \text{if } (x, y, z) \text{ is in } E \\ 0 & \text{if } (x, y, z) \text{ is in } B \text{ but not in } E \end{cases}$$

Then we define

$$\iiint_E f(x, y, z) dV = \iiint_B F(x, y, z) dV$$

(continued)

CHAPTER 15 CONCEPT CHECK ANSWERS (continued)

- (d) What is a type 1 solid region? How do you evaluate $\iiint_E f(x, y, z) dV$ if E is such a region?

A region E is of type 1 if it lies between the graphs of two continuous functions of x and y , that is,

$$E = \{(x, y, z) \mid (x, y) \in D, u_1(x, y) \leq z \leq u_2(x, y)\}$$

where D is the projection of E onto the xy -plane. Then

$$\iiint_E f(x, y, z) dV = \iint_D \left[\int_{u_1(x, y)}^{u_2(x, y)} f(x, y, z) dz \right] dA$$

- (e) What is a type 2 solid region? How do you evaluate $\iiint_E f(x, y, z) dV$ if E is such a region?

A type 2 region is of the form

$$E = \{(x, y, z) \mid (y, z) \in D, u_1(y, z) \leq x \leq u_2(y, z)\}$$

where D is the projection of E onto the yz -plane. Then

$$\iiint_E f(x, y, z) dV = \iint_D \left[\int_{u_1(y, z)}^{u_2(y, z)} f(x, y, z) dx \right] dA$$

- (f) What is a type 3 solid region? How do you evaluate $\iiint_E f(x, y, z) dV$ if E is such a region?

A type 3 region is of the form

$$E = \{(x, y, z) \mid (x, z) \in D, u_1(x, z) \leq y \leq u_2(x, z)\}$$

where D is the projection of E onto the xz -plane. Then

$$\iiint_E f(x, y, z) dV = \iint_D \left[\int_{u_1(x, z)}^{u_2(x, z)} f(x, y, z) dy \right] dA$$

8. Suppose a solid object occupies the region E and has density function $\rho(x, y, z)$. Write expressions for each of the following.

- (a) The mass:

$$m = \iiint_E \rho(x, y, z) dV$$

- (b) The moments about the coordinate planes:

$$M_{yz} = \iiint_E x \rho(x, y, z) dV$$

$$M_{xz} = \iiint_E y \rho(x, y, z) dV$$

$$M_{xy} = \iiint_E z \rho(x, y, z) dV$$

- (c) The coordinates of the center of mass:

$$(\bar{x}, \bar{y}, \bar{z}), \text{ where } \bar{x} = \frac{M_{yz}}{m}, \bar{y} = \frac{M_{xz}}{m}, \bar{z} = \frac{M_{xy}}{m}$$

- (d) The moments of inertia about the axes:

$$I_x = \iiint_E (y^2 + z^2) \rho(x, y, z) dV$$

$$I_y = \iiint_E (x^2 + z^2) \rho(x, y, z) dV$$

$$I_z = \iiint_E (x^2 + y^2) \rho(x, y, z) dV$$

CHAPTER 15 CONCEPT CHECK ANSWERS (continued)

9. (a) How do you change from rectangular coordinates to cylindrical coordinates in a triple integral?

$$\iiint_E f(x, y, z) dV = \int_{\alpha}^{\beta} \int_{h_1(\theta)}^{h_2(\theta)} \int_{u_1(r \cos \theta, r \sin \theta)}^{u_2(r \cos \theta, r \sin \theta)} f(r \cos \theta, r \sin \theta, z) r dz dr d\theta$$

where

$$E = \{(r, \theta, z) \mid \alpha \leq \theta \leq \beta, h_1(\theta) \leq r \leq h_2(\theta), u_1(r \cos \theta, r \sin \theta) \leq z \leq u_2(r \cos \theta, r \sin \theta)\}$$

Thus we replace x by $r \cos \theta$, y by $r \sin \theta$, dV by $r dz dr d\theta$, and use appropriate limits of integration.

- (b) How do you change from rectangular coordinates to spherical coordinates in a triple integral?

$$\iiint_E f(x, y, z) dV = \int_c^d \int_{\alpha}^{\beta} \int_{g_1(\theta, \phi)}^{g_2(\theta, \phi)} f(\rho \sin \phi \cos \theta, \rho \sin \phi \sin \theta, \rho \cos \phi) \rho^2 \sin \phi d\rho d\theta d\phi$$

where

$$E = \{(\rho, \theta, \phi) \mid \alpha \leq \theta \leq \beta, c \leq \phi \leq d, g_1(\theta, \phi) \leq \rho \leq g_2(\theta, \phi)\}$$

Thus we replace x by $\rho \sin \phi \cos \theta$, y by $\rho \sin \phi \sin \theta$, z by $\rho \cos \phi$, dV by $\rho^2 \sin \phi d\rho d\theta d\phi$, and use appropriate limits of integration.

- (c) In what situations would you change to cylindrical or spherical coordinates?

We may want to change from rectangular to cylindrical or spherical coordinates in a triple integral if the region E of integration is more easily described in cylindrical or spherical coordinates. Regions that involve symmetry about the z -axis are often simpler to describe using cylindrical coordinates, and regions that are symmetrical about the origin are often simpler in spherical coordinates. Also, sometimes the integrand is easier to integrate using cylindrical or spherical coordinates.

10. (a) If a transformation T is given by $x = g(u, v)$, $y = h(u, v)$, what is the Jacobian of T ?

$$\frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \frac{\partial x}{\partial u} \frac{\partial y}{\partial v} - \frac{\partial x}{\partial v} \frac{\partial y}{\partial u}$$

- (b) How do you change variables in a double integral?

We change from an integral in x and y to an integral in u and v by expressing x and y in terms of u and v and writing

$$dA = \left| \frac{\partial(x, y)}{\partial(u, v)} \right| du dv$$

Thus, under the appropriate conditions,

$$\iint_R f(x, y) dA = \iint_S f(x(u, v), y(u, v)) \left| \frac{\partial(x, y)}{\partial(u, v)} \right| du dv$$

where R is the image of S under the transformation.

- (c) How do you change variables in a triple integral?

Similarly to the case of two variables in part (b),

$$\iiint_R f(x, y, z) dV = \iiint_S f(x(u, v, w), y(u, v, w), z(u, v, w)) \left| \frac{\partial(x, y, z)}{\partial(u, v, w)} \right| du dv dw$$

where

$$\frac{\partial(x, y, z)}{\partial(u, v, w)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} & \frac{\partial z}{\partial w} \end{vmatrix}$$

is the Jacobian.

CHAPTER 16 CONCEPT CHECK ANSWERS

- 1.** What is a vector field? Give three examples that have physical meaning.

A vector field is a function that assigns a vector to each point in its domain.

A vector field can represent, for example, the wind velocity at any location in space, the speed and direction of the ocean current at any location, or the force vector of the earth's gravitational field at a location in space.

- 2. (a) What is a conservative vector field?**

A conservative vector field \mathbf{F} is a vector field that is the gradient of some scalar function f , that is, $\mathbf{F} = \nabla f$.

- (b) What is a potential function?**

The function f in part (a) is called a potential function for \mathbf{F} .

- 3. (a) Write the definition of the line integral of a scalar function f along a smooth curve C with respect to arc length.**

If C is given by the parametric equations $x = x(t)$, $y = y(t)$, $a \leq t \leq b$, we divide the parameter interval $[a, b]$ into n subintervals $[t_{i-1}, t_i]$ of equal width. The i th subinterval corresponds to a subarc of C with length Δs_i . Then

$$\int_C f(x, y) ds = \lim_{n \rightarrow \infty} \sum_{i=1}^n f(x_i^*, y_i^*) \Delta s_i$$

where (x_i^*, y_i^*) is any sample point in the i th subarc.

- (b) How do you evaluate such a line integral?**

$$\int_C f(x, y) ds = \int_a^b f(x(t), y(t)) \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2} dt$$

Similarly, if C is a smooth space curve, then

$$\begin{aligned} \int_C f(x, y, z) ds &= \int_a^b f(x(t), y(t), z(t)) \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2} dt \end{aligned}$$

- (c) Write expressions for the mass and center of mass of a thin wire shaped like a curve C if the wire has linear density function $\rho(x, y)$.**

The mass is $m = \int_C \rho(x, y) ds$.

The center of mass is (\bar{x}, \bar{y}) , where

$$\bar{x} = \frac{1}{m} \int_C x \rho(x, y) ds$$

$$\bar{y} = \frac{1}{m} \int_C y \rho(x, y) ds$$

- (d) Write the definitions of the line integrals along C of a scalar function f with respect to x , y , and z .**

$$\int_C f(x, y, z) dx = \lim_{n \rightarrow \infty} \sum_{i=1}^n f(x_i^*, y_i^*, z_i^*) \Delta x_i$$

$$\int_C f(x, y, z) dy = \lim_{n \rightarrow \infty} \sum_{i=1}^n f(x_i^*, y_i^*, z_i^*) \Delta y_i$$

$$\int_C f(x, y, z) dz = \lim_{n \rightarrow \infty} \sum_{i=1}^n f(x_i^*, y_i^*, z_i^*) \Delta z_i$$

(We have similar results when f is a function of two variables.)

- (e) How do you evaluate these line integrals?**

$$\int_C f(x, y, z) dx = \int_a^b f(x(t), y(t), z(t)) x'(t) dt$$

$$\int_C f(x, y, z) dy = \int_a^b f(x(t), y(t), z(t)) y'(t) dt$$

$$\int_C f(x, y, z) dz = \int_a^b f(x(t), y(t), z(t)) z'(t) dt$$

- 4. (a) Define the line integral of a vector field \mathbf{F} along a smooth curve C given by a vector function $\mathbf{r}(t)$.**

If \mathbf{F} is a continuous vector field and C is given by a vector function $\mathbf{r}(t)$, $a \leq t \leq b$, then

$$\int_C \mathbf{F} \cdot d\mathbf{r} = \int_a^b \mathbf{F}(\mathbf{r}(t)) \cdot \mathbf{r}'(t) dt = \int_C \mathbf{F} \cdot \mathbf{T} ds$$

- (b) If \mathbf{F} is a force field, what does this line integral represent?**

It represents the work done by \mathbf{F} in moving a particle along the curve C .

- (c) If $\mathbf{F} = \langle P, Q, R \rangle$, what is the connection between the line integral of \mathbf{F} and the line integrals of the component functions P , Q , and R ?**

$$\int_C \mathbf{F} \cdot d\mathbf{r} = \int_C P dx + Q dy + R dz$$

- 5. State the Fundamental Theorem for Line Integrals.**

If C is a smooth curve given by $\mathbf{r}(t)$, $a \leq t \leq b$, and f is a differentiable function whose gradient vector ∇f is continuous on C , then

$$\int_C \nabla f \cdot d\mathbf{r} = f(\mathbf{r}(b)) - f(\mathbf{r}(a))$$

- 6. (a) What does it mean to say that $\int_C \mathbf{F} \cdot d\mathbf{r}$ is independent of path?**

$\int_C \mathbf{F} \cdot d\mathbf{r}$ is independent of path if the line integral has the same value for any two curves that have the same initial points and the same terminal points.

- (b) If you know that $\int_C \mathbf{F} \cdot d\mathbf{r}$ is independent of path, what can you say about \mathbf{F} ?**

We know that \mathbf{F} is a conservative vector field, that is, there exists a function f such that $\nabla f = \mathbf{F}$.

(continued)

CHAPTER 16 CONCEPT CHECK ANSWERS

(continued)

7. State Green's Theorem.

Let C be a positively oriented, piecewise-smooth, simple closed curve in the plane and let D be the region bounded by C . If P and Q have continuous partial derivatives on an open region that contains D , then

$$\int_C P \, dx + Q \, dy = \iint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dA$$

8. Write expressions for the area enclosed by a curve C in terms of line integrals around C .

$$A = \oint_C x \, dy = -\oint_C y \, dx = \frac{1}{2} \oint_C x \, dy - y \, dx$$

9. Suppose \mathbf{F} is a vector field on \mathbb{R}^3 .

(a) Define $\text{curl } \mathbf{F}$.

$$\text{curl } \mathbf{F} = \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z} \right) \mathbf{i} + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x} \right) \mathbf{j} + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) \mathbf{k}$$

$$= \nabla \times \mathbf{F}$$

(b) Define $\text{div } \mathbf{F}$.

$$\text{div } \mathbf{F} = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z} = \nabla \cdot \mathbf{F}$$

(c) If \mathbf{F} is a velocity field in fluid flow, what are the physical interpretations of $\text{curl } \mathbf{F}$ and $\text{div } \mathbf{F}$?

At a point in the fluid, the vector $\text{curl } \mathbf{F}$ aligns with the axis about which the fluid tends to rotate, and its length measures the speed of rotation; $\text{div } \mathbf{F}$ at a point measures the tendency of the fluid to flow away (diverge) from that point.

10. If $\mathbf{F} = P \mathbf{i} + Q \mathbf{j}$, how do you determine whether \mathbf{F} is conservative? What if \mathbf{F} is a vector field on \mathbb{R}^3 ?

If P and Q have continuous first-order derivatives and $\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}$, then \mathbf{F} is conservative.

If \mathbf{F} is a vector field on \mathbb{R}^3 whose component functions have continuous partial derivatives and $\text{curl } \mathbf{F} = \mathbf{0}$, then \mathbf{F} is conservative.

11. (a) What is a parametric surface? What are its grid curves?

A parametric surface S is a surface in \mathbb{R}^3 described by a vector function

$$\mathbf{r}(u, v) = x(u, v) \mathbf{i} + y(u, v) \mathbf{j} + z(u, v) \mathbf{k}$$

of two parameters u and v . Equivalent parametric equations are

$$x = x(u, v) \quad y = y(u, v) \quad z = z(u, v)$$

The grid curves of S are the curves that correspond to holding either u or v constant.

(b) Write an expression for the area of a parametric surface.

If S is a smooth parametric surface given by

$$\mathbf{r}(u, v) = x(u, v) \mathbf{i} + y(u, v) \mathbf{j} + z(u, v) \mathbf{k}$$

where $(u, v) \in D$ and S is covered just once as (u, v) ranges throughout D , then the surface area of S is

$$A(S) = \iint_D |\mathbf{r}_u \times \mathbf{r}_v| dA$$

(c) What is the area of a surface given by an equation $z = g(x, y)$?

$$A(S) = \iint_D \sqrt{1 + \left(\frac{\partial z}{\partial x} \right)^2 + \left(\frac{\partial z}{\partial y} \right)^2} dA$$

12. (a) Write the definition of the surface integral of a scalar function f over a surface S .

We divide S into “patches” S_{ij} . Then

$$\iint_S f(x, y, z) dS = \lim_{m, n \rightarrow \infty} \sum_{i=1}^m \sum_{j=1}^n f(P_{ij}^*) \Delta S_{ij}$$

where ΔS_{ij} is the area of the patch S_{ij} and P_{ij}^* is a sample point from the patch. (S is divided into patches in such a way that ensures that $\Delta S_{ij} \rightarrow 0$ as $m, n \rightarrow \infty$.)

(b) How do you evaluate such an integral if S is a parametric surface given by a vector function $\mathbf{r}(u, v)$?

$$\iint_S f(x, y, z) dS = \iint_D f(\mathbf{r}(u, v)) |\mathbf{r}_u \times \mathbf{r}_v| dA$$

where D is the parameter domain of S .

(c) What if S is given by an equation $z = g(x, y)$?

$$\begin{aligned} \iint_S f(x, y, z) dS \\ = \iint_D f(x, y, g(x, y)) \sqrt{\left(\frac{\partial z}{\partial x} \right)^2 + \left(\frac{\partial z}{\partial y} \right)^2 + 1} dA \end{aligned}$$

(d) If a thin sheet has the shape of a surface S , and the density at (x, y, z) is $\rho(x, y, z)$, write expressions for the mass and center of mass of the sheet.

The mass is

$$m = \iint_S \rho(x, y, z) dS$$

The center of mass is $(\bar{x}, \bar{y}, \bar{z})$, where

$$\bar{x} = \frac{1}{m} \iint_S x \rho(x, y, z) dS$$

$$\bar{y} = \frac{1}{m} \iint_S y \rho(x, y, z) dS$$

$$\bar{z} = \frac{1}{m} \iint_S z \rho(x, y, z) dS$$

(continued)

CHAPTER 16 CONCEPT CHECK ANSWERS (continued)

- 13. (a)** What is an oriented surface? Give an example of a non-orientable surface.

An oriented surface S is one for which we can choose a unit normal vector \mathbf{n} at every point so that \mathbf{n} varies continuously over S . The choice of \mathbf{n} provides S with an orientation.

A Möbius strip is a nonorientable surface. (It has only one side.)

- (b)** Define the surface integral (or flux) of a vector field \mathbf{F} over an oriented surface S with unit normal vector \mathbf{n} .

$$\iint_S \mathbf{F} \cdot d\mathbf{S} = \iint_S \mathbf{F} \cdot \mathbf{n} \, dS$$

- (c)** How do you evaluate such an integral if S is a parametric surface given by a vector function $\mathbf{r}(u, v)$?

$$\iint_S \mathbf{F} \cdot d\mathbf{S} = \iint_D \mathbf{F} \cdot (\mathbf{r}_u \times \mathbf{r}_v) \, dA$$

We multiply by -1 if the opposite orientation of S is desired.

- (d)** What if S is given by an equation $z = g(x, y)$?

If $\mathbf{F} = \langle P, Q, R \rangle$,

$$\iint_S \mathbf{F} \cdot d\mathbf{S} = \iint_D \left(-P \frac{\partial g}{\partial x} - Q \frac{\partial g}{\partial y} + R \right) dA$$

for the upward orientation of S ; we multiply by -1 for the downward orientation.

- 14. State Stokes' Theorem.**

Let S be an oriented piecewise-smooth surface that is bounded by a simple, closed, piecewise-smooth boundary curve C with positive orientation. Let \mathbf{F} be a vector field whose components have continuous partial derivatives on an open region in \mathbb{R}^3 that contains S . Then

$$\int_C \mathbf{F} \cdot d\mathbf{r} = \iint_S \operatorname{curl} \mathbf{F} \cdot d\mathbf{S}$$

- 15. State the Divergence Theorem.**

Let E be a simple solid region and let S be the boundary surface of E , given with positive (outward) orientation. Let \mathbf{F} be a vector field whose component functions have continuous partial derivatives on an open region that contains E . Then

$$\iint_S \mathbf{F} \cdot d\mathbf{S} = \iiint_E \operatorname{div} \mathbf{F} \, dV$$

- 16. In what ways are the Fundamental Theorem for Line Integrals, Green's Theorem, Stokes' Theorem, and the Divergence Theorem similar?**

In each theorem, we integrate a “derivative” over a region, and this integral is equal to an expression involving the values of the original function only on the *boundary* of the region.

CHAPTER 17 CONCEPT CHECK ANSWERS

- 1. (a)** Write the general form of a second-order homogeneous linear differential equation with constant coefficients.

$$ay'' + by' + cy = 0$$

where a , b , and c are constants and $a \neq 0$.

- (b)** Write the auxiliary equation.

$$ar^2 + br + c = 0$$

- (c)** How do you use the roots of the auxiliary equation to solve the differential equation? Write the form of the solution for each of the three cases that can occur.

If the auxiliary equation has two distinct real roots r_1 and r_2 , the general solution of the differential equation is

$$y = c_1 e^{r_1 x} + c_2 e^{r_2 x}$$

If the roots are real and equal, the solution is

$$y = c_1 e^{rx} + c_2 x e^{rx}$$

where r is the common root.

If the roots are complex, we can write $r_1 = \alpha + i\beta$ and $r_2 = \alpha - i\beta$, and the solution is

$$y = e^{\alpha x}(c_1 \cos \beta x + c_2 \sin \beta x)$$

- 2. (a)** What is an initial-value problem for a second-order differential equation?

An initial-value problem consists of finding a solution y of the differential equation that also satisfies given conditions $y(x_0) = y_0$ and $y'(x_0) = y_1$, where y_0 and y_1 are constants.

- (b)** What is a boundary-value problem for such an equation?

A boundary-value problem consists of finding a solution y of the differential equation that also satisfies given boundary conditions $y(x_0) = y_0$ and $y(x_1) = y_1$.

- 3. (a)** Write the general form of a second-order nonhomogeneous linear differential equation with constant coefficients.

$ay'' + by' + cy = G(x)$, where a , b , and c are constants and G is a continuous function.

- (b)** What is the complementary equation? How does it help solve the original differential equation?

The complementary equation is the related homogeneous equation $ay'' + by' + cy = 0$. If we find the general solution y_c of the complementary equation and y_p is any particular solution of the nonhomogeneous differential equation, then the general solution of the original differential equation is $y(x) = y_p(x) + y_c(x)$.

- (c)** Explain how the method of undetermined coefficients works.

To determine a particular solution y_p of $ay'' + by' + cy = G(x)$, we make an initial guess that y_p is a general function of the same type as G . If $G(x)$

is a polynomial, choose y_p to be a general polynomial of the same degree. If $G(x)$ is of the form Ce^{kx} , choose $y_p(x) = Ae^{kx}$. If $G(x)$ is $C \cos kx$ or $C \sin kx$, choose $y_p(x) = A \cos kx + B \sin kx$. If $G(x)$ is a product of functions, choose y_p to be a product of functions of the same type. Some examples are:

$G(x)$	$y_p(x)$
x^2	$Ax^2 + Bx + C$
e^{2x}	Ae^{2x}
$\sin 2x$	$A \cos 2x + B \sin 2x$
xe^{-x}	$(Ax + B)e^{-x}$

We then substitute y_p , y'_p , and y''_p into the differential equation and determine the coefficients.

If y_p happens to be a solution of the complementary equation, then multiply the initial trial solution by x (or x^2 if necessary).

If $G(x)$ is a sum of functions, we find a particular solution for each function and then y_p is the sum of these.

The general solution of the differential equation is

$$y(x) = y_p(x) + y_c(x)$$

- (d)** Explain how the method of variation of parameters works.

We write the solution of the complementary equation $ay'' + by' + cy = 0$ as $y_c(x) = c_1 y_1(x) + c_2 y_2(x)$, where y_1 and y_2 are linearly independent solutions. We then take $y_p(x) = u_1(x)y_1(x) + u_2(x)y_2(x)$ as a particular solution, where $u_1(x)$ and $u_2(x)$ are arbitrary functions. After computing y'_p , we impose the condition that

$$u'_1 y_1 + u'_2 y_2 = 0 \quad (1)$$

and then compute y''_p . Substituting y_p , y'_p , and y''_p into the original differential equation gives

$$(u'_1 y'_1 + u'_2 y'_2) = G \quad (2)$$

We then solve equations (1) and (2) for the unknown functions u'_1 and u'_2 . If we are able to integrate these functions, then a particular solution is $y_p(x) = u_1(x)y_1(x) + u_2(x)y_2(x)$ and the general solution is $y(x) = y_p(x) + y_c(x)$.

- 4. Discuss two applications of second-order linear differential equations.**

The motion of an object with mass m at the end of a spring is an example of simple harmonic motion and is described by the second-order linear differential equation

$$m \frac{d^2x}{dt^2} + kx = 0$$

(continued)

CHAPTER 17 CONCEPT CHECK ANSWERS (continued)

where k is the spring constant and x is the distance the spring is stretched (or compressed) from its natural length. If there are external forces acting on the spring, then the differential equation is modified.

Second-order linear differential equations are also used to analyze electrical circuits involving an electromotive force, a resistor, an inductor, and a capacitor in series.

See the discussion in Section 17.3 for additional details.

5. How do you use power series to solve a differential equation?

We first assume that the differential equation has a power series solution of the form

$$y = \sum_{n=0}^{\infty} c_n x^n = c_0 + c_1 x + c_2 x^2 + c_3 x^3 + \dots$$

Differentiating gives

$$y' = \sum_{n=1}^{\infty} n c_n x^{n-1} = \sum_{n=0}^{\infty} (n+1) c_{n+1} x^n$$

and

$$y'' = \sum_{n=2}^{\infty} n(n-1) c_n x^{n-2} = \sum_{n=0}^{\infty} (n+2)(n+1) c_{n+2} x^n$$

We substitute these expressions into the differential equation and equate the coefficients of x^n to find a recursion relation involving the constants c_n . Solving the recursion relation gives a formula for c_n and then

$$y = \sum_{n=0}^{\infty} c_n x^n$$

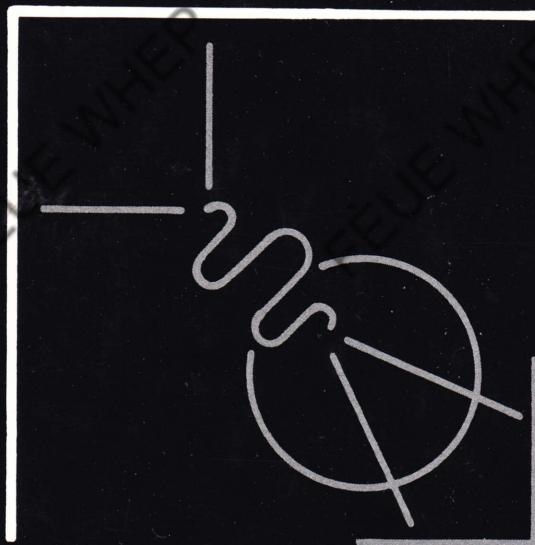
is the solution of the differential equation.

This is an electronic version of the print textbook. Due to electronic rights restrictions, some third party content may be suppressed. Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. The publisher reserves the right to remove content from this title at any time if subsequent rights restrictions require it. For valuable information on pricing, previous editions, changes to current editions, and alternate formats, please visit www.cengage.com/highered to search by ISBN, author, title, or keyword for materials in your areas of interest.

Important notice: Media content referenced within the product description or the product text may not be available in the eBook version.

QUARKS & LEPTONS:

An Introductory Course in
Modern Particle Physics

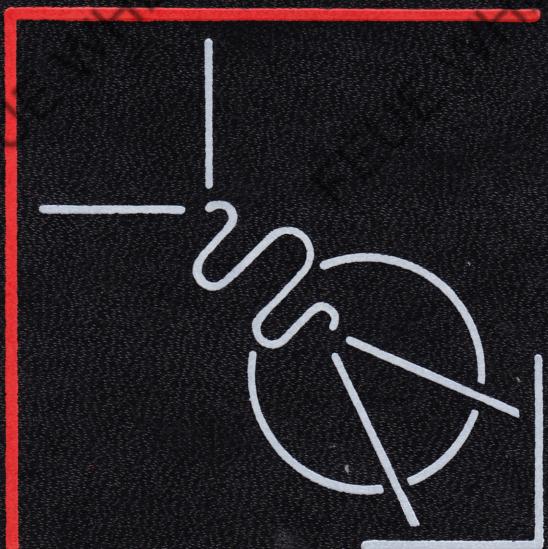


Francis Halzen
Alan D. Martin

9802

QUARKS & LEPTONS.

An Introductory Course in
Modern Particle Physics



Francis Halzen
Alan D. Martin
9803

Quarks and Leptons is a new and stimulating introductory text on the revolution in particle physics that has taken place during the past two decades. Two active researchers in the field offer a clear, logical and accessible treatment of quarks and leptons, as well as other recent developments in this vitally important area.

Featuring a careful, pedagogically sound approach to the subject, *Quarks and Leptons* thoroughly covers the fundamental concepts and theories that make quarks and leptons understandable. The text describes how these fundamental particles interact, and introduces—in steps of gradually increasing complexity—the basic techniques required to perform physical calculations.

Coverage in individual chapters includes:

- a preview of particle physics
- symmetries and quarks
- antiparticles
- electrodynamics of spinless particles
- the Dirac Equation
- quantum electrodynamics
- loops, renormalization, and running coupling constants
- the structure of hadrons
- partons
- quantum chromodynamics
- $e^+ e^-$ annihilation and QCD
- weak interactions
- electroweak interactions
- gauge symmetries
- the Weinberg-Salam model
- plus references, supplementary readings, answers and comments on the exercises, and a detailed index.

In providing students with a balanced and informed overview of the entire subject, *Quarks and Leptons* includes such recent advances as quark model and hadron spectroscopy...quantum electrodynamics and quantum chromodynamics...electroweak theory...gauge theories...and much more.

(continued on back flap)

QUARKS AND LEPTONS:

An Introductory Course

in Modern Particle Physics

9805

9806

QUARKS AND LEPTONS:

An Introductory Course

in Modern Particle Physics

Francis Halzen

*University of Wisconsin
Madison, Wisconsin*

Alan D. Martin

*University of Durham
Durham, England*

JOHN WILEY & SONS,

New York • Chichester • Brisbane • Toronto • Singapore

Copyright © 1984, by John Wiley & Sons, Inc.

All rights reserved. Published simultaneously in Canada.

Reproduction or translation of any part of
this work beyond that permitted by Sections
107 and 108 of the 1976 United States Copyright
Act without the permission of the copyright
owner is unlawful. Requests for permission
or further information should be addressed to
the Permissions Department, John Wiley & Sons.

Library of Congress Cataloging in Publication Data:

Halzen, Francis.

Quarks and leptons.

Includes index.

1. Quarks. 2. Leptons (Nuclear physics)
I. Martin, Alan D. (Alan Douglas) II. Title.
QC793.5.Q2522H34 1984 539.7'21 83-14649
ISBN 0-471-88741-2

Printed in the United States of America

10 9 8 7 6 5 4 3 2

9808

To
Nelly and Penny
Rebecca, Robert, Rachel, and David

9810

Preface

Dramatic progress has been made in particle physics during the past two decades. A series of important experimental discoveries has firmly established the existence of a subnuclear world of quarks and leptons. The protons and neutrons (“nucleons”), which form nuclei, are no longer regarded as elementary particles but are found to be made of quarks. That is, in the sequence molecules → atoms → nuclei → nucleons, there is now known to be another “layer in the structure of matter.” However, the present euphoria in particle physics transcends this remarkable discovery. The excitement is due to the realization that the dynamics of quarks and leptons can be described by an extension of the sort of quantum field theory that proved successful in describing the electromagnetic interactions of charged particles. To be more precise, the fundamental interactions are widely believed to be described by quantum field theories possessing local gauge symmetry. One of the aims of this book is to transmit a glimpse of the amazing beauty and power of these gauge theories. We discuss quarks and leptons, and explain how they interact through the exchange of gauge field quanta (photons, gluons, and weak bosons).

We are very conscious that this book has been written at a crucial time when pertinent questions regarding the existence of the weak bosons and the stability of the proton may soon be decided experimentally. Some sections of the book should therefore be approached with a degree of caution, bearing in mind that the promising theory of today may only be the effective phenomenology of the theory of tomorrow. But no further apology will be made for our enthusiasm for gauge theories.

We have endeavored to provide the reader with sufficient background to understand the relevance of the present experimental assault upon the nature of matter and to appreciate contemporary theoretical speculations. The required core of knowledge is the standard *electroweak model*, which describes the electromagnetic and weak interactions of leptons and quarks; and *quantum chromodynamics* (QCD), which describes the strong interactions of quarks and gluons. The primary purpose of this book is to introduce these ideas in the simplest possible way. We assume only a basic knowledge of nonrelativistic quantum mechanics and the theory of special relativity. We spend considerable time introducing quantum electrodynamics (QED) and try to establish a working

familiarity with the Feynman rules. These techniques are subsequently generalized and applied to quantum chromodynamics and to the theory of weak interactions.

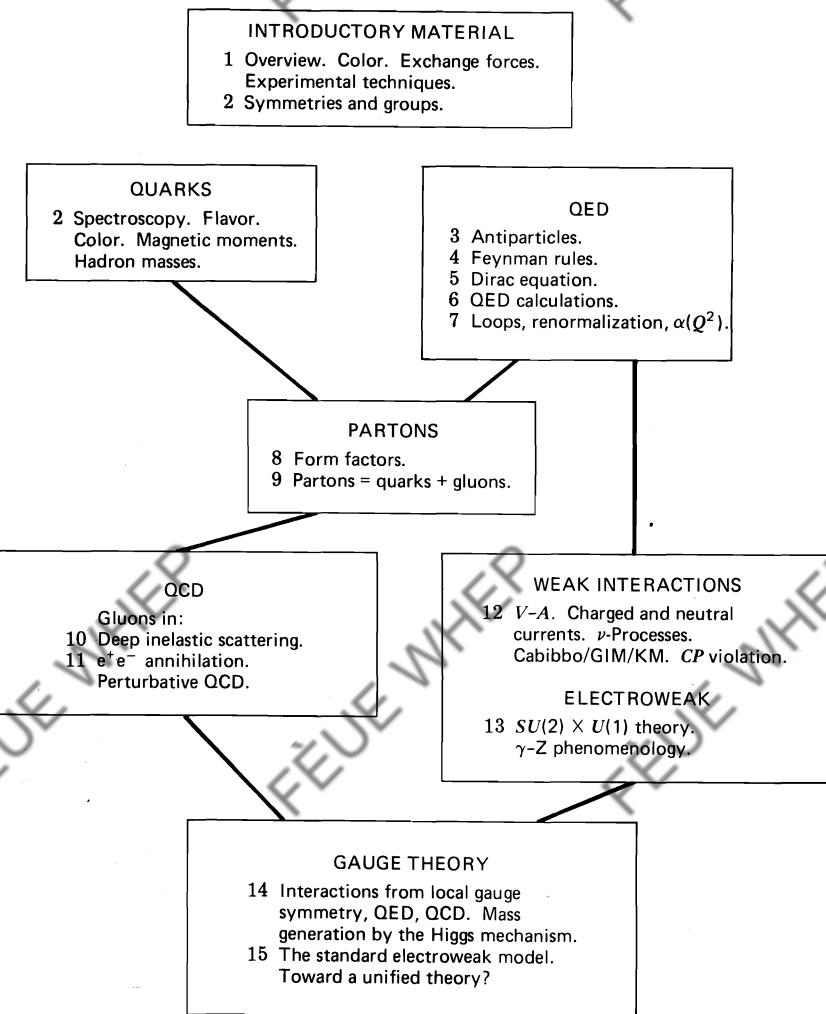
The emphasis of the book is pedagogical. This has several implications. No attempt is made to cover each subject completely. Examples are chosen solely on pedagogical merit and not because of their historic importance. The book does not contain the references to the original scientific papers. However, we do refer to books and appropriate review articles whenever possible, and of course no credit for original discovery is implied by our choice. A supplementary reading list can be found at the end of the book, and we also encourage the students to read the original papers mentioned in these articles. A deliberate effort is made to present material which will be of immediate interest to the student, irrespective of his experimental or theoretical bias. It is possible that aspiring theorists may feel that an injustice has been done to the subtle beauty of the formalism, while experimentalists may justifiably argue that the role of experimental discoveries is insufficiently emphasized. Fortunately, the field is rich in excellent books and review articles covering such material, and we hope that our guidance toward alternative presentations will remedy these defects.

Although the book is primarily written as an introductory course in particle physics, we list several other *teaching options*. The accompanying flow diagram gives some idea of the material covered in the various chapters.

- A One teaching option is based on the belief that because of its repeated phenomenological successes, modern particle physics, or at least some aspects of it, is suitable material for an advanced quantum mechanics course alongside the more traditional subjects such as atomic physics. For this purpose, we suggest Chapters 3 through 6, with further examples from Chapter 12, together perhaps with parts of Chapter 14.
- B An undergraduate course on the introduction of the Feynman rules for QED could be based on Chapters 3 through 6.
- C The sequence of Chapters 3 through 11 could serve as an introduction to quantum chromodynamics.
- D A course on weak and electromagnetic interactions could cover Chapters 3 through 6 and 12 and 13, perhaps supplemented with parts of Chapters 14 and 15.
- E For a standard introductory particle physics course, it may not be possible to cover the full text in depth, and Chapters 7, 10, 11, 14, and 15 can be partially or completely omitted.

Exercises are provided throughout the text, and several of the problems are an integral part of the discussion. Outline solutions to selected problems are given at the end of the book, particularly when the exercise provides a crucial link in the text.

This book was developed and written with the encouragement of students and friends at the Universities of Durham and Wisconsin. Many colleagues have given



us valuable assistance. In particular, we acknowledge our special debt to Peter Collins and Paul Stevenson. They read through the entire manuscript and suggested countless improvements. We also thank our other colleagues for their valuable comments on parts of the manuscript, especially D. Bailin, V. Barger, U. Camerini, C. Goebel, K. Hagiwara, G. Karl, R. March, C. Michael, M. Pennington, D. Reeder, G. Ross, D. Scott, T. Shimada, T. D. Spearman, and B. Webber.

We thank Vicky Kerr, and also Linda Dolan, for excellent typing of a difficult manuscript. By so doing, they made our task that much easier.

Francis Halzen and Alan D. Martin

Durham, England
January 12th, 1983

9814

Contents

1 A Preview of Particle Physics	1
1.1 What is the World Made of?, 1	
1.2 Quarks and Color, 4	
1.3 Color: The Charge of Nuclear Interactions, 7	
1.4 Natural Units, 12	
1.5 Alpha (α) is not the Only Charge Associated with Particle Interactions, 14	
1.6 There Are Weak Interactions, Too, 21	
1.7 Down Mendeleev's Path: More Quarks and Leptons, 26	
1.8 Gravity, 27	
1.9 Particles: The Experimentalist's Point of View, 28	
1.10 Particle Detectors, 30	
2 Symmetries and Quarks	33
<i>Symmetries and Groups</i>	
2.1 Symmetries in Physics: An Example, 33	
2.2 Symmetries and Groups: A Brief Introduction, 35	
2.3 The Group $SU(2)$, 39	
2.4 Combining Representations, 40	
2.5 Finite Symmetry Groups: P and C , 41	
2.6 $SU(2)$ of Isospin, 41	
2.7 Isospin for Antiparticles, 42	
2.8 The Group $SU(3)$, 43	
2.9 Another Example of an $SU(3)$ Group: Isospin and Strangeness, 44	

Quark “Atoms”

- 2.10 Quark–Antiquark States: Mesons, 46
- 2.11 Three-Quark States: Baryons, 50
- 2.12 Magnetic Moments, 55
- 2.13 Heavy Quarks: Charm and Beyond, 57
- 2.14 Hadron Masses, 63
- 2.15 Color Factors, 67

3 Antiparticles 70

- 3.1 Nonrelativistic Quantum Mechanics, 71
- 3.2 Lorentz Covariance and Four-Vector Notation, 72
- 3.3 The Klein–Gordon Equation, 74
- 3.4 Historical Interlude, 75
- 3.5 The Feynman–Stückelberg Interpretation of $E < 0$ Solutions, 77
- 3.6 Nonrelativistic Perturbation Theory, 79
- 3.7 Rules for Scattering Amplitudes in the Feynman–Stückelberg Approach, 82

4 Electrodynamics of Spinless Particles 84

- 4.1 An “Electron” in an Electromagnetic Field A^μ , 84
- 4.2 “Spinless” Electron–Muon Scattering, 86
- 4.3 The Cross Section in Terms of the Invariant Amplitude \mathcal{M} , 88
- 4.4 The Decay Rate in Terms of \mathcal{M} , 92
- 4.5 “Spinless” Electron–Electron Scattering, 92
- 4.6 Electron–Positron Scattering: An Application of Crossing, 93
- 4.7 Invariant Variables, 94
- 4.8 The Origin of the Propagator, 97
- 4.9 Summary, 99

5 The Dirac Equation 100

- 5.1 Covariant Form of the Dirac Equation. Dirac γ -Matrices, 102
- 5.2 Conserved Current and the Adjoint Equation, 103

5.3	Free-Particle Spinors,	104
5.4	Antiparticles,	107
5.5	Normalization of Spinors and the Completeness Relations,	110
5.6	Bilinear Covariants,	111
5.7	Zero-Mass Fermions. The Two-Component Neutrino,	114

6 Electrodynamics of Spin- $\frac{1}{2}$ Particles 117

6.1	An Electron Interacting with an Electromagnetic Field A^μ ,	117
6.2	Møller Scattering $e^-e^- \rightarrow e^-e^-$,	119
6.3	The Process $e^-\mu^- \rightarrow e^-\mu^-$,	121
6.4	Trace Theorems and Properties of γ -Matrices,	123
6.5	$e^-\mu^-$ Scattering and the Process $e^+e^- \rightarrow \mu^+\mu^-$,	123
6.6	Helicity Conservation at High Energies,	126
6.7	Survey of $e^-e^+ \rightarrow e^-e^+, \mu^-\mu^+$,	129
6.8	$e^-\mu^- \rightarrow e^-\mu^-$ in the Laboratory Frame. Kinematics Relevant to the Parton Model,	130
6.9	Photons. Polarization Vectors,	132
6.10	More on Propagators. The Electron Propagator,	135
6.11	The Photon Propagator,	137
6.12	Massive Vector Particles,	138
6.13	Real and Virtual Photons,	139
6.14	Compton Scattering $\gamma e^- \rightarrow \gamma e^-$,	141
6.15	Pair Annihilation $e^+e^- \rightarrow \gamma\gamma$,	144
6.16	The $+ie$ Prescription for Propagators,	145
6.17	Summary of the Feynman Rules for QED,	150

7 Loops, Renormalization, Running Coupling Constants, and All That 152

7.1	Scattering Electrons Off a Static Charge,	152
7.2	Higher-Order Corrections,	154
7.3	The Lamb Shift,	158
7.4	More Loops: The Anomalous Magnetic Moment,	159
7.5	Putting the Loops Together: Ward Identities,	162

7.6 Charge Screening and $e^- \mu^-$ Scattering,	163
7.7 Renormalization,	163
7.8 Charge Screening in QED: The Running Coupling Constant,	167
7.9 Running Coupling Constant for QCD,	169
7.10 Summary and Comments,	171
8 The Structure of Hadrons	172
8.1 Probing a Charge Distribution with Electrons. Form Factors,	172
8.2 Electron–Proton Scattering. Proton Form Factors,	175
8.3 Inelastic Electron–Proton Scattering $ep \rightarrow eX$,	179
8.4 Summary of the Formalism for Analyzing ep Scattering,	183
8.5 Inelastic Electron Scattering as a (Virtual) Photon–Proton Total Cross Section,	184
9 Partons	188
9.1 Bjorken Scaling,	188
9.2 Partons and Bjorken Scaling,	191
9.3 The Quarks Within the Proton,	196
9.4 Where Are the Gluons?,	202
10 Quantum Chromodynamics	205
10.1 The Dual Role of Gluons,	205
10.2 Embedding γ^* -Parton Processes in Deep Inelastic Scattering,	208
10.3 The Parton Model Revisited,	209
10.4 The Gluon Emission Cross Section,	210
10.5 Scaling Violations. The Altarelli–Parisi Equation,	215
10.6 Including Gluon Pair Production,	219
10.7 Complete Evolution Equations for the Parton Densities,	220
10.8 Physical Interpretation of the P Functions,	221
10.9 The Altarelli–Parisi Techniques Also Apply to Leptons and Photons: The Weizsäcker–Williams Formula,	224

11 e⁺e⁻ Annihilation and QCD	226
11.1 e ⁻ e ⁺ Annihilation Into Hadrons: $e^-e^+ \rightarrow q\bar{q}$, 226	
11.2 Fragmentation Functions and Their Scaling Properties, 230	
11.3 A Comment on Heavy Quark Production, 234	
11.4 Three-Jet Events: $e^-e^+ \rightarrow q\bar{q}g$, 234	
11.5 An Alternative Derivation of the $e^-e^+ \rightarrow q\bar{q}g$ Cross Section, 238	
11.6 A Discussion of Three-Jet Events, 240	
11.7 QCD Corrections to $e^-e^+ \rightarrow$ Hadrons, 243	
11.8 Perturbative QCD, 245	
11.9 A Final Example: The Drell–Yan Process, 247	
12 Weak Interactions	251
12.1 Parity Violation and the $V-A$ Form of the Weak Current, 252	
12.2 Interpretation of the Coupling G , 256	
12.3 Nuclear β -Decay, 258	
12.4 Further Trace Theorems, 261	
12.5 Muon Decay, 261	
12.6 Pion Decay, 264	
12.7 Charged Current Neutrino–Electron Scattering, 267	
12.8 Neutrino–Quark Scattering, 271	
12.9 First Observation of Weak Neutral Currents, 276	
12.10 Neutral Current Neutrino–Quark Scattering, 277	
12.11 The Cabibbo Angle, 279	
12.12 Weak Mixing Angles, 283	
12.13 CP Invariance?, 287	
12.14 CP Violation: The Neutral Kaon System, 289	
13 Electroweak Interactions	292
13.1 Weak Isospin and Hypercharge, 292	
13.2 The Basic Electroweak Interaction, 295	
13.3 The Effective Current–Current Interaction, 298	
13.4 Feynman Rules for Electroweak Interactions, 299	
13.5 Neutrino–Electron Scattering, 302	

13.6 Electroweak Interference in $e^+ e^-$ Annihilation, 305
13.7 Other Observable Electroweak Interference Effects, 308

14 Gauge Symmetries	311
----------------------------	------------

14.1 The Lagrangian and Single-Particle Wave Equations, 311
14.2 Noether's Theorem: Symmetries and Conservation Laws, 314
14.3 $U(1)$ Local Gauge Invariance and QED, 316
14.4 Non-Abelian Gauge Invariance and QCD, 317
14.5 Massive Gauge Bosons?, 320
14.6 Spontaneous Symmetry Breaking: "Hidden" Symmetry, 321
14.7 Spontaneous Breaking of a Global Gauge Symmetry, 324
14.8 The Higgs Mechanism, 326
14.9 Spontaneous Breaking of a Local $SU(2)$ Gauge Symmetry, 327

15 The Weinberg–Salam Model and Beyond	331
---	------------

15.1 Electroweak Interactions Revisited, 331
15.2 Choice of the Higgs Field, 334
15.3 Masses of the Gauge Bosons, 335
15.4 Masses of the Fermions, 338
15.5 The Standard Model: The Final Lagrangian, 341
15.6 Electroweak Theory is Renormalizable, 342
15.7 Grand Unification, 344
15.8 Can the Proton Decay?, 349
15.9 The Early Universe as a High-Energy Physics Experiment, 352
15.10 "Grander" Unification?, 354

Answers and Comments on the Exercises	355
--	------------

Supplementary Reading	377
------------------------------	------------

References	381
-------------------	------------

Index	385
--------------	------------

QUARKS AND LEPTONS:

An Introductory Course

in Modern Particle Physics

FEUE WHEP

FEUE WHEP

FEUE WHEP

9822

1

A Preview of Particle Physics

1.1 What is the World Made of?

Present-day particle physics research represents man's most ambitious and most organized effort to answer this question. Earlier answers to this riddle included the solution proposed by Anaximenes of Miletus, shown in Fig. 1.1. Everyone is familiar with the answer Mendeleev came up with 25 centuries later: the periodic table, a sort of extended version of Fig. 1.1, which now contains well over 100 chemical elements. Anaximenes's model of the fundamental structure of matter is clearly conceptually superior because of its simplicity and economy in number of building blocks. It has one fatal problem: it is wrong! Mendeleev's answer is right, but it is too complicated to represent the "ultimate" or fundamental solution. The proliferation of elements and the apparent systematics in the organization of the table strongly suggests a substructure. We know now that the elements in Mendeleev's table are indeed built up of the more fundamental electrons and nuclei.

Our current answer to the question what the world is made of is displayed in Table 1.1. It shares the conceptual simplicity of Anaximenes's solution; it is, however, just like Mendeleev's proposal, truly quantitative and in agreement with experimental facts. The answer of Table 1.1 was actually extracted step by step from a series of experiments embracing the fields of atomic, nuclear, cosmic-ray, and high-energy physics. This experimental effort originated around the turn of this century, but it was a sequence of very important discoveries in the last decade that directly guided us to a world of quarks, leptons, and gauge bosons. Previously, this picture had just been one of many competing suggestions for solving the puzzle of the basic structure of matter.

The regularities in Mendeleev's table were a stepping-stone to nuclei and to particles called protons and neutrons (collectively labeled nucleons), which are "glued" together with a strong or nuclear force to form the nuclei. These subsequently bind with electrons through the electromagnetic force to produce the atoms of the chemical elements. Conversion of neutrons into protons by so-called weak interactions is responsible for the radioactive β -decay of nuclei, including the slow decay of the neutron into a proton accompanied by an electron

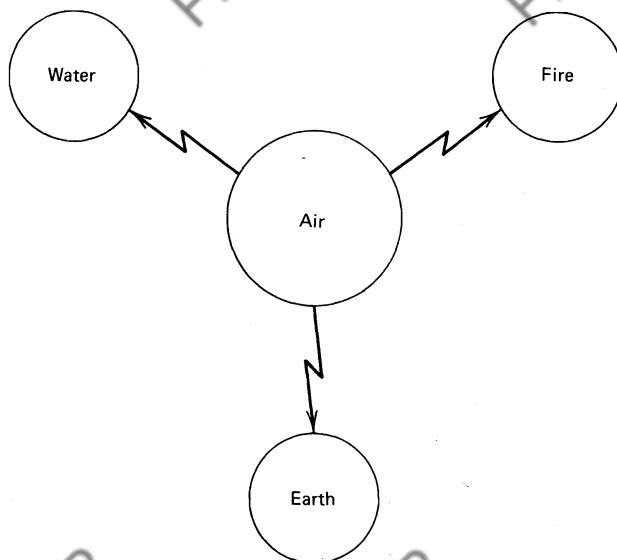


Fig. 1.1 In the original version of the theory, all forms of matter are obtained by condensing or rarefying air. Later, a “chemistry” was constructed using the four elements shown in the figure.

and an antineutrino. At this point, the world looked very much like Table 1.1 but with the nucleons p and n playing the role of the quarks u and d.

But the neutron and proton were not alone. They turned out to be just the lightest particles in a spectrum of strongly interacting fermion states, called *baryons*, numbering near 100 at the latest count. An equally numerous sequence of strongly interacting bosons, called *mesons*, has also been discovered, the pion being the lightest. Fermions (bosons) refer to particle states with spin $J = n(\hbar/2)$, where n is an odd (even) integer. All the particles which undergo strong interactions, baryons and mesons, are collectively called “*hadrons*.”

This proliferation of so-called “elementary” particles pointed the way to the substructure of the nucleons (the quarks) in a rather straightforward replay of the arguments for composite atoms based on Mendeleev’s table. Also, the π -meson and all other hadrons are made of quarks. The electron and neutrino do not experience strong interactions and so are not hadrons. They form a separate group of particles known as *leptons*. The neutrino participates exclusively in the weak interactions, but the charged electron can of course also experience electromagnetic interactions. Leptons have not proliferated like hadrons and so are entered directly into Table 1.1 as elementary point-like particles along with the quarks. The pion, neutron, proton... are not part of the ultimate pieces of the puzzle; they join nuclei and atoms as one more manifestation of bound-state structures that exist in a world made of quarks and leptons.

TABLE 1.1Building Blocks of Elementary Particles and Some of Their Quantum Numbers^a

Name	Spin	Baryon Number <i>B</i>	Lepton Number <i>L</i>	Charge <i>Q</i>
Quarks				
u (up)	$\frac{1}{2}$	$\frac{1}{3}$	0	$+\frac{2}{3}$
d (down)	$\frac{1}{2}$	$-\frac{1}{3}$	0	$-\frac{1}{3}$
Leptons				
e (electron)	$\frac{1}{2}$	0	1	-1
ν (neutrino)	$\frac{1}{2}$	0	1	0
Gauge bosons				
γ (photon)	1	0	0	0
W^\pm, Z (weak bosons)	1	0	0	$\pm 1, 0$
g_i ($i = 1, \dots, 8$ gluons)	1	0	0	0

^aThe spin is given in units of \hbar . The charge units are defined in such a way that the electron charge is -1. Not listed in the table are the antiparticles of the quarks and leptons [\bar{u} , \bar{d} (antiquarks), e^+ (positron), $\bar{\nu}$ (antineutrino)]. Although they will not be carefully defined until Chapter 3, they are for present purposes identical to the corresponding particles except for the reversal of the sign of their *B*, *L*, and *Q* quantum numbers. For example, the \bar{u} has $B = -\frac{1}{3}$, $L = 0$, $Q = -\frac{2}{3}$, and the e^+ has $B = 0$, $L = -1$, and $Q = +1$.

A theoretical framework was needed that could translate these conceptual developments into a quantitative calculational scheme. Clearly, Schrödinger's equation could not handle the creation and annihilation of particles as observed in neutron decay and was furthermore unable to describe highly relativistic particles as encountered in routine cosmic ray experiments. In the early 1930s, a theory emerged describing the electromagnetic interactions of electrons and photons (quantum electrodynamics) that encompassed these desired features: it was quantized and relativistically invariant. Even though it has become essential to include quarks, as well as leptons, and to consider other interactions besides electromagnetism, relativistic quantum field theory, of which quantum electrodynamics is the prototype, stands unchanged as the calculational framework of particle physics. The most recent developments in particle physics, however, have revealed the relevance of a special class of such theories, called "gauge" theories; quantum electrodynamics itself is the simplest example of such a theory. The weak and strong interactions of quarks and leptons are both believed to be described by gauge theories: the unified electroweak model and quantum chromodynamics.

The interplay of models and ideas, formulated in the general framework of gauge theories, with new experimental information has been the breeding ground

for repeated success. The purpose of this book is to introduce the reader to this form of research which is entering new energy domains with the completion and imminent construction of a new generation of particle accelerators. These accelerators, together with very sophisticated particle detectors, will probe matter to previously unexplored submicroscopic distances.

1.2 Quarks and Color

The overwhelming experimental evidence that the nucleons of nuclear physics are made of particles called *quarks* is reviewed in Chapter 2. The *baryons* are bound states of three quarks; the *mesons* are composed of a quark and an antiquark. The proton is a uud bound state; the additive quantum numbers of the two kinds of quarks, u and d, listed in Table 1.1, correctly match the fact that the proton is a baryon ($B = 1 = \frac{1}{3} + \frac{1}{3} + \frac{1}{3}$) and not a lepton ($L = 0$) and that it has total charge 1 ($Q = \frac{2}{3} + \frac{1}{3} - \frac{1}{3}$). By analogous arguments, the neutron is obtained as a udd bound state. The π^+ -meson is a ud state; it is a meson, in the sense that $B = L = 0$, and its charge is indeed 1 [$Q = \frac{2}{3} - (-\frac{1}{3})$]. In passing, we should note that the charge Q , the baryon number B , and the lepton number L are more than simply labels; they are *conserved* additive quantum numbers. That is, a particle reaction (e.g., $\pi^- p \rightarrow \pi^0 n$ or $n \rightarrow p e^- \bar{\nu}$) can only occur if the sum of the B values in the initial state equals that in the final state; similarly for L and Q .

By the conventional rules of addition of angular momenta, the total spin $J = \frac{1}{2}$ for the nucleon and $J = 0$ for the π -meson can be constructed from their $J = \frac{1}{2}$ constituents. The quark scheme naturally accommodates the observed separation of *hadrons* into baryons (three quark fermion states) and mesons (quark-antiquark boson states).

An immediate success of the quark model is theoretical in nature. Protons and neutrons are relatively complicated objects, with a size and a rich internal quark structure. Quantum field theory, on the other hand, deals with point-like elementary particles, that is, with structureless objects like the electron, for example. The structureless quarks, rather than the nucleons, are the fundamental entities described by quantum field theory. Their introduction enables us to explore the other interactions with the same powerful theoretical techniques that were so successful in describing the properties and electromagnetic interactions of electrons (*quantum electrodynamics*).

When implementing the quark scheme, however, one runs into trouble at the next logical step:

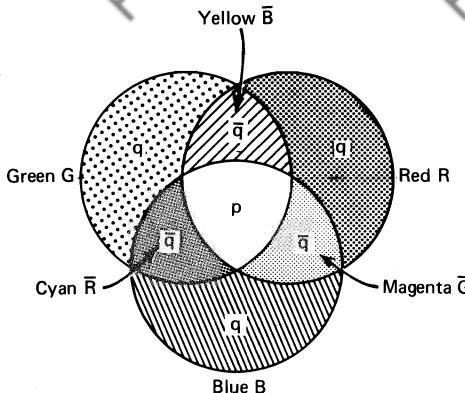
$$\begin{aligned} p &= uud \\ n &= udd \\ \Delta^{++} &= uuu. \end{aligned} \tag{1.1}$$

The uuu configuration correctly matches the properties of the doubly charged

Δ^{++} -baryon (the $\pi^+ p$ resonance originally discovered by Fermi and collaborators in 1951). Its spin, $J = \frac{3}{2}$, is obtained by combining three identical $J = \frac{1}{2}$ u quarks in their ground state. That is, the quark scheme forces us to combine three identical fermions u in a completely symmetric ground state uuu in order to accommodate the known properties of the Δ -particle. Such a state is of course forbidden by Fermi statistics.

Even ignoring the statistics fiasco, this naive quark model is clearly unsatisfactory: it is true that the qqq , $\bar{q}\bar{q}\bar{q}$, and $q\bar{q}$ states reproduce the observed sequence of baryon, antibaryon, and meson states, as will be demonstrated in the next chapter, but what about all the other possibilities such as qq , $\bar{q}\bar{q}, \dots$, or single quarks themselves? No uu particles with charge $\frac{4}{3}$ have ever been observed. Both problems can be resolved by introducing a new property or quantum number for quarks (not for leptons!): “color.” We suppose that quarks come in three primary colors: red, green, and blue, denoted symbolically by R, G, and B, respectively. “Color” has of course no relation to the real colors of everyday life; the terminology is just based on the analogy with the way all real colors are made up of three primary colors. If we then rewrite the quark wavefunction for the Δ -state in (1.1) as $u_R u_G u_B$, we have clearly overcome the statistics problem by disposing of the identical quarks. The three quarks that make up the Δ -state are now distinguishable by their color quantum number. This solution might appear very contrived; we request, however, that you reserve judgement until, well, maybe Chapters 10 and 14, where the color degree of freedom will acquire a “physical” meaning in the context of a gauge theory. There remains a more immediate problem: if $u_R u_G u_B$ is Fermi’s Δ^{++} , then we appear to have many candidate states for the proton: $u_R u_G d_B$, $u_R u_G d_G$, $u_B u_R d_R$, and so on. Yet only one proton state exists; we have to introduce our color quantum number without proliferating the number of states, since this would lead us to a direct conflict with observation. The way this is done is to assert that all particle states observed in nature are “colorless” or “white” (or, to be more precise, unchanged by rotations in R, G, B space). It is easy to visualize the color quantum number by associating the three possible colors of a quark with the three spots of primary red, green, and blue light focussed on a screen, as shown in Fig. 1.2. The antiquarks are assigned the complementary colors: cyan (\bar{R}), magenta (\bar{G}), and yellow (\bar{B}). If you do not know color theory, it may be more helpful to just think of the complementary colors as antired, antigreen, and antiblue. The colors assigned to the antiquarks appear in Fig. 1.2 in those parts of the screen where two and only two primary beams overlap. There is now a unique set of ways to obtain colorless (white) combinations by mixing colors (quarks) and complementary colors (antiquarks):

- Equal mixture of red, green, and blue (RGB).
- Equal mixture of cyan, magenta, and yellow ($\bar{R}\bar{G}\bar{B}$).
- Equal mixtures of color and complementary color ($R\bar{R}$, $G\bar{G}$, $B\bar{B}$).



$$p = "RGB" \quad \bar{p} = "\bar{R}\bar{G}\bar{B}"$$

$$\pi = "R\bar{R} + B\bar{B} + G\bar{G}"$$

Fig. 1.2 Color composition of hadrons.

These possibilities correspond respectively to the particle states observed in nature: baryons, antibaryons, and mesons. For example,

$$\begin{aligned} p &= "RGB" \\ \bar{p} &= "\bar{R}\bar{G}\bar{B}" \\ \pi &= "R\bar{R} + B\bar{B} + G\bar{G}". \end{aligned} \tag{1.2}$$

In other words, the proton is still a uud state as in (1.1), but with the specific color assignments to the quarks exhibited in (1.2). The quotation marks remind us that these wavefunctions eventually have to be properly symmetrized and normalized. We will do so in Chapter 2, and indeed show that the combination “RGB” is antisymmetric under interchange of a pair of color labels as required by the Fermi statistics of the quarks.

The analogy we have developed between the color quantum number and color is not perfect. The three $q\bar{q}$ states $\bar{R}R$, $\bar{G}G$, and $\bar{B}B$ are colorless, but it is only the combination $\bar{R}R + \bar{G}G + \bar{B}B$, unchanged by rotations in R, G, B color space, which can form an observed meson. In other words, we use “colorless” to mean a singlet representation of the color group.

Let us briefly recapitulate. We have assigned a “hidden” color quantum number (no connection with hidden variables implied!) to quarks; it is hidden from the world in the sense that all the particles or quark bound states that hit the experimentalists’ detectors are colorless (color singlets). It solves the embarrassment that our successful (see Chapter 2) quark model appears to violate Fermi statistics, but it does much more. Notice indeed that, for example, qq states of the

type RG, GB... are necessarily colored and therefore cannot occur according to our dogma that only colorless bound states exist. The color scheme explains the exclusive role played in nature by qqq , $\bar{q}\bar{q}\bar{q}$, and $q\bar{q}$ quark combinations. The quarks themselves are colored and therefore hidden from our sight. But, as we discuss later on, there are nevertheless a multitude of ways to infer experimentally their existence inside hadrons. We are now ready to introduce, however, the most profound implication of the color concept.

1.3 Color: The Charge of Nuclear Interactions

In Maxwell's theory of electromagnetism, charged particles such as electrons interact through their electromagnetic fields. However, for many years it was difficult to conceive how such action-at-a-distance between charges came about. That is, how can charged particles interact without some tangible connection? In quantum field theory, we have such a tangible connection: all the forces of nature are a result of particle exchange. Consider first the event taking place at point A in Fig. 1.3. An electron emits a photon (the quantum of the electromagnetic field) and as a result recoils in order to conserve momentum. It is clearly impossible to conserve energy as well, so the emitted photon is definitely not a real photon. It is a photon with "not quite the right energy"; we call it a "*virtual*" photon (Chapters 4 and 6). An electron can nevertheless emit such a photon as long as it is sufficiently quickly reabsorbed. Because of the uncertainty inherent in quantum mechanics, the photon can in fact live for a time $\Delta t \lesssim (\hbar/\Delta E)$, where ΔE is the "borrowed" or missing energy. However, suppose that instead of being reabsorbed by the same electron, the photon is absorbed by another electron, as in Fig. 1.3. The latter electron will recoil in the act of absorbing the virtual photon at point B. The net result is a repulsive force between the two electrons. In quantum field theory, such exchanges are responsible for the Coulomb repulsion of two like charges!

The dramatized picture of the force sketched in Fig. 1.3 might lead one to believe that only repulsive forces can be described by the exchange of particles. This is not so; the impulse of a virtual particle can have either sign in quantum field theory because its momentum vector does not necessarily have the orienta-

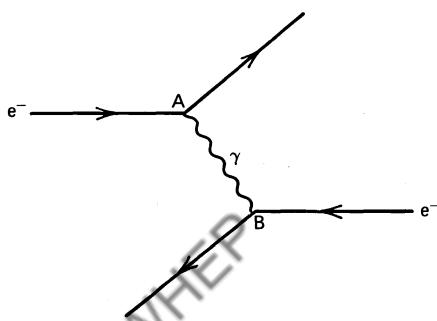


Fig. 1.3 Electrons repel by exchanging a photon.

tion prescribed by classical physics. This should not be too surprising, as clearly the whole discussion of the exchange force of Fig. 1.3 is classically impossible. The exchanged “virtual” photons are in other respects different from freely propagating real photons encountered in, for example, radio transmission, where the energy of the oscillating current in the transmitter is carried away by the radio waves. Virtual photons cannot live an existence independent of the charges that emit or absorb them. They can only travel a distance allowed by the uncertainty principle, $c \Delta t$, where c is the velocity of light.

From now on, electromagnetic interactions of charged particles are represented by pictures (*Feynman diagrams*, see Chapter 3 onward) like the one shown in Fig. 1.4a. The charged quarks will also interact by photon exchange. Although we are familiar with the fact that electromagnetic interactions bind positronium ($e^- e^+$), it is clear that the electromagnetic interaction cannot bind quarks into hadrons. A “*strong*” force, overruling the electromagnetic repulsion of the three (same-charge) u quarks in the Δ^{++} -particle, must be invoked to bind quarks into hadrons. In fact, the color “charge” endows quarks with a new color field making this strong binding possible. The interaction of two quarks by the exchange of a virtual “gluon” is shown in Fig. 1.4b. Gluons are the quanta of the color field that bind quarks in nucleons and also nucleons into nuclei. It is useful to follow the flow of color in the diagram of Fig. 1.4b. This is shown in Fig. 1.4c. The red quark moving in from the left of the page switches color with the blue quark coming from the right. Quarks interact strongly by exchanging color. Now match Figs. 1.4b and 1.4c. The gluon, shown as a curly line in Fig. 1.4b, must itself be colored: it is in fact a bicolored object, labeled $(\bar{B}\bar{R})$ in Fig. 1.4c. Based on the analogy pictured in Fig. 1.4, one can construct a theory of the color, strong, or nuclear force, whatever you prefer to call it, by copying the quantized version of Maxwell’s theory (quantum electrodynamics, or QED). The theory thus obtained is called *quantum chromodynamics*, or QCD (Chapters 10 and 11).

This theory is fortunately sufficiently like QED to share its special property of being a *renormalizable* (i.e., calculable) gauge theory. Although we postpone explanation of this statement to much later (Chapter 14), it should preempt any other statement in a presentation of QCD. We are equally fortunate that it is sufficiently different from QED to play the role of the strong force. First, nine bicolored states of the type depicted in Fig. 1.4c exist: $R\bar{R}$, $R\bar{G}$, $R\bar{B}$, $G\bar{R}$, $G\bar{G}$, $G\bar{B}$, $B\bar{R}$, $B\bar{G}$, and $B\bar{B}$. Notice that the gluon depicted in Fig. 1.4c is labeled $\bar{B}\bar{R}$, not $B\bar{R}$. Indeed, when in Chapter 3 an operational definition is given to these diagrams, it will be clear that the different directions of the arrows on the color lines of the exchanged quantum in Fig. 1.4c imply that they represent a color-anticolor combination. One of the nine combinations, $R\bar{R} + G\bar{G} + B\bar{B}$, is a color singlet [see (1.2)] which lacks any net color charge and therefore cannot play the role of a gluon carrying color from one quark to another. Chromodynamics is therefore a theory like electromagnetism, but with eight gluons instead of a single photon. Since the gluons themselves carry a color charge, they can directly interact with other gluons, as depicted in Fig. 1.4d. This possibility is not

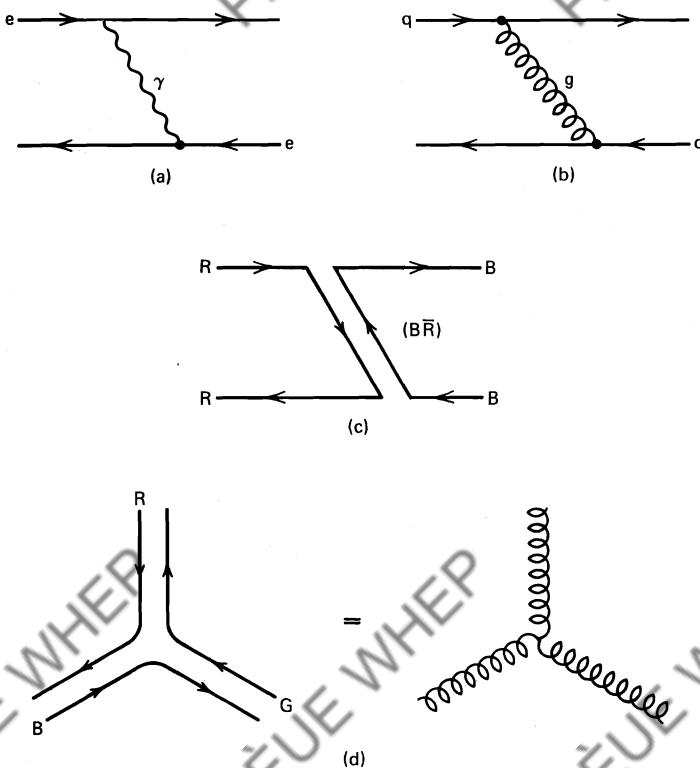
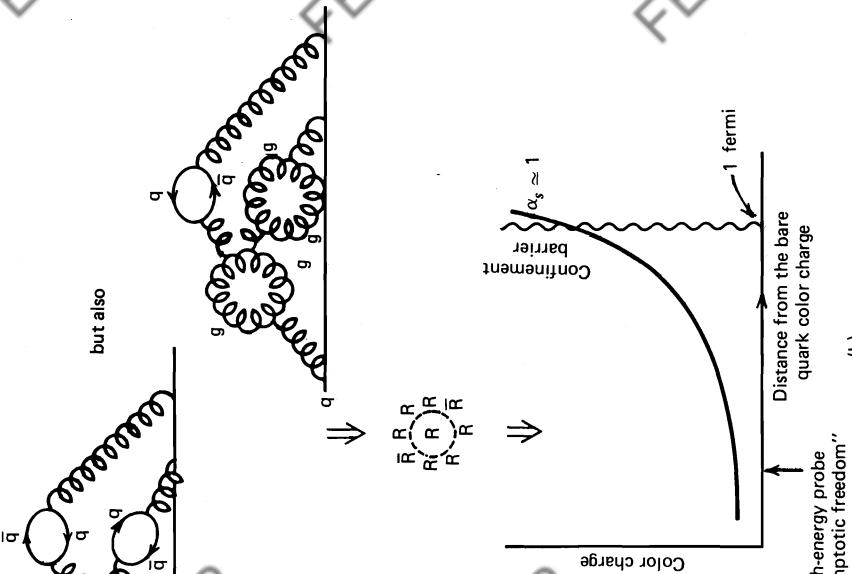


Fig. 1.4 (a) Electromagnetic interaction by photon exchange. (b) Strong interaction by gluon exchange. (c) Flow of color in (b). (d) Self-coupling of gluons.

available in electrodynamics, as photons do not have electric charge. Theories in which field quanta may interact directly are called “*non-Abelian*” (Chapter 14).

The existence of this direct coupling of gluons has dramatic implications that become evident if one contrasts the effects of *charge screening* in both QED and QCD. Screening of the electric charge in electrodynamics is illustrated in Fig. 1.5. In quantum field theory, an electron is not just an electron—it can suddenly emit a photon, or it can emit a photon that subsequently annihilates into an electron–positron pair, and so on. In other words, an electron in quantum field theory exhibits itself in many disguises, one of which we show in Fig. 1.5a. Note that the original electron is surrounded by $e^- e^+$ pairs and, because opposite charges attract, the positrons will be preferentially closer to the electron. Therefore, the electron is surrounded by a cloud of charges which is polarized in such a way that the positive charges are closer to the electron; the negative charge of the electron is thus screened, as shown in more detail in Fig. 1.6. Suppose that we want to determine the charge of the electron in Fig. 1.6 by measuring the Coulomb force experienced by a test charge. The result will depend on where we

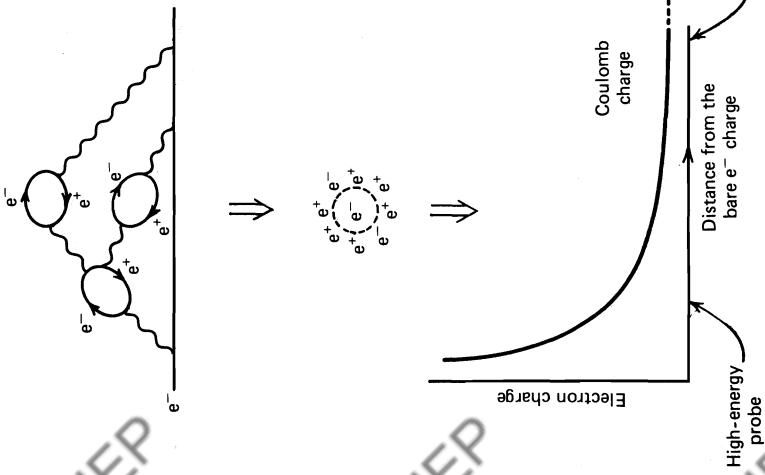
Quantum chromodynamics (QCD)



(b)

High-energy probe
"Asymptotic freedom"

Quantum electrodynamics (QED)



(a)

Fig. 1.5 Screening of the (a) electric and (b) color charge in quantum field theory.

place the test charge; when moving the test charge closer to the electron, we penetrate the cloud of positrons that screens the electron's charge. Therefore, the closer one approaches the electron, the larger is the charge one measures. In quantum field theory, the vacuum surrounding the electron has become a polarizable medium. The situation is analogous to that of a negative charge in a dielectric medium: the electron–positron pairs in Fig. 1.6 respond to the presence of the electron like the polarized molecules do in the dielectric. This effect is known as charge screening; as a result, the “measured charge” depends on the distance one is probing the electron; the result is shown pictorially in Fig. 1.5a. In QED, this variation of the charge is calculable by considering all possible configurations of the electron's charge cloud, only one of which is shown in Fig. 1.5a (see Chapter 7).

One can carry through the same calculation for the color charge of a quark. Color screening would be a carbon copy of charge screening if it were not for the new configurations involving gluons turning into pairs of gluons, as shown in Fig. 1.5b. The gluons, themselves carriers of color, also spread out the effective color charge of the quark. It turns out that the additional diagrams reverse the familiar result of quantum electrodynamics: a red charge is preferentially surrounded by other red charges, as shown in Fig. 1.5b. We now repeat the experiment of Fig. 1.6 for color charges. By moving our test probe closer to the original red quark, the probe penetrates a sphere of predominantly red charge and the amount of red

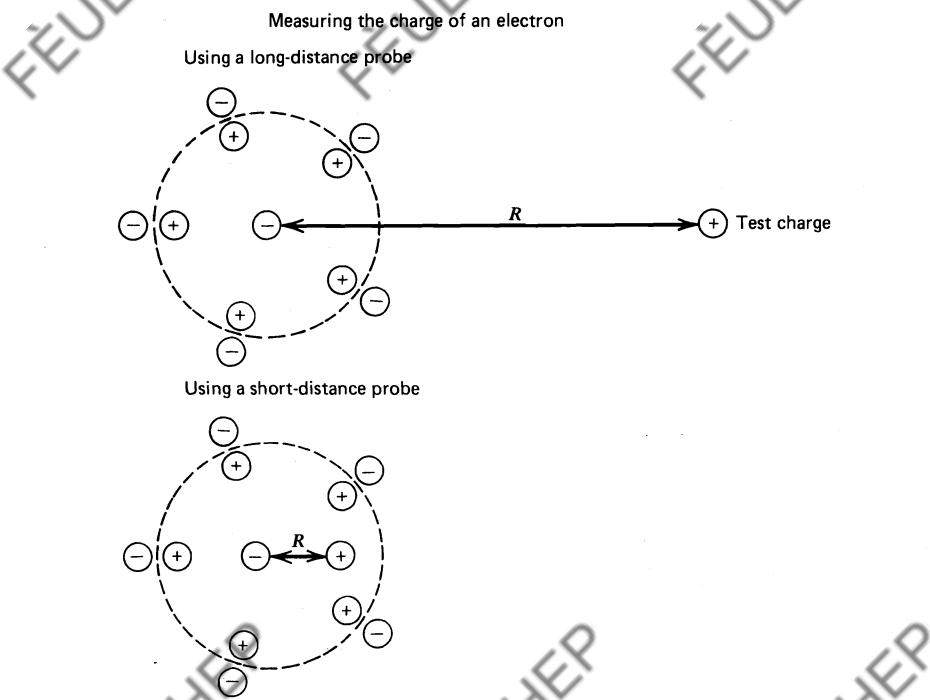


Fig. 1.6 Measuring the charge of an electron.

charge measured decreases. The resulting “antiscreening” of the red color is pictured in Fig. 1.5b and is referred to as “*asymptotic freedom*” (Chapter 7). Asymptotically, two red quarks interact (i.e., for very small separations) through color fields of reduced strength and approach a state where they behave as essentially free, noninteracting particles.

You might wonder why we have immediately emphasized this rather exotic property of color theory. As we see further on, it is asymptotic freedom that turns QCD into a quantitative calculational scheme (Chapters 10 and 11).

1.4 Natural Units

At this stage, it is necessary to break the flow of the physics discussion and to introduce units appropriate to particle physics. The two fundamental constants of relativistic quantum mechanics are Planck's constant, \hbar , and the velocity of light *in vacuo*, c :

$$\hbar \equiv \frac{h}{2\pi} = 1.055 \times 10^{-34} \text{ J sec}$$

$$c = 2.998 \times 10^8 \text{ msec}^{-1}.$$

It is convenient to use a system of units in which \hbar is one unit of action (ML^2/T) and c is one unit of velocity (L/T). Our system of units will be completely defined if we now specify, say, our unit of energy (ML^2/T^2). In particle physics, it is common to measure quantities in units of GeV (1 GeV $\equiv 10^9$ electron volts), a choice motivated by the fact that the rest energy of the proton is roughly 1 GeV.

By choosing units with $\hbar = c = 1$, it becomes unnecessary to write \hbar and c explicitly in the formulas, thus saving a lot of time and trouble. We can always use dimensional analysis to work out unambiguously where the \hbar 's and c 's enter any formula. Hence, with a slight but permissible laziness, it is customary to speak of mass (m), momentum (mc), and energy (mc^2) all in terms of GeV, and to measure length (\hbar/mc) and time (\hbar/mc^2) in units of GeV^{-1} . Table 1.2a displays the connection between GeV units and mks units and Table 1.2b lists some useful conversion formulae.

EXERCISE 1.1 Cross sections are often expressed in millibarns, where 1 mb = 10^{-3} b = 10^{-27} cm 2 . Using GeV units, show that

$$1 \text{ GeV}^{-2} = 0.389 \text{ mb}.$$

So far, we have not considered the elementary charge e , which measures how strongly electrons, say, interact electromagnetically with each other. To obtain a dimensionless measure of the strength of this interaction, we compare the electrostatic energy of repulsion between two electrons one natural unit of length apart with the rest mass energy of an electron:

$$\alpha = \frac{1}{4\pi} \frac{e^2}{(\hbar/mc)} / mc^2 = \frac{e^2}{4\pi\hbar c} \approx \frac{1}{137}. \quad (1.3)$$

TABLE 1.2a

Conventional Mass, Length, Time Units, and Positron Charge in Terms of $\hbar = c = 1$ Energy Units

Conversion Factor	$\hbar = c = 1$ Units	Actual Dimension
$1 \text{ kg} = 5.61 \times 10^{26} \text{ GeV}$	GeV	$\frac{\text{GeV}}{c^2}$
$1 \text{ m} = 5.07 \times 10^{15} \text{ GeV}^{-1}$	GeV^{-1}	$\frac{\hbar c}{\text{GeV}}$
$1 \text{ sec} = 1.52 \times 10^{24} \text{ GeV}^{-1}$	GeV^{-1}	$\frac{\hbar}{\text{GeV}}$
$e = \sqrt{4\pi\alpha}$	—	$(\hbar c)^{1/2}$

TABLE 1.2b

Some Useful Conversion Factors

$$1 \text{ TeV} = 10^3 \text{ GeV} = 10^6 \text{ MeV} = 10^9 \text{ KeV} = 10^{12} \text{ eV}$$

$$1 \text{ fermi} \equiv 1 \text{ F} = 10^{-13} \text{ cm} = 5.07 \text{ GeV}^{-1}$$

$$(1 \text{ F})^2 = 10 \text{ mb} = 10^4 \mu\text{b} = 10^7 \text{ nb} = 10^{10} \text{ pb}$$

$$(1 \text{ GeV})^{-2} = 0.389 \text{ mb}$$

In (1.3), we have adopted the rationalized Heaviside–Lorentz system of electromagnetic units. That is, the 4π factors appear in the force equations rather than in the Maxwell equations, and ϵ_0 is set equal to unity. This choice, which is conventional in particle physics, reduces Maxwell's equations to their simplest possible form. The value of α is, of course, the same in all systems of units, but the numerical value of e is different.

For historical reasons, α is known as the fine structure constant. Unfortunately, this name conveys a false impression. We have seen that the charge of an electron is not strictly constant but varies with distance because of quantum effects; hence α must be regarded as a variable, too. The value $\frac{1}{137}$ is the asymptotic value of α shown in Fig. 1.5a.

EXERCISE 1.2 Show that, in $\hbar = c = 1$ units, the Compton wavelength of an electron is m^{-1} , the Bohr radius of the hydrogen atom is $(\alpha m)^{-1}$, and the velocity of an electron in its lowest Bohr orbit is simply α ; m is the mass of the electron or, more precisely, the reduced mass $m_e m_p / (m_e + m_p)$.

EXERCISE 1.3 Justify the statements, that due to the weakness of the electromagnetic interaction ($\alpha \approx \frac{1}{137}$), the hydrogen atom is a loosely bound extended structure and that the nonrelativistic Schrödinger equation is adequate to describe the gross structure of atomic energy levels.

1.5 Alpha (α) is not the Only Charge Associated with Particle Interactions

In Fig. 1.3, we pictorially associated the electromagnetic interaction of two charges with the emission and reabsorption of a field quantum γ . We can somewhat “quantify” this picture by interpreting some familiar examples of classical electromagnetic phenomena in terms of this language. If you feel uneasy about the vagueness of the arguments below, we would like to point out that an exact operational definition of these so-called Feynman diagrams is forthcoming in Chapters 3 through 7. There they will represent the probability amplitudes for the process pictured and will be calculated using relativistic quantum mechanics. As a first example, consider the diagram of Fig. 1.7a, representing the Thomson scattering of photons off electrons. For long-wavelength photons, the cross section is given by

$$\sigma_{TH} = \frac{8\pi}{3} \left(\frac{\alpha}{m_e} \right)^2 = \frac{2}{3} \alpha^2 (4\pi R_e^2), \quad (1.4)$$

where R_e is the Compton wavelength of the electron,

$$R_e = \frac{\hbar}{m_e c} = \frac{1}{m_e} \quad (1.5)$$

in natural units. Not surprisingly, the Thomson cross section can be obtained classically as well as from quantum electrodynamics. In fact, it is this long-wavelength limit that is used to define the charge $-e$ of the electron. It is the value of the charge (or α) when probed with a low-energy probe from a large distance, see Fig. 1.5a.

Another example of an electromagnetic interaction is the Rutherford scattering of an electron of energy E on a nucleus of charge Ze , for which the differential cross section is

$$\frac{d\sigma_R}{d\Omega} = \frac{Z^2 \alpha^2}{4E^2} \frac{1}{\sin^4(\theta/2)}. \quad (1.6)$$

This process is pictured by the Feynman diagram of Fig. 1.7b. We now want to use the two processes pictured in Fig. 1.7 to illustrate the fact that α is a measure

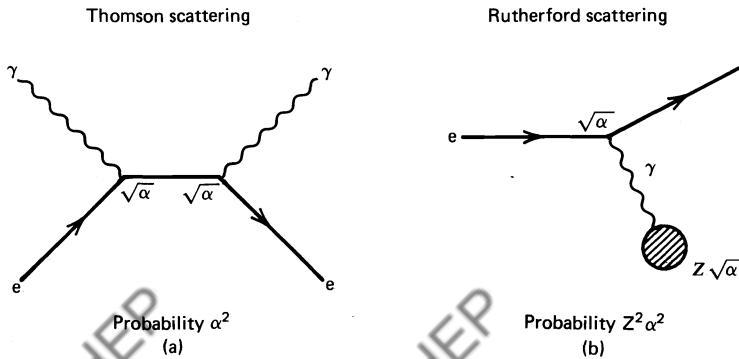
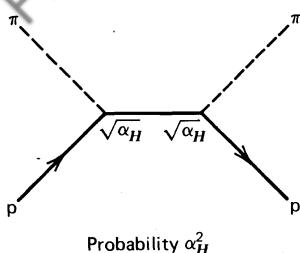


Fig. 1.7 (a) Thomson scattering. (b) Rutherford scattering.

Fig. 1.8 πp scattering.

of the strength of the electromagnetic interaction or, in particle exchange language, the probability for emitting or absorbing a photon. As we eventually want the Feynman diagrams to represent probability *amplitudes*, a factor e or $\sqrt{\alpha}$ is associated with each absorption or emission of a photon by a charge e . Inspection of the diagrams in Figs. 1.7a and 1.7b makes it clear that two such factors $(\sqrt{\alpha} \sqrt{\alpha})^2$ should appear in the cross section (1.4) and a factor $(\sqrt{\alpha} Z \sqrt{\alpha})^2$ in (1.6).

Armed with this concept, we now return to Thomson scattering but replace the photon beam by π -mesons. First, we note that, in analogy to (1.4), the Thomson cross section for the scattering of long-wavelength photons off a proton target is

$$\sigma_{TH} = \frac{2}{3} \alpha^2 (4\pi R_p^2), \quad (1.7)$$

where $R_p = 1/m_p$ in natural units. Our analyzer photon beam sees an effective area $(4\pi R_p^2)$ of the target proton, and α^2 is, as before, the probability of absorbing and emitting a photon. The factor $\frac{2}{3}$ is immaterial for present considerations; it actually reflects the fact that for a photon, one of the three polarization states expected for a particle of spin 1 is missing.

If we now repeat the measurement of the proton's radius using a π -meson instead of a photon beam (Fig. 1.8), we find that the coupling (emission or absorption) of π -mesons to nucleons cannot possibly be electromagnetic. The cross section for the process of Fig. 1.8 is readily found to be

$$\sigma_T(\pi p) = \alpha_H^2 (4\pi R_p^2), \quad (1.8)$$

where, as before, the factor $4\pi R_p^2$ is the effective area of the target; $\sigma_T(\pi p)$ is measured to be well over 1 mb. From a comparison of this result with the much smaller measured values of σ_{TH} , given by (1.7), we are forced to conclude that α_H , the probability for absorbing and emitting π -mesons, exceeds α by two to three orders of magnitude. That is,

$$\alpha_H \sim 1 \text{ to } 10. \quad (1.9)$$

The conclusion that a new “charge” and a new field has to be invoked to explain πN interaction cross sections is therefore inevitable. A substantial effort was made to construct an appropriate quantum field theory. More quantitative analyses of the type presented above actually show that the nucleon is not a simple object with classical radius m_N^{-1} but a complicated structure with a size

$$\langle r^2 \rangle = \frac{1}{m_\pi^2}.$$

This was for a long time taken as a hint that π -mesons are in fact the field quanta associated with this new charge. These historic attempts failed for several reasons. First, the π -meson itself possesses just as rich a substructure as the nucleon. It therefore does not fit well into its role as the “photon” of the nuclear interactions. The other problem is that actually $\alpha_H \approx 15$. This immediately precludes any hope of repeating the successes of quantum electrodynamics, which are anchored in the use of perturbation theory exploiting the numerical smallness of α to ensure that the perturbation series converges rapidly.

Nowadays, we associate the structure of the proton with quarks; the quark color charge is the “true” charge of strong interactions. Gluons are the “photons” of strong interactions. Just as the probability for photon emission by a charged particle is given by $\alpha = e^2/4\pi$, so the probability for gluon emission by a colored quark is characterized by α_s , where α_s is essentially the square of the color charge divided by 4π (see Fig. 1.9). We no longer regard α_H as a fundamental number but as an “epiphenomenon” that reflects the structure of pions and nucleons as well as α_s .

Scattering of low-energy π -mesons probes in some way the color charge at the surface of the nucleon, that is, over a typical distance $1/m_\pi \approx 1.4$ F. Here, even $\alpha_s \gtrsim 1$, as can be seen from Fig. 1.5b; α_s is too large to permit the use of perturbation theory. Using the modern technology of very high-energy particle beams, we can however probe the color charge of individual quarks deep inside the nucleon where α_s is smaller (see again Fig. 1.5b). A large variety of experimental probes has been invented to reach energies and distances where $\alpha_s \approx 0.2$; perturbation theory then is a workable approximation. Asymptotic freedom of QCD is the key to the use of perturbation theory (see Chapters 10 and 11).

Examples of experimental probes are shown in Figs. 1.10 and 1.11. Fig. 1.10a depicts the familiar technique of mapping the structure of an atom by scattering a beam of electrons off it. In elastic collisions, where the exchanged photon in Fig. 1.10a carries very little momentum and has therefore a relatively long wavelength,

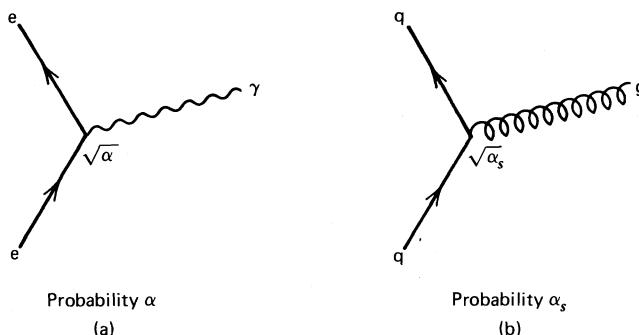


Fig. 1.9 (a) The probability amplitude for emitting a photon is proportional to e or $\sqrt{\alpha}$. (b) The probability amplitude for emitting a gluon is proportional to $\sqrt{\alpha_s}$.

the “optics” is such that we are illuminating the complete atom. In such collisions, one is therefore measuring the size $\langle r^2 \rangle$ of the electron cloud. But for very inelastic interactions of the electron beam, the exchanged photon now acquires a large momentum. The photon in Fig. 1.10a has turned into a short-wavelength probe illuminating the atom with high resolution. The cross section takes the form of (1.6). Figure 1.10c shows the result of such an experiment scattering α -particles on a gold target. The large cross section observed at large angles is due to α -particles rebounding off nuclei deep inside the gold atoms. More than 60 years after these pioneering atomic structure experiments, a very similar phenomenon was observed in high-energy proton–proton scattering. This time, the enhancement (Fig. 1.10c) is indicative of a hard-core substructure inside

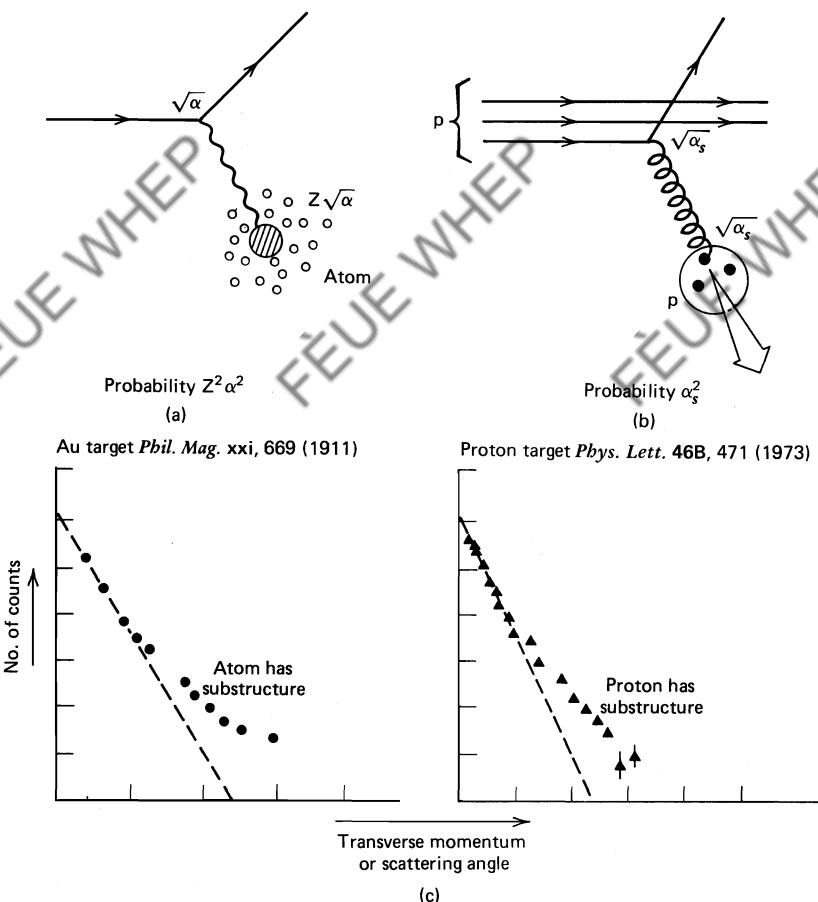


Fig. 1.10 (a) Inelastically scattered charged particle beam reveals the substructure of the atom. (b) Inelastically scattered proton beam reveals the quark structure of the proton target. (c) Experimental results.

the proton. A direct analogy between the atomic and hadronic situation is made in Fig. 1.10b. In inelastic collisions of very high-energy protons, individual quarks in the beam and the target “Rutherford scatter” off one another. As the interactions involve color charges over distances smaller than the size of a nucleon, α_s is small, and quantitative predictions can be made. An analysis of data based on the diagram shown in Fig. 1.10b shows indeed that quarks follow the Rutherford formula of (1.6):

$$\frac{d\sigma_R}{d\Omega} \sim \frac{\alpha_s^2}{4E^2} \frac{1}{\sin^4(\theta/2)}, \quad (1.10)$$

with $\alpha_s \approx 0.2$. The presence of gluons inside the nucleon have to be taken into account, and this complicates the analysis; but in principle, these experiments are exact analogues of Rutherford scattering.

This experimental approach for resolving quarks inside nucleons is not unique. There are other techniques that allow us to illuminate hadrons with photons of very large momentum Q and therefore small Compton wavelength $\lambda = 1/Q$. Imitating the atomic physics experiments, such photons are often “prepared” by the inelastic scattering of high-energy electron and muon beams off nuclear targets (see Fig. 1.11a). Colliding very high-energy electron and positron beams also provides us with an extremely “clean” technique to probe quarks (see Fig. 1.11b). The photon, produced at rest by e^+ and e^- beams colliding head on, can decay into a $q\bar{q}$ pair as shown in the picture (Fig. 1.11b). When the pair separates by distances of order 1 fm, α_s becomes large; that is, the color interactions between the quark and antiquark become truly strong, and these violent forces decelerate the quarks. The decelerated quarks radiate hadrons (mostly light π -mesons) just

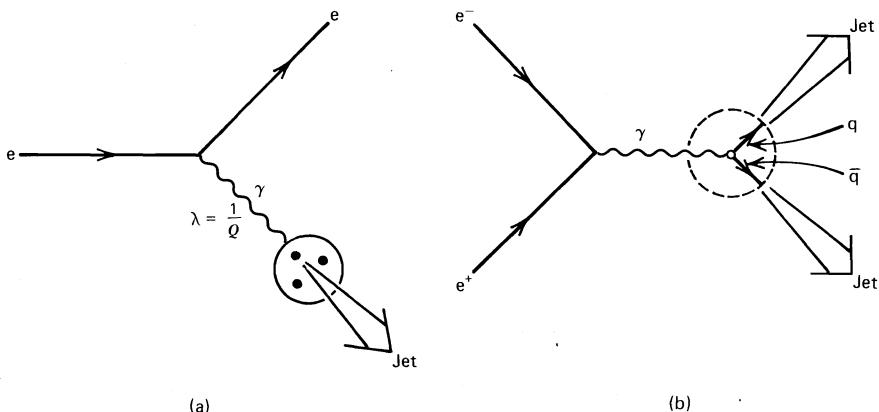


Fig. 1.11 (a) Virtual photons with a short wavelength resolve quarks in the proton target. (b) Virtual photons, obtained by annihilating e^- and e^+ beams, decay into quark-antiquark pairs.

like a decelerated charge emits photons by bremsstrahlung. The original quark is never seen in its “free” state; only these π -mesons and other (colorless) hadrons hit the experimentalist’s detector. The widening arrows in Figs. 1.10b and 1.11 indicate the showers or jets of hadrons radiated from, and traveling more or less in the direction of, the struck quark. The quark thus “escapes,” but only as a constituent of one of the radiated hadrons.

The central role played by the property of color screening (Fig. 1.5b) becomes more and more crucial. But is the argument that quarks (or color) are confined within hadrons any more than an excuse for our failure to observe free quarks? The answer is that this picture makes unique and dramatic predictions; as the virtual photon is produced at rest in Fig. 1.11b, the quark and antiquark should emerge in opposite directions to conserve momentum. Therefore, the two sprays of hadrons (jets) should be observed on opposite sides of the annihilation point where the photon is produced. The verification of this prediction was a real boost for the quark theory; any pre-quark theory of elementary particles had led us to expect a uniform, isotropic distribution of the emerging hadrons. The experimental verdict is unambiguous; Fig. 1.12 shows an example of a two-jet event seen by the struck wires of a detector. Notice the back-to-back orientation of the two jets.

The most striking (and certainly far from obvious) aspect of two-jet events is that the bremsstrahlung products, and never the original quark, reach the detector. Even with our very qualitative insight of color theory, it is possible to “rationalize” this experimental fact. It is once more crucial to recall the difference between QED and QCD recorded in Fig. 1.5. When a quark and antiquark separate, their color interaction becomes stronger. Through the interaction of gluons with one another, the color field lines of force between the quark and the antiquark are squeezed into a tube-like region as shown in Fig. 1.13a. This has to be contrasted with the Coulomb field where nothing prevents the lines of force from spreading out. There is no self-coupling of the photons to contain them. This is yet another way to visualize the different screening properties of QED and QCD. If the color tube has a constant energy density per unit length, the potential energy between the quark and the antiquark will increase with separation, $V(r) \sim \lambda r$, and so the quarks (and gluons) can never escape. This so-called “infrared slavery” is believed to be the origin of the total *confinement* of quarks to colorless hadrons. But how do they materialize as hadron jets? The answer is shown in Fig. 1.14. The separating $q\bar{q}$ pair stretches the color lines of force until the increasing potential energy is sufficient to create another $q\bar{q}$ pair. These act as the end points for the lines of force, which thus break into two shorter tubes with lower net energy despite the penalty of providing the extra $q\bar{q}$ mass. The outgoing quark and antiquark continue on their way (remember that they originally carried the momentum of the colliding e^- and e^+), further stretching the color lines. More $q\bar{q}$ pairs are produced until eventually their kinetic energy has degraded into clusters of quarks and gluons, each of which has zero net color and low internal momentum and therefore very strong color coupling. This coupling turns them

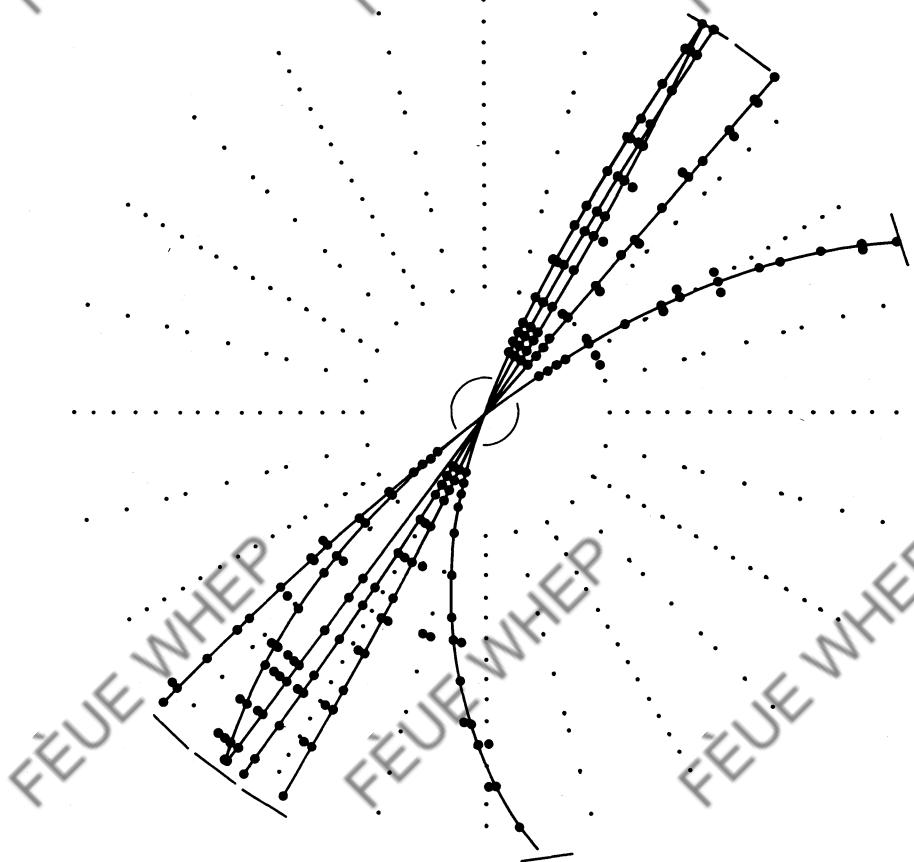


Fig. 1.12 Tracks of charged particles in a quark and antiquark jet. The TASSO detector at PETRA observes the products of a very high-energy e^- and e^+ head-on collision in the center of the picture.

into the hadrons forming two jets of particles traveling more or less in the direction of the original quark and antiquark (see Fig. 1.14).

In summary, we have argued that quarks have color as well as electric charge. The experimental evidence is compelling. Color is the same property previously introduced in an *ad hoc* way for solving a completely unrelated Fermi-statistics problem. We have argued that the theory of color (QCD) and of electrodynamic (QED) interactions are much alike in that the massless gluons exchanged between colored quarks are very similar to the massless photon exchanged between charged electrons. The crucial difference is the magnitude and screening properties of α_s and α (the strengths of the respective interactions) as indicated by the sketches in Figs. 1.5 and 1.13.

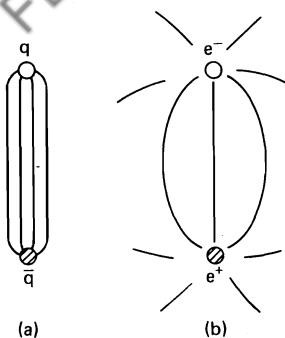


Fig. 1.13 The $q\bar{q}$ color field with $V(r) \sim r$, and the e^+e^- Coulomb field with $V(r) \sim 1/r$.

Where does all this leave the short-range nuclear force that binds neutrons and protons to form nuclei, which was in fact the original motivation for introducing the strong interaction? The answer is best seen in analogy with chemical binding due to the electromagnetic forces between two neutral atoms as they approach each other. As the atoms are electrically neutral, very little force is experienced until their electron clouds start to overlap. The force, known as the van der Waals force, increases gradually at first, and then very rapidly as the interpenetration increases. It is responsible for molecular binding and involves the exchange of electrons between the atoms. In itself it is not a fundamental interaction, but is instead a complicated manifestation of the basic electromagnetic interactions between two extended charged systems. In the same manner, the nucleon–nucleon force can be viewed as a complicated manifestation of the basic interaction between colored quarks. The nucleons, which are color neutral, only experience strong interactions at short distances when the quarks in one nucleon can “sense” the presence of the quarks in the other.

1.6 There are Weak Interactions, too

The Δ^{++} -particle, which gave us the crucial hint to introduce the color quantum number, was discovered by Fermi and collaborators by bombarding π^+ -mesons

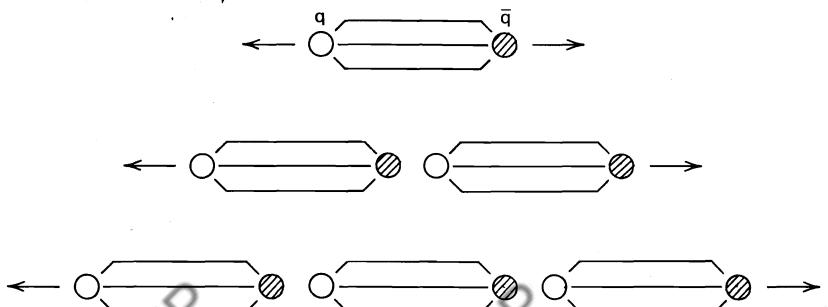


Fig. 1.14 Jet formation when a quark and antiquark separate.

on protons. The doubly charged Δ^{++} only lives for about 10^{-23} sec, then decays back into a π^+ and a proton. In the quark picture, one could imagine a decay mechanism of the type shown in Fig. 1.15a. Strong interactions, with range about 1 F, are responsible for the decay of the Δ^{++} ; therefore, the typical decay time is the time it takes the π^+ and the proton to separate by a distance $R \approx 1$ F:

$$\tau = \frac{R}{c} \approx 10^{-23} \text{ sec.} \quad (1.11)$$

Other “resonant” baryon states formed in the scattering of π -mesons and protons also have lifetimes of order 10^{-23} sec. In contrast, the proton itself is known to

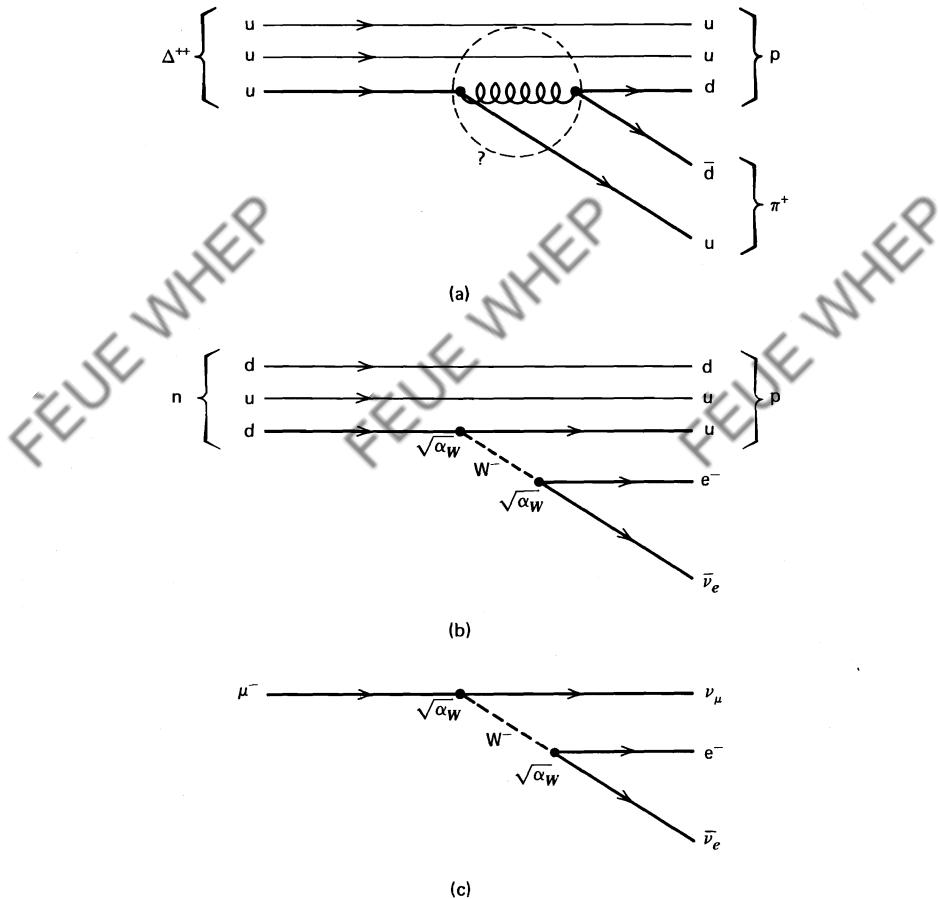


Fig. 1.15 (a) Decay of $\Delta^{++} \rightarrow \pi^+ p$. The decay mechanism is only symbolic; it reminds us that the Δ^{++} decays by strong interactions (with range about 1 F). (b) Neutron β -decay is mediated by weak interactions. A massive weak boson W is emitted and absorbed with probability α_W . (c) Muon decay.

have a lifetime in excess of 10^{30} years. To cover this special case, we have to rely on baryon number conservation. Without it, the proton could be unstable and decay; for example,

$$p \rightarrow e^+ \pi^0.$$

This decay is forbidden (but see Chapter 15). Indeed, using Table 1.1, we notice that the total baryon number before decay is $3 \times \frac{1}{3}$ for the uud quarks in the proton and zero after decay because of the cancellation of the q and \bar{q} baryon number making up the π^0 . The electron, the lightest charged particle, is of course stable by charge conservation.

The neutron's decay is familiar from radioactivity:

$$n \rightarrow p + e^- + \bar{\nu}.$$

This transition is energetically allowed and is not excluded by baryon number conservation. But the neutron lives 15 minutes! This is a puzzle. We can imagine that some particles have longer lifetimes than the 10^{-23} sec obtained in (1.11) because they decay exclusively through the weaker electromagnetic interactions and not through the color interaction sketched in Fig. 1.15a. The lifetime of the π^0 , which decays electromagnetically via $\pi^0 \rightarrow \gamma\gamma$, is 10^{-16} sec. This is what we expect, as

$$\frac{\tau_{\text{electromagnetic}}}{\tau_{\text{strong}}} \simeq \left(\frac{\alpha_s}{\alpha} \right)^2 = 10^4 \sim 10^6, \quad (1.12)$$

and therefore electromagnetic lifetimes should be about $10^4 \sim 10^6$ times longer than the typical 10^{-23} sec lifetime of particles decaying via the strong interaction. To obtain (1.12), we took the hint from Fig. 1.15 that the hadronic decay probability is proportional to $(\sqrt{\alpha_s} \sqrt{\alpha_s})^2$. However, in our world with two scales, α and α_s , nothing can account for lifetimes substantially longer than 10^{-16} sec. The proton and electron are “protected” by conservation laws, but what about the 15-minute lifetime of the neutron?

The problem is not limited to the neutron. The charged π^- decays into $\pi^- \rightarrow e^- + \bar{\nu}$ in 10^{-12} sec, and a series of “strange” particles exists with lifetimes of order 10^{-10} sec, for example, $\Sigma^+ \rightarrow n + \pi^+$. The $\Sigma \rightarrow n + \pi$ decay demonstrates the puzzle in a very striking way. Energetically, the decays $\Sigma \rightarrow n + \pi$ and $\Delta \rightarrow n + \pi$ are almost identical. The available phase space for $\Sigma \rightarrow n + \pi$ is 0.12 GeV kinetic energy, approximately the same as that for $\Delta \rightarrow n + \pi$. Nevertheless, their lifetimes differ by about 13 orders of magnitude. We are forced to introduce a new scale α_w and to invent a new “weak” interaction, arguing in analogy with (1.12) that

$$\frac{\tau(\Delta \rightarrow n + \pi)}{\tau(\Sigma \rightarrow n + \pi)} \simeq \frac{10^{-23} \text{ sec}}{10^{-10} \text{ sec}} \simeq \left(\frac{\alpha_w}{\alpha_s} \right)^2 \quad (1.13)$$

where α_w is the probability for emitting or absorbing some “weak quantum” W . Neutron β -decay is then represented by the Feynman diagram shown in Fig.

1.15b. Equation (1.13) implies that

$$\alpha_W \simeq 10^{-6} \quad (1.14)$$

compared to $\alpha_s \simeq 1$ and $\alpha \simeq 10^{-2}$.

Implicit in the above discussion is that the weak field quanta (unlike gluons) couple with equal strength, α_W , to leptons (e, ν) and quarks (u, d) (see Fig. 1.15b). There is direct experimental support for this from the comparison of β -decay with purely leptonic weak processes, such as $\mu^- \rightarrow e^- \bar{\nu}_e \nu_\mu$ (compare Figs. 1.15b and 15c). The weak interaction has another novel feature: it changes a d quark into a u quark (Fig. 1.15b) and a muon into a neutrino. We say that weak interactions change quark and lepton “flavor.” The weak field quanta W in Fig. 1.15 have electric charge, unlike the photon and gluons. In fact, they exist with positive and negative, as well as neutral, electric charge (see Chapters 12 and 13).

We can now summarize the exchanged bosons and the associated charges by which quarks and leptons interact. This is done in Tables 1.3 and 1.4.

Although we had no choice but to introduce a new interaction to explain radioactive β -decay and other “weak” phenomena, this need not imply the existence of a new field with its own *ad hoc* coupling strength α_W . There is a more pleasing and more appropriate (Chapter 15) interpretation for (1.14). One can insist that the coupling of the W in the diagrams of Figs. 1.15b and 1.15c is in fact electromagnetic in strength. The probability of emitting a W is then essentially the same as that for emitting a photon, namely, α . The slow rate of weak decays is achieved instead by giving the W a large mass. The probability of exchanging a W is small compared to that for exchanging a photon, not because it is less likely to be emitted, but because it is massive. Equation (1.14) is then reinterpreted to read

$$\alpha_W = \frac{\alpha}{(M_W/m_p)^2} \simeq 10^{-6}. \quad (1.15)$$

The qualitative statement expressed by (1.15) is already clear. The large mass M_W suppresses the coupling strength α of the weak bosons to a given “effective” coupling strength α_W . For dimensional reasons, M_W has to be expressed in terms of some reference mass; our choice is the proton mass. That the suppression is quadratic is not obvious at all but depends on the details of the “electroweak” theory we construct. An equation looking like (1.15) is derived in Chapter 15. Since $\alpha \simeq 10^{-2}$, we have the prediction that

$$M_W \simeq 10^2 m_p. \quad (1.16)$$

This result also implies that the weak interactions have a short range. A minimum energy $M_W c^2$ is necessary to emit a virtual W . It can therefore only live a time $\Delta t \lesssim \hbar/M_W c^2$, after which it has to be reabsorbed (remember the discussion of Fig. 1.3). During that time, it can travel at most a distance $c \Delta t = \hbar/M_W c \simeq 10^{-3}$ F, using (1.16). This range is much smaller than the 1-F range of a strong interaction. The important observation is, however, that no new charge need

TABLE 1.3

Diagrams Showing Typical Interactions

Interaction	Charge	Quarks	Leptons
Strong	Color		-
Electromagnetic	Electric charge (e)		
Weak	Weak charge (g), giving $u \rightarrow d$ or $\nu \rightarrow e^-$ flavor-changing transitions		

necessarily be introduced to accommodate weak interactions. Only the electric charge appears in (1.15). One might question the conceptual superiority of introducing a new mass scale instead of a new charge. A discussion of *spontaneous symmetry breaking* in Chapters 14 and 15 addresses this question. Of course, no matter how aesthetically satisfying the idea, experiment will be the final arbiter.

The previous discussion should be somewhat familiar. Maxwell's theory of electromagnetism also unifies two interactions: electricity and magnetism. The force on a charged particle moving with velocity v is

$$\mathbf{F} = e\mathbf{E} + e_M\mathbf{v} \times \mathbf{B}.$$

In Maxwell's theory one does not introduce a new charge to accommodate magnetic interactions. It *unifies* the two by stating that $e = e_M$. At low velocities,

TABLE 1.4

Summary of Chapter 1

Interaction	Range	Typical Lifetime (sec)	Typical Cross Section (mb)	Typical Coupling α_i
Strong	$1/F \approx \frac{1}{m_\pi}$ Color confinement range ^a	10^{-23}	10	1
Electromagnetic	∞	$10^{-20} \sim 10^{-16}$ e.g., $\pi^0 \rightarrow \gamma\gamma$ $\Sigma \rightarrow \Lambda\gamma$	10^{-3} e.g., $\gamma p \rightarrow p\pi^0$	10^{-2}
Weak	$\frac{1}{M_W}$ with $M_W \approx 100 m_p$	10^{-12} or longer e.g., $\Sigma^- \rightarrow n\pi^-$ $\pi^- \rightarrow \mu^-\bar{\nu}$	10^{-11} e.g., $\nu p \rightarrow \nu p$ $\nu p \rightarrow \mu^- p\pi^+$	10^{-6}

^a"van der Waals" manifestation of massless gluon exchange (see end of Section 1.5).

the magnetic forces are very weak; but for high velocities, the electric and magnetic forces play a comparable role. Unification of the two forces introduces a scale in the theory: the velocity of light. The velocity of light, c , is the scale which governs the relative strengths of the two forces. We also introduced a scale when unifying weak and electromagnetic interactions. It is an energy scale, M_W .

1.7 Down Mendeleev's Path: More Quarks and Leptons

The discovery of so-called "strange" particles such as the Σ , providing us with a striking example of a weak decay, has a more significant implication. They do not fit into the scheme of colorless qqq and $q\bar{q}$ excited states of u and d quarks. The

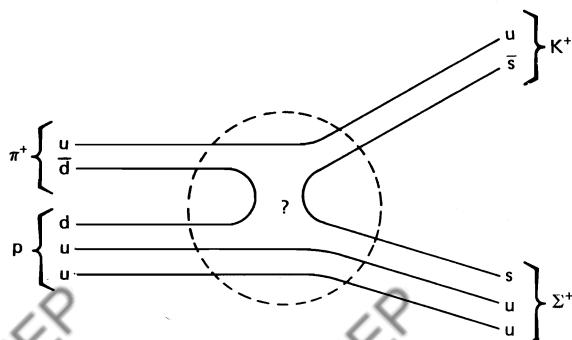


Fig. 1.16 The associated production of strange particles.

TABLE 1.5
Present Proliferation of Quarks and Leptons^a

Quarks		
u (up)	c (charm)	t (top)?
d (down)	s (strange)	b (beauty)
Leptons		
e (electron)	μ (muon)	τ (tau)
ν_e (e neutrino)	ν_μ (μ -neutrino)	ν_τ (τ -neutrino)

^aThe masses increase from left to right, and no direct evidence for the heaviest entry (the t quark) exists at present.

experimental observation that strange particles are produced in pairs in the strong interactions of nonstrange hadrons is a clear hint that they contain a new quark. Figure 1.16 illustrates this point using the interaction

$$\pi^+ + p \rightarrow K^+ + \Sigma^+$$

as an example. The blob in the center represents the complicated (1-F range) color interactions before the pair of strange particles emerges. Turning our attention to Table 1.1, one concludes from inspection of such data that the new strange quark has not only the electric charge (see Fig. 1.16), but also the other quantum numbers of the d quark. The measured masses of the strange particles imply that the s quark is heavier: the $K(u\bar{s})$ meson is heavier than the $\pi(u\bar{d})$, the $\Sigma(uus)$ is heavier than the $p(uud)$. Known look-alikes of the d quark now also include the even heavier b quark. The u quark has its heavier partner: the c quark. The history of heavier look-alikes to the quarks and leptons of Table 1.1 goes back to the muon, which appears in all respects to be an electron except for its mass, which is 200 times larger. The present evidence is that Table 1.1 should be expanded to Table 1.5.

Nuclei, atoms, and molecules, that is, “our world,” is built up out of the first column of Table 1.5. The particles in adjacent columns are identical in all properties. They differ only in mass. Why our world is doubled, tripled,... is one of the major unanswered questions. A sentence from the beginning of this chapter, referring to Mendeleev’s table, comes to mind: “Proliferation of elements and the apparent systematics in the organization of the table strongly suggests a substructure of more fundamental building blocks.” But with five (six?) quarks and leptons, is it too soon to worry?

1.8 Gravity

It may be surprising that the force most evident in everyday life, the gravitational force, has not been mentioned. The reason is simply that it is by far the weakest

force known. As a result, it has no measurable effects on a subatomic scale and no manifestations that can guide us to a quantum field theory. Then why is it so evident? That is because the gravitational force is cumulative, unlike the other long-range interaction, the electromagnetic force, and we live adjacent to an astronomical body: the Earth! Otherwise, gravity would only be apparent to astronomers. Large bodies are usually electrically neutral so that their electromagnetic forces cancel, whereas the gravitational pull of two bodies is the cumulative sum of the attractions between their constituent masses. The reader can however imagine that constructing a gauge theory for gravity has become one of the prime goals of particle physics. The problem is not easy. "Gravitons," unlike the spin 1 quanta in Table 1.1, have spin 2. The discussions are beyond the scope of this book.

1.9 Particles: The Experimentalist's Point of View

If electrons emitted from a vacuum tube, for example, are subjected to electric or magnetic fields, they follow trajectories as given by classical electrodynamics applied to a particle of charge $-e$ and mass m . This is also true for nucleons or α -particles emitted by a radioactive source, even though from the high-energy physicist's point of view they are complicated structures with a spatial extension. The experimental physicist's working definition of a particle is therefore an object to which he can assign a well-defined charge and mass and which behaves, for all practical purposes, like a point particle in the macroscopic electromagnetic fields of his accelerators and detectors.

Suppose we accelerate an electron through a potential difference of 1 volt between a cathode and an anode. If extracted from the vacuum tube "accelerator," it will emerge with an energy of 1 eV. In high-energy physics, we require, however, beams of energy $E \approx 10^9$ eV, or 1 GeV, so that our "electron microscope" can achieve a spatial resolution of order

$$\Delta x \sim \frac{\hbar c}{E}$$

$$\sim \frac{(6.6 \times 10^{-16} \text{ eV sec}) \times (3 \times 10^8 \text{ m/sec})}{10^9 \text{ eV}}$$

$$\sim 10^{-15} \text{ m, or 1 fermi.}$$

Obviously, a vacuum tube will not do! We cannot create sufficiently great potential differences. There are two solutions to this problem: either put a (very) large number of them in series, one perfectly lined up behind the other (linear accelerators), or place a number of them in a circle so that a circulating electron can be repeatedly passed through them and subjected to their acceleration (synchrocyclotrons).

The first solution was adopted for the construction of the 2-mile linear accelerator at the Stanford Linear Accelerator Laboratory where electrons reach

energies of up to 20 GeV. Electrons emerging from an ion source are accelerated by a series of radio-frequency (rf) cavities. These contain an oscillating electric field which is parallel to the electron's velocity when it enters each cavity. The electron therefore "rides on the crest of a wave" along the 2-mile accelerator. In order to confine it to a straight line, further electric or magnetic fields are usually needed to focus the beam, in the same way that lenses guide or focus a beam of light.

The circular accelerator solution was chosen for the construction of synchrotrons at Fermilab near Chicago (U.S.A.) and at CERN in Geneva (Switzerland). These are 2 km in diameter, and protons are accelerated up to 500 GeV. The physical principles involved are familiar from the old-fashioned cyclotron (see Fig. 1.17). In a cyclotron, a proton in a magnetic field B describes an orbit with radius

$$R = \frac{mv}{eB} \quad (1.17)$$

which increases as the velocity v increases. The proton thus spirals (see, e.g., Halliday and Resnick; see also Omnes). At relativistic speeds,

$$m = \gamma m_0 = \frac{m_0}{\sqrt{1 - (v^2/c^2)}}, \quad (1.18)$$

where m_0 is the proton rest mass, and the rotation frequency ω diminishes:

$$\omega = \frac{eB}{m}. \quad (1.19)$$

Remember that ω is a constant at nonrelativistic velocities. To achieve perfect acceleration, it is therefore sufficient to tune the alternating E field in Fig. 1.17 to the fixed frequency ω .

A synchrocyclotron, using a series of accelerating rf cavities placed in a circle, differs from a cyclotron in two ways: (1) the B field, supplied by dipole bending magnets interspaced between the rf cavities, is not constant but is continuously adjusted to force the particles into a fixed orbit; (2) the frequency of the oscillating electric field in the rf cavities is "synchronized" with the changing

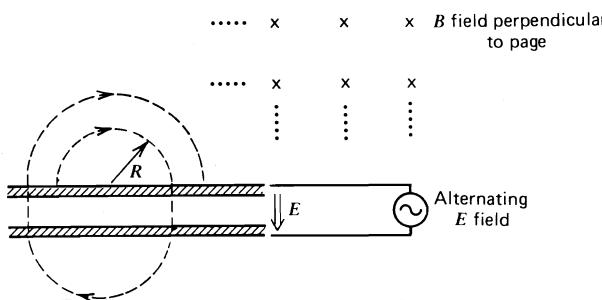


Fig. 1.17 A particle orbiting by means of the Lorentz force is repeatedly accelerated by an electric field E .

rotation frequency ω of the protons given by (1.19) (hence the name synchrocyclotron). Only protons entering the cavity at the right moment are accelerated. Thus, we are forced to accelerate “bunches” of protons rather than a continuous beam. In the CERN machine, 4600 proton bunches rotate with a revolution time of 23 μ sec. The machine operates therefore at 200 MHz. To create the necessary 5 million volts per revolution requires a radio-frequency power of 2 MW. During acceleration, the particles circulate roughly 10^5 times around the machine; this takes a total of 2 sec. They cover 500,000 km but are nevertheless kept in their orbit to a precision of 1 mm. To achieve this accuracy, quadrupole focusing magnets are interspaced with the rf cavities and bending magnets.

An increasingly popular technique is to accelerate two beams of particles which have opposite charges inside the same circular accelerator. The beams, circulating in opposite directions, are made to intersect at specific interaction regions, leading to violent head-on collisions (see Exercise 3.3). The products of a collision can be viewed with detectors surrounding the interaction region. Electron–positron colliders with beam energies of up to 20 GeV are in operation at Hamburg (West Germany) and at SLAC (Stanford, California, U.S.A.). The same technique is used in a proton–antiproton collider at CERN where the beams circulate with an energy of 270 GeV. Designs to reach higher energies are under study at CERN and Fermilab, and a new generation of accelerators is forthcoming, with Europe, the United States, Japan, and the Soviet Union all engaged in their construction.

Electron or proton beams can also be extracted from accelerators and directed onto external hydrogen or nuclear targets to study particle–nucleon interactions. Alternatively, we can focus the charged pions or kaons or antiprotons produced in these interactions into secondary beams, which in turn can be used to bombard a nuclear target, and so study their interactions with nucleons.

In a secondary π^+ -beam, some particles will decay in flight via $\pi^+ \rightarrow \mu^+ + \nu$. Therefore, the beam becomes “contaminated” with muons and neutrinos. By sending the beam through shielding material, the π^+ -component in the beam will be absorbed, because the π^+ -particles, unlike muons and neutrinos, interact strongly with the absorber. In so doing, we have constructed a μ^+ -beam “contaminated” with neutrinos. By increasing further the amount of absorber, eventually only the weakly interacting neutrinos survive. Our π^+ -beam has thus become a neutrino beam!

We should not forget that interactions of cosmic rays with the nitrogen and oxygen nuclei in the atmosphere have been observed with collision energies exceeding those reached in the laboratory by more than five orders of magnitude. The low flux of these high-energy cosmic rays makes a systematic study of the interactions a challenging, but also a very intriguing, mission.

1.10 Particle Detectors

Detection of charged particles is based on a simple physics principle: matter is ionized when traversed by charged particles. The electric field associated with a

charged particle moving through matter accelerates the outer electrons of nearby atoms and so ionizes them. The charged particle will thus leave behind a trail of ionized atoms which can be used to infer the trajectory it has followed. A common technique is to place a dielectric, usually a gas, between condenser plates. A potential close to breakdown is applied across the plates. Then, when a charged particle enters the detector, it ionizes the dielectric and the condenser will discharge. This is the principle on which the Geiger counter is based. *Spark*, *streamer*, or *flash* chambers used in high-energy physics experiments are basically a large collection of such counters. By observing the sequence of discharges in the multicomponent detector, we can reconstruct the particle track. The same technique is used in *proportional* or *drift* chambers. They are ionization counters operated in the "sub-Geiger" mode; that is, the voltage is maintained below breakdown of the dielectric. The ionization along the particle's trajectory is observed via electric pulses on the anode wires which collect the electrons that result from the ionization (see Fig. 1.18). The positive ions will drift toward cathode planes and also contribute to the detected current. It is not unusual to construct counters of $5\text{ m} \times 5\text{ m}$ using large numbers of anode wires and cathode strips.

The *bubble chamber* illustrates an alternative method of displaying the ionization which signals the passage of a charged particle. A similar method is also used in *cloud chambers* and *photographic emulsions*. In a large vessel a few meters in diameter, a liquid is kept under $5 \sim 20$ atmospheres of overpressure. Through a sudden decompression, the liquid is superheated and the boiling will start with the formation of bubbles along the particle track. It is the ionized atoms which catalyze the formation of these bubbles. The bubbles are allowed to grow for 10 msec and are then recorded by stereo cameras.

In some materials the atoms are excited, rather than ionized, by high-energy charged particles. The excited levels decay with the emission of light which can be observed by photomultiplier tubes. Such detectors, called *scintillation counters*,

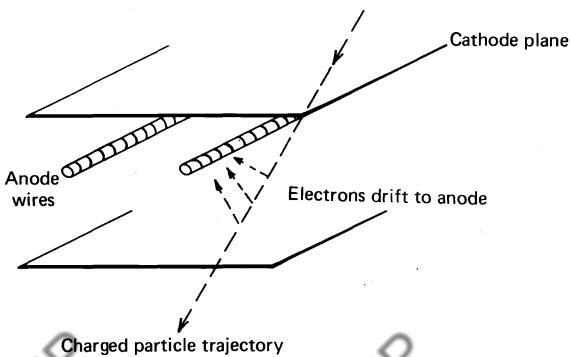


Fig. 1.18 The principle of a drift chamber.

can be built from an organic crystal scintillator, or organic materials where the particle excites the molecular levels.

Detectors exist which can discriminate between π -mesons, K-mesons, and nucleons by measuring their mass. In these experiments, particles move close to the speed of light. Their velocity can nevertheless be measured by sending the particles through a dielectric and observing the Čerenkov light emitted by excited atoms. Čerenkov light is emitted when a particle's velocity exceeds the speed of light in the dielectric medium ($v_{th} = c/n$, where n is the refractive index). The phenomenon is similar to the sonic boom radiated by an aircraft flying faster than the speed of sound in air. This *velocity* threshold can be used to distinguish particles with the same *momentum* but different masses.

Clearly, neutral particles cannot be observed by these techniques. However, we can detect their presence by observing their charged decay products, for example, $\pi^0 \rightarrow \gamma\gamma$ followed by $\gamma \rightarrow e^-e^+$. The last reaction is forbidden by energy conservation, but it is possible in the electric field of a nucleus $\gamma N \rightarrow e^-e^+N'$. When a photon or electron enters a high-Z material, repeated bremsstrahlung and pair production will create an avalanche of such particles. Their number increases exponentially with depth inside the detector until eventually the cascade has used up the total energy of the incident particle. Such detectors are called *calorimeters*.

A modern particle detector is usually a hybrid system made up from several of the above detectors. It will usually contain a magnet which deflects the particle tracks and allows a determination of their momenta. Its iron will absorb the hadrons and so identify surviving muons. Detectors often consist of over 1000 tons of magnetized iron. A detailed discussion of these multipurpose detectors and of the sophisticated electronics required to collect and digest the information from its various components is outside the scope of this book. For further reading, see, for example, Chapter 2 of Perkins (1982) and Kleinknecht (1982).

2

Symmetries and Quarks

SYMMETRIES AND GROUPS

2.1 Symmetries in Physics: An Example

A glance at a table of particle masses shows that the proton and neutron masses are amazingly close in value. Nuclear physicists took this as a hint that they are in fact two manifestations of one and the same particle called the “nucleon,” in very much the same way as the two states of an electron with spin up and down are thought of as one, not two, particles. Indeed, the mathematical structure used to discuss the similarity of the neutron and the proton is almost a carbon copy of spin, and is called “isospin” (see Fig. 2.1).

This concept is very useful. As a simple illustration, consider the description of the two-nucleon system. Each nucleon has spin $\frac{1}{2}$ (with spin states \uparrow and \downarrow), and so, following the rules for the addition of angular momenta, the composite system may have total spin $S = 1$ or $S = 0$. We are using units with $\hbar = 1$ (see Section 1.4). The composition of these spin triplet and spin singlet states is

$$\begin{cases} |S = 1, M_S = 1\rangle = \uparrow\uparrow \\ |S = 1, M_S = 0\rangle = \sqrt{\frac{1}{2}} (\uparrow\downarrow + \downarrow\uparrow) \\ |S = 1, M_S = -1\rangle = \downarrow\downarrow \end{cases} \quad (2.1)$$

$$|S = 0, M_S = 0\rangle = \sqrt{\frac{1}{2}} (\uparrow\downarrow - \downarrow\uparrow).$$

Each nucleon is similarly postulated to have isospin $I = \frac{1}{2}$, with $I_3 = \pm \frac{1}{2}$ for protons and neutrons, respectively. $I = 1$ and $I = 0$ states of the nucleon–nucleon system can be constructed in exact analogy to spin:

$$\begin{cases} |I = 1, I_3 = 1\rangle = pp \\ |I = 1, I_3 = 0\rangle = \sqrt{\frac{1}{2}} (pn + np) \\ |I = 1, I_3 = -1\rangle = nn \end{cases} \quad (2.2)$$

$$|I = 0, I_3 = 0\rangle = \sqrt{\frac{1}{2}} (pn - np).$$

EXERCISE 2.1 Justify the decomposition shown in (2.1) by either (1) considering the symmetry of the states under interchange of the nucleons or (2) using the angular momentum “step-down” operator.

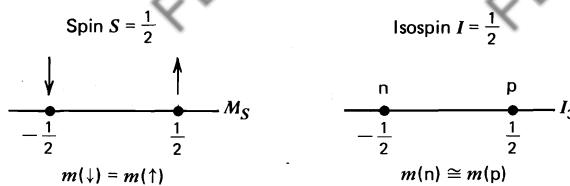


Fig. 2.1 Spin and isospin doublets.

EXERCISE 2.2 If the nucleons are in a state of relative orbital angular momentum $L = 0$, use the Pauli exclusion principle to show that $S + I$ must be an odd integer.

There is much evidence to show that the nuclear force is invariant under isospin transformations [for example, that it is independent of the value of I_3 in the $I = 1$ multiplet of (2.2)]. For instance, consider the three nuclei ${}^6\text{He}$, ${}^6\text{Li}$, and ${}^6\text{Be}$, which can be regarded respectively as an nn, np, and pp system attached to a ${}^4\text{He}$ core of $I = 0$. After correcting for the Coulomb repulsion between the protons and for the neutron-proton mass difference, the observed nuclear masses are as sketched in Fig. 2.2. Furthermore, isospin invariance requires that the same nuclear physics should be obtained for each of the three $I = 1$ states ($I_3 = -1, 0, 1$), just as rotational invariance ensures that the $2J + 1$ substates of an isolated system of total angular momentum J describe exactly equivalent physical systems.

EXERCISE 2.3 Use isospin invariance to show that the reaction cross sections σ must satisfy

$$\frac{\sigma(pp \rightarrow \pi^+ d)}{\sigma(np \rightarrow \pi^0 d)} = 2,$$

given that the deuteron d has isospin $I = 0$ and the π has isospin $I = 1$.

Hint You may assume that the reaction rate is

$$\sigma \sim |\text{amplitude}|^2 \sim \sum_I |\langle I', I'_3 | A | I, I_3 \rangle|^2$$

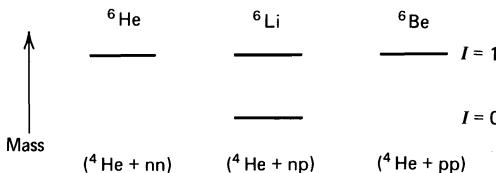


Fig. 2.2 Nuclear energy levels. The states reflect those of eq. (2.2). The two states of ${}^6\text{Li}$ are separated by about 2 MeV. After correcting for electromagnetic effects, the excited state of ${}^6\text{Li}$ and the ground states of ${}^6\text{He}$ and ${}^6\text{Be}$ are found to be degenerate in mass.

where I and I' are the total isospin quantum numbers of the initial and final states, respectively, and $I = I'$ and $I_3 = I'_3$.

In Sections 2.2 to 2.9, we amplify these ideas and take the opportunity to introduce some useful group theory concepts. You may prefer to skip these sections and, instead, refer to the appropriate results as and when necessary.

2.2 Symmetries and Groups: A Brief Introduction

Group theory is the branch of mathematics that underlies the treatment of symmetry. Although we shall not need the formal machinery of group theory, it is useful to introduce some of the concepts and the terminology which belongs to the jargon of particle physics. We take the rotation group as an illustrative example.

The set of rotations of a system form a group, each rotation being an element of the group. Two successive rotations R_1 followed by R_2 (written as the “product” $R_2 R_1$) are equivalent to a single rotation (that is, to another group element). The set of rotations is closed under “multiplication.” There is an identity element (no rotation), and every rotation has an inverse (rotate back again). The “product” is not necessarily commutative, $R_1 R_2 \neq R_2 R_1$, but the associative law $R_3(R_2 R_1) = (R_3 R_2)R_1$ always holds. The rotation group is a continuous group in that each rotation can be labeled by a set of continuously varying parameters ($\alpha_1, \alpha_2, \alpha_3$). These can be regarded as the components of a vector α directed along the axis of rotation with magnitude given by the angle of rotation.

The rotation group is a Lie group. The crucial property here is that every rotation can be expressed as the product of a succession of infinitesimal rotations (rotations arbitrarily close to the identity). The group is then completely defined by the “neighborhood of the identity.”

We do not want an experimental result to depend on the specific laboratory orientation of the system we are measuring. Rotations must therefore form a symmetry group of a system. They are a subset of the Lorentz transformations that can be performed on a system, namely, those transformations that leave it at rest. By definition, the physics is unchanged by a symmetry operation. In particular, these operations leave the transition probabilities of the system invariant. For example, suppose that under a rotation R the states of a system transform as

$$|\psi\rangle \rightarrow |\psi'\rangle = U|\psi\rangle. \quad (2.3)$$

The probability that a system described by $|\psi\rangle$ will be found in state $|\phi\rangle$ must be unchanged by R ,

$$|\langle\phi|\psi\rangle|^2 = |\langle\phi'|\psi'\rangle|^2 = |\langle\phi|U^\dagger U|\psi\rangle|^2, \quad (2.4)$$

and so U must be a unitary operator. The operators $U(R_1), U(R_2), \dots$ form a

group with exactly the same structure as the original group R_1, R_2, \dots ; they are said to form a unitary representation of the rotation group.

Moreover, the Hamiltonian is unchanged by a symmetry operation R of the system, and the matrix elements are preserved:

$$\langle \phi' | H | \psi' \rangle = \langle \phi | U^\dagger H U | \psi \rangle = \langle \phi | H | \psi \rangle,$$

so that

$$H = U^\dagger H U \quad \text{or} \quad [U, H] \equiv UH - HU = 0. \quad (2.5)$$

The transformation U has no explicit time dependence, and the equation of motion,

$$i \frac{d}{dt} |\psi(t)\rangle = H |\psi(t)\rangle, \quad (2.6)$$

is unchanged by the symmetry operation. As a consequence, the expectation value of U is a constant of the motion:

$$i \frac{d}{dt} \langle \psi(t) | U | \psi(t) \rangle = \langle \psi(t) | UH - HU | \psi(t) \rangle = 0. \quad (2.7)$$

All the group properties follow from considering infinitesimal rotations in the neighborhood of the identity. As an example, consider a rotation through an infinitesimal angle ϵ about the 3- (or z) axis. We may write, to first order in ϵ ,

$$U = 1 - i\epsilon J_3. \quad (2.8)$$

The operator J_3 is called the generator of rotations about the 3-axis. Now,

$$\begin{aligned} 1 &= U^\dagger U = (1 + i\epsilon J_3^\dagger)(1 - i\epsilon J_3) \\ &= 1 + i\epsilon(J_3^\dagger - J_3) + O(\epsilon^2). \end{aligned}$$

Therefore, J_3 is hermitian and hence is a (quantum mechanical) observable. The i was introduced in (2.8) to make this so.

To identify the observable J_3 , we consider the effect of a rotation on the wave function $\psi(\mathbf{r})$ describing the system. First, we must distinguish between two points of view. Either we may rotate the axes and keep the physical system fixed (the passive viewpoint) or we can keep the axes fixed and rotate the system (the active viewpoint). The viewpoints are equivalent; a rotation of the axes through an angle θ is the same as a rotation of the physical system by $-\theta$. We adopt the active viewpoint and rotate the physical system. The wave function ψ' describing the rotated state at \mathbf{r} is then equal to the original function ψ at the point $R^{-1}\mathbf{r}$, which is transformed into \mathbf{r} under the rotation R , that is,

$$\psi'(\mathbf{r}) = \psi(R^{-1}\mathbf{r}). \quad (2.9)$$

This specifies the one-to-one correspondence between ψ' and ψ , which we have written [cf. (2.3)]

$$\psi' = U\psi. \quad (2.10)$$

For an infinitesimal rotation ϵ about the z axis, (2.9) and (2.10) give

$$\begin{aligned} U\psi(x, y, z) &= \psi(R^{-1}\mathbf{r}) \approx \psi(x + \epsilon y, y - \epsilon x, z) \\ &\approx \psi(x, y, z) + \epsilon \left(y \frac{\partial \psi}{\partial x} - x \frac{\partial \psi}{\partial y} \right) \\ &= (1 - i\epsilon(xp_y - yp_x))\psi. \end{aligned} \quad (2.11)$$

Comparing (2.11) with (2.8), that is, with

$$U\psi = (1 - i\epsilon J_3)\psi,$$

we identify the *generator*, J_3 , of rotations about the 3- (or z) axis with the third-component of the *angular momentum operator*.

From (2.7), we see that the eigenvalues of the observable J_3 are constants of the motion. They are conserved quantum numbers. A symmetry of the system has led to a conservation law. The fact that experiments performed with different orientations of an apparatus give the same physics results (rotational symmetry) has led to the conservation of angular momentum.

A rotation through a finite angle θ may be built up from a succession of n infinitesimal rotations

$$U(\theta) = (U(\epsilon))^n = \left(1 - i\frac{\theta}{n}J_3\right)^n \xrightarrow{n \rightarrow \infty} e^{-i\theta J_3}. \quad (2.12)$$

We may introduce similar hermitian generators of rotations about the 1- and 2-axes, J_1 and J_2 , respectively. The commutator algebra of the generators is (see Exercise 2.4)

$$[J_j, J_k] = i\epsilon_{jkl}J_l \quad (2.13)$$

where $\epsilon_{jkl} = +1(-1)$ if jkl are a cyclic (anticyclic) permutation of 1 2 3 and $\epsilon_{jkl} = 0$ otherwise. Relations (2.13) completely define the group properties; the ϵ_{jkl} coefficients are called the structure constants of the group. The J 's are said to form a Lie algebra. Since no two J 's commute with each other, only the eigenvalues of one generator, say J_3 , are useful quantum numbers.

EXERCISE 2.4 Show that the four successive infinitesimal rotations (ϵ about the 1-axis, followed by η about the 2-axis, then $-\epsilon$ about the 1-axis, and finally $-\eta$ about the 2-axis) are equivalent to the second-order rotation $e\eta$ about the 3-axis. Hence, show that the generators satisfy

$$[J_1, J_2] = iJ_3.$$

Nonlinear functions of the generators which commute with all the generators are called invariants or Casimir operators. For the rotation group,

$$J^2 = J_1^2 + J_2^2 + J_3^2 \quad (2.14)$$

is the only Casimir operator,

$$[J^2, J_i] = 0 \quad \text{with } i = 1, 2, 3. \quad (2.15)$$

It follows that we can construct simultaneous eigenstates $|jm\rangle$ of J^2 and one of the generators, say J_3 . Using only (2.13), it is possible to show that

$$\begin{aligned} J^2|jm\rangle &= j(j+1)|jm\rangle \\ J_3|jm\rangle &= m|jm\rangle \end{aligned} \quad (2.16)$$

with $m = -j, -j+1, \dots, j$, and where j can take one of the values $0, \frac{1}{2}, 1, \frac{3}{2}, \dots$.

EXERCISE 2.5 Verify (2.16). To do this, it is useful to form the so-called “step-up” and “step-down” operators

$$J_{\pm} = J_1 \pm iJ_2. \quad (2.17)$$

First, show that

$$J_{\pm}|jm\rangle = (C - m(m \pm 1))^{1/2}|j, m \pm 1\rangle, \quad (2.18)$$

that is, J_{\pm} step m up and down by one unit, respectively. Show that $C = j(j+1)$.

A state $|jm\rangle$ is transformed under a rotation through an angle θ about the 2-axis into a linear combination of the $2j+1$ states $|jm'\rangle$, with $m' = -j, -j+1, \dots, j$:

$$e^{-i\theta J_2}|jm\rangle = \sum_{m'} d_{m'm}^j(\theta)|jm'\rangle, \quad (2.19)$$

where the coefficients $d_{m'm}^j$ are written in conventional notation and are frequently called rotation matrices. From (2.19), we see the states having the same j but all possible m values transform among themselves under rotations. In fact, all the $2j+1$ states are mixed by rotations. They form the basis of a $(2j+1)$ -dimensional irreducible representation of the rotation group. The set of states is called a multiplet.

EXERCISE 2.6 Show that the rotation matrices

$$d_{m'm}^j(\theta) = \langle jm'|e^{-i\theta J_2}|jm\rangle$$

for $j = \frac{1}{2}$ and $j = 1$ are

$$j = \frac{1}{2} \begin{cases} d_{++} = d_{--} = \cos \frac{1}{2}\theta \\ d_{-+} = -d_{+-} = \sin \frac{1}{2}\theta \end{cases} \quad (2.20)$$

where \pm denote $m = \pm \frac{1}{2}$, respectively, and

$$j = 1 \begin{cases} d_{01} = -d_{10} = -d_{0-1} = d_{-10} = \sqrt{\frac{1}{2}} \sin \theta \\ d_{11} = d_{-1-1} = \frac{1}{2}(1 + \cos \theta) \\ d_{-11} = d_{1-1} = \frac{1}{2}(1 - \cos \theta) \\ d_{00} = \cos \theta. \end{cases} \quad (2.21)$$

2.3 The Group $SU(2)$

In the lowest-dimension nontrivial representation of the rotation group ($j = \frac{1}{2}$), the generators may be written

$$J_i = \frac{1}{2}\sigma_i \quad \text{with } i = 1, 2, 3, \quad (2.22)$$

where σ_i are the Pauli matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (2.23)$$

The basis (or set of base states) for this representation is conventionally chosen to be the eigenvectors of σ_3 , that is, the column vectors

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

describing a spin- $\frac{1}{2}$ particle of spin projection up ($m = +\frac{1}{2}$ or \uparrow) and spin projection down ($m = -\frac{1}{2}$ or \downarrow) along the 3-axis, respectively.

The Pauli matrices σ_i are hermitian, and the transformation matrices

$$U(\theta_i) = e^{-i\theta_i \sigma_i / 2} \quad (2.24)$$

are unitary. The set of all unitary 2×2 matrices is known as the group $U(2)$. However, $U(2)$ is larger than the group of matrices $U(\theta_i)$, since the generators σ_i all have zero trace. Now, for any hermitian traceless matrix σ , we can show that

$$\det(e^{i\sigma}) = e^{iTr(\sigma)} = 1. \quad (2.25)$$

Since the unit determinant is preserved in matrix multiplication, the set of traceless unitary 2×2 matrices form a subgroup, $SU(2)$, of $U(2)$. $SU(2)$ denotes the special unitary group in two dimensions. The set of transformation matrices $U(\theta_i)$ therefore form an $SU(2)$ group. The $SU(2)$ algebra is just the algebra of the generators J_i , relations (2.13). There are thus $1, 2, 3, 4, \dots$ dimensional representations of $SU(2)$ corresponding to $j = 0, \frac{1}{2}, 1, \frac{3}{2}, \dots$, respectively. The two-dimensional representation is, of course, just the σ -matrices themselves. It is called the fundamental representation of $SU(2)$, the representation from which all other representations can be built, as we will now show.

EXERCISE 2.7 Show that the rotation of a spin- $\frac{1}{2}$ system through a finite angle θ about the 2-axis corresponds to the unitary transformation

$$e^{-i\theta\sigma_2/2} = \cos\frac{\theta}{2} - i\sigma_2 \sin\frac{\theta}{2}. \quad (2.26)$$

2.4 Combining Representations

A composite system formed from two systems having angular momentum j_A and j_B may be described in terms of the basis

$$|j_A j_B m_A m_B\rangle \equiv |j_A m_A\rangle |j_B m_B\rangle.$$

However, the combined operator

$$\mathbf{J} = \mathbf{J}_A + \mathbf{J}_B \quad (2.27)$$

also satisfies the Lie algebra of (2.13), and it is the eigenvalues $J(J+1)$, M of J^2 , J_3 which are the conserved quantum numbers. In fact, the “product” of the two irreducible representations of dimension $2j_A + 1$ and $2j_B + 1$ may be decomposed into the sum of irreducible representations of dimension $2J + 1$ with

$$J = |j_A - j_B|, |j_A - j_B| + 1, \dots, j_A + j_B, \quad (2.28)$$

with basis $|j_A j_B JM\rangle$, where

$$M = m_A + m_B. \quad (2.29)$$

The last equality follows directly from the third component of (2.27). One basis may be expressed in terms of the other by

$$|j_A j_B JM\rangle = \sum_{m_A, m_B} C(m_A m_B; JM) |j_A j_B m_A m_B\rangle, \quad (2.30)$$

where the coefficients C are called Clebsch–Gordan coefficients; they have been tabulated, for example, in the *Review of Particle Properties* (1982). These coefficients are readily calculated by repeatedly applying the step-down operator [cf. (2.17)]

$$J_- = (J_A)_- + (J_B)_-$$

to the “fully stretched” state

$$|j_A j_B J, M = J\rangle = |j_A j_B, m_A = j_A, m_B = j_B\rangle$$

and using orthogonality when necessary.

Equation (2.1) is a simple example of (2.30). The composite system of two spin- $\frac{1}{2}$ particles $j_A = j_B = \frac{1}{2}$ may have spin $J = 1$ or 0. We may write the decomposition symbolically as

$$2 \otimes 2 = 3 \oplus 1, \quad (2.31)$$

using the dimensions (that is, the size of the multiplet) to label the irreducible representations. We can readily extend this procedure. Combining a third spin- $\frac{1}{2}$

particle, we have

$$\begin{aligned}(2 \otimes 2) \otimes 2 &= (3 \otimes 2) \oplus (1 \otimes 2) \\ &= 4 \oplus 2 \oplus 2.\end{aligned}\quad (2.32)$$

That is, three spin- $\frac{1}{2}$ particles group together into a quartet of spin $\frac{3}{2}$ and two doublets of spin $\frac{1}{2}$.

2.5 Finite Symmetry Groups: *P* and *C*

A finite group is one which contains only a finite number of elements. In particle physics, we encounter a very simple symmetry group containing just two elements, the identity e and an element g satisfying $g^2 = e$. For example, g may be space inversion or particle–antiparticle conjugation. Invariance of the physics under g means that g is represented by a unitary (or antiunitary) operator $U(g)$ which satisfies [see (2.5)]

$$[U, H] = 0, \quad (2.33)$$

Time-reversal invariance is the only symmetry requiring an antiunitary operator [see, for example, Schiff (1968), Martin and Spearman (1970), and Messiah (1962)], and so here we take U to be unitary. For our two-element group, we have

$$U^2 = 1, \quad (2.34)$$

and, since U is unitary, it must also be hermitian. Thus, U itself is an observable conserved quantity [cf. (2.7)], and its eigenvalues are conserved quantum numbers. If p is an eigenvalue of U corresponding to eigenvector $|p\rangle$, then

$$U^2|p\rangle = p^2|p\rangle. \quad (2.35)$$

From (2.34), $p^2 = 1$, and so the allowed eigenvalues are $p = \pm 1$. Invariance of the system under the symmetry operation g (for example, space inversion or particle–antiparticle conjugation) means that if the system is in an eigenstate of U (with $U = P$ or C), then transitions can only occur to eigenstates with the same eigenvalue. We see that the eigenvalues of U are *multiplicative* quantum numbers. By contrast, the eigenvalues of the commuting generators of $SU(n)$ are *additive* quantum numbers.

Strong and electromagnetic interactions are invariant under both P and C , whereas weak interactions do not respect these symmetries. However, to a good approximation, weak interactions are invariant under the product transformation CP (see Chapter 12).

2.6 *SU(2) of Isospin*

Isospin arises because the nucleon may be viewed as having an internal degree of freedom with two allowed states, the proton and the neutron, which the nuclear interaction does not distinguish. We therefore have an $SU(2)$ symmetry in which

the (n, p) form the fundamental representation. It is a mathematical copy of spin in that the isospin generators satisfy

$$[I_j, I_k] = i\epsilon_{jkl}I_l \quad (2.36)$$

[cf. (2.13)]. In the fundamental representation, the generators are denoted $I_i \equiv \frac{1}{2}\tau_i$, where

$$\tau_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \tau_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \tau_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (2.37)$$

are the isospin version of the Pauli matrices (2.23). They act on the proton and neutron states represented by

$$p = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad n = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

In general, the most positively charged particle is chosen to have the maximum value of I_3 .

2.7 Isospin for Antiparticles

The construction of antiparticle isospin multiplets requires care. It is well illustrated by a simple example. Consider a particular isospin transformation of the nucleon doublet, a rotation through π about the 2-axis. We obtain [see (2.26)]

$$\begin{pmatrix} p' \\ n' \end{pmatrix} = e^{-i\pi(\tau_2/2)} \begin{pmatrix} p \\ n \end{pmatrix} = -i\tau_2 \begin{pmatrix} p \\ n \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} p \\ n \end{pmatrix}. \quad (2.38)$$

We define antinucleon states using the particle–antiparticle conjugation operator C ,

$$Cp = \bar{p}, \quad Cn = \bar{n}. \quad (2.39)$$

Applying C to (2.38) therefore gives

$$\begin{pmatrix} \bar{p}' \\ \bar{n}' \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \bar{p} \\ \bar{n} \end{pmatrix}. \quad (2.40)$$

However, we want the antiparticle doublet to transform in exactly the same way as the particle doublet, so that we can combine particle and antiparticle states using the same Clebsch–Gordan coefficients, and so on. We must therefore make two changes. First, we must reorder the doublet so that the most positively charged particle has $I_3 = +\frac{1}{2}$, and then we must introduce a minus sign to keep the matrix transformation identical to (2.38). We obtain

$$\begin{pmatrix} -\bar{n}' \\ \bar{p}' \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} -\bar{n} \\ \bar{p} \end{pmatrix}. \quad (2.41)$$

That is, the antiparticle doublet $(-\bar{n}, \bar{p})$ transforms exactly as the particle doublet (p, n) . This is a special property of $SU(2)$; it is not possible, for example, to arrange an $SU(3)$ triplet of antiparticles so that it transforms as the particle triplet.

A composite system of a nucleon–antinucleon pair has isospin states [compare (2.2)]:

$$\begin{cases} |I = 1, I_3 = 1\rangle = -p\bar{n} \\ |I = 1, I_3 = 0\rangle = \sqrt{\frac{1}{2}}(p\bar{p} - n\bar{n}) \\ |I = 1, I_3 = -1\rangle = n\bar{p} \end{cases} \quad (2.42)$$

$$|I = 0, I_3 = 0\rangle = \sqrt{\frac{1}{2}}(p\bar{p} + n\bar{n}).$$

2.8 The Group $SU(3)$

The set of unitary 3×3 matrices with $\det U = 1$ form the group $SU(3)$. The generators may be taken to be any $3^2 - 1 = 8$ linearly independent traceless hermitian 3×3 matrices. Since it is possible to have only two of these traceless matrices diagonal, this is the maximum number of mutually commuting generators. This number is called the rank of the group, so that $SU(3)$ has rank 2 and $SU(2)$ has rank 1. It can be shown that the number of Casimir operators is equal to the rank of the group.

The fundamental representation of $SU(3)$ is a triplet. The three color charges of a quark, R, G, and B of Section 1.2, form the fundamental representation of an $SU(3)$ symmetry group. In this representation, the generators are 3×3 matrices. They are traditionally denoted λ_i , with $i = 1, \dots, 8$, and the diagonal matrices are taken to be

$$\lambda_3 = \begin{pmatrix} 1 & & \\ & -1 & \\ & & 0 \end{pmatrix}, \quad \lambda_8 = \sqrt{\frac{1}{3}} \begin{pmatrix} 1 & & \\ & 1 & \\ & & -2 \end{pmatrix} \quad (2.43)$$

with simultaneous eigenvectors

$$R = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad G = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

These base states R, G, B are plotted in Fig. 2.3 in terms of their λ_3, λ_8 eigenvalues. The figure also shows how the remaining six generators give the analogues of the “step-up” and “step-down” operators of $SU(2)$. With this numbering of the λ_i matrices, $\lambda_1, \lambda_2, \lambda_3$ correspond to the three Pauli matrices and thus they exhibit explicitly one $SU(2)$ subgroup of $SU(3)$. The λ_i are known as the Gell-Mann matrices.

EXERCISE 2.8 Obtain the matrix representations of the λ_i of Fig. 2.3.
Show that

$$\left[\frac{\lambda_i}{2}, \frac{\lambda_j}{2} \right] = i \sum_k f_{ijk} \frac{\lambda_k}{2} \quad (2.44)$$

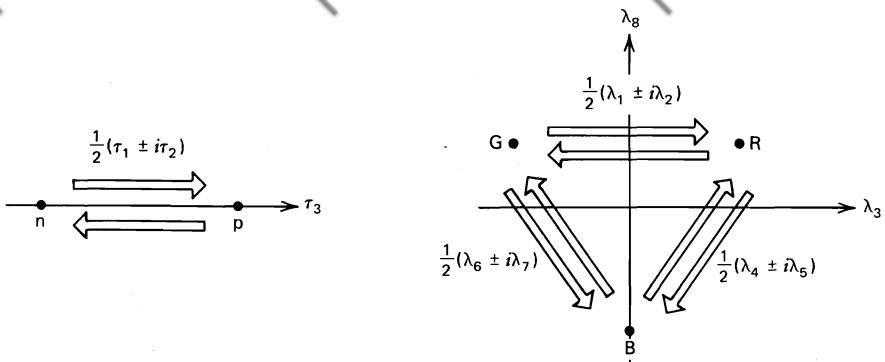


Fig. 2.3 The action of the generators (τ_i and λ_i) on fundamental representations of $SU(2)$ of isospin and $SU(3)$ of color, respectively.

where the $SU(3)$ structure constants f_{ijk} are fully antisymmetric under interchange of any pair of indices (see exercise 14.9) and the nonvanishing values are permutations of

$$f_{123} = 1, \quad f_{458} = f_{678} = \sqrt{3}/2,$$

$$f_{147} = f_{165} = f_{246} = f_{257} = f_{345} = f_{376} = \frac{1}{2}.$$

2.9 Another Example of an $SU(3)$ Group: Isospin and Strangeness

In 1947, the pion was discovered, and from that date on, the nucleon lost its unique role in particle physics. Subsequently, many more strongly interacting particles (hadrons) have been identified. Some of the new particles were surprisingly long-lived on the time scale of strong interactions, despite being massive enough to decay into lighter objects without violating the conservation of charge or baryon number. For instance, a Σ^- is readily produced by the strong interaction $\pi^- p \rightarrow K^+ \Sigma^-$ and yet decays only weakly via $\Sigma^- \rightarrow n \pi^-$. The contrast to a typical strong decay, $\Delta \rightarrow n \pi$, was emphasized in Section 1.6. Gell-Mann and, independently, Nishijima, took this as a manifestation of a new additive quantum number, which was called “strangeness,” S . They assigned to each hadron an integer value of strangeness,

$$\begin{aligned} S = 0: & \quad \pi, N, \Delta, \dots, \\ S = 1: & \quad K^+, \dots, \\ S = -1: & \quad \Lambda, \Sigma, \dots. \end{aligned} \tag{2.45}$$

with $-S$ for their antiparticles, and they asserted that strong and electromagnetic interactions are forbidden unless S is conserved by the reaction. Gell-Mann and Nishijima’s proposal immediately accounts for the strong production and the

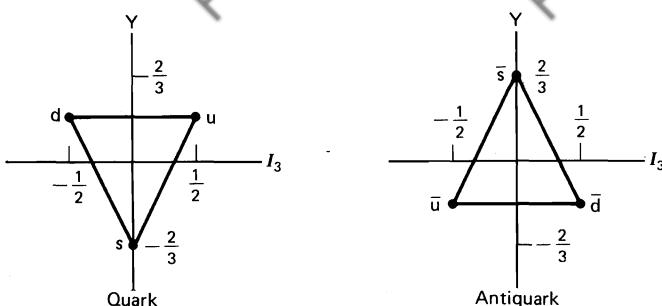
weak decay of the Σ . Indeed, in the reaction $\pi^- p \rightarrow K^+ \Sigma^-$ both the initial and final states have a total strangeness $S = 0$ [see (2.45)]. The Σ -particle can therefore be produced by the strong interaction. It could also decay via the strong interaction $\Sigma^- \rightarrow \Lambda \pi^-$ if it were not for the fact that the Λ is too heavy so that the strangeness conserving decay is kinematically forbidden. The Σ^- can only decay by the strangeness-violating weak interaction $\Sigma^- \rightarrow n \pi^-$, thus “explaining” its long lifetime (see Section 1.6). The Gell-Mann and Nishijima scheme was confirmed by observations of the properties of the large number of strange particles that were subsequently discovered.

With the existence of a second additive quantum number S , in addition to I_3 , it was natural to attempt to enlarge isospin symmetry to a larger group, namely, a group of rank 2. This new symmetry group had to naturally fit the hadrons with similar properties into its multiplet representations. This task was relatively easy for the $SU(2)$ group of isospin: the neutron and proton, which are almost identical in mass, are nicely accommodated in an $SU(2)$ doublet. However, no strange particles exist that are close in mass to the nucleon, so the appropriate grouping is difficult to identify and the choice of group far from obvious. $SU(3)$ was originally proposed in 1961: we shall see (Fig. 2.8) that it groups the $n, p, \Sigma^+, \Sigma^0, \Sigma^-, \Lambda, \Xi^0$, and Ξ^- , with a mass spread of nearly 400 MeV, into an $SU(3)$ octet representation. Moreover, the lightest mesons are also fitted into an octet, with the K -meson belonging to the same representation as the much lighter π ($m_K > 3m_\pi$!). It is clear that the extra symmetry linking strange and non-strange particles is much more approximate than is isospin. Not until 1964 was $SU(3)$ symmetry firmly established. The $SU(3)$ multiplet structure of the so-called “elementary” particles was reminiscent of the grouping of chemical elements in Mendeleev’s table. Like the periodic table, the $SU(3)$ classification strongly hinted at the existence of a substructure. The role of the $SU(3)$ group of isospin and strangeness is mostly historical: it set the scene for the entry of quarks into particle physics.

With hindsight, we now realize that the success of $SU(2)$ isospin symmetry is due to the essentially equal mass of the u, d quark constituents. However, $SU(3)$ incorporating the heavier s quark is not such a good symmetry; rather, we use it to enumerate the hadronic states. We call this “flavor $SU(3)$,” u, d, s being the three lightest flavors of quark. It is completely unrelated to “color $SU(3)$,” which is believed to be an exact symmetry of fundamental origin (see Chapter 14).

QUARK “ATOMS”

According to the quark model, all hadrons are made up of a small variety of more basic entities, called quarks, bound together in different ways. The fundamental representation of $SU(3)$, the multiplet from which all other multiplets can be built, is a triplet. This basic quark multiplet is given in Fig. 2.4; also shown is the antiquark multiplet in which the signs of the additive quantum numbers are

Fig. 2.4 $SU(3)$ quark and antiquark multiplets; $Y \equiv B + S$.

reversed. Each quark is assigned spin $\frac{1}{2}$ and baryon number $B = \frac{1}{3}$. Baryons are made of three quarks (qqq) and the mesons of a quark–antiquark pair ($q\bar{q}$). In Fig. 2.4, the new additive quantum number is shown as the “hypercharge,”

$$Y \equiv B + S, \quad (2.46)$$

rather than the strangeness S . This choice has no physical significance; it simply centers the multiplets on the origin. The charge, Qe , is

$$Q = I_3 + \frac{Y}{2}.$$

The quantum numbers of the quarks are listed in Table 2.1. Baryon conservation means it is impossible to destroy or to make a single quark, but we can annihilate or create a quark–antiquark pair (a meson). Moreover, quarks retain their identity under strong or electromagnetic transitions; that is, transmutations such as $s \rightarrow u + \text{leptons}$, $s \rightarrow u + d\bar{u}$ occur only by weak interactions.

2.10 Quark–Antiquark States: Mesons

In the quark model, mesons are made of a quark and an antiquark bound together. Let us start with two flavors, $q = u$ or d . The $q\bar{q}$ bound-state wave functions are readily obtained by making the substitutions $p \rightarrow u$ and $n \rightarrow d$ in

TABLE 2.1
Quantum Numbers of the Quarks ($Y = B + S$, $Q = I_3 + Y/2$)^a

Quark	Spin	B	Q	I_3	S	Y
u	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{1}{2}$	0	$\frac{1}{3}$
d	$\frac{1}{2}$	$\frac{1}{3}$	$-\frac{1}{3}$	$-\frac{1}{2}$	0	$\frac{1}{3}$
s	$\frac{1}{2}$	$\frac{1}{3}$	$-\frac{1}{3}$	0	-1	$-\frac{2}{3}$

^aHere, S denotes the strangeness.

(2.42). We thus obtain an isotriplet and an isosinglet of mesons

$$\begin{cases} |I=1, I_3=1\rangle = -u\bar{d} \\ |I=1, I_3=0\rangle = \sqrt{\frac{1}{2}}(u\bar{u} - d\bar{d}) \\ |I=1, I_3=-1\rangle = d\bar{u}, \\ |I=0, I_3=0\rangle = \sqrt{\frac{1}{2}}(u\bar{u} + d\bar{d}). \end{cases} \quad (2.47)$$

For three flavors of quarks, $q = u, d$, or s , there are nine possible $q\bar{q}$ combinations. The resulting multiplet structure is shown in Fig. 2.5b and is readily obtained by superimposing the center of gravity of the antiquark multiplet on top of every site of the quark multiplet, Fig. 2.5a. The nine states divide into an $SU(3)$ octet and an $SU(3)$ singlet; that is, under operations of the $SU(3)$ group, the eight states transform among themselves but do not mix with the singlet state. This is the extension to $SU(3)$ of the familiar separation (2.31) or (2.47) of $SU(2)$.

Of the nine $q\bar{q}$ states, we note that three, labeled A, B, and C on Fig. 2.5, have $I_3 = Y = 0$. These are linear combinations of $u\bar{u}$, $d\bar{d}$, and $s\bar{s}$ states. The singlet combination, C, must contain each quark flavor on an equal footing; and so, after normalization, we have

$$C = \sqrt{\frac{1}{3}}(u\bar{u} + d\bar{d} + s\bar{s}). \quad (2.48)$$

State A is taken to be a member of the isospin triplet ($d\bar{u}$, A, $-u\bar{d}$) and so

$$A = \sqrt{\frac{1}{2}}(u\bar{u} - d\bar{d}), \quad (2.49)$$

see (2.47). By requiring orthogonality to both A and C , the isospin singlet state B is found to be

$$\mathbf{B} = \sqrt{\frac{1}{6}} (\mathbf{u}\bar{\mathbf{u}} + \mathbf{d}\bar{\mathbf{d}} - 2\mathbf{s}\bar{\mathbf{s}}). \quad (2.50)$$

Like any quantum-mechanical bound system, the $q\bar{q}$ pair will have a discrete energy level spectrum corresponding to the different modes of $q\bar{q}$ excitations,

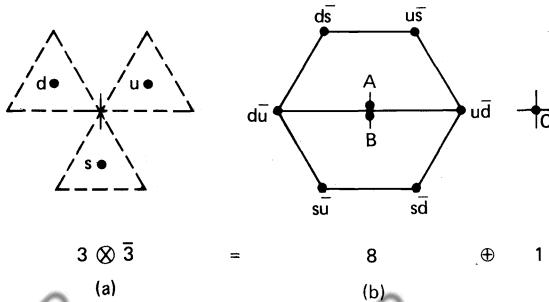


Fig. 2.5 The quark content of the meson nonet, showing the $SU(3)$ decomposition in the I_3 , Y plane.

rotations, vibrations, and so on. These must correspond to the observed meson states. Even in the absence of knowledge about the potential which binds the quark to the antiquark, the model is very predictive. Recall that the quark has spin $\frac{1}{2}$, and so the total intrinsic spin of the $q\bar{q}$ pair can be either $S = 0$ or 1 . The spin J of the composite meson is the vector sum of this spin S and the relative orbital angular momentum L of the q and \bar{q} . Moreover, the parity of the meson is

$$P = -(-1)^L \quad (2.51)$$

where the minus sign arises because the q and \bar{q} have opposite intrinsic parity [cf. (5.63)], and $(-1)^L$ arises from the space inversion replacements $\theta \rightarrow \pi - \theta$, $\phi \rightarrow \phi + \pi$ in the angular part of the $q\bar{q}$ wavefunction $Y_{LM}(\theta, \phi)$. A neutral $q\bar{q}$ system is an eigenstate of the particle–antiparticle conjugation operator C . The value of C can be deduced by $q \leftrightarrow \bar{q}$ and then interchanging their positions and spins. The combined operation gives

$$C = -(-1)^{S+1}(-1)^L = (-1)^{L+S} \quad (2.52)$$

where the minus sign arises from interchanging fermions, the $(-1)^{S+1}$ from the symmetry of the $q\bar{q}$ spin states [see (2.1)], and the $(-1)^L$ is as before. Here S is the total intrinsic spin of the $q\bar{q}$ pair.

The allowed sets of quantum numbers for the ground ($L = 0$) and first ($L = 1$) excited states are shown in Table 2.2, together with the observed candidate meson states. In each nonet, there are two isospin doublets (see Fig. 2.5). For example, in the $J^P = 0^-$ nonet, we have

$$\begin{array}{lll} K^0(d\bar{s}), & K^+(u\bar{s}) & \text{with } Y = 1 \\ K^-(s\bar{u}), & \bar{K}^0(s\bar{d}) & \text{with } Y = -1. \end{array} \quad (2.53)$$

These pseudoscalar mesons form an octet along with the $Y = 0$ isotriplet (the π^+, π^0, π^- states) and the $I = 0$ state (the η meson). The $SU(3)$ singlet state is identified with η' meson, see Table 2.2.

TABLE 2.2

Quantum Numbers of Observed Mesons Composed of u, d, and s Quarks

		Observed Nonet			Typical Mass (MeV)	
Orbital Ang. Mom.	$q\bar{q}$ Spin	J^{PC}	$I = 1$	$I = \frac{1}{2}$	$I = 0$	
$L = 0$	$S = 0$	0^{-+}	π	K	η, η'	500
	$S = 1$	1^{--}	ρ	K^*	ω, ϕ	800
$L = 1$	$S = 0$	1^{+-}	B	Q_2	$H, ?$	1250
	$S = 1$	2^{++}	A_2	K^*	f, f'	1400
		$\begin{cases} 1^{++} \\ 0^{++} \end{cases}$	A_1	Q_1	$D, ?$	1300
			δ	κ	ϵ, S^*	1150

In the 1^- and 2^+ nonets, the two neutral octet and singlet $I = 0$ states, (2.50) and (2.48), are found to mix with one another so that to a good approximation the physical particles are those made entirely from strange and nonstrange quarks. For example, in the vector (1^-) nonet,

$$\phi \approx s\bar{s}, \quad \omega \approx \sqrt{\frac{1}{2}}(u\bar{u} + d\bar{d}). \quad (2.54)$$

Note that, in general, the physical neutral $q\bar{q}$ states correspond to orthogonal quantum-mechanical superpositions of the singlet and the $I = 0$ neutral octet states, which have indeed identical quantum numbers. For example, although we said above that the observed η is an octet state, there is in fact a small singlet admixture.

With reference to Table 2.2, we should add that we would not have expected L and the spin S to be good quantum numbers. However, parity conservation forbids the mixing of even and odd L states, and then C conservation requires the spin S to be unique. This leaves the possibility of mixing only for $S = 1$ states for which L differs by two units.

The $L = 1$ states of Table 2.2 are examples of orbital excitations. The typical excitation energy is about 600 MeV. Just as in positronium, we would also expect radial excitations; for example, a repeat of the $L = 0$ nonets at a higher mass.

The random names attributed to the observed states of Table 2.2 are relics of the past, when experimenters had the difficult task of identifying mesons and determining their quantum numbers. However, there is no doubt that the success of the quark model predictions is impressive; all established mesons lie within the expected $q\bar{q}$ multiplets. Before about 1971, when the first good data with high enough resolution to provide direct evidence for quarks became available, tests of this type were the principal basis for accepting the quark hypothesis.

EXERCISE 2.9 Make use of (2.54) and (2.53) to predict the decay modes and branching ratios of the ϕ -meson (mass 1020 MeV). Comment on the width of the resonance.

EXERCISE 2.10 Explain why a particularly good way of identifying mesons coupled to the $\pi\pi$ -channel is to study the reaction $\pi N \rightarrow (\pi\pi)N$ at high energies. Show that $I + J$ must be an even integer for these mesons.

In passing, we should note that particles that decay by strong interactions do not live long enough to leave tracks in an experimentalist's detector. Rather, they are identified by tracking their decay products. The mass of the decaying particle is the total energy of these products as measured in its rest frame. Due to its short lifetime, the uncertainty in its mass ($\sim \hbar/\Delta t$) is sufficiently large to be directly observable. For example, the Δ is formed and rapidly decays in πN scattering, $\pi N \rightarrow \Delta \rightarrow \pi N$. Such an unstable particle decays according to the exponential

law

$$|\psi(t)|^2 = |\psi(0)|^2 e^{-\Gamma t}, \quad (2.55)$$

where $\tau \equiv 1/\Gamma$ is called the lifetime of the state. Thus, the time dependence of $\psi(t)$ for an unstable state must include the decay factor $\Gamma/2$; that is,

$$\psi(t) \sim e^{-iMt} e^{-\Gamma t/2}$$

where M is the rest mass energy of the state. As a function of the center-of-mass energy E of the πN system, the state is described by the Fourier transform

$$\begin{aligned} \chi(E) &= \int \psi(t) e^{iEt} dt \\ &\sim \frac{1}{E - M + (i\Gamma/2)}. \end{aligned} \quad (2.56)$$

The experimenter thus sees a πN reaction rate of the form

$$|\chi(E)|^2 = \frac{A}{(E - M)^2 + (\Gamma/2)^2}. \quad (2.57)$$

This function has a sharp peak centered at M with a width determined by Γ . Equation (2.57) is called a Breit–Wigner resonance form, and M and Γ are known as the mass and width of the resonance, respectively. In a detailed resonance analysis, the form (2.55) will include kinematic factors. For instance, resonance production and decay near threshold are inhibited by phase space; its observed width is suppressed by kinematic factors.

2.11 Three-Quark States: Baryons

The flavor $SU(3)$ decomposition of the 27 possible qqq combinations is more involved than that for mesons; nevertheless, the quark content of baryons can be readily obtained using the same techniques. We first combine two of the quarks. Figure 2.6 shows that the nine qq combinations arrange themselves into two $SU(3)$ multiplets,

$$3 \otimes 3 = 6 \oplus \bar{3}, \quad (2.58)$$

where the 6 is symmetric and the $\bar{3}$ is antisymmetric under interchange of the two

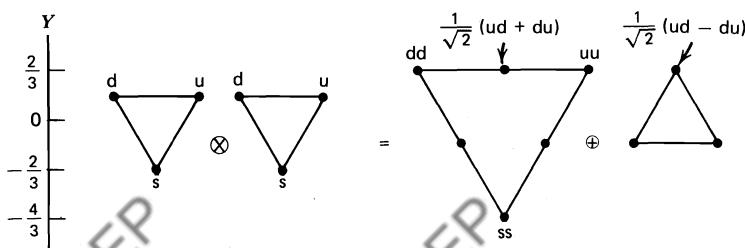


Fig. 2.6 The qq $SU(3)$ multiplets; $3 \otimes 3 = 6 \oplus \bar{3}$.

quarks. The quark content of the nonstrange sector is found by combining the two (d, u) I -spin doublets. It is explicitly shown on the figure and simply repeats eqs. (2.2), with the substitution $n \rightarrow d$, $p \rightarrow u$. The quark content of the other states is found in exactly the same way. For example, we combine two (s, d) doublets to obtain the states in the s, d sector. The state on the line linking the dd and ss states in Fig. 2.6 is $(ds + sd)/\sqrt{2}$. We talk of combining U -spin doublets, and in the u, s sector we speak of V -spin. The mathematics of I , U , and V spin are identical, all based on the $SU(2)$ group, which also underlies the description of ordinary spin. In fact, we shall see that almost all the $SU(3)$ structure that we require can be obtained simply by successive application of $SU(2)$.

We are now ready to add the third quark triplet. The final decomposition,

$$\begin{aligned} 3 \otimes 3 \otimes 3 &= (6 \otimes 3) \oplus (\bar{3} \otimes 3) \\ &= 10 \oplus 8 \oplus 8 \oplus 1, \end{aligned} \quad (2.59)$$

is displayed in Fig. 2.7. As an example, we form the three “uud” combinations which are denoted Δ , p_S , and p_A on the figure. Combining the nonstrange member of the $\bar{3}$ (see Fig. 2.6) with the u quark of the 3, we have immediately

$$p_A = \sqrt{\frac{1}{2}} (ud - du)u. \quad (2.60)$$

The decuplet states are totally symmetric under interchange of quarks, as evidenced by the uuu, ddd, and sss members. The symmetric combination of “uud” is

$$\Delta = \sqrt{\frac{1}{3}} [uud + (ud + du)u]. \quad (2.61)$$

Requiring orthogonality of the remaining “uud” state to both p_A and Δ gives

$$p_S = \sqrt{\frac{1}{6}} [(ud + du)u - 2uud]. \quad (2.62)$$

The states p_S and p_A have mixed symmetry; however, the subscripts are to remind us that they have symmetry and antisymmetry, respectively, under interchange of the first two quarks. The quark structure of the other states can be readily obtained in a similar way (by application of either U or V spin).

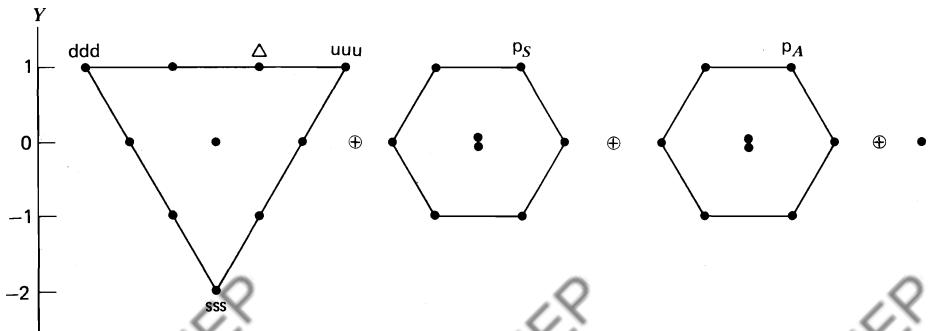


Fig. 2.7 The qqq $SU(3)$ multiplets; $3 \otimes 3 \otimes 3 = 10 \oplus 8 \oplus 8 \oplus 1$.

EXERCISE 2.11 Write down the quark composition of the three “dds” states.

EXERCISE 2.12 Determine the structure of the six “uds” states. In particular, show that the $SU(3)$ singlet state is the completely antisymmetric combination

$$(qqq)_{\text{singlet}} = \sqrt{\frac{1}{6}} (\text{uds} - \text{usd} + \text{sud} - \text{sdu} + \text{dsu} - \text{dus}). \quad (2.63)$$

In the ground state, the baryon spin is found simply by the addition of three spin- $\frac{1}{2}$ angular momenta. Writing the decomposition in terms of the multiplicities of the spin states, we found [see (2.32)]

$$2 \otimes 2 \otimes 2 = (3 \oplus 1) \otimes 2 = 4 \oplus 2 \oplus 2, \quad (2.64)$$

$$\begin{array}{ccc} S & A & \\ \text{---} & \text{---} & \end{array} \quad \begin{array}{ccc} S & M_S & M_A \\ \text{---} & \text{---} & \end{array}$$

or, in other words, baryon spin multiplets with $S = \frac{3}{2}, \frac{1}{2}, \frac{1}{2}$. The subscripts on the mixed symmetry doublets (M_S, M_A) indicate that the spin states are symmetric or antisymmetric under interchange of the first two quarks. The four $S = \frac{3}{2}$ spin states are totally symmetric.

Note that in deriving eqs. (2.60)–(2.62), we were working in the $SU(2)$ isospin sector of $SU(3)$. We can therefore apply these results directly to $SU(2)$ spin if we make the replacements $u \rightarrow \uparrow$ and $d \rightarrow \downarrow$. Using this analogy, we immediately obtain the composition of the spin “up” state belonging to each of the three spin multiplets

$$\begin{aligned} \chi(S) &= \sqrt{\frac{1}{3}} (\uparrow\uparrow\downarrow + \uparrow\downarrow\uparrow + \downarrow\uparrow\uparrow) \\ \chi(M_S) &= \sqrt{\frac{1}{6}} (\uparrow\downarrow\uparrow + \downarrow\uparrow\uparrow - 2\uparrow\uparrow\downarrow) \\ \chi(M_A) &= \sqrt{\frac{1}{2}} (\uparrow\downarrow\uparrow - \downarrow\uparrow\uparrow). \end{aligned} \quad (2.65)$$

To enumerate the baryons expected in the quark model, we must combine the $SU(3)$ flavor decomposition of (2.59) with the $SU(2)$ spin decomposition of (2.64),

$$(10 + 8 + 8 + 1), \quad (4 + 2 + 2) \quad (2.66)$$

$$\begin{array}{cccccc} S & M_S & M_A & A & S & M_S & M_A. \end{array}$$

Considering the product symmetries, we are led to assign the ($SU(3)$, $SU(2)$) multiplets to the following categories:

$$S: (10, 4) + (8, 2)$$

$$M_S: (10, 2) + (8, 4) + (8, 2) + (1, 2)$$

$$M_A: (10, 2) + (8, 4) + (8, 2) + (1, 2) \quad (2.67)$$

$$A: (1, 4) + (8, 2)$$

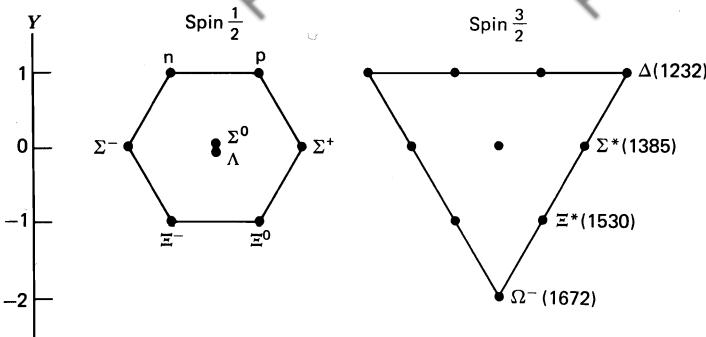


Fig. 2.8 Ground-state baryons: $(8, 2) + (10, 4)$.

where, for example, the totally symmetric (S) octet arises from the combination

$$\sqrt{\frac{1}{2}} [(8, 2) + (8, 2)]. \quad (2.68)$$

$M_S, M_S \quad M_A, M_A$

The lowest-mass baryons fit neatly into the symmetric spin- $\frac{3}{2}$ decuplet $(10, 4)$ and the spin- $\frac{1}{2}$ octet $(8, 2)$ (see Fig. 2.8).

This symmetry of the ground state poses a problem, however. For example, a Δ^{++} of $J_3 = \frac{3}{2}$ is described by the symmetric wave function

$$u \uparrow u \uparrow u \uparrow, \quad (2.69)$$

whereas we expect antisymmetry under the exchange of identical fermion quarks. As noted in Chapter 1, the explanation is that the quarks possess an additional attribute, called color, which can take three possible values, R, G, or B. The quarks form a fundamental triplet of an $SU(3)$ color symmetry which, unlike $SU(3)$ flavor, is believed to be exact. All hadrons are postulated to be colorless; that is, they belong to singlet representations of the $SU(3)$ color group. The color wavefunction for a baryon is therefore [compare (2.63)]

$$(qqq)_{\text{col. singlet}} = \sqrt{\frac{1}{6}} (\text{RGB} - \text{RBG} + \text{BRG} - \text{BGR} + \text{GBR} - \text{GRB}). \quad (2.70)$$

The required antisymmetric character of the total wavefunction is achieved; it is overall symmetric in space, spin, and flavor structure and antisymmetric in color. As the color structure of (2.70) is common to all baryons, we suppress it from now on, but remember to select only overall symmetric representations of space \times spin \times flavor.

A relevant example of an explicit quark model wavefunction is that for a spin-up proton. From (2.68),

$$|p \uparrow\rangle = \sqrt{\frac{1}{2}} (p_S \chi(M_S) + p_A \chi(M_A)),$$

where the flavor and spin components are given by (2.60), (2.62), and (2.65). Thus (omitting an irrelevant overall minus sign), we have

$$\begin{aligned} |p\uparrow\rangle &= \sqrt{\frac{1}{18}} [uud(\uparrow\downarrow\uparrow + \downarrow\uparrow\uparrow - 2\uparrow\uparrow\downarrow) + udu(\uparrow\uparrow\downarrow + \downarrow\uparrow\uparrow - 2\uparrow\downarrow\uparrow) \\ &\quad + duu(\uparrow\downarrow\uparrow + \uparrow\uparrow\downarrow - 2\downarrow\uparrow\uparrow)] \\ &= \sqrt{\frac{1}{18}} [u\uparrow u\downarrow d\uparrow + u\downarrow u\uparrow d\uparrow - 2u\uparrow u\uparrow d\downarrow + \text{permutations}]. \end{aligned} \quad (2.71)$$

EXERCISE 2.13 Construct the quark model wavefunctions $|p\downarrow\rangle$, $|n\uparrow\rangle$, and $|n\downarrow\rangle$. The charge operator is defined as $Q = \sum Q_i$, where Q_i are the charges of the quarks in units of the proton charge e . The sum is over the constituent quarks of the hadron. Show that

$$\langle p\uparrow|Q|p\uparrow\rangle = \langle p\downarrow|Q|p\downarrow\rangle = 1$$

$$\langle n\uparrow|Q|n\uparrow\rangle = \langle n\downarrow|Q|n\downarrow\rangle = 0.$$

EXERCISE 2.14 Express the π^+ -wavefunction in terms of the spin, flavor, and color of the component quarks.

EXERCISE 2.15 Convince yourself that the photon is a U -spin scalar; that is, $U = 0$. By inspection of Fig. 2.8, show that if $SU(3)$ flavor symmetry were exact, the electromagnetic decay $\Sigma^*(1385)^-\rightarrow\Sigma^-\gamma$ is forbidden, whereas $\Sigma^*(1385)^+\rightarrow\Sigma^+\gamma$ is allowed.

The three quarks have zero orbital angular momentum in ground-state baryons. That is, $l = l' = 0$ in Fig. 2.9, and so the parity of the state, $(-1)^{l+l'}$, is therefore positive. The first excited state has either $l = 1$, $l' = 0$, or $l = 0$, $l' = 1$; in fact, it is a combination of the two which, when combined with the mixed symmetry multiplets of (2.67), gives a totally symmetric space, spin, and flavor wavefunction. That is, the first excited state is predicted to contain $(1 + 8 + 10)$ flavor multiplets of $S = \frac{1}{2}$ baryons and an octet of $S = \frac{3}{2}$ baryons [see (2.67)]. These spins combine with $L = 1$ to give

Multiplets 1, 8, 10 with $J^P = \frac{1}{2}^-$ and $J^P = \frac{3}{2}^-$

Three octets with $J^P = \frac{1}{2}^-, \frac{3}{2}^-, \frac{5}{2}^-$.

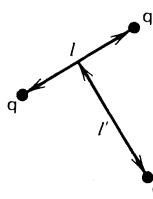


Fig. 2.9 Orbital angular momenta, l and l' , of qq and $(qq)q$ systems, respectively.

Impressive agreement with the observed baryons with masses around 1600 MeV is again found.

EXERCISE 2.16 The $Y = 1$ baryons are most easily identified as resonances in πN elastic scattering. Show that the relative orbital angular momentum, L' , between the π and the N is a good quantum number, and that it is even for resonances of negative parity. Use the quark model to list the isospin, spin, and L' of the πN states expected in the first excited level. Identify these resonances in the particle data tables.

2.12 Magnetic Moments

The calculation of the charge in Exercise 2.13 can be repeated for the magnetic moments of the hadrons. The magnetic moment operator is $\sum_i \mu_i (\sigma_3)_i$, where the summation is again over the constituent quarks. According to the accepted convention, the operator is evaluated between states with $M_J = +J$. Now the magnetic moment of a point-like spin- $\frac{1}{2}$ particle of charge e is $e/2m$ (see Chapter 5). Thus, a structureless quark of charge $Q_i e$ and mass m_i has magnetic moment

$$\mu_i = Q_i \left(\frac{e}{2m_i} \right) \quad (2.72)$$

Hence, in the nonrelativistic approximation, we may write the magnetic moment of the proton as

$$\mu_p = \sum_{i=1}^3 \langle p \uparrow | \mu_i (\sigma_3)_i | p \uparrow \rangle.$$

Using the explicit wavefunction (2.71), we obtain

$$\mu_p = \frac{1}{18} \{ (\mu_u - \mu_u + \mu_d) + (-\mu_u + \mu_u + \mu_d) + 4(2\mu_u - \mu_d) \} \times 3,$$

where the factor 3 takes care of the “permutations.” The magnetic moment of the proton in terms of the component moments is therefore

$$\mu_p = \frac{1}{3} (4\mu_u - \mu_d). \quad (2.73)$$

The neutron magnetic moment is obtained by the interchange of u and d,

$$\mu_n = \frac{1}{3} (4\mu_d - \mu_u).$$

In the limit that $m_u = m_d$, we have, from (2.72),

$$\mu_u = -2\mu_d \quad (2.74)$$

and so the quark model prediction is

$$\frac{\mu_n}{\mu_p} = -\frac{2}{3}. \quad (2.75)$$

This agrees quite well with experiment:

$$\frac{\mu_n}{\mu_p} = -0.68497945 \pm 0.00000058.$$

EXERCISE 2.17 Determine the magnetic moments of the other members of the $J^P = \frac{1}{2}^+$ baryon octet in terms of μ_p and compare with the measured values.

EXERCISE 2.18 The spin-flavor wavefunctions of the ground-state baryons are symmetric, and color was invoked to recover the required antisymmetric character. You should notice, however, and some people did, that we can construct a totally antisymmetric proton wavefunction, for example,

$$|p\uparrow\rangle = \sqrt{\frac{1}{2}} [p_A \chi(M_S) - p_S \chi(M_A)],$$

and forget about color! Write this function in an explicit form, comparable to (2.71). Obtain $|n\uparrow\rangle$, and hence show that

$$\frac{\mu_n}{\mu_p} = -2.$$

So this option is ruled out by experiment. In fact, glancing at your derivation, you will notice that μ_p is negative. It is measured to be positive. Long live color.

EXERCISE 2.19 Prove that the quark model relations for the magnetic moments of the ρ^\pm mesons are

$$\mu_{\rho^+} = -\mu_{\rho^-} = \mu_\rho.$$

EXERCISE 2.20 Use the quark model to calculate the amplitude for the radiative decay $\omega \rightarrow \pi^0 \gamma$. The ω and π^0 belong to the $J^P = 1^-$, $S = 1$ and

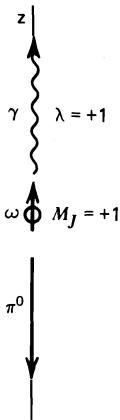


Fig. 2.10 The radiative decay $\omega \rightarrow \pi^0 \gamma$.

the $J^P = 0^-, S = 0$ nonets, respectively. We therefore require a quark spin flip (magnetic dipole) transition. This will involve the quark magnetic moment operator.

First, assume (2.54) and obtain the spin-flavor wavefunctions for an ω with $M_J = 1$ and for a π^0 . If the z axis is chosen as in Fig. 2.10, show that the required amplitude is

$$\sum_{i=1,2} \langle \pi^0 | \mu_i \sigma_i \cdot \epsilon_R^* | \omega(M_J = 1) \rangle = -\sqrt{2} \sum_{i=1,2} \langle \pi^0 | \mu_i (\sigma_-)_i | \omega(M_J = 1) \rangle \\ = \mu_d - \mu_u$$

where $\epsilon_R \equiv -\sqrt{\frac{1}{2}}(1, i, 0)$ is the polarization vector of the emitted (helicity-one) photon and $\sigma_- \equiv \frac{1}{2}(\sigma_1 - i\sigma_2)$ is the operator which “steps down” or “flips” the quark spin.

EXERCISE 2.21 Assuming (2.54), show that the quark model forbids the decay $\phi \rightarrow \pi^0 \gamma$, and predicts that

$$\frac{\text{Rate}(\omega \rightarrow \pi^0 \gamma)}{\text{Rate}(\rho \rightarrow \pi^0 \gamma)} = \left(\frac{\mu_d - \mu_u}{\mu_d + \mu_u} \right)^2 = 9.$$

2.13 Heavy Quarks: Charm and Beyond

The discovery in November 1974 of a very narrow resonance, called ψ , in e^+e^- annihilation near a center-of-mass energy of 3.1 GeV, followed two weeks later by the appearance of a second narrow resonance, ψ' , at 3.7 GeV, can rightly be called the “November revolution.” Independently, a group of experimentalists discovered the ψ -particle by producing it in proton–proton collisions. They called it J , and it is therefore often referred to in the literature as the J/ψ particle.

The ψ and ψ' were immediately interpreted as the lowest bound states of a new quark and its antiquark, $c\bar{c}$. This new charmed quark, c , had been much heralded. As we shall see in Chapter 12, the existence of another quark of charge $+\frac{2}{3}$ had long been needed in the theory of the weak interactions of hadrons (cf. the GIM mechanism).

As the total energy of the colliding e^+ and e^- beams was increased beyond 3.7 GeV, the cross section for $e^+e^- \rightarrow$ hadrons showed complicated resonance structure. A peep ahead at the column marked e^+e^- in Fig. 2.13 shows the four resonances identified below about 4 GeV. Above 3.7 GeV, the widths of the resonances become larger and more typical of hadronic decays. This phenomenon is a replay of ϕ -decay (see Exercise 2.9). The narrow width (4 MeV) of the ϕ -meson arises because it is an $s\bar{s}$ bound state just above the $K\bar{K}$ threshold, K being the lightest strange meson. The decay

$$\phi(s\bar{s}) \rightarrow K(q\bar{s}) + \bar{K}(\bar{q}s) \quad (2.76)$$

with $q = u, d$, is inhibited by lack of phase space, while $\phi \rightarrow \pi\pi\pi$ has plenty of phase space but requires annihilation of the $s\bar{s}$ pair. Nevertheless, the dominant decay of the ϕ is through the $K\bar{K}$ mode. This result is pictured in terms of quark lines in Fig. 2.11 and is an example of the Zweig or OZI rule which asserts that disconnected quark-line diagrams (Fig. 2.11a) are highly suppressed relative to connected ones (Fig. 2.11b). Similarly, in the case of the $\psi(3.1)$ and $\psi'(3.68)$, the very small widths (69 and 225 keV, respectively) arise because they are $c\bar{c}$ bound states below the $D\bar{D}$ threshold, where D is the lightest charmed meson. Hence, their hadronic decays $\psi \rightarrow \pi\pi\pi$, and so forth, require annihilation of the $c\bar{c}$ pair (Fig. 2.11c). The larger hadronic-type widths of the higher ψ -states are attributed to the allowed decay

$$\psi(c\bar{c}) \rightarrow D(\bar{q}q) + \bar{D}(q\bar{c}) \quad (2.77)$$

with $q = u, d$ (see Fig. 2.11d). Thus, we expect the kinematic threshold for $D\bar{D}$ production to lie somewhere between the $\psi'(3.68)$ and the $\psi''(3.77)$. We therefore predict the mass of the D meson to be $m(D) \geq 3.7/2 = 1.85$ GeV.

Just like the introduction of strangeness S , we assign an additive quantum number $C = \pm 1$ to the c, \bar{c} quarks, respectively, and $C = 0$ to the lighter quarks. The c quark has charge $Q = +\frac{2}{3}$ and isospin $I = 0$, and so we should update the relations of Table 2.1 to read

$$Y = B + S + C, \quad Q = I_3 + \frac{1}{2}Y. \quad (2.78)$$

We show the basic quark multiplet in Fig. 2.12, together with the antiquark tetrahedron. Now that we have basic building blocks, we can repeat our proce-

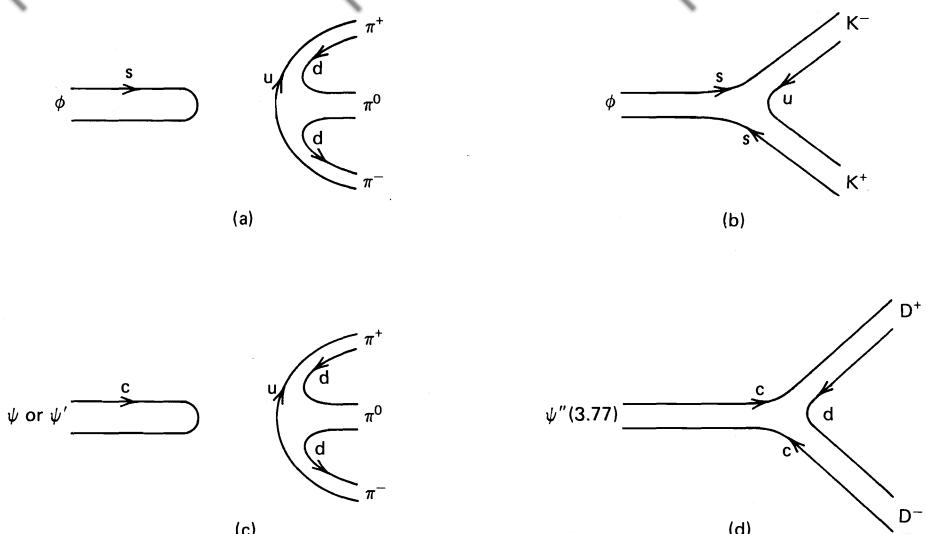


Fig. 2.11 Suppressed decay modes $\phi(s\bar{s}), \psi(c\bar{c}) \rightarrow \pi\pi\pi$ and allowed decay modes $\phi \rightarrow K\bar{K}$, $\psi'' \rightarrow D\bar{D}$.

ture of combining them to form hadrons. The $q\bar{q}$ states, the mesons, are constructed as in Fig. 2.12; the mesons shown in parentheses are members of the lowest-lying ($J^P = 0^-$) multiplet. Charmed members have been observed, with masses

$$m(D) = 1.86 \text{ GeV}, \quad m(F) = 1.97 \text{ GeV}. \quad (2.79)$$

Charmed mesons are also required to complete the other multiplets listed in Table 2.2. The charmed states of the $J^P = 1^-$ multiplet are, not surprisingly, called D^* and F^* . The observed masses are

$$m(D^*) = 2.01 \text{ GeV}, \quad m(F^*) = 2.14 \text{ GeV}?$$

EXERCISE 2.22 According to weak interaction theory, the dominant hadronic weak decay proceeds via the quark transmutations $c \rightarrow s$ and/or $u \leftrightarrow d$ (see Chapter 12). For example, an allowed charmed meson decay is $c\bar{u} \rightarrow s\bar{d}(u\bar{u})$.

Assuming that these, and only these, transmutations can occur, show that

$$D^0 \rightarrow K^-\pi^+ \quad \text{and} \quad K^-\pi^+\pi^+\pi^-$$

are allowed decay modes, but that

$$D^0 \rightarrow \pi^+\pi^-, K^+K^-, K^+\pi^-, \quad \text{and} \quad K^+\pi^-\pi^+\pi^-$$

are all forbidden. Further, show that $D^+ \rightarrow K^-\pi^+\pi^+$ is an allowed weak decay, but that $D^+ \rightarrow K^+\pi^+\pi^-$ is forbidden. This distinctive feature of D^+ decays was in fact convincing evidence in the first ever observation of a charmed particle in 1976, some 18 months after the revolutionary discovery of the “hidden” charm state $\psi(c\bar{c})$.

Each meson multiplet contains a state, $c\bar{c}$, of “hidden” charm. For the $J^P = 0^-$ and 1^- multiplets, it is $\eta_c(2.98)$ and the original $\psi(3.1)$, respectively. The states of the bound $c\bar{c}$ system can be compared with those of positronium e^+e^- . We speak of “charmonium.” It is a particularly clean system to study and has revo-

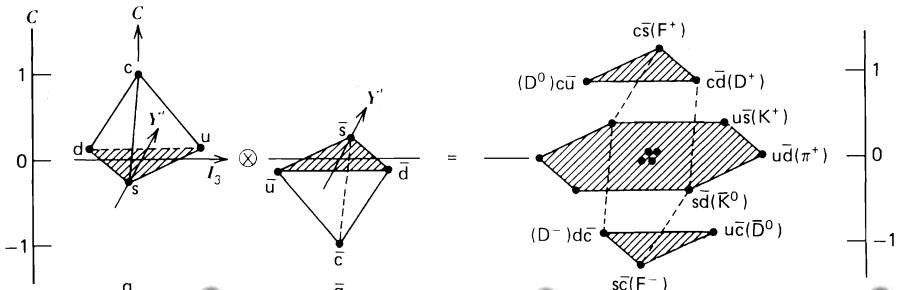


Fig. 2.12 The 16 meson states made from u, d, s, c quarks, plotted in (I_3, Y', C) space with $Y' = Y - \frac{4}{3}C$. Some members of the $J^P = 0^-$ multiplet are indicated.

lutionized meson spectroscopy. States with $J^{PC} = 1^{--}$ can be directly produced ($e^+e^- \rightarrow$ virtual $\gamma \rightarrow c\bar{c}$); and, via their decays, other charmonium states can be identified. The observed states are shown in Fig. 2.13, labeled in the conventional spectroscopic manner ${}^{2S+1}L_J$, where S , L , and J are, respectively, the total intrinsic spin, orbital angular momentum, and total angular momentum of the $c\bar{c}$ system. This is, of course, a nonrelativistic classification; it is the heavy mass of the c quark which makes it possible to use a nonrelativistic picture. We also show the J^{PC} values of the states and note that the observations coincide with quark model expectations. The six J^{PC} values listed in Table 2.2 are reproduced, except that the 1^{+-} (or 1P_1) state still awaits discovery. As in positronium, radial as well as orbital excitations are expected. In fact, the $2\ {}^3S$ and $3\ {}^3S$ excitations are seen directly as resonances in the cross section for $e^+e^- \rightarrow$ hadrons (see Fig. 2.13).

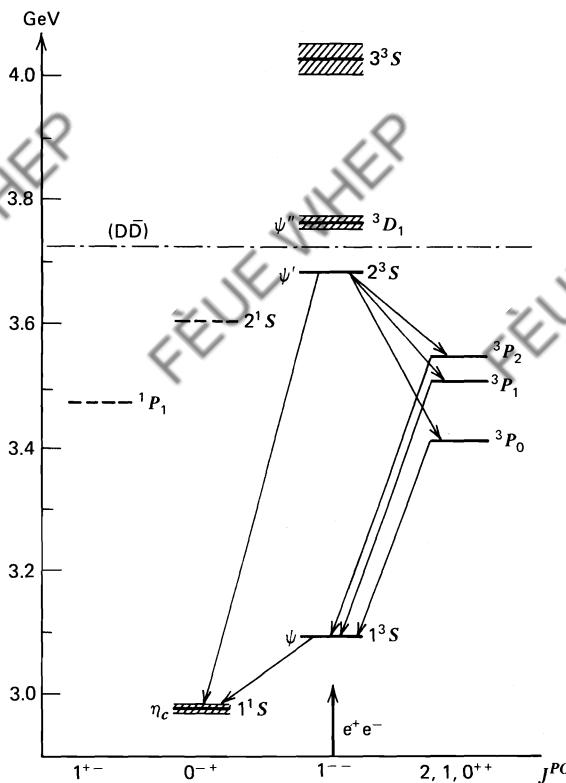


Fig. 2.13 The observed charmonium spectrum. The transitions shown have all been observed. The 1P_1 and 2S states await discovery. The particle widths are shown by shaded bands. The dot-dash line shows the $D\bar{D}$ threshold; states below this line cannot decay into charmed mesons. The states with $J^{PC} = 1^{--}$ can be directly produced by e^+e^- collisions.

EXERCISE 2.23 The decay

$$\psi'(3.7) \rightarrow \psi(3.1) + \text{hadrons}$$

is observed. What are the hadrons?

EXERCISE 2.24 Mark on Fig. 2.13 the expected radiative transitions between the levels, indicating which are electric and which are magnetic dipole transitions.

Justify that the rates for the radiative transitions from $\psi'(3.7)$ to the three χ -levels, 3P_J , with $J = 2, 1, 0$, are proportional to $(2J + 1)k^3$, where k is the momentum of the emitted photon. Hence, show that the branching ratios of these decay modes of ψ' are approximately equal.

EXERCISE 2.25 The leptonic decay of neutral vector ($J^{PC} = 1^{--}$) mesons can be pictured as proceeding via a virtual photon,

$$V(q\bar{q}) \rightarrow \gamma \rightarrow e^+ e^- \quad (2.80)$$

The technique for calculating such amplitudes will be explained in succeeding chapters. Here, it suffices to note that the $V-\gamma$ coupling is proportional to the charge of the quark q . Neglecting a possible dependence on the vector meson mass, show that the leptonic decay widths are in the ratios

$$\rho : \omega : \phi : \psi = 9 : 1 : 2 : 8.$$

EXERCISE 2.26 Comment on the rate you would expect for the e^+e^- decay mode of the 3D_1 state as compared to the $\psi'(3.7)$ state. Can these two states mix?**EXERCISE 2.27** The hadronic decay widths of η_c and $\psi(3.1)$ are estimated using

$$\begin{aligned} \eta_c(c\bar{c}) &\rightarrow n g \rightarrow \text{hadrons} \\ \psi(c\bar{c}) &\rightarrow n' g \rightarrow \text{hadrons}, \end{aligned} \quad (2.81)$$

where g is a gluon and n and n' are integers. These are QCD analogues of the QED process of (2.80). Show that the minimum values of n and n' are 2 and 3, respectively.

Properties of the potential between the c and \bar{c} can be inferred from the charmonium spectrum. In Chapter 1, we noted that at small c and \bar{c} separations, QCD predicts a Coulomb-type potential $-\alpha_s/r$, but that at large separation r , we

expect a confining potential which increases with r . A glance at the $1S$, $2S$, and the “center of gravity” of the P levels of Fig. 2.13 shows that the potential is in fact somewhere between Coulomb (which has $2S$ and P degenerate) and an oscillator potential $V \sim r^2$ (which has the P level halfway between $1S$ and $2S$). A naive potential, which is phenomenologically rather satisfactory, is

$$V(r) = -\frac{4}{3} \frac{\alpha_s}{r} + ar \quad (2.82)$$

where a is a constant parameter and $\frac{4}{3}$ is the color factor associated with the quark-gluon coupling α_s [see (2.98)].

Let us now repeat the steps of constructing baryons, but this time include the c quark. Combining three basic quark multiplets, we find that the analogue of (2.59) is

$$4 \otimes 4 \otimes 4 = 20 \oplus 20 \oplus 20 \oplus \bar{4}. \quad (2.83)$$

$$S \quad M_S \quad M_A \quad A$$

Rather than to derive this decomposition, it is better at this stage to use the elegant techniques of group theory (Young tableaux); see, for example, Close (1979). Including spin, (2.64), we can as before form the required symmetric spin-flavor ground state in two ways: either the symmetric 20 with a symmetric spin $\frac{3}{2}$ or a mixed-symmetry 20 with spin $\frac{1}{2}$ constructed in exact analogy to (2.68). Extracting the flavor multiplets from a superposition of three basic (quark) tetrahedra leads to the ground-state baryons of Fig. 2.14a.

The spin- $\frac{1}{2}$ multiplet can be viewed as three $SU(3)$ octets propping each other up and based on the edges of a fourth $SU(3)$ octet. In fact, we do not need the elegance of group theory to enumerate the states. For example, the $C = 1$ spin- $\frac{1}{2}$ baryons are cqq composites with $q = u, d$, or s . The qq decomposition is given in (2.58), namely,

$$3 \otimes 3 = 6 \oplus \bar{3},$$

and the states are shown in Fig. 2.14b. The lightest charmed baryons are the Σ_c isospin triplet and Λ_c^+ . The observed masses are

$$m(\Lambda_c) = 2.28 \text{ GeV}, \quad m(\Sigma_c) = 2.44 \text{ GeV}. \quad (2.84)$$

EXERCISE 2.28 Determine the flavor wavefunctions of the Λ_c and Σ_c baryons. Give an observable decay sequence of Σ_c^{++} .

The c quark was desired theoretically. The same cannot be said of the b quark. Evidence for this fifth quark came in a replay of the charmonium phenomenon in the e^+e^- energy region around 10 GeV. Four e^+e^- resonances were quickly identified: $T(1S)$, $T(2S)$, $T(3S)$, and $T(4S)$ with masses of 9.46, 10.02, 10.35, and 10.57 GeV, respectively. The first three states are narrow and the fourth is much wider. The lightest meson ($b\bar{u}$ or $b\bar{d}$) with explicit beauty is therefore expected to have mass $m(D_b) \approx 10.4/2 = 5.2$ GeV [cf. (2.77)].

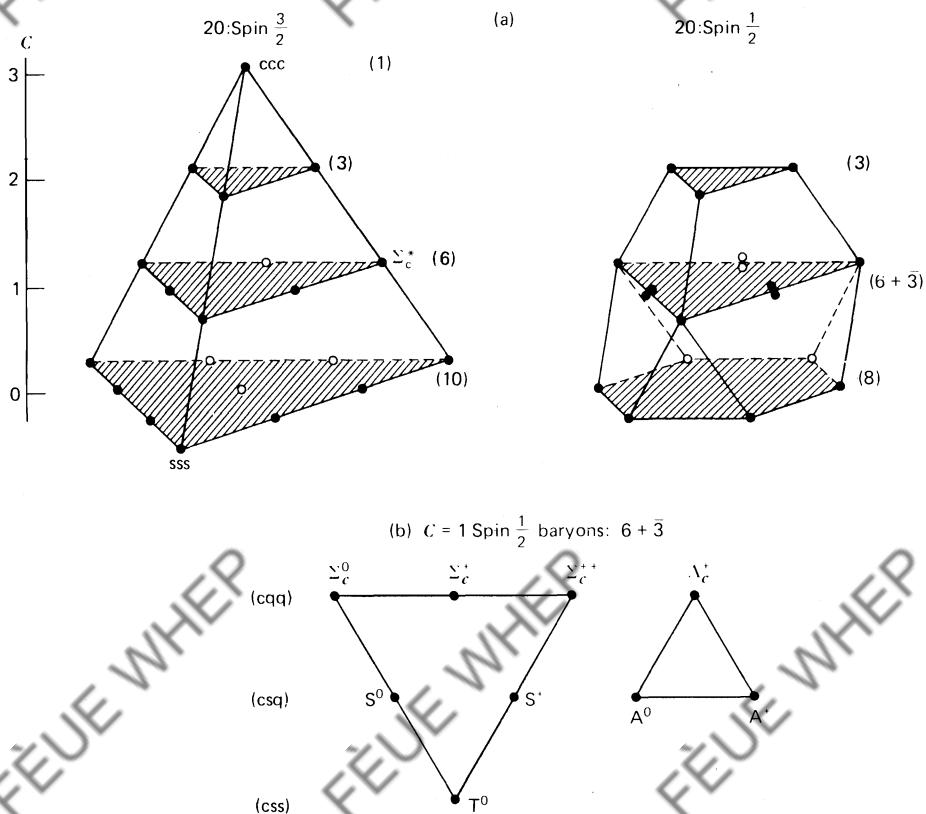


Fig. 2.14 (a) The spin- $\frac{1}{2}$ and spin- $\frac{3}{2}$ ground-state baryons made from u,d,s,c quarks, with the $SU(3)$ multiplicities of the $C = 0, 1, 2, 3$ states shown in parentheses. (b) The $C = 1$ spin- $\frac{1}{2}$ baryons ($6 + \bar{3}$) and their quark content with $q = u, d$.

2.14 Hadron Masses

If $SU(4)$ flavor symmetry were exact, all members of a given $SU(4)$ multiplet would have the same mass. This is manifestly not the case. For example, within the 1^- meson multiplet, we have

$$\begin{aligned}
 m_\omega &\approx m_\rho(u\bar{u}) = 0.78 \text{ GeV} \\
 m_\phi(s\bar{s}) &= 1.02 \text{ GeV} \\
 m_{K^*}(s\bar{u}) &= 0.89 \text{ GeV} \\
 m_{D^*}(c\bar{u}) &= 2.01 \text{ GeV} \\
 m_{F^*}(c\bar{s}) &= 2.14 \text{ GeV (?)} \\
 m_\psi(c\bar{c}) &= 3.1 \text{ GeV.}
 \end{aligned} \tag{2.85}$$

It is true that members of an $SU(2)$ isospin multiplet have the same mass to within about 5 MeV. However, u, d, s $SU(3)$ flavor symmetry is broken by mass differences of the order of 100 MeV, and $SU(4)$ flavor symmetry by considerably greater than 1 GeV. Indeed, if we attribute the mass of a hadron to simply the sum of the masses of the constituent quarks, then eqs. (2.85) imply

$$\begin{aligned} m_u &\approx m_d \approx 0.39 \text{ GeV} \\ m_s &\approx 0.51 \text{ GeV} \\ m_c &\approx 1.6 \text{ GeV}. \end{aligned} \tag{2.86}$$

We may use the observed magnetic moments of the baryons to obtain an alternative estimate of the quark masses. For example, we use the magnetic moment of the proton and the Λ to estimate m_u and m_s , respectively. From (2.73) and (2.74), we have

$$\mu_u = \frac{2}{3} \mu_p = \frac{2}{3} (2.79) \frac{e}{2m_p}$$

and so, noting (2.72),

$$m_u = \frac{m_p}{2.79} = 0.34 \text{ GeV}. \tag{2.87}$$

Similarly, the quark model prediction

$$\mu_s = \mu_\Lambda = -0.61 \frac{e}{2m_p}$$

gives $m_s = 0.51$ GeV. The agreement with (2.86) increases our confidence that quarks are indeed point-like constituents with Dirac magnetic moments.

Equations (2.86) are to be regarded as the effective masses of quarks bound within (color singlet) hadrons. We speak of constituent quark masses. It is useful to think of the constituent masses of the quark and antiquark as their zero-point energy when they are bound by some potential like (2.82) with an energy spectrum that corresponds to the masses of the observed mesons. For charm, and heavier quarks, it appears that the total zero-point energy is not much different from the masses of the lowest-lying meson states. These $c\bar{c}$ or $b\bar{b}$ states can therefore be regarded as essentially nonrelativistic bound states of the quark and antiquark.

The success of simple quark counting in explaining the gross features of the baryon and meson masses leads us to attempt to understand more detailed properties of the mass spectra. Why, for instance, is the Δ (spin $\frac{3}{2}$) heavier than the N (spin $\frac{1}{2}$), and the ρ (spin 1) heavier than the π (spin 0), despite having the same quark content? How do we account for the different masses of the neutral Λ , Σ (spin $\frac{1}{2}$), and Σ^* (spin $\frac{3}{2}$) baryons, even though they are each made of uds quarks? The differing spin configurations of the quarks offer a clue. In QED, we know that the forces are spin dependent. Should we therefore not expect an analogous result in QCD?

We begin by recalling (see, for example, Bethe and Salpeter, 1957) that the spin-spin, or magnetic moment, interaction leads to hyperfine splitting of the ground-state level of the hydrogen atom (or of positronium):

$$\Delta E_{hf} = -\frac{2}{3}\mu_1 \cdot \mu_2 |\psi(0)|^2 = \frac{2\pi\alpha}{3} \frac{\sigma_1 \cdot \sigma_2}{m_1 m_2} |\psi(0)|^2, \quad (2.88)$$

where the magnetic moment $\mu_i = e_i \sigma / 2m_i$ and $e_1 e_2 = -e^2 = -4\pi\alpha$. This is a contact interaction; it involves the square of the relative wavefunction evaluated at zero separation and so only applies to $L = 0$ states. For the hydrogen atom, it is truly a hyperfine splitting; but for positronium, we see that it is enhanced by a factor m_p/m_e .

EXERCISE 2.29 Verify that the spin 1 level (3S_1) is higher than the spin 0 level (1S_0).

The QED result, (2.88), can be taken over directly to QCD, provided we replace the electromagnetic coupling $e_1 e_2$ by the product of color charges. For mesons and baryons, the substitutions are

$$-\alpha \rightarrow \begin{cases} -\frac{4}{3}\alpha_s & \text{for } (\bar{q}\bar{q}) \\ -\frac{2}{3}\alpha_s & \text{for } (qqq) \end{cases} \quad (2.89)$$

where $\frac{4}{3}$ and $\frac{2}{3}$ are the appropriate color factors. We show how to compute these factors in a moment.

We can now make a model for the ground-state hadron masses. We assume (1) that quark confinement, which is operative at large separations, is independent of the spins and of the masses of the quarks; (2) that at near-separation, α_s is small enough for QCD hyperfine splitting to be relevant; and (3) that the only symmetry breaking arises from the different constituent masses assigned to the quarks of different flavors. In this scheme, the meson and baryon masses are therefore

$$m(q_1 \bar{q}_2) = m_1 + m_2 + [a(\sigma_1 \cdot \sigma_2)/m_1 m_2] \quad (2.91)$$

$$m(q_1 q_2 q_3) = m_1 + m_2 + m_3 + \left[\frac{a'}{2} \sum_{i>j} (\sigma_i \cdot \sigma_j)/m_i m_j \right] \quad (2.92)$$

where a and a' are positive constants [see (2.88)–(2.90)].

EXERCISE 2.30 For the π (spin 0) and the K^* (spin 1), show that (2.91) gives

$$m(\pi) = m_u + m_d - (3a/m_u m_d)$$

$$m(K^*) = m_u + m_s + (a/m_u m_s).$$

Calculate the masses of all the members of the 0^- and 1^- meson multiplets (Fig. 2.12) using

$$m_u = m_d = 0.31, \quad m_s = 0.48, \quad m_c = 1.65, \quad a/m_u^2 = 0.16,$$

all in units of GeV. Compare your predictions with the meson masses listed in the particle data tables.

Check that

$$(\rho - \pi) = \frac{m_s}{m_u} (K^* - K) = \frac{m_c}{m_u} (D^* - D) = \frac{m_c m_s}{m_u^2} (F^* - F),$$

where the meson names are used to denote their masses.

EXERCISE 2.31 Show that the model gives the Δ heavier than the nucleon. Further, show that if $a = a'$ in (2.91) and (2.92), then

$$m(\Delta) - m(N) = \frac{1}{2} [m(\rho) - m(\pi)].$$

EXERCISE 2.32 Use (2.92) to study the relative masses of the Λ , Σ , Σ^* , Λ_c , Σ_c , and Σ_c^* baryons of Figs. 2.8 and 2.14. Each baryon is a qqQ composite, where $q = u$ or d and $Q = s$ or c . For the Λ and Σ baryons, show that the qq have isospin $I = 0$ and $I = 1$, respectively, and hence spin 0 and spin 1, respectively. Use this result to evaluate $\mathbf{S}_1 \cdot \mathbf{S}_2$, where $\mathbf{S}_i \equiv \frac{1}{2} \boldsymbol{\sigma}_i$. By considering $(\mathbf{S}_1 + \mathbf{S}_2 + \mathbf{S}_3)^2$, show that

$$(\mathbf{S}_1 + \mathbf{S}_2) \cdot \mathbf{S}_3 = 0, -1, \text{ and } +\frac{1}{2} \quad \text{for } \Lambda, \Sigma, \text{ and } \Sigma^*, \text{ respectively.}$$

Thus, confirm that (2.92) gives

$$m(\Lambda_Q) = m_0 - \frac{3a'}{2m_u^2}$$

$$m(\Sigma_Q) = m_0 + \frac{2a'}{m_u^2} \left(\frac{1}{4} - \frac{m_u}{m_Q} \right)$$

$$m(\Sigma_Q^*) = m_0 + \frac{a'}{m_u^2} \left(\frac{1}{2} + \frac{m_u}{m_Q} \right)$$

where $m_0 = 2m_u + m_Q$. Verify that [see (2.84)]

$$[m(\Sigma_c) - m(\Lambda_c)] = \frac{m_s}{m_c} \frac{(m_c - m_u)}{(m_s - m_u)} [m(\Sigma) - m(\Lambda)] \simeq 0.16 \text{ GeV},$$

The masses of the other $\frac{1}{2}^+$ and $\frac{3}{2}^+$ baryons can also be calculated from (2.92) in terms of a' and the quark masses.

Considering the crude nature of the model, the quantitative agreement between the predictions and the observed masses is impressive. Indeed, all the observed features are reproduced. It is straightforward to enlarge the calculation to include hadrons containing b quarks.

2.15 Color Factors

We have already mentioned some of the evidence for color. We saw that there are good reasons to believe that each of the N flavors (u, d, ...) of quark comes in three colors which we called R, G, and B. To be precise, the quarks are assigned to a triplet of an $SU(3)$ color group (see Fig. 2.3). Unlike $SU(N)$ flavor symmetry, $SU(3)$ color symmetry is expected to be exactly conserved. A glance back at Fig. 1.4 reminds us that the gluons, which mediate the QCD force between color charges, come in eight different color combinations:

$$R\bar{G}, R\bar{B}, G\bar{R}, G\bar{B}, B\bar{R}, B\bar{G}, \sqrt{\frac{1}{2}}(R\bar{R} - G\bar{G}), \sqrt{\frac{1}{6}}(R\bar{R} + G\bar{G} - 2B\bar{B}). \quad (2.93)$$

In other words, the gluons belong to an $SU(3)$ color octet [recall the $SU(3)$ flavor analogy of (2.49), (2.50) and Fig. 2.5]. The remaining combination, the $SU(3)$ color singlet,

$$\sqrt{\frac{1}{3}}(R\bar{R} + G\bar{G} + B\bar{B}), \quad (2.94)$$

does not carry color and cannot mediate between color charges.

In QED, the strength of the electromagnetic coupling between two quarks is given by $e_1 e_2 \alpha$, where e_i is the electric charge in units of e (that is, $e_i = +\frac{2}{3}$ or $-\frac{1}{3}$) and α is the fine structure constant. Similarly, in QCD, the strength of the (strong) coupling for single-gluon exchange between two color charges is $\frac{1}{2} c_1 c_2 \alpha_s$, where c_1 and c_2 are the color coefficients associated with the vertices. It has become conventional to call

$$C_F \equiv \frac{1}{2} |c_1 c_2| \quad (2.95)$$

the color factor (although, in fact, it would have been more natural to absorb the factor $\frac{1}{2}$ in a redefinition of the strong coupling α_s and to just let the product $|c_1 c_2|$ be known as the color factor).

As a first example, we calculate the color factor for the interaction between two quarks of the same color, say B. Out of the eight gluons, only the one containing the $B\bar{B}$ combination can be exchanged. The product $c_1 c_2$ is therefore $\frac{2}{3}$ (see Fig. 2.15a). On the other hand, the interaction between colored R quarks can be mediated by two different gluons (see Fig. 2.15b). Nevertheless, the total $c_1 c_2 = \frac{1}{6} + \frac{1}{2} = \frac{2}{3}$ is the same, as indeed it has to be from color symmetry.

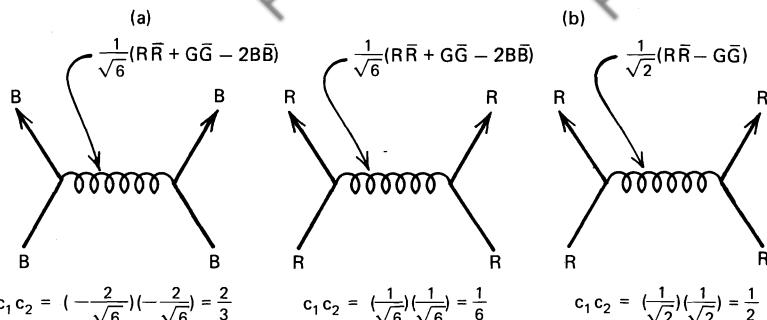


Fig. 2.15 The product of color couplings for (a) the B–B quark interaction and (b) the R–R quark interaction.

What about the interaction between two quarks of different color, say R and B? Here again, two different gluons are allowed with $c_1 c_2 = -\frac{1}{3}$ and $+1$ (see Fig. 2.16). Do we add or subtract these two (indistinguishable) amplitudes? The answer depends on the symmetry of the color wavefunction under interchange of the quarks. For a symmetric (antisymmetric) state, we sum (subtract) to give a factor $+\frac{2}{3}$ ($-\frac{4}{3}$). Indeed, we already followed this prescription when we added the two amplitudes describing the R–R interaction.

All the results so far can be concisely summarized by

$$c_1 c_2 = P - \frac{1}{3} \quad (2.96)$$

where $P = \pm 1$ according to whether the two quarks are in a color symmetric or antisymmetric state.

It is relevant to compute the color factor for gluon exchange between a quark and an antiquark in the color singlet state

$$\sqrt{\frac{1}{3}}(R\bar{R} + G\bar{G} + B\bar{B}), \quad (2.97)$$

that is, between a $q\bar{q}$ pair in a meson. Here, all colors occur on an equal footing, and so it is sufficient to consider, say, the $B\bar{B}$ interaction. There are three possible diagrams (see Fig. 2.17). In computing $c_1 c_2$, we insert a minus sign at the antiparticle vertex, just as in QED, where the antiparticle has opposite charge to

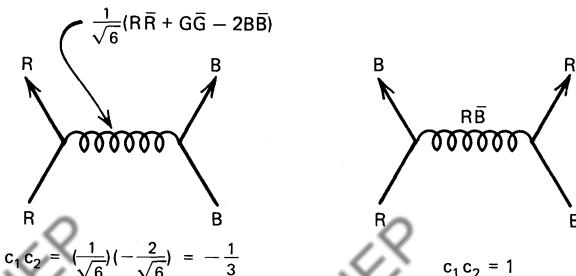


Fig. 2.16 The two diagrams describing the R–B quark interaction.

the particle. In a color singlet meson, each initial and final state of Fig. 2.17 has a factor $\sqrt{\frac{1}{3}}$ [see (2.97)]. The total factor for the $q\bar{q}$ interaction, via single-gluon exchange, in a meson is therefore

$$c_1 c_2 = 3 \sqrt{\frac{1}{3}} \sqrt{\frac{1}{3}} \left(-\frac{2}{3} - 1 - 1 \right) = -\frac{8}{3},$$

where the first factor of 3 allows for the contributions when $R\bar{R}$ and $G\bar{G}$ are the initial states. That is, the color factor

$$C_F = \frac{4}{3}, \quad (2.98)$$

a result we used in (2.89).

It is instructive to also derive (2.90), that is, to calculate the color factor for two quarks exchanging a gluon within a baryon. Now remembering that $3 \otimes 3 = 6 \oplus \bar{3}$, we see that every quark pair in a baryon is in a color $\bar{3}$. The reason is that the pair must be coupled to a third quark (a color 3) to give an overall color singlet. The alternative, $6 \otimes 3$, does not contain a singlet. The $\bar{3}$ is an antisymmetric color state [compare (2.58)], and so we must use (2.96) with $P = -1$. The required color factor, (2.90), is therefore

$$C_F = \frac{2}{3}. \quad (2.99)$$

The message of this chapter is that all the observed strongly interacting particles (hadrons) are bound states of quarks. Historically, it was flavor $SU(3)$ that led to the discovery of this fundamental fact, but its role in particle physics is now superseded by the quark model. Later, we develop a dynamical theory for the interactions of the (constituent) quarks which is based on color $SU(3)$ (gauge) symmetry.

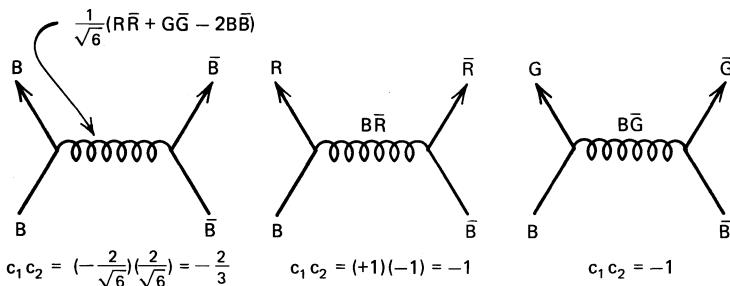


Fig. 2.17 Diagrams describing the $B-\bar{B}$ interaction.

3

Antiparticles

Quantum electrodynamics (QED) is the theory describing the electromagnetic interaction of quarks and leptons. Some of the high-energy physicist's favorite QED processes are $e^+e^- \rightarrow \mu^+\mu^-$, $eq \rightarrow eq$, $\gamma q \rightarrow (e^+e^-)q$, and so on, where q denotes a quark. The list reveals the technical problems to be faced in any computation of their transition rates: (1) we are concerned with a many-particle situation, (2) we are dealing with a relativistic problem. Indeed, not only are experiments routinely performed using beams of highly relativistic particles, ruling out any nonrelativistic approach, but also antiparticles occur. They are of course not required by nonrelativistic theory.

The problem is not as formidable as it looks; perturbation theory will save us. We obtain the solutions of the one-particle wave equations for *free* leptons (or quarks) and then study the scattering of one particle by another by treating the interaction as a perturbation.

At first sight, it is very surprising that single-particle wave equations can be used to describe interactions in which particles can be created and annihilated. The crucial observation is that relativistic wave equations have negative energy solutions which can be exploited so as to introduce antiparticles into the theory.

The final formalism is a covariant copy of nonrelativistic perturbation theory using only solutions to single-particle wave equations. As a calculational scheme, it is most readily implemented by summing the relevant "Feynman diagrams" that can be drawn for the process under study, where the diagrams are evaluated using a set of well-established rules: the Feynman rules. This heuristic, but very intuitive, approach is due to Feynman and has the practical advantage that we can calculate transition rates and cross sections at an early stage in the development of the formalism. We thus avoid having to develop the formal machinery of quantum field theory, which eventually yields the Feynman rules from a Lagrangian. (For an introduction to quantum field theory, see, for example, Mandl (1966) or Sakurai (1967).)

The spin of the quarks and leptons complicates to some extent the essential simplicity of Feynman's approach. We therefore introduce the calculational scheme using the unphysical example of "spinless leptons" (Chapters 3 and 4) and subsequently introduce their spin as a technical complication (Chapters 5 and 6). Chapter 7 offers a glimpse of the accuracy that can be achieved in QED

calculations and describes computations which are verified by experiment to, for example, one part per million in the case of the magnetic moment of the electron.

3.1 Nonrelativistic Quantum Mechanics

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

$$E = \frac{p^2}{2m} \quad (3.1)$$

the differential operators

$$E \rightarrow i\hbar \frac{\partial}{\partial t}, \quad \mathbf{p} \rightarrow -i\hbar \nabla. \quad (3.2)$$

The resulting operator equation is understood to act on a (complex) wavefunction $\psi(\mathbf{x}, t)$. That is (with $\hbar \equiv 1$),

$$i \frac{\partial \psi}{\partial t} + \frac{1}{2m} \nabla^2 \psi = 0, \quad (3.3)$$

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, \mathbf{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, that is,

$$-\frac{\partial}{\partial t} \int_V \rho dV = \int_S \mathbf{j} \cdot \hat{\mathbf{n}} dS = \int_V \nabla \cdot \mathbf{j} dV$$

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

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{j} = 0. \quad (3.4)$$

To determine the flux, we first form $\partial \rho / \partial t$ by subtracting the wave equation, (3.3), multiplied by $-i\psi^*$ from the complex conjugate equation multiplied by $-i\psi$. We then obtain

$$\frac{\partial \rho}{\partial t} - \frac{i}{2m} (\psi^* \nabla^2 \psi - \psi \nabla^2 \psi^*) = 0. \quad (3.5)$$

Comparing this with (3.4), we identify the probability flux density as

$$\mathbf{j} = -\frac{i}{2m}(\psi^* \nabla \psi - \psi \nabla \psi^*). \quad (3.6)$$

For example, a solution of (3.3),

$$\psi = N e^{i \mathbf{p} \cdot \mathbf{x} - i E t}, \quad (3.7)$$

which describes a free particle of energy E and momentum \mathbf{p} , has

$$\rho = |N|^2, \quad \mathbf{j} = \frac{\mathbf{p}}{m} |N|^2. \quad (3.8)$$

3.2 Lorentz Covariance and Four-Vector Notation

A cornerstone of modern physics is that the fundamental laws have the same form in all Lorentz frames; that is, in reference frames which have a uniform relative velocity. The fundamental equations are said to be *Lorentz covariant*. Recall that the theory of special relativity is based on the premise that the velocity of light, c , is the same in all Lorentz frames. A Lorentz transformation relates the coordinates in two such frames. The basic invariant is $c^2 t^2 - \mathbf{x}^2$.

EXERCISE 3.1 Consider a Lorentz transformation in which the new frame (primed coordinates) moves with velocity v along the z axis of the original frame (unprimed coordinates). For such a Lorentz “boost”, show that

$$ct' = \cosh \theta ct - \sinh \theta z,$$

$$z' = -\sinh \theta ct + \cosh \theta z,$$

with x and y unchanged; here, $\tanh \theta = v/c$. As $\cos i\theta = \cosh \theta$ and $\sin i\theta = i \sinh \theta$, we see that the Lorentz transformation may be regarded as a rotation through an imaginary angle $i\theta$ in the ict - z plane.

By definition, any set of four quantities which transform like (ct, \mathbf{x}) under Lorentz transformations is called a *four-vector*. We use the notation

$$(ct, \mathbf{x}) \equiv (x^0, x^1, x^2, x^3) \equiv x^\mu. \quad (3.9)$$

According to the theory of special relativity, the total energy E and the momentum \mathbf{p} of an isolated system transform as the components of a four-vector

$$\left(\frac{E}{c}, \mathbf{p} \right) \equiv (p^0, p^1, p^2, p^3) = p^\mu$$

with the basic invariant $(E^2/c^2) - \mathbf{p}^2$. The simplest system is a free particle, for which

$$\frac{E^2}{c^2} - \mathbf{p}^2 = m^2 c^2, \quad (3.10)$$

where m is the rest mass of the particle. From now on, we revert back to the use of natural units with $c \equiv 1$ (see Section 1.4).

Just as in three-dimensional space, we may introduce the scalar product of two four-vectors $A^\mu \equiv (A^0, \mathbf{A})$ and $B^\mu \equiv (B^0, \mathbf{B})$

$$\mathbf{A} \cdot \mathbf{B} \equiv A^0 B^0 - \mathbf{A} \cdot \mathbf{B},$$

which is left invariant under Lorentz transformations. Due to the minus sign, it is convenient to introduce a new type of four-vector, $A_\mu \equiv (A^0, -\mathbf{A})$, so that the scalar product is

$$\mathbf{A} \cdot \mathbf{B} = A_\mu B^\mu = A^\mu B_\mu = g_{\mu\nu} A^\mu B^\nu = g^{\mu\nu} A_\mu B_\nu. \quad (3.11)$$

Here, we have introduced the (metric) tensor $g_{\mu\nu}$, which is defined by

$$g_{00} = 1, \quad g_{11} = g_{22} = g_{33} = -1, \quad \text{other components} = 0$$

(and similarly for $g^{\mu\nu}$). A summation over repeated indices is implied in (3.11). Upper (lower) index vectors are called contravariant (covariant) vectors. The rule for forming Lorentz invariants is to make the upper indices balance the lower indices. If an equation is Lorentz covariant, we must ensure that all unrepeatable indices (upper and lower separately) balance on either side of the equation, and that all repeated indices appear once as an upper and once as a lower index.

EXERCISE 3.2 Show that $g_{\mu\nu} g^{\mu\nu} = 4$.

Examples of scalar products are

$$p^\mu x_\mu \equiv p \cdot x = Et - \mathbf{p} \cdot \mathbf{x}$$

$$p^\mu p_\mu \equiv p \cdot p \equiv p^2 = E^2 - \mathbf{p}^2.$$

These quantities are Lorentz invariants. For a free particle, we have $p^2 = m^2$, see (3.10). We say that the particle is on its mass shell.

EXERCISE 3.3 The collision of two particles, each of mass M , is viewed in a Lorentz frame in which they hit head-on with momenta equal in magnitude but opposite in direction. We speak of this as the “center-of-mass” frame (though the name “center-of-momentum” would be more appropriate). The total energy of the system is E_{cm} . Show that the Lorentz invariant

$$s \equiv (p_1 + p_2)_\mu (p_1 + p_2)^\mu \equiv (p_1 + p_2)^2 = E_{cm}^2. \quad (3.12)$$

If the collision is viewed in the “laboratory” frame where one of the particles is at rest, then show, by evaluating the invariant s , that the other has energy

$$E_{lab} = \frac{E_{cm}^2}{2M} - M.$$

We can see from this result that colliding-beam accelerators have an enormous advantage over fixed-target accelerators in achieving a given total center-of-mass energy \sqrt{s} . List some advantages of fixed-target accelerators.

Note that the space-like components of A^μ and A_μ are \mathbf{A} and $-\mathbf{A}$, respectively. The exception is

$$\partial^\mu = \left(\frac{\partial}{\partial t}, -\nabla \right) \quad \text{and} \quad \partial_\mu = \left(\frac{\partial}{\partial t}, \nabla \right), \quad (3.13)$$

which can be shown to transform like $x^\mu = (t, \mathbf{x})$ and $x_\mu = (t, -\mathbf{x})$, respectively. Thus, the covariant form of (3.2) is

$$p^\mu \rightarrow i\partial^\mu. \quad (3.14)$$

From ∂_μ and ∂^μ we can form the invariant (D'Alembertian) operator

$$\square^2 \equiv \partial_\mu \partial^\mu. \quad (3.15)$$

3.3 The Klein–Gordon Equation

Wave equation (3.3) violates Lorentz covariance and is not suitable for a particle moving relativistically. It is tempting to repeat the steps of Section 3.2, but starting from the relativistic energy–momentum relation, (3.10),

$$E^2 = \mathbf{p}^2 + m^2.$$

Making the operator substitutions (3.2), we obtain

$$-\frac{\partial^2 \phi}{\partial t^2} + \nabla^2 \phi = m^2 \phi, \quad (3.16)$$

which is known as the Klein–Gordon equation (but could, more correctly, have been called the relativistic Schrödinger equation). Multiplying the Klein–Gordon equation by $-i\phi^*$ and the complex conjugate equation by $-i\phi$, and subtracting, gives the relativistic analogue of (3.5)

$$\underbrace{\frac{\partial}{\partial t} \left[i \left(\phi^* \frac{\partial \phi}{\partial t} - \phi \frac{\partial \phi^*}{\partial t} \right) \right]}_{\rho} + \nabla \cdot \underbrace{[-i(\phi^* \nabla \phi - \phi \nabla \phi^*)]}_{\mathbf{j}} = 0. \quad (3.17)$$

By comparison with (3.4), we identify the probability and the flux densities with the terms in square brackets. For example, for a free particle of energy E and momentum \mathbf{p} , described by the Klein–Gordon solution

$$\phi = Ne^{i\mathbf{p} \cdot \mathbf{x} - iEt},$$

we find from (3.17) that [see (3.8)].

$$\begin{aligned} \rho &= i(-2iE)|N|^2 = 2E|N|^2 \\ \mathbf{j} &= -i(2i\mathbf{p})|N|^2 = 2\mathbf{p}|N|^2. \end{aligned} \quad (3.18)$$

We see that the probability density is proportional to E , the relativistic energy of the particle. (We defer the explanation of this for a moment.)

It is advantageous to express these results in four-vector notation. Not only are they then more concise, but also the covariance becomes explicit. Using the D'Alembertian operator, (3.15), the Klein–Gordon equation becomes

$$(\square^2 + m^2)\phi = 0. \quad (3.19)$$

Moreover, the probability and the flux densities form a four-vector

$$j^\mu = (\rho, \mathbf{j}) = i(\phi^* \partial^\mu \phi - \phi \partial^\mu \phi^*) \quad (3.20)$$

which satisfies the (covariant) continuity relation

$$\partial_\mu j^\mu = 0. \quad (3.21)$$

Taking the free particle solution

$$\phi = Ne^{-ip \cdot x}, \quad (3.22)$$

we have [see (3.18)]

$$j^\mu = 2p^\mu |N|^2. \quad (3.23)$$

We noted that the probability density ρ is the time-like component of a four-vector; ρ is proportional to E . This result may be anticipated since under a Lorentz boost of velocity \mathbf{v} , a volume element suffers a Lorentz contraction $d^3x \rightarrow d^3x\sqrt{1-v^2}$; and so, to keep ρd^3x invariant, we require ρ to transform as the time-like component of a four-vector $\rho \rightarrow \rho/\sqrt{1-v^2}$.

So far, so good; but what are the energy eigenvalues of the Klein–Gordon equation? Substitution of (3.22) into (3.19) gives

$$E = \pm(\mathbf{p}^2 + m^2)^{1/2}. \quad (3.24)$$

Thus, in addition to the acceptable $E > 0$ solutions, we have negative energy solutions. This looks at first like a total disaster, because transitions can occur to lower and lower (more negative) energies. A second problem is that the $E < 0$ solutions are associated with a negative probability density from (3.18). To summarize, the difficulties are

$E < 0$ solutions with $\rho < 0$.

It is clear that this problem cannot be simply ignored. 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.

3.4 Historical Interlude

In 1927, in an attempt to avoid these problems, Dirac devised a relativistic wave equation linear in $\partial/\partial t$ and ∇ . He succeeded in overcoming the problem of the negative probability density, with the unexpected bonus that the equation de-

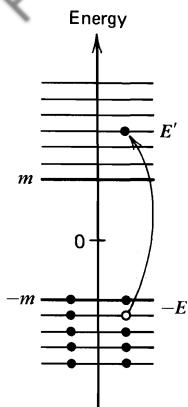


Fig. 3.1 Energy level spectrum for the electron. Dirac's picture of the vacuum has all the negative energy states occupied. We show two states per level to account for the two spin states of the electron.

scribed spin- $\frac{1}{2}$ particles. However, $E < 0$ solutions still occurred, as can be seen in the energy spectrum for a free Dirac electron sketched in Fig. 3.1. Dirac sidestepped the negative energy solutions by invoking the exclusion principle. He postulated that all the negative energy states are occupied and regarded the vacuum as an infinite sea of $E < 0$ electrons. Now the positive energy electrons cannot collapse into the lower (negative) energy levels, as this is prevented by the exclusion principle. One can, however, create a “hole” in the sea by excitation of an electron from a negative energy ($-E$) state to a positive energy (E') state, as shown. The absence of an electron of charge $-e$ and energy $-E$ is interpreted as the presence of an antiparticle (a positron) of charge $+e$ and energy $+E$. Thus, the net effect of this excitation is the production of a pair of particles

$$e^-(E') + e^+(E),$$

which clearly requires energy $E + E' \geq 2m$ (see diagram). Until 1934, the Dirac equation was considered to be the only acceptable relativistic wave equation.

In 1934, Pauli and Weisskopf revived the Klein-Gordon equation by inserting the charge $-e$ into j^μ and interpreting it as the charge-current density of the electron,

$$j^\mu = -ie(\phi^* \partial^\mu \phi - \phi \partial^\mu \phi^*). \quad (3.25)$$

Now, $\rho = j^0$ 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 shall make clear in a moment, the $E < 0$ solutions may then be regarded as $E > 0$ solutions for particles of opposite charge (antiparticles). Unlike “hole theory,” this interpretation is applicable to bosons as well as fermions. One cannot fill up the Dirac sea with bosons, as there is no exclusion principle operative to stack the particles. To develop the antiparticle idea and to introduce Feynman diagrams, it is useful to first ignore the complications due to the spin of the electrons. We therefore begin by obtaining the Feynman rules for “spinless” electrons and use them to calculate the scattering amplitudes and cross sections for interacting

particles. Not until then do we turn to the Dirac equation and to the Feynman rules for the physically realistic case of electromagnetic interactions of spin- $\frac{1}{2}$ electrons.

3.5 The Feynman–Stückelberg Interpretation of $E < 0$ Solutions

The prescription we use for handling negative energy states was proposed by Stückelberg (1941) and by Feynman (1948). Expressed most simply, the idea is that a negative energy solution describes a particle which propagates backward in time or, equivalently, a positive energy antiparticle propagating forward in time. It is crucial to master this idea, as it lies at the heart of our approach to Feynman diagrams. We try to make it plausible in the following way.

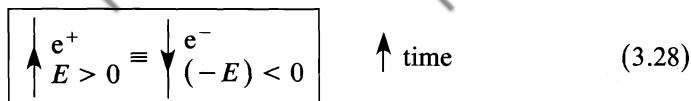
Consider an electron of energy E , three-momentum \mathbf{p} , and charge $-e$. From (3.25) and (3.22), we know that the electromagnetic four-vector current is

$$j^\mu(e^-) = -2e|N|^2(E, \mathbf{p}). \quad (3.26)$$

Now take an antiparticle, a positron, with the same E, \mathbf{p} . Since its charge is $+e$,

$$\begin{aligned} j^\mu(e^+) &= +2e|N|^2(E, \mathbf{p}) \\ &= -2e|N|^2(-E, -\mathbf{p}), \end{aligned} \quad (3.27)$$

which is exactly the same as the current j^μ for an electron with $-E, -\mathbf{p}$. Thus, as far as a system is concerned, the emission of a positron with energy E is the same as the absorption of an electron of energy $-E$. Pictorially, we have



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

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

The single-particle (e^-) wavefunction formalism not only handles antiparticles but can even describe many-particle situations. As an example, we consider the double scattering of an electron in a potential. We picture this in the space-time (Feynman) diagrams of Fig. 3.2. The crucial observation is that there are two pictures corresponding to the same observation. There are two different time orderings of the two interactions with the potential that lead to the same observable event. Indeed, note that the (observable) path of the electron before and after the double scattering is the same in the two diagrams. The second picture is only possible because of the antiparticle prescription. At time t_2 , the electron scatters backward in time (with $E < 0$). This electron is interpreted as a

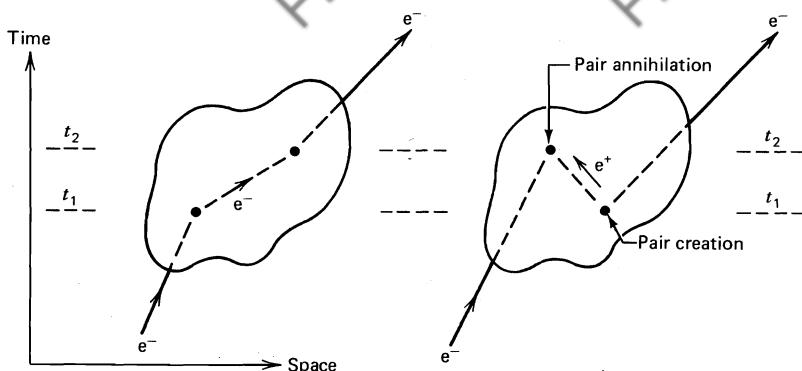


Fig. 3.2 Different time orderings of the double scattering of an electron.

positron (with $E > 0$) going forward in time. Then, the events in the diagram can be viewed as follows: first, at time t_1 , an e^-e^+ pair is created; then, at a later time t_2 , the e^+ is annihilated by the incident e^- . Therefore, between t_1 and t_2 , the electron trajectory drawn in the second diagram actually describes three particles: the initial and final electrons and a positron! As both double scatterings lead to the same observed final electron, they both have to be included in computing the probability of this event. Note that, just as in hole theory, the vacuum has become a very complex environment: e^-e^+ pairs can pop out of it and disappear into it as a result of the antiparticle prescription!

All possible processes can be described with the interpretation of a single-particle (e^-) wavefunction; the antiparticle (e^+) states are never used. For example, for a single e^+ scattering, Fig. 3.3, we use negative-energy e^- solutions with the exit and entrance states interchanged.

Our objective is to calculate transition rates and cross sections. Yet, so far, we have only the wavefunctions for free particles. How are interactions to be included? As implied by the above discussion, with its mention of “single” and “double” scattering, perturbation theory will be the method that we shall use to calculate scattering amplitudes. It is therefore appropriate to recall the main results of perturbation theory we shall need.

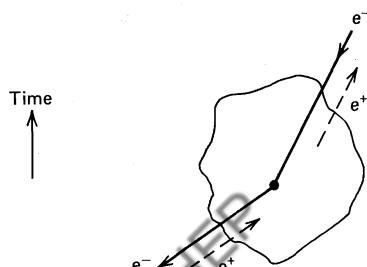


Fig. 3.3 Positron scattering by a potential.

3.6 Nonrelativistic Perturbation Theory

Suppose we know the solutions to the free-particle Schrödinger equation

$$H_0 \phi_n = E_n \phi_n \quad \text{with} \int_V \phi_n^* \phi_n d^3x = \delta_{mn} \quad (3.29)$$

where H_0 , the Hamiltonian, is time independent. For simplicity, we have normalized the solutions to one particle in a box of volume V . The objective is to solve Schrödinger's equation

$$(H_0 + V(\mathbf{x}, t))\psi = i \frac{\partial \psi}{\partial t} \quad (3.30)$$

for a particle moving in the presence of an interaction potential $V(\mathbf{x}, t)$.

Any solution of (3.30) can be expressed in the form

$$\psi = \sum_n a_n(t) \phi_n(\mathbf{x}) e^{-iE_n t}. \quad (3.31)$$

Now, to find the unknown coefficients $a_n(t)$, we substitute (3.31) into (3.30) and obtain

$$i \sum_n \frac{da_n}{dt} \phi_n(\mathbf{x}) e^{-iE_n t} = \sum_n V(\mathbf{x}, t) a_n \phi_n(\mathbf{x}) e^{-iE_n t}.$$

Multiplying by ϕ_j^* , integrating over the volume, and using the orthonormality relation (3.29) leads to the following coupled linear differential equations for the a_n coefficients:

$$\frac{da_f}{dt} = -i \sum_n a_n(t) \int \phi_f^* V \phi_n d^3x e^{i(E_f - E_n)t}. \quad (3.32)$$

Suppose that before the potential V acts the particle is in an eigenstate i of the unperturbed Hamiltonian, that is, at time $t = -T/2$:

$$\begin{aligned} a_i(-T/2) &= 1, \\ a_n(-T/2) &= 0 \quad \text{for } n \neq i, \end{aligned} \quad (3.33)$$

and

$$\frac{da_f}{dt} = -i \int d^3x \phi_f^* V \phi_i e^{i(E_f - E_i)t}. \quad (3.34)$$

Now, provided that the potential is small and transient, we can, as a first approximation, assume that these initial conditions remain true at all times. Then, integrating (3.34), we obtain

$$a_f(t) = -i \int_{-T/2}^t dt' \int d^3x \phi_f^* V \phi_i e^{i(E_f - E_i)t'} \quad (3.35)$$

and, in particular, at time $t = +T/2$ after the interaction has ceased,

$$T_{fi} \equiv a_f(T/2) = -i \int_{-T/2}^{T/2} dt \int d^3x [\phi_f(\mathbf{x}) e^{-iE_f t}]^* V(\mathbf{x}, t) [\phi_i(\mathbf{x}) e^{-iE_i t}] \quad (3.36)$$

which we may write in the covariant form

$$T_{fi} = -i \int d^4x \phi_f^*(x) V(x) \phi_i(x). \quad (3.37)$$

Of course, the expression for $a_f(t)$ is only a good approximation if $a_f(t) \ll 1$, as this has been assumed in obtaining the result.

We are tempted to interpret $|T_{fi}|^2$ as the probability that the particle is scattered from an initial state i to a final state f . Is this interpretation valid? Consider the case when $V(\mathbf{x}, t) = V(\mathbf{x})$ is time independent; then, (3.36) can be written as

$$\begin{aligned} T_{fi} &= -iV_{fi} \int_{-\infty}^{\infty} dt e^{i(E_f - E_i)t} \\ &= -2\pi i V_{fi} \delta(E_f - E_i) \end{aligned} \quad (3.38)$$

with

$$V_{fi} \equiv \int d^3x \phi_f^*(\mathbf{x}) V(\mathbf{x}) \phi_i(\mathbf{x}). \quad (3.39)$$

The δ -function in (3.38) expresses the fact that the energy of the particle is conserved in the transition $i \rightarrow f$. By the uncertainty principle, this means that an infinite time separates the states i and f , and $|T_{fi}|^2$ is therefore not a meaningful quantity. We define instead a transition probability per unit time

$$W = \lim_{T \rightarrow \infty} \frac{|T_{fi}|^2}{T}. \quad (3.40)$$

Squaring (3.38),

$$\begin{aligned} W &= \lim_{T \rightarrow \infty} 2\pi \frac{|V_{fi}|^2}{T} \delta(E_f - E_i) \int_{-T/2}^{+T/2} dt e^{i(E_f - E_i)t} \\ &= \lim_{T \rightarrow \infty} 2\pi \frac{|V_{fi}|^2}{T} \delta(E_f - E_i) \int_{-T/2}^{+T/2} dt \\ &= 2\pi |V_{fi}|^2 \delta(E_f - E_i). \end{aligned} \quad (3.41)$$

This equation can only be given physical meaning after integrating over a set of initial and final states. In particle physics, we usually deal with situations where we start with a specified initial state and end up in one of a set of final states. Let $\rho(E_f)$ be the density of final states; that is, $\rho(E_f) dE_f$ is the number of states in the energy interval E_f to $E_f + dE_f$. We integrate over this density, imposing energy conservation, and obtain the transition rate

$$\begin{aligned} W_{fi} &= 2\pi \int dE_f \rho(E_f) |V_{fi}|^2 \delta(E_f - E_i) \\ &= 2\pi |V_{fi}|^2 \rho(E_i). \end{aligned} \quad (3.42)$$

This is Fermi's Golden Rule.

Clearly, we can improve on the above approximation by inserting the result for $a_n(t)$, (3.35), in the right-hand side of (3.32):

$$\frac{da_f}{dt} = \dots + (-i)^2 \left[\sum_{n \neq i} V_{ni} \int_{-T/2}^t dt' e^{i(E_n - E_i)t'} \right] V_{fn} e^{i(E_f - E_n)t} \quad (3.43)$$

where the dots represent the first-order result. The correction to T_{fi} is

$$T_{fi} = \dots - \sum_{n \neq i} V_{fn} V_{ni} \int_{-\infty}^{\infty} dt e^{i(E_f - E_n)t} \int_{-\infty}^t dt' e^{i(E_n - E_i)t'}.$$

To make the integral over dt' meaningful, we must include a term in the exponent involving a small positive quantity ϵ which we let go to zero after integration

$$\int_{-\infty}^t dt' e^{i(E_n - E_i - i\epsilon)t} = i \frac{e^{i(E_n - E_i - i\epsilon)t}}{E_i - E_n + i\epsilon}.$$

The second-order correction to T_{fi} is therefore

$$T_{fi} = \dots - 2\pi i \sum_{n \neq i} \frac{V_{fn} V_{ni}}{E_i - E_n + i\epsilon} \delta(E_f - E_i). \quad (3.44)$$

EXERCISE 3.4 Show that the rate for the $i \rightarrow f$ transition is given by (3.42) with the replacement

$$V_{fi} \rightarrow V_{fi} + \sum_{n \neq i} V_{fn} \frac{1}{E_i - E_n + i\epsilon} V_{ni} + \dots \quad (3.45)$$

Obtain the form in the next correction.

Equation (3.45) is the perturbation series for the amplitude with terms to first, second, ... order in V . The Feynman diagrams of Fig. 3.4 represent the first two terms in the nonrelativistic perturbation series. For each interaction vertex, we

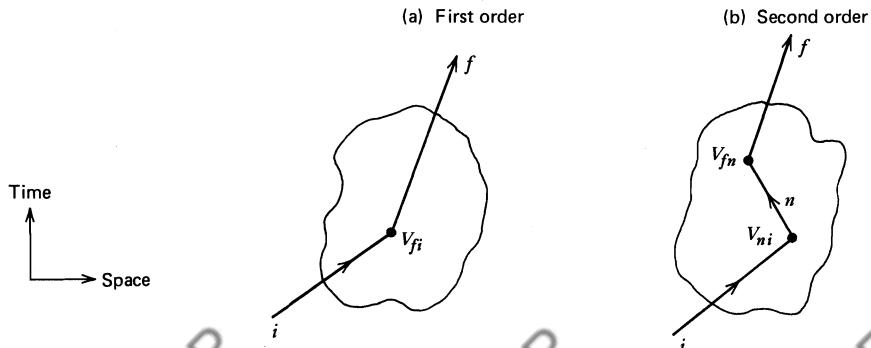


Fig. 3.4 First- and second-order contributions to the $i \rightarrow f$ transition.

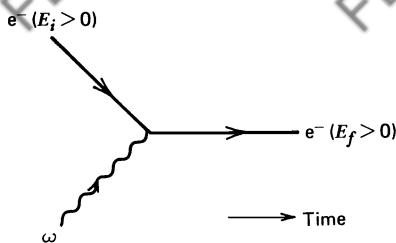


Fig. 3.5 e^- Scattering, with time increasing from left to right (rather than upward).

have a factor like V_{ni} , and for the propagation of each intermediate state, we have a “propagator” factor like $1/(E_i - E_n)$. The intermediate states are “virtual” in the sense that energy is not conserved, $E_n \neq E_i$, but there is of course energy conservation between the initial and final states, $E_f = E_i$, as indicated by the delta function $\delta(E_f - E_i)$. The central problem is to generalize this scheme to handle relativistic particles, including their antiparticles. This is the topic of the next chapter.

3.7 Rules for Scattering Amplitudes in the Feynman–Stückelberg Approach

How are we to form scattering amplitudes T_{fi} involving antiparticles if antiparticles are to be regarded as negative-energy particle solutions going backward in time? Clearly, our antiparticle prescription will have to be consistent with energy conservation.

The diagrams of Fig. 3.4 represent a noncovariant situation; they refer to scattering from a fixed, static potential. However, we are interested in the scattering of one particle by another, and to do this we will take one particle to be moving in the electromagnetic potential V due to the other. In Chapter 1, we described how the electromagnetic interaction between electrons is due to the emission and absorption of photons. Consider now energy conservation at the vertex of Fig. 3.5, in which a photon is absorbed by an electron. The form of the (covariant) interaction V is derived in the next chapter, but it is clear that V

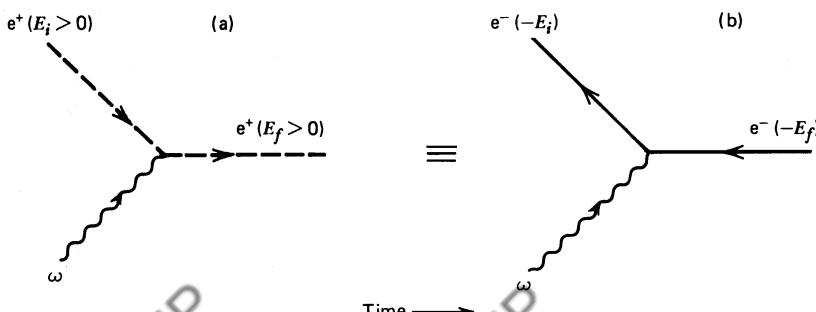


Fig. 3.6 e^+ Scattering, pictured as negative-energy e^- scattering backward in time.

has a time dependence $e^{-i\omega t}$ for an incoming photon of energy ω . Thus, the transition amplitude T_{fi} , (3.36), is proportional to

$$\int (e^{-iE_f t})^* e^{-i\omega t} e^{-iE_i t} dt = 2\pi \delta(E_f - \omega - E_i)$$

so that $E_f = E_i + \omega$.

Obviously, exactly the same argument will follow for antiparticle (positron) scattering in Fig. 3.6a. However, as before, we wish to formulate the matrix element in terms of electron states alone, as shown in Fig. 3.6b. Now, the ingoing state describes an electron of (negative) energy $-E_i$, and the transition amplitude contains the factor

$$\int (e^{-i(-E_i)t})^* e^{-i\omega t} e^{-i(-E_f)t} dt = 2\pi \delta(-E_i - \omega + E_f)$$

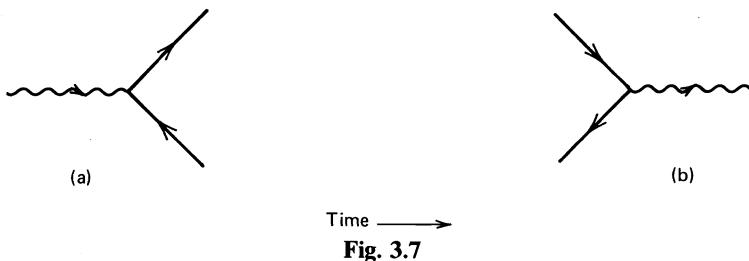
so again, $E_f = E_i + \omega$, as required.

The rule is to form the matrix element

$$\int \phi_{\text{outgoing}}^* V \phi_{\text{ingoing}} d^4x,$$

where ingoing and outgoing always refer to the arrows on the *particle* (electron) lines.

EXERCISE 3.5 Check that the rule satisfies the conservation of energy for (a) e^-e^+ pair creation, and (b) e^-e^+ annihilation of Fig. 3.7. Following the same idea, use the space part of the matrix element to show that the expected three-momentum conservation laws are obtained.



We have now set up a formalism based on perturbation theory which can handle interactions of particles and antiparticles. It can even describe multiparticle situations. The next task is to cast it in a relativistically covariant form. Note that (3.37) is already covariant.

4

Electrodynamics of Spinless Particles

The title of this chapter requires some explanation. 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, which appear throughout this chapter, are completely fictitious objects. The goal of this chapter is to find out how to use perturbation theory in a covariant way. To illustrate this, we have to choose particles and an interaction. For simplicity, we choose the particles to be “spinless” charged leptons. Clearly, it is desirable to begin by avoiding the complications of their spin. For the interaction, we choose the electromagnetic force. Electromagnetic interactions are of fundamental importance in particle physics. Quantum electrodynamics is the simplest example of a gauge theory in the sense that it has only one gauge particle, the photon. Gauge theories are now thought to be capable of describing all interactions. Chromodynamics and weak interactions are described by gauge theories that copy QED. This is demonstrated in Chapters 14 and 15. So, although the “spinless” leptons in this chapter are fictitious, the interaction is not. Understanding this is essential for further progress. The complications that arise from the spin of the leptons are revealed in Chapters 5 and 6.

4.1 An “Electron” in an Electromagnetic Field A^μ

A free “spinless” electron satisfies the Klein–Gordon equation, (3.19):

$$(\partial_\mu \partial^\mu + m^2)\phi = 0.$$

In classical electrodynamics, the motion of a particle of charge $-e$ in an electromagnetic potential $A^\mu = (A^0, \mathbf{A})$ is obtained by the substitution

$$p^\mu \rightarrow p^\mu + eA^\mu \tag{4.1}$$

(see a standard text such as Goldstein or Jackson). The corresponding quantum-

mechanical substitution is therefore

$$i\partial^\mu \rightarrow i\partial^\mu + eA^\mu, \quad (4.2)$$

see (3.14); and the Klein–Gordon equation becomes

$$(\partial_\mu \partial^\mu + m^2)\phi = -V\phi, \quad (4.3)$$

where the (electromagnetic) perturbation is

$$V = -ie(\partial_\mu A^\mu + A^\mu \partial_\mu) - e^2 A^2. \quad (4.4)$$

The sign of V in (4.3) is chosen to be in accord with the relative sign of the kinetic and potential energy terms of the Schrödinger equation.

Everything is happening in (4.2)! This is quantum electrodynamics. (In Chapter 14, we shall see that this fundamental prescription emerges naturally from insisting that physics is unchanged under gauge, or phase, transformations.)

The potential, (4.4), is characterized by the parameter e , which (in natural units) is related to the fine structure constant α by

$$\alpha = \frac{e^2}{4\pi} \simeq \frac{1}{137}, \quad (4.5)$$

recall (1.3). The smallness of the electromagnetic coupling means that it is sensible to make a perturbation expansion of V in powers of α . The lowest-order (in α) contribution to a scattering amplitude should be a good approximation.

Working to lowest order, we omit the $e^2 A^2$ term in (4.4). The amplitude, (3.37), for the scattering of a “spinless” electron from a state ϕ_i to ϕ_f off an electromagnetic potential A_μ , which we represent by Fig. 4.1, is

$$\begin{aligned} T_{fi} &= -i \int \phi_f^*(x) V(x) \phi_i(x) d^4x \\ &= i \int \phi_f^* ie (\partial^\mu \partial_\mu + \partial_\mu A^\mu) \phi_i d^4x. \end{aligned} \quad (4.6)$$

The derivative, in the second term, which acts on both A^μ and ϕ_i , can be turned around by integration by parts, so that it acts on ϕ_f^* :

$$\int \phi_f^* \partial_\mu (A^\mu \phi_i) d^4x = - \int \partial_\mu (\phi_f^*) A^\mu \phi_i d^4x, \quad (4.7)$$

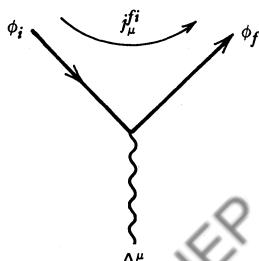


Fig. 4.1 A “spinless” electron interacting with A^μ .

where we have omitted the surface term as the potential is taken to vanish as $|x|$, $t \rightarrow \pm \infty$. We may therefore rewrite the amplitude T_{fi} in the suggestive form

$$T_{fi} = -i \int j_\mu^{fi} A^\mu d^4x, \quad (4.8)$$

where

$$j_\mu^{fi}(x) \equiv -ie(\phi_f^*(\partial_\mu \phi_i) - (\partial_\mu \phi_f^*) \phi_i) \quad (4.9)$$

which, by comparison with (3.25), can be regarded as the electromagnetic current for the $i \rightarrow f$ electron transition. If the ingoing spinless electron has four-momentum p_i , we have

$$\phi_i(x) = N_i e^{-ip_i \cdot x}, \quad (4.10)$$

where N_i is the normalization constant. Using a similar expression for ϕ_f , it follows that

$$j_\mu^{fi} = -eN_i N_f (p_i + p_f)_\mu e^{i(p_f - p_i) \cdot x}. \quad (4.11)$$

4.2 “Spinless” Electron–Muon Scattering

Using the results for the scattering of an electron off an electromagnetic potential A^μ , shown in Fig. 4.1, we are able to calculate the scattering of the same electron off another charged particle, say, another electron or a muon. Let us choose a muon to avoid dealing with identical particles. The Feynman diagram corresponding to the process is shown in Fig. 4.2. It suggests how to approach the problem. The calculation is an extension of the previous one; we just have to identify the electromagnetic potential A^μ with its source, the charged “spinless” muon. This identification is done using Maxwell’s equations

$$\square^2 A^\mu = j_{(2)}^\mu \quad (4.12)$$

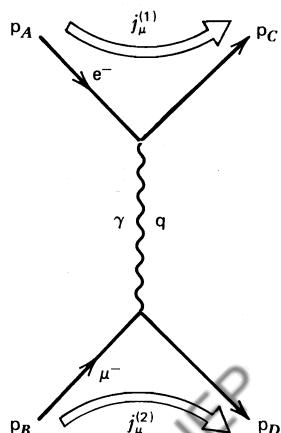


Fig. 4.2 Electron–muon scattering showing the particle four-momenta.

which determine the electromagnetic field A^μ associated with the current $j_{(2)}^\mu$ of the muon. (If you are unfamiliar with (4.12), that is, with the covariant form of Maxwell’s equations in the Lorentz gauge, work through Exercises 6.9 and 6.10.) Now what do we take for the current in (4.12)? Again, Fig. 4.2 suggests the answer. The current associated with a spinless muon has the same form as that for the electron, which is given by (4.11). Thus, we have

$$j_{(2)}^\mu = -eN_B N_D (p_D + p_B)^\mu e^{i(p_D - p_B) \cdot x}, \quad (4.13)$$

where the momenta are defined in Fig. 4.2. Since

$$\square^2 e^{iq \cdot x} = -q^2 e^{iq \cdot x}, \quad (4.14)$$

the solution of (4.12) is

$$A^\mu = -\frac{1}{q^2} j_{(2)}^\mu \quad \text{with } q = p_D - p_B. \quad (4.15)$$

Inserting this field due to the muon into (4.8), we find that the (lowest-order) amplitude for electron–muon scattering is

$$T_{fi} = -i \int j_\mu^{(1)}(x) \left(-\frac{1}{q^2} \right) j_{(2)}^\mu(x) d^4x. \quad (4.16)$$

Inserting (4.13), together with the corresponding expression for the electron current, (4.11), and carrying out the x integration, we find

$$T_{fi} = -iN_A N_B N_C N_D (2\pi)^4 \delta^{(4)}(p_D + p_C - p_B - p_A) \mathcal{M} \quad (4.17)$$

with

$$-i\mathcal{M} = (ie(p_A + p_C)^\mu) \left(-i \frac{g_{\mu\nu}}{q^2} \right) (ie(p_B + p_D)^\nu). \quad (4.18)$$

A consistency check on result (4.18) is that we get the same amplitude if we take the muon to be moving in the field A_μ produced by the electron. \mathcal{M} , as defined by (4.17), is known as the *invariant* amplitude. The delta function expresses energy–momentum conservation for the process.

In order to catalogue the different terms in the perturbative expansion of T_{fi} in nonrelativistic perturbation theory, we drew pictures like Fig. 3.4. Furthermore, the different factors in T_{fi} as given by (3.44) were associated with interaction vertices and particle propagators in Fig. 3.4b. It is useful to draw the same pictures for the covariant form of the perturbation series. For example, Fig. 4.3 represents spinless electron–muon scattering to order e^2 (or α), and the amplitude is given by (4.17) and (4.18). This is the lowest-order Feynman diagram. The wavy line represents a photon exchanged between the leptons, and the associated factor $-ig_{\mu\nu}/q^2$ is called the *photon propagator*; it carries Lorentz indices because the photon is a spin-1 particle (see Chapter 6). The four-momentum q of the photon is determined by four-momentum conservation at the vertices. We see that

$q^2 \neq 0$, and we say the photon is “virtual” or “off-mass shell.” At each of the vertices of the diagram, we associate the factor shown. Each *vertex factor* contains the electromagnetic coupling e and a four-vector index to connect with the photon index. The particular distribution of minus signs and factors i has been made to give the correct results for higher-order diagrams. Note that the multiplication of the three factors gives $-i\mathcal{M}$.

Whenever the same vertex or internal line occurs in a Feynman diagram, the corresponding factor will contribute multiplicatively to the amplitude $-i\mathcal{M}$ for that diagram. We may thus start to draw up a table of Feynman rules for quantum electrodynamics, which will, when complete, allow us to quickly write down the expression for the amplitude, $-i\mathcal{M}$, for any Feynman diagram. The table is given in Section 6.17.

4.3 The Cross Section in Terms of the Invariant Amplitude \mathcal{M}

To relate these calculations to experimental observables, we need to fix the normalization N of our free particle wavefunctions:

$$\phi = Ne^{-ip \cdot x}. \quad (4.19)$$

Recall from (3.18) that the probability density ρ of particles described by ϕ is

$$\rho = 2E|N|^2.$$

The proportionality of ρ to E was just what we needed to compensate for the Lorentz contraction of the volume element d^3x and to keep the number of particles ρd^3x unchanged. Let us therefore work in a volume V and normalize to $2E$ particles in V ,

$$\int_V \rho dV = 2E. \quad (4.20)$$

That is, we adopt a covariant normalization

$$N = \frac{1}{\sqrt{V}}. \quad (4.21)$$

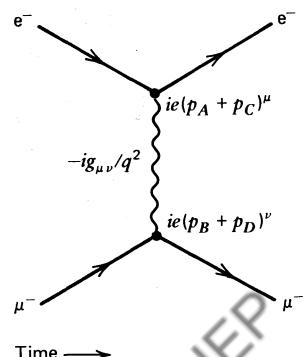


Fig. 4.3 The vertex factors and propagator for “spinless” electron–muon scattering.

Now, the $A + B \rightarrow C + D$ transition rate per unit volume is

$$W_{fi} = \frac{|T_{fi}|^2}{TV}$$

where T is the time interval of the interaction, and the transition amplitude is

$$T_{fi} = -iN_A N_B N_C N_D (2\pi)^4 \delta^{(4)}(p_C + p_D - p_A - p_B) \mathcal{M},$$

see (4.17). On squaring, one delta function remains, and $(2\pi)^4$ times the other gives TV [the calculation is identical to the one that led to (3.41)]. Thus, making use of (4.21), we obtain

$$W_{fi} = (2\pi)^4 \frac{\delta^{(4)}(p_C + p_D - p_A - p_B) |\mathcal{M}|^2}{V^4}. \quad (4.22)$$

Experimental results on $AB \rightarrow CD$ scattering are usually quoted in the form of a “cross section.” It is related to the transition rate by

$$\text{Cross section} = \frac{W_{fi}}{(\text{initial flux})} (\text{number of final states}), \quad (4.23)$$

where the factors in brackets allow for the “density” of the incoming and outgoing states. We first carefully define these factors and then show how the cross section, so defined, may be regarded as the effective area over which particles A, B interact to produce C, D.

For a single particle, quantum theory restricts the *number of final states* in a volume V with momenta in element $d^3 p$ to be $V d^3 p / (2\pi)^3$ (see Exercise 4.1). But we have $2E$ particles in V , and so

$$\text{No. of final states/particle} = \frac{V d^3 p}{(2\pi)^3 2E}. \quad (4.24)$$

Thus, for particles C, D scattered into momentum elements $d^3 p_C, d^3 p_D$,

$$\text{No. of available final states} = \frac{V d^3 p_C}{(2\pi)^3 2E_C} \frac{V d^3 p_D}{(2\pi)^3 2E_D}. \quad (4.25)$$

EXERCISE 4.1 Working in a box of volume $V = L^3$, show that the number of allowed states of momentum with x component in the range p_x to $p_x + dp_x$ is $(L/2\pi) dp_x$. Convince yourself that you need to impose periodic boundary conditions on the wavefunction and its derivative to ensure no net particle flow out of the volume.

Turning now to the *initial flux*, we find that it is easiest to calculate it in the laboratory frame. The number of beam particles passing through unit area per unit time is $|v_A| 2E_A / V$, and the number of target particles per unit volume is $2E_B / V$. To obtain a normalization-independent measure of the ingoing “density,”

we therefore take

$$\text{Initial flux} = |\mathbf{v}_A| \frac{2E_A}{V} \frac{2E_B}{V}. \quad (4.26)$$

Inserting (4.22), (4.25), and (4.26) into (4.23), we arrive at a differential cross section $d\sigma$ for scattering into $d^3 p_C d^3 p_D$:

$$d\sigma = \frac{V^2}{|\mathbf{v}_A| 2E_A 2E_B} \frac{1}{V^4} |\mathcal{M}|^2 \frac{(2\pi)^4}{(2\pi)^6} \delta^{(4)}(\mathbf{p}_C + \mathbf{p}_D - \mathbf{p}_A - \mathbf{p}_B) \frac{d^3 p_C}{2E_C} \frac{d^3 p_D}{2E_D} V^2. \quad (4.27)$$

The arbitrary normalization volume V cancels, as indeed it must. From now on, we drop V and work in unit volume. That is, we normalize to $2E$ particles/unit volume, and the normalization factor (4.21) of the wavefunction is

$$N = 1. \quad (4.28)$$

This is the origin of the multiplicative factors $N = 1$ which are associated with the external lines of spinless particles (see the table of Feynman rules in Section 6.17).

What is the physical interpretation of the cross section as defined by (4.23) [and (4.27)]? The appearance of the number of final states in conjunction with the transition rate W_{fi} is familiar, see (3.42); it represents the number of scatters, n_s , per unit time. The flux is introduced in (4.23) to make the rate independent of the number of particles present in the beam or target used in a particular experimental setup. That is, we want the cross section to represent an intrinsic scattering probability, that is, the intrinsic strength of the $AB \rightarrow CD$ interaction. We therefore divide in (4.23) by the number of particles in the target (n_t) and the flux of the beam ($n_b v_b$) which counts the number of beam particles traversing a unit area perpendicular to the beam velocity per unit time (v_b is the relative velocity of beam and target if the latter is not stationary). Thus, (4.23) can be symbolically written as

$$n_s = (n_b v_b) n_t \sigma.$$

The counting rate n_s is always proportional to the (beam flux \times n_t); it is the proportionality constant σ which contains the physics, that is, the intrinsic scattering probability. It has the units of area. This can be checked by realizing that the left-hand side has units $(\text{time})^{-1}$ and that $(n_b v_b) n_t$ has the units of flux, namely, $(\text{area} \times \text{time})^{-1}$. We can interpret σ intuitively as the effective area of the beam seen by a target particle, or the area over which A, B interact to produce C, D.

We may write the differential cross section, (4.27), in the symbolic form

$$d\sigma = \frac{|\mathcal{M}|^2}{F} dQ, \quad (4.29)$$

where dQ is the Lorentz invariant phase space factor (sometimes written as $d\text{Lips}$)

$$dQ = (2\pi)^4 \delta^{(4)}(p_C + p_D - p_A - p_B) \frac{d^3 p_C}{(2\pi)^3 2E_C} \frac{d^3 p_D}{(2\pi)^3 2E_D} \quad (4.30)$$

(recall $d^3 p/E$ is a Lorentz invariant quantity), and the incident flux in the laboratory is

$$F = |\mathbf{v}_A| 2E_A \cdot 2E_B, \quad (4.31)$$

with $\mathbf{v}_A = \mathbf{p}_A/E_A$. For a general collinear collision between A and B,

$$\begin{aligned} F &= |\mathbf{v}_A - \mathbf{v}_B| \cdot 2E_A \cdot 2E_B \\ &= 4(|\mathbf{p}_A|E_B + |\mathbf{p}_B|E_A) \\ &= 4((p_A \cdot p_B)^2 - m_A^2 m_B^2)^{1/2}, \end{aligned} \quad (4.32)$$

which is manifestly invariant.

Equation (4.29) is the final result. We see that the physics resides in the invariant amplitude \mathfrak{M} . To relate the observable rate $d\sigma$ to this universal measure of the interaction, we need to include the “bookkeeping” factors dQ and F .

EXERCISE 4.2 In the center-of-mass frame for the process $AB \rightarrow CD$, show that

$$dQ = \frac{1}{4\pi^2} \frac{p_f}{4\sqrt{s}} d\Omega \quad (4.33)$$

$$F = 4p_i\sqrt{s}, \quad (4.34)$$

and hence that the differential cross section is

$$\boxed{\frac{d\sigma}{d\Omega} \Big|_{cm} = \frac{1}{64\pi^2 s} \frac{p_f}{p_i} |\mathfrak{M}|^2}, \quad (4.35)$$

where $d\Omega$ is the element of solid angle about \mathbf{p}_C , $s = (E_A + E_B)^2$, $|\mathbf{p}_A| = |\mathbf{p}_B| = p_i$ and $|\mathbf{p}_C| = |\mathbf{p}_D| = p_f$.

EXERCISE 4.3 Use (4.18) to show that for very high-energy “spinless” electron–muon scattering,

$$\frac{d\sigma}{d\Omega} \Big|_{cm} = \frac{\alpha^2}{4s} \left(\frac{3 + \cos\theta}{1 - \cos\theta} \right)^2$$

where θ is the scattering angle and $\alpha = e^2/4\pi$. Neglect the particle masses.

4.4 The Decay Rate in Terms of \mathcal{M}

The derivation of the formula for particle decay rates proceeds along similar lines. The differential rate for the decay $A \rightarrow 1 + 2 + \dots + n$ into momentum elements $d^3 p_1, \dots, d^3 p_n$ of the final state particles is

$$d\Gamma = \frac{1}{2E_A} |\mathcal{M}|^2 \frac{d^3 p_1}{(2\pi)^3 2E_1} \dots \frac{d^3 p_n}{(2\pi)^3 2E_n} (2\pi)^4 \delta^{(4)}(p_A - p_1 - \dots - p_n). \quad (4.36)$$

The formula has the form of (4.29) and (4.30). $2E_A$ is the number of decaying particles per unit volume and \mathcal{M} is the invariant amplitude which has been computed from the relevant Feynman diagram. One common application is the calculation of the integrated rate for the decay mode $A \rightarrow 1 + 2$, that is, we integrate (4.36) over all possible momenta $\mathbf{p}_1, \mathbf{p}_2$. In the rest frame of A, we find, using (4.33),

$$\Gamma(A \rightarrow 1 + 2) = \frac{p_f}{32\pi^2 m_A^2} \int |\mathcal{M}|^2 d\Omega. \quad (4.37)$$

The total decay rate, Γ , is the sum of the rates for all the decay channels. Clearly, the rate

$$\Gamma = -\frac{dN_A}{dt}/N_A, \quad (4.38)$$

which leads to the exponential decay law for the number of A particles

$$N_A(t) = N_A(0) e^{-\Gamma t}. \quad (4.39)$$

We say Γ^{-1} is the lifetime of particle A.

4.5 “Spinless” Electron–Electron Scattering

We return to the application of the Feynman rules to some sample processes. For electron–electron scattering, the new feature is that we have identical particles in the initial and the final states, and so the amplitude should be symmetric under interchange of particle labels C \leftrightarrow D (and A \leftrightarrow B). Consequently, in addition to the Feynman diagram of Fig. 4.4a, we have a second diagram, Fig. 4.4b, which, to maintain the order of A, B, C, and D, is drawn as Fig. 4.4c.

There is no way to experimentally distinguish whether electron C came from A or B, so we must add amplitudes (rather than probabilities). Thus, the invariant amplitude for the scattering of spinless electrons is, to lowest order, the sum of the amplitudes for diagrams (a) and (c):

$$-i\mathcal{M}_{e^-e^-} = -i \left(-\frac{e^2(p_A + p_C)_\mu (p_B + p_D)^\mu}{(p_D - p_B)^2} - \frac{e^2(p_A + p_D)_\mu (p_B + p_C)^\mu}{(p_C - p_B)^2} \right). \quad (4.40)$$

The first term follows directly from (4.18), and the second is simply that with

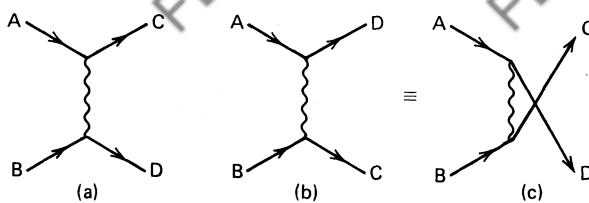


Fig. 4.4 The two (lowest-order) Feynman diagrams for electron–electron scattering.

$p_C \leftrightarrow p_D$. Note that symmetry under $p_C \leftrightarrow p_D$ ensures that \mathcal{M} is also symmetric under $p_A \leftrightarrow p_B$.

4.6 Electron–Positron Scattering: An Application of Crossing

Again, here we have two possible Feynman diagrams, Figs. 4.5a and 4.5c. We are working only in terms of particle (electron) states, and so we must use the antiparticle prescription, (3.28), to translate these to diagrams (b) and (d), respectively. We can use the Feynman rules (obtained above and collected in Section 6.17) to calculate the (lowest-order) $e^-e^+ \rightarrow e^-e^+$ amplitude

$$-i\mathcal{M}_{e^-e^+} = -i \left(-e^2 \frac{(p_A + p_C)_\mu (-p_D - p_B)^\mu}{(p_D - p_B)^2} - e^2 \frac{(p_A - p_B)_\mu (-p_D + p_C)^\mu}{(p_C + p_D)^2} \right). \quad (4.41)$$

Note that, for example, the factor at the B, D vertex of diagram (b) has simply changed sign from that of Fig. 4.4(a), which is just what we expect in going from

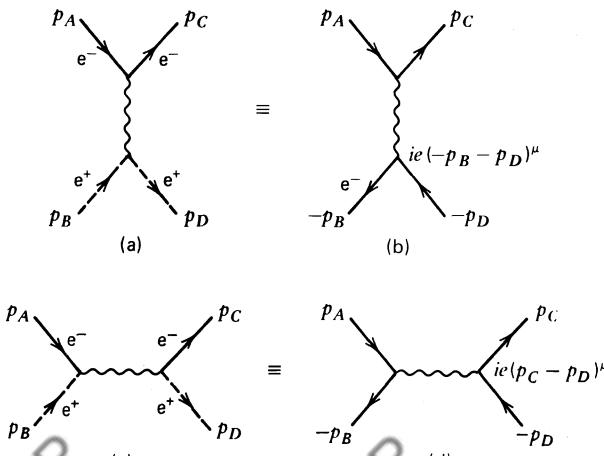


Fig. 4.5 The two (lowest-order) Feynman diagrams, (b) and (d), for spinless $e^-e^+ \rightarrow e^-e^+$ scattering.

a charge $-e$ to a $+e$ vertex. We also observe that \mathcal{M} is symmetric under $p_C \leftrightarrow -p_B$, that is, under the interchange of the two “outgoing” electrons.

But we need not do this; to obtain $\mathcal{M}_{e^-e^+}$, we can simply use the antiparticle prescription to “cross” the result we derived for $\mathcal{M}_{e^-e^-}$ (see Fig. 4.6). In this way, we obtain

$$\mathcal{M}_{e^-e^+\rightarrow e^-e^+}(p_A, p_B, p_C, p_D) = \mathcal{M}_{e^-e^-\rightarrow e^-e^-}(p_A, -p_D, p_C, -p_B). \quad (4.42)$$

Indeed, we see that (4.41) is simply (4.40) with $p_D \leftrightarrow -p_B$.

EXERCISE 4.4 Use the Feynman rules to obtain the invariant amplitudes for the “spinless” processes $e^- \mu^+ \rightarrow e^- \mu^+$ and $e^- e^+ \rightarrow \mu^- \mu^+$. Check your answers by appropriately crossing the $e^- \mu^- \rightarrow e^- \mu^-$ amplitude of Section 4.2.

4.7 Invariant Variables

For a scattering process of the form $AB \rightarrow CD$, we expect two independent kinematic variables; for example, the incident energy and the scattering angle. It is however possible, and desirable, to express the invariant amplitude \mathcal{M} as a function of variables invariant under Lorentz transformations. We have at our disposal the particle four-momenta, and so possible invariant variables are the scalar products $p_A \cdot p_B$, $p_A \cdot p_C$, $p_A \cdot p_D$. Since $p_i^2 = m_i^2$ [see (3.10)], and since $p_A + p_B = p_C + p_D$ due to energy-momentum conservation, only two of the three variables are independent. Rather than these, it is conventional to use the related (Mandelstam) variables

$$\begin{aligned} s &= (p_A + p_B)^2, \\ t &= (p_A - p_C)^2, \\ u &= (p_A - p_D)^2. \end{aligned} \quad (4.43)$$

EXERCISE 4.5 Show that

$$s + t + u = m_A^2 + m_B^2 + m_C^2 + m_D^2, \quad (4.44)$$

where m_i is the rest mass of particle i .

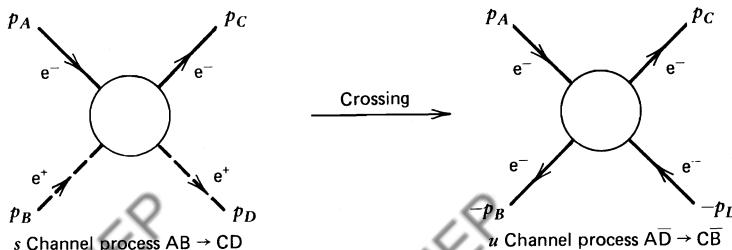


Fig. 4.6 The crossing (or interchange) of particles B and D.

To display the kinematic (or physical) regions of processes related by crossing, we construct a two-dimensional plot which maintains the symmetry of s , t , u . The three axes s , t , $u = 0$ are drawn (Fig. 4.7) to form an equilateral triangle of height Σm_i^2 . From any point inside and also outside (if attention is paid to the signs of s , t , u) the triangle, the sum of the perpendicular distances to the axes is equal to the height of the triangle [see (4.44)].

It is easy to show that s is the square of the total center-of-mass energy of the process $AB \rightarrow CD$ [see (3.12) or Exercise 4.6]. It is conventional to take the reaction under study to be the s channel process. In our last example, this was $e^-e^+ \rightarrow e^-e^+$ scattering. The crossed reactions $\bar{A}\bar{D} \rightarrow C\bar{B}$ and $\bar{D}\bar{B} \rightarrow C\bar{A}$ are called the u and t channels, respectively, since u and t are equal to the square of the total center-of-mass energy in the respective channels (see Exercise 4.7).

EXERCISE 4.6 Taking $e^-e^+ \rightarrow e^-e^+$ to be the s channel process, verify that

$$s = 4(k^2 + m^2)$$

$$t = -2k^2(1 - \cos \theta) \quad (4.45)$$

$$u = -2k^2(1 + \cos \theta)$$

where θ is the center-of-mass scattering angle and $k = |\mathbf{k}_i| = |\mathbf{k}_f|$, where \mathbf{k}_i and \mathbf{k}_f are, respectively, the momenta of the incident and scattered electrons

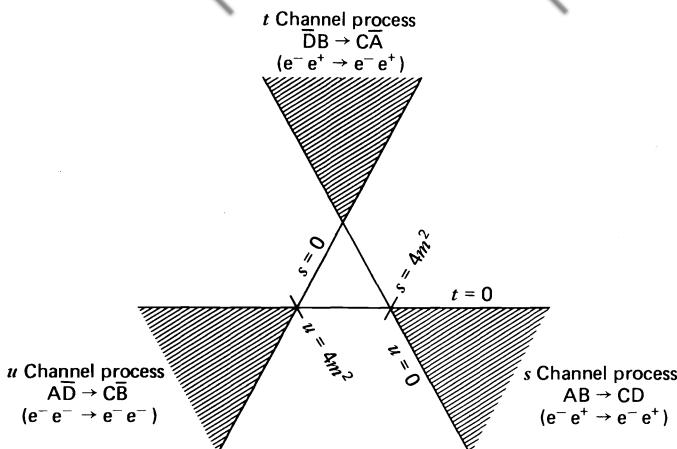


Fig. 4.7 The (Mandelstam) s , t , u plot showing the physical regions for $e^-e^+ \rightarrow e^-e^+$ and the crossed reactions. For scattering between particles of unequal masses, the boundaries of the physical regions are more complicated, but the general result of three nonoverlapping regions holds true (see Exercise 4.8).

in the center-of-mass frame. Show that the process is physically allowed provided $s \geq 4m^2$, $t \leq 0$, and $u \leq 0$. The physical region is shown shaded on Fig. 4.7. Note that $t = 0$ ($u = 0$) corresponds to forward (backward) scattering.

EXERCISE 4.7 For the crossed reaction $A\bar{D} \rightarrow C\bar{B}$ ($e^-e^- \rightarrow e^-e^-$), show that u becomes the square of the total center-of-mass energy and that this process would become physical in a different kinematic region: $u \geq 4m^2$, $t \leq 0$, and $s \leq 0$. (Note that, for example, $-p_D = (E, \mathbf{p})$, where E and \mathbf{p} refer to the incoming \bar{D}).

EXERCISE 4.8 If the s channel process is $e^- \mu^- \rightarrow e^- \mu^-$, show that the boundaries of the physical regions of this and the crossed channel reactions are given by

$$t = 0, \quad su = (M^2 - m^2)^2$$

where m and M are the electron and muon masses, respectively. Construct the Mandelstam plot.

EXERCISE 4.9 Verify that crossing relation (4.42) is of the form

$$\mathcal{M}_{e^-e^+}(s, t, u) = \mathcal{M}_{e^-e^-}(u, t, s) \quad (4.46)$$

EXERCISE 4.10 Show that the invariant amplitude, (4.41), for “spinless” electron–positron scattering can be written as

$$\mathcal{M}_{e^-e^+}(s, t, u) = e^2 \left(\frac{s-u}{t} + \frac{t-u}{s} \right). \quad (4.47)$$

Comment on the symmetry of \mathcal{M} under $s \leftrightarrow t$.

Let us look back at the amplitude for “spinless” electron–electron scattering. The amplitude (4.40) is derived taking, of course, the process to be $AB \rightarrow CD$,

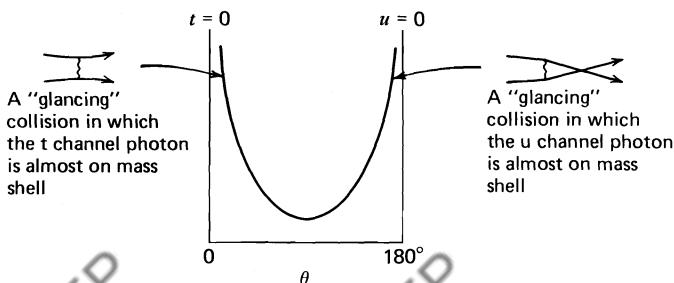


Fig. 4.8 The differential cross section, $d\sigma/d\Omega$, for electron–electron scattering.

that is, the s channel process. In terms of invariant variables, (4.40) becomes

$$\mathcal{M}_{e^- e^-} = e^2 \left(\frac{u-s}{t} + \frac{t-s}{u} \right).$$

The resulting cross section is sketched in Fig. 4.8, and the origin of the forward and backward peaks is identified; $-t$ and $-u$ are the squares of the three-momentum transferred in Figs. 4.4a and 4.4c, respectively, that is, of the momentum carried by the virtual photon. When the photon has a very small momentum squared ($-q^2$), that is, almost on its mass shell, then by the uncertainty principle the range of the interaction is very large. Interactions with small deflections therefore occur with large cross sections.

4.8 The Origin of the Propagator

We saw earlier (Section 4.2) that a virtual photon line in a Feynman diagram corresponds to a propagator $1/q^2$, where q is the four-momentum carried by the virtual photon. For example, the photon propagator in the annihilation process $e^- e^+ \rightarrow \gamma \rightarrow e^- e^+$ of Fig. 4.9 is of the form $1/q^2$, where $q = p_A + p_B$ is given by four-momentum conservation.

In general (aside from complications of spin), a particle of mass m has a propagator $1/(p^2 - m^2)$. Feynman (1962) has given a nice explanation of why this is so. For instance, for the photon propagator of Fig. 4.9, the argument goes as follows. There are two interaction vertices, and so we must be able to interpret the result as the relativistic generalization of the second-order term in perturbation series (3.44),

$$T_{fi}^{(2)} = -i \sum_{n \neq i} V_{fn} \frac{1}{E_i - E_n} V_{ni} 2\pi \delta(E_f - E_i). \quad (4.48)$$

Here, we are not interested in the precise details [a detailed discussion is given by Aitchison (1972)], but rather to simply see how the propagator generalizes:

$$\frac{1}{E_i - E_n} \rightarrow \frac{1}{(p_A + p_B)^2} ? \quad (4.49)$$

We take the energy difference to refer to relativistic energies.

A Feynman diagram is the sum of all possible time-ordered diagrams. There are two possible time-ordered diagrams corresponding to Fig. 4.9. These are

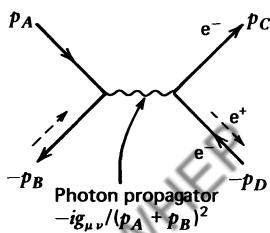


Fig. 4.9 Annihilation diagram $e^- e^+ \rightarrow \gamma \rightarrow e^- e^+$, drawn using only particle (electron) lines.

shown in Fig. 4.10. The resulting amplitude is thus of the form

$$\mathcal{M} \sim V_{fn} \frac{1}{E_i - E_\gamma} V_{ni} + V_{fn} \frac{1}{E_i - 2E_i - E_\gamma} V_{ni} = V_{fn} \frac{2E_\gamma}{E_i^2 - E_\gamma^2} V_{ni}, \quad (4.50)$$

where the factor $2E_\gamma$ has to do with normalization. This method of calculating the amplitude is now often referred to as old-fashioned perturbation theory, OFPT. In OFPT, three-momentum is conserved at a vertex, but not energy, as pointed out in Chapter 3 (clearly a noninvariant situation); moreover, particles stay on mass shell. To determine the propagator, we calculate

$$E_i^2 = (p_A + p_B)^2 + (\mathbf{p}_A + \mathbf{p}_B)^2,$$

$$E_\gamma^2 = m_\gamma^2 + \mathbf{p}^2.$$

Now, since $\mathbf{p} = \mathbf{p}_A + \mathbf{p}_B$, we obtain

$$\frac{1}{E_i^2 - E_\gamma^2} = \frac{1}{(p_A + p_B)^2 - m_\gamma^2} = \frac{1}{q^2}. \quad (4.51)$$

We have displayed the photon mass ($m_\gamma = 0$) until the last equality so as to be able to compare with propagators for $m \neq 0$ particles. The relativistic generalization of the propagator for a spinless particle of mass m is [see (4.51)]

$$\frac{1}{(p_A + p_B)^2 - m^2} = \frac{1}{p^2 - m^2}. \quad (4.52)$$

Each of the two time-ordered diagrams of Fig. 4.10 (considered separately) is not invariant; but by including the second term together with the standard nonrelativistic result, we have obtained an invariant expression.

Another example is the electron propagator in the process $\gamma e^- \rightarrow \gamma e^-$. In this case, we make a pictorial comparison of OFPT and covariant perturbation theory.

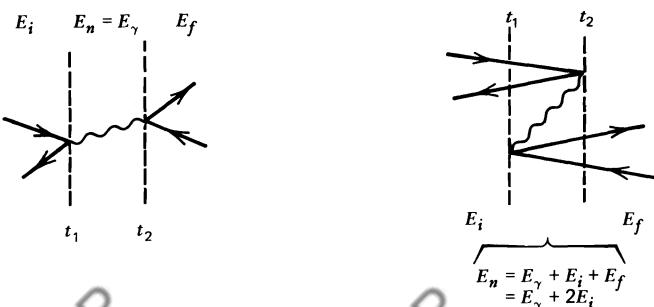
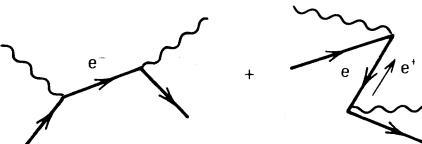
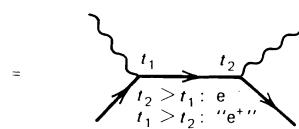


Fig. 4.10 Time-ordered diagrams for $e^- e^+ \rightarrow \gamma \rightarrow e^- e^+$.

OFPT
Time-ordered diagrams



Covariant perturbation theory
Feynman diagram



Clearly not an invariant separation, but the sum is invariant

Three-momentum is conserved at each vertex, but not energy. The intermediate particle is on mass shell, $p^2 = m^2$.

Four-momentum is conserved at each vertex, see (4.15). The intermediate particle is not on mass shell, $p^2 \neq m^2$.

The single Feynman diagram embodies the possibility that the intermediate particle is an electron ($t_2 > t_1$) or a positron ($t_1 > t_2$). How this “magic” comes about is explained in Section 6.16; there, we also discuss how to handle the singularity in the propagator at $p^2 = m^2$.

4.9 Summary

We now have shown how to do nonrelativistic perturbation theory in a covariant way. The crucial step was to exploit the fact that the expression (3.37) for T_{fi} was already covariant. The invariant amplitude \mathcal{M} [related to T_{fi} by (4.22)] is calculated by identifying the covariant replacements for the vertex factors V_{ni} and the propagators $1/(E_n - E_i)$. We explicitly derived their form for the electromagnetic interaction of spinless charged particles. An important difference with the nonrelativistic formalism should be remembered: energy, as well as three-momentum, is conserved at each vertex. Finally, the procedure for relating the invariant amplitude to observables was discussed.

This in fact concludes the discussion of relativistic perturbation theory and Feynman diagrams. The rest of the work is purely technical in nature. Electrons, muons, and quarks have spin $\frac{1}{2}$. They therefore satisfy the Dirac equation, not the Klein–Gordon equation. Repeating the procedure described in this chapter for particles satisfying the Dirac equation will clearly lead to modified results for vertex factors and propagators. The calculational procedure, however, is the same.

5

The Dirac Equation

At this point, we would have mastered the computational techniques to attack the particle physics problems presented in the rest of the book if it were not for the fact that quarks and leptons are spin- $\frac{1}{2}$ particles. We constructed the Feynman rules for particles (and antiparticles) described by wavefunctions ϕ that satisfy the Klein–Gordon equation. These wavefunctions do not have the required two-component structure to accommodate, for instance, the spins of the electron and positron. We are looking for a relativistic equation with solutions that have two-component structure for both particle and antiparticle. For some time, it was thought that the Klein–Gordon equation was the only relativistic generalization of the Schrödinger equation until Dirac discovered an alternative one. His goal was to write an equation which, unlike the Klein–Gordon equation, was linear in $\partial/\partial t$. In order to be covariant, it must then also be linear in ∇ and has therefore the general form

$$H\psi = (\alpha \cdot \mathbf{P} + \beta m)\psi. \quad (5.1)$$

The four coefficients β and α_i ($i = 1, 2, 3$) are determined by the requirement that a free particle must satisfy the relativistic energy–momentum relation (3.10),

$$H^2\psi = (\mathbf{P}^2 + m^2)\psi. \quad (5.2)$$

Equations (5.1) and (5.2) represent the Dirac equation. We will show that its solutions have sufficiently rich structure to describe spin- $\frac{1}{2}$ particles and antiparticles.

The historical impact of Dirac's suggestion is most profound and goes far beyond providing us with a relativistic equation to describe fermions, which is our current interest. Its study led to developments ranging from quantum field theory to semiconductors and beyond. Note, however, that Dirac's original motivation for linearizing the Klein–Gordon equation in $\partial/\partial t$ was not to explain spin but to remove negative probability densities. The appearance of $\partial/\partial t$ in the probability (3.17) is indeed at the root of this “problem.” However, for us this feature of the Klein–Gordon equation is no longer a problem. It is a bonus that allows the correct treatment of antiparticles, at least when they have no spin!

Let us forget history and study how (5.1) describes leptons (or quarks) with spin. From (5.1), we have

$$H^2\psi = (\alpha_i P_i + \beta m)(\alpha_j P_j + \beta m)\psi = (\underbrace{\alpha_i^2 P_i^2}_{\downarrow} + \underbrace{(\alpha_i \alpha_j + \alpha_j \alpha_i) P_i P_j}_{0} + \underbrace{(\alpha_i \beta + \beta \alpha_i) P_i m + \beta^2 m^2}_{0})\psi,$$

where we sum over repeated indices, with the condition $i > j$ on the second term. Comparing with (5.2), we see that

- $\alpha_1, \alpha_2, \alpha_3, \beta$ all anticommute with each other,
 - $\alpha_1^2 = \alpha_2^2 = \alpha_3^2 = \beta^2 = 1$.
- (5.3)

Since the coefficients α_i and β do not commute, they cannot simply be numbers, and we are led to consider matrices operating on a wavefunction ψ , which is a multicomponent column vector.

EXERCISE 5.1 Prove that the α_i and β are hermitian, traceless matrices of even dimensionality, with eigenvalues ± 1 .

The lowest dimensionality matrices satisfying all these requirements are 4×4 . The choice of the four matrices (α, β) is not unique. The Dirac–Pauli representation is most frequently used:

$$\alpha = \begin{pmatrix} 0 & \sigma \\ \sigma & 0 \end{pmatrix}, \quad \beta = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} \quad (5.4)$$

where I denotes the unit 2×2 matrix (which is frequently written as 1) and where σ are the Pauli matrices:

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (5.5)$$

Another possible representation, the Weyl representation, is

$$\alpha = \begin{pmatrix} -\sigma & 0 \\ 0 & \sigma \end{pmatrix}, \quad \beta = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}. \quad (5.6)$$

Most of the results are independent of the choice of representation. Certainly, all the physics depends only on the properties listed in (5.3). In fact, not until we exhibit explicit solutions of the Dirac equation in Section 5.3 will we use a particular representation. Unless stated otherwise, we shall always choose the Dirac–Pauli representation, (5.4).

A four-component column vector ψ which satisfies the Dirac equation (5.1) is called a Dirac spinor. We might have anticipated two independent solutions (particles and antiparticles), but instead we have four!

Maybe the surprise should not have been total. We know at least one other example where a field with more components appears when linearizing the

equation. The covariant Maxwell equations $\square^2 A_\mu = 0$ are second-order but can be written in a linear form $\partial_\mu F^{\mu\nu} = 0$ by introducing the field strength $F_{\mu\nu}$, which has more components than A_μ .

5.1 Covariant Form of the Dirac Equation. Dirac γ -Matrices

On multiplying Dirac's equation, (5.1), by β from the left, we obtain

$$i\beta \frac{\partial \psi}{\partial t} = -i\beta \alpha \cdot \nabla \psi + m\psi,$$

which may be rewritten

$$(i\gamma^\mu \partial_\mu - m)\psi = 0, \quad (5.7)$$

where we have introduced four Dirac γ -matrices

$$\gamma^\mu \equiv (\beta, \beta\alpha). \quad (5.8)$$

Equation (5.7) is called the covariant form of the Dirac equation. We must wait until Section 5.2 to understand the sense in which the four 4×4 matrices $\gamma^0, \gamma^1, \gamma^2, \gamma^3$ are to be regarded as a four-vector. The Dirac equation is really four differential equations which couple the four components of a single column vector ψ :

$$\sum_{k=1}^4 \left[\sum_\mu i(\gamma^\mu)_{jk} \partial_\mu - m\delta_{jk} \right] \psi_k = 0.$$

Using (5.3) and (5.8), it is straightforward to show that the Dirac γ -matrices satisfy the anticommutation relations

$$\gamma^\mu \gamma^\nu + \gamma^\nu \gamma^\mu = 2g^{\mu\nu}. \quad (5.9)$$

Moreover, since $\gamma^0 = \beta$, we have

$$\gamma^{0\dagger} = \gamma^0, \quad (\gamma^0)^2 = I \quad (5.10)$$

and

$$\left. \begin{aligned} \gamma^{k\dagger} &= (\beta\alpha^k)^\dagger = \alpha^k\beta = -\gamma^k \\ (\gamma^k)^2 &= \beta\alpha^k\beta\alpha^k = -I \end{aligned} \right\} \quad k = 1, 2, 3. \quad (5.11)$$

Note that the hermitian conjugation results can be summarized by

$$\gamma^{\mu\dagger} = \gamma^0 \gamma^\mu \gamma^0.$$

5.2 Conserved Current and the Adjoint Equation

In order to construct the currents, we proceed as for the Klein–Gordon equation, see (3.17), except that, as we now have a matrix equation, we must consider the hermitian, rather than the complex, conjugate equation. The hermitian conjugate

of Dirac's equation,

$$i\gamma^0 \frac{\partial \psi}{\partial t} + i\gamma^k \frac{\partial \psi}{\partial x^k} - m\psi = 0 \quad (5.12)$$

where $k = 1, 2, 3$, is

$$-i \frac{\partial \psi^\dagger}{\partial t} \gamma^0 - i \frac{\partial \psi^\dagger}{\partial x^k} (-\gamma^k) - m\psi^\dagger = 0. \quad (5.13)$$

To restore the covariant form, we need to remove the minus sign of $-\gamma^k$ while leaving the first term unchanged. Since $\gamma^0 \gamma^k = -\gamma^k \gamma^0$, this can be done by multiplying (5.13) from the right by γ^0 . Introducing the adjoint (row) spinor

$$\bar{\psi} \equiv \psi^\dagger \gamma^0, \quad (5.14)$$

we obtain

$$i\partial_\mu \bar{\psi} \gamma^\mu + m\bar{\psi} = 0. \quad (5.15)$$

We can now derive a continuity equation, $\partial_\mu j^\mu = 0$, by multiplying (5.7) from the left by $\bar{\psi}$ and (5.15) from the right by ψ , and adding. We find

$$\bar{\psi} \gamma^\mu \partial_\mu \psi + (\partial_\mu \bar{\psi}) \gamma^\mu \psi = \partial_\mu (\bar{\psi} \gamma^\mu \psi) = 0.$$

Thus, we see that

$$j^\mu = \bar{\psi} \gamma^\mu \psi$$

satisfies the continuity equation, which suggests that we should identify j^μ with the probability and flux densities, ρ and \mathbf{j} . The probability density

$$\rho \equiv j^0 = \bar{\psi} \gamma^0 \psi = \psi^\dagger \psi = \sum_{i=1}^4 |\psi_i|^2 \quad (5.16)$$

is now positive definite. As previously remarked, this result historically motivated Dirac's work.

However, from the Pauli–Weisskopf prescription in Chapter 3, we saw that $j^\mu = (\rho, \mathbf{j})$ should be identified with the charge current density. We therefore insert the charge $-e$ in j^μ ,

$$j^\mu = -e \bar{\psi} \gamma^\mu \psi, \quad (5.17)$$

and from now on regard j^μ as the electron (four-vector) current density. Recall that the reason for $-e$ is that the electron (rather than the positron) is regarded as the particle.

For the covariance of the continuity equation $\partial_\mu j^\mu = 0$, it is necessary that j^μ transforms as a four-vector. This can indeed be proved; that is, using the four Dirac matrices γ^μ , we can form a four-vector $\bar{\psi} \gamma^\mu \psi$ (cf. Section 5.6).

5.3 Free-Particle Spinors

EXERCISE 5.2 Operate on (5.7) with $\gamma^\nu \partial_\nu$, and show that each of the four components ψ_i satisfies the Klein–Gordon equation

$$(\square^2 + m^2)\psi_i = 0.$$

For a free particle, we can therefore seek four-momentum eigensolutions of Dirac's equation of the form

$$\psi = u(\mathbf{p})e^{-ip\cdot x}, \quad (5.18)$$

where u is a four-component spinor independent of x . Substituting in (5.7), we have

$$(\gamma^\mu p_\mu - m)u(\mathbf{p}) = 0 \quad (5.19)$$

or, using the abbreviated notation $A \equiv \gamma^\mu A_\mu$ for any four-vector A_μ ,

$$(\not{p} - m)u = 0. \quad (5.20)$$

Since we are seeking energy eigenvectors, it is easier to use the original form, (5.1),

$$Hu = (\alpha \cdot \mathbf{p} + \beta m)u = Eu. \quad (5.21)$$

There are four independent solutions of this equation, two with $E > 0$ and two with $E < 0$. This is particularly easy to see in the Dirac–Pauli representation of α and β . First, take the particle at rest, $\mathbf{p} = 0$. Using (5.4), we have

$$Hu = \beta mu = \begin{pmatrix} mI & 0 \\ 0 & -mI \end{pmatrix}u$$

with eigenvalues $E = m, m, -m, -m$, and eigenvectors

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}. \quad (5.22)$$

As we have just mentioned, the electron, charge $-e$, is regarded as the particle. The first two solutions therefore describe an $E > 0$ electron. The $E < 0$ particle solutions are to be interpreted, as before, as describing an $E > 0$ antiparticle (positron) (see Section 3.6).

For $\mathbf{p} \neq 0$, (5.21) becomes, using (5.4),

$$Hu = \begin{pmatrix} m & \sigma \cdot \mathbf{p} \\ \sigma \cdot \mathbf{p} & -m \end{pmatrix} \begin{pmatrix} u_A \\ u_B \end{pmatrix} = E \begin{pmatrix} u_A \\ u_B \end{pmatrix}, \quad (5.23)$$

where u has been divided into two two-component spinors, u_A and u_B . This

reduces to

$$\begin{aligned}\boldsymbol{\sigma} \cdot \mathbf{p} u_B &= (E - m) u_A, \\ \boldsymbol{\sigma} \cdot \mathbf{p} u_A &= (E + m) u_B.\end{aligned}\quad (5.24)$$

For the two $E > 0$ solutions, we may take $u_A^{(s)} = \chi^{(s)}$, where

$$\chi^{(1)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \chi^{(2)} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (5.25)$$

The corresponding lower components of u are then specified by using the second equation of (5.24),

$$u_B^{(s)} = \frac{\boldsymbol{\sigma} \cdot \mathbf{p}}{E + m} \chi^{(s)}, \quad (5.26)$$

and so the positive-energy four-spinor solutions of Dirac's equation are

$$u^{(s)} = N \begin{pmatrix} \chi^{(s)} \\ \frac{\boldsymbol{\sigma} \cdot \mathbf{p}}{E + m} \chi^{(s)} \end{pmatrix}, \quad E > 0 \quad (5.27)$$

with $s = 1, 2$, where N is the normalization constant. For the $E < 0$ solutions, we take $u_B^{(s)} = \chi^{(s)}$, and so, from (5.24),

$$u_A^{(s)} = \frac{\boldsymbol{\sigma} \cdot \mathbf{p}}{E - m} u_B^{(s)} = -\frac{\boldsymbol{\sigma} \cdot \mathbf{p}}{|E| + m} \chi^{(s)}. \quad (5.28)$$

Hence, we obtain

$$u^{(s+2)} = N \begin{pmatrix} -\frac{\boldsymbol{\sigma} \cdot \mathbf{p}}{|E| + m} \chi^{(s)} \\ \chi^{(s)} \end{pmatrix}, \quad E < 0. \quad (5.29)$$

For an electron of given momentum \mathbf{p} , we have four solutions: $u^{(1,2)}$, corresponding to positive energy, and $u^{(3,4)}$, corresponding to negative energy. We can readily verify that the four solutions are orthogonal:

$$u^{(r)\dagger} u^{(s)} = 0 \quad r \neq s,$$

with $r, s = 1, 2, 3, 4$.

The bonus embodied in the Dirac equation, for instance, (5.23), is the extra twofold degeneracy. This means that there must be another observable which commutes with H and \mathbf{P} , whose eigenvalues can be taken to distinguish the states. On inspection, we see that the operator

$$\Sigma \cdot \hat{\mathbf{p}} \equiv \begin{pmatrix} \boldsymbol{\sigma} \cdot \hat{\mathbf{p}} & 0 \\ 0 & \boldsymbol{\sigma} \cdot \hat{\mathbf{p}} \end{pmatrix} \quad (5.30)$$

commutes with H and \mathbf{P} ; $\hat{\mathbf{p}}$ is the unit vector pointing in the direction of the momentum, $\mathbf{p}/|\mathbf{p}|$. The "spin" component in the direction of motion, $\frac{1}{2}\boldsymbol{\sigma} \cdot \hat{\mathbf{p}}$, is therefore a "good" quantum number and can be used to label the solutions. We

call this quantum number the *helicity* of the state. The possible eigenvalues λ of the helicity operator $\frac{1}{2}\boldsymbol{\sigma} \cdot \hat{\mathbf{p}}$ are

$$\lambda = \begin{cases} +\frac{1}{2} \text{ positive helicity,} \\ -\frac{1}{2} \text{ negative helicity.} \end{cases}$$

We see that no other component of $\boldsymbol{\sigma}$ has eigenvalues which are good quantum numbers.

With the above choice (5.25), of the spinors $\chi^{(s)}$, it is appropriate to choose \mathbf{p} along the z axis, $\mathbf{p} = (0, 0, p)$. Then

$$\frac{1}{2}\boldsymbol{\sigma} \cdot \hat{\mathbf{p}}\chi^{(s)} = \frac{1}{2}\sigma_3\chi^{(s)} = \lambda\chi^{(s)}$$

with $\lambda = \pm \frac{1}{2}$ corresponding to $s = 1, 2$, respectively.

EXERCISE 5.3 Calculate the $\lambda = +\frac{1}{2}$ helicity eigenspinor of an electron of momentum $\mathbf{p}' = (p \sin \theta, 0, p \cos \theta)$.

EXERCISE 5.4 Confirm the desired result that the Dirac equation describes “intrinsic” angular momentum (\equiv spin)- $\frac{1}{2}$ particles.

Hint We are clearly interested in angular momentum, so first we should explore the commutation of the orbital angular momentum $\mathbf{L} = \mathbf{r} \times \mathbf{P}$ with the Hamiltonian. Use $[x_i, P_j] = i\delta_{ij}$ to show that

$$[H, \mathbf{L}] = -i(\boldsymbol{\alpha} \times \mathbf{P}).$$

So \mathbf{L} is not conserved! There must be some other angular momentum. Show that

$$[H, \boldsymbol{\Sigma}] = +2i(\boldsymbol{\alpha} \times \mathbf{P}), \quad \text{where } \boldsymbol{\Sigma} \equiv \begin{pmatrix} \sigma & 0 \\ 0 & \sigma \end{pmatrix}.$$

Clearly, neither \mathbf{L} nor $\boldsymbol{\Sigma}$ are conserved. The combination

$$\mathbf{J} = \mathbf{L} + \frac{1}{2}\boldsymbol{\Sigma},$$

which is nothing other than the total angular momentum, is however conserved, as now

$$[H, \mathbf{J}] = 0.$$

The eigenvalues of $\frac{1}{2}\boldsymbol{\Sigma}$ are $\pm \frac{1}{2}$.

EXERCISE 5.5 For a nonrelativistic electron of velocity v , use (5.24) to show that u_A is larger than u_B by a factor of the order v/c . In nonrelativistic problems, ψ_A and ψ_B are referred to as the “large” and “small” components of the electron wavefunction ψ .

In the nonrelativistic limit, show that the Dirac equation for an electron (charge $-e$) in an electromagnetic field $A^\mu = (A^0, \mathbf{A})$ reduces to the Schrödinger–Pauli equation

$$\left(\frac{1}{2m} (\mathbf{P} + e\mathbf{A})^2 + \frac{e}{2m} \boldsymbol{\sigma} \cdot \mathbf{B} - eA^0 \right) \psi_A = E_{NR} \psi_A, \quad (5.31)$$

where the magnetic field $\mathbf{B} = \nabla \times \mathbf{A}$ and $E_{NR} = E - m$. Assume $|eA^0| \ll m$.

Hint Make the substitution $P^\mu \rightarrow P^\mu + eA^\mu$ in eqs. (5.24) written in terms of ψ_A, ψ_B , eliminate ψ_B , and use

$$\mathbf{P} \times \mathbf{A} - \mathbf{A} \times \mathbf{P} = -i\nabla \times \mathbf{A}$$

where $\mathbf{P} = -i\nabla$.

We see from the form (5.31) of the nonrelativistic reduction of the Dirac equation for an electron in an electromagnetic field that we may associate with the electron an intrinsic (or spin) magnetic moment

$$\boldsymbol{\mu} = -\frac{e}{2m} \boldsymbol{\sigma} \equiv -g \frac{e}{2m} \mathbf{S}, \quad (5.32)$$

where the gyromagnetic ratio g is 2 and the spin angular momentum operator \mathbf{S} is $\frac{1}{2}\boldsymbol{\sigma}$. We speak of the Dirac moment $-e/2m$ of the electron. Experimentally, $g = 2.00232$. The prediction $g = 2$ is a triumph of the Dirac equation. The small difference from 2 is discussed in Chapter 7.

5.4 Antiparticles

The first two solutions of the Dirac equation,

$$u^{(1,2)}(\mathbf{p}) e^{-i\mathbf{p} \cdot \mathbf{x}},$$

clearly describe a free electron of energy E and momentum \mathbf{p} . The two negative-energy electron solutions $u^{(3,4)}$ are to be associated with the antiparticle, the positron. Indeed, using the antiparticle prescription of Section 3.6, a positron of energy E , momentum \mathbf{p} is described by one of the $-E, -\mathbf{p}$ electron solutions, namely,

$$u^{(3,4)}(-\mathbf{p}) e^{-i[-\mathbf{p}] \cdot \mathbf{x}} \equiv v^{(2,1)}(\mathbf{p}) e^{i\mathbf{p} \cdot \mathbf{x}}, \quad (5.33)$$

where $p^0 \equiv E > 0$. The “positron” spinors, v , are introduced for notational convenience. Recall that the Dirac equation for $u(\mathbf{p})$ is [see (5.20)]

$$(\not{p} - m) u(\mathbf{p}) = 0.$$

What does this imply for $v(\mathbf{p})$? For an electron of energy $-E$ and momentum

$-\mathbf{p}$, we have

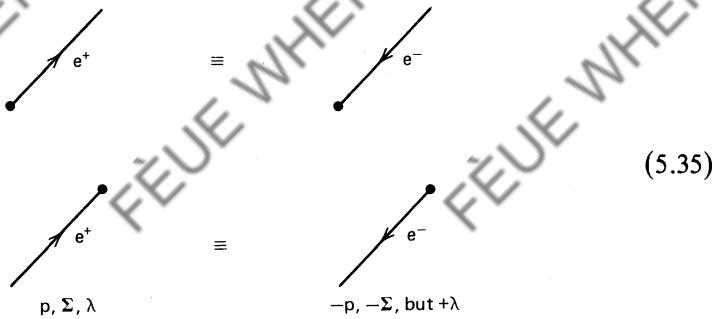
$$(-\not{p} - m) u(-\mathbf{p}) = 0,$$

and so

$$(\not{p} + m) v(\mathbf{p}) = 0. \quad (5.34)$$

We emphasize that here that $p^0 \equiv E > 0$.

As before, we continue to draw the Feynman diagrams entirely in terms of particle (electron) states. For example, an incoming positron of energy E is drawn as an outgoing electron of energy $-E$ [see (5.35)]. The only new feature in the antiparticle-particle correspondence is particle spin. Note the identification of the spinor labels 1, 2 with the negative energy 4, 3 states in (5.33). One way to anticipate this reverse order is to observe that, in the rest frame, the absence of spin up along a certain axis is equivalent to the presence of spin down along that axis. The detailed connection is made in Exercise 5.6. As both spin and momentum are reversed, the helicity, $\frac{1}{2}\boldsymbol{\sigma} \cdot \hat{\mathbf{p}}$, is unchanged. We may summarize these results by



The Dirac equation for an electron (charge $-e$) in an electromagnetic field is [see (4.2)]

$$[\gamma^\mu (i\partial_\mu + eA_\mu) - m] \psi = 0 \quad (5.36)$$

Now, there must be an equivalent Dirac equation for the positron ($+e$):

$$[\gamma^\mu (i\partial_\mu - eA_\mu) - m] \psi_C = 0. \quad (5.37)$$

There must be a one-to-one correspondence between ψ_C and ψ . To relate ψ_C to ψ , we first take the complex conjugate of (5.36),

$$[-\gamma^{\mu*} (i\partial_\mu - eA_\mu) - m] \psi^* = 0. \quad (5.38)$$

Thus, if we can find a matrix, denoted $(C\gamma^0)$, which satisfies

$$-(C\gamma^0)\gamma^{\mu*} = \gamma^\mu (C\gamma^0),$$

then (5.38) can be written in the form of (5.37), namely,

$$[\gamma^\mu(i\partial_\mu - eA_\mu) - m](C\gamma^0\psi^*) = 0,$$

and we can satisfy

$$\psi_C = C\gamma^0\psi^* = C\bar{\psi}^T,$$

where T denotes the transpose of a matrix.

EXERCISE 5.6 In representation (5.4) and (5.8) of the γ -matrices, show that a possible choice of C is

$$C\gamma^0 = i\gamma^2 = \begin{pmatrix} & & & 1 \\ & -1 & -1 & \\ 1 & & & \end{pmatrix}.$$

Try this operation out on a particular spinor, and show, for instance, that

$$\psi_C^{(1)} = i\gamma^2 [u^{(1)}(\mathbf{p}) e^{-ip \cdot x}]^* = u^{(4)}(-\mathbf{p}) e^{ip \cdot x} = v^{(1)}(\mathbf{p}) e^{ip \cdot x}.$$

Further, show that in this representation

$$\begin{aligned} C^{-1}\gamma^\mu C &= (-\gamma^\mu)^T, \\ C &= -C^{-1} = -C^\dagger = -C^T, \\ \bar{\psi}_C &= -\psi^T C^{-1}. \end{aligned} \tag{5.39}$$

We have seen that the electron current is

$$j^\mu = -e\bar{\psi}\gamma^\mu\psi.$$

The current associated with the charge conjugate field is therefore

$$\begin{aligned} j_C^\mu &= -e\bar{\psi}_C\gamma^\mu\psi_C \\ &= +e\psi^T C^{-1}\gamma^\mu C\bar{\psi}^T \\ &= -e\psi^T(\gamma^\mu)^T\bar{\psi}^T \\ &= -(-)e\bar{\psi}\gamma^\mu\psi \end{aligned} \tag{5.40}$$

[see (5.39)]. The origin of the extra minus sign introduced in the last line is subtle but important. It is clearly necessary for a physically meaningful result, that is, if j_C^μ is to be the positron current. The minus sign is related to the connection between spin and statistics; in field theory, it occurs because of the antisymmetric nature of the fermion fields. In field theory, the charge conjugation operator C changes a positive-energy electron into a positive-energy positron, and the for-

110 The Dirac Equation

malism is completely $e^+ \leftrightarrow e^-$ symmetric. However, in a single-particle (electron) theory, positron states are not allowed; rather, C changes a positive-energy electron state into a negative-energy electron state. As a result, we can show that we must add to our Feynman rules the requirement that we insert by hand an extra minus sign for every negative-energy electron in the final state of the process. The C invariance of electromagnetic interactions then follows:

$$j_\mu^C(A^\mu)^C = (-j_\mu)(-A^\mu) = j_\mu A^\mu.$$

5.5 Normalization of Spinors and the Completeness Relations

For fermions, we choose the covariant normalization in which we have $2E$ particles/unit volume, just as we did for bosons. That is,

$$\int_{\text{unit vol.}} \rho dV = \int \psi^\dagger \psi dV = u^\dagger u = 2E,$$

where we have used (5.16) and (5.18). Thus, we have the orthogonality relations

$$u^{(r)\dagger} u^{(s)} = 2E\delta_{rs}, \quad v^{(r)\dagger} v^{(s)} = 2E\delta_{rs}, \quad (5.41)$$

with $r, s = 1, 2$. Now, using (5.27), we calculate

$$u^{(s)\dagger} u^{(s)} = |N|^2 \left[1 + \left(\frac{\mathbf{q} \cdot \mathbf{p}}{E + m} \right)^2 \right] = |N|^2 \frac{2E}{E + m},$$

and so we may take the normalization constant to be

$$N = \sqrt{E + m}, \quad (5.42)$$

and the same for $v^{(s)}$.

To obtain the Dirac equation for $\bar{u} \equiv u^\dagger \gamma^0$, we need the hermitian conjugate of (5.19):

$$u^\dagger \gamma^{\mu\dagger} p_\mu - mu^\dagger = 0.$$

If we multiply from the right by γ^0 and note from (5.9)–(5.11) that

$$\gamma^{\mu\dagger} \gamma^0 = \gamma^0 \gamma^\mu, \quad (5.43)$$

this reduces to

$$\bar{u}(\not{p} - m) = 0. \quad (5.44)$$

Similarly, (5.35) gives

$$\bar{v}(\not{p} + m) = 0. \quad (5.45)$$

EXERCISE 5.7 Use (5.41) to show that

$$\bar{u}^{(s)} u^{(s)} = 2m, \quad \bar{v}^{(s)} v^{(s)} = -2m. \quad (5.46)$$

EXERCISE 5.8 Show that $(\sigma \cdot p)^2 = |\mathbf{p}|^2$.

EXERCISE 5.9 Derive the completeness relations

$$\boxed{\begin{aligned} \sum_{s=1,2} u^{(s)}(p) \bar{u}^{(s)}(p) &= \not{p} + m, \\ \sum_{s=1,2} v^{(s)}(p) \bar{v}^{(s)}(p) &= \not{p} - m. \end{aligned}} \quad (5.47)$$

These 4×4 matrix relations are used extensively in the evaluation of Feynman diagrams (see Chapter 6).

EXERCISE 5.10 Show that $\not{p}\not{p} = p^2$.

EXERCISE 5.11 Show that

$$\Lambda_+ = \frac{\not{p} + m}{2m}, \quad \Lambda_- = \frac{-\not{p} + m}{2m} \quad (5.48)$$

project over positive and negative energy states, respectively. Recall that the projection operators must satisfy

$$\Lambda_\pm^2 = \Lambda_\pm \quad \text{and} \quad \Lambda_+ + \Lambda_- = 1.$$

5.6 Bilinear Covariants

In order to construct the most general form of currents consistent with Lorentz covariance, we need to tabulate bilinear quantities of the form

$$(\bar{\psi})(4 \times 4)(\psi)$$

which have definite properties under Lorentz transformations, where the 4×4 matrix is a product of γ -matrices. To simplify the notation, we introduce

$$\gamma^5 \equiv i\gamma^0\gamma^1\gamma^2\gamma^3. \quad (5.49)$$

It follows that

$$\gamma^{5\dagger} = \gamma^5, \quad (\gamma^5)^2 = I, \quad \gamma^5\gamma^\mu + \gamma^\mu\gamma^5 = 0. \quad (5.50)$$

In the Dirac–Pauli representation, (5.4),

$$\gamma^0 = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}, \quad \gamma = \begin{pmatrix} 0 & \sigma \\ -\sigma & 0 \end{pmatrix}, \quad \gamma^5 = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}. \quad (5.51)$$

We are interested in the behavior of bilinear quantities under proper Lorentz transformations (that is, rotations, boosts) and under space inversion (the parity

112 The Dirac Equation

operation). An exhaustive list of the possibilities is

	No. of Compts.	Space Inversion, P
Scalar	$\bar{\psi}\psi$	1 + under P
Vector	$\bar{\psi}\gamma^\mu\psi$	4 Space compts.: - under P
Tensor	$\bar{\psi}\sigma^{\mu\nu}\psi$	6
Axial Vector	$\bar{\psi}\gamma^5\gamma^\mu\psi$	4 Space compts.: + under P
Pseudoscalar	$\bar{\psi}\gamma^5\psi$	1 - under P

(5.52)

Due to the anticommutation relations, (5.9), the tensor is antisymmetric:

$$\sigma^{\mu\nu} = \frac{i}{2}(\gamma^\mu\gamma^\nu - \gamma^\nu\gamma^\mu). \quad (5.53)$$

The list is arranged in increasing order of the number of γ^μ matrices that are sandwiched between $\bar{\psi}$ and ψ . The pseudoscalar is the product of four matrices [see (5.49)]. If five matrices were used, at least two would be the same, in which case the product would reduce to three and be already included in the axial vector.

It is useful to see how the above properties are established. To do this, we must consider Dirac's equation in two frames (x and x') related by a Lorentz transformation Λ . From (5.7), we have

$$i\gamma^\mu \frac{\partial \psi(x)}{\partial x^\mu} - m \psi(x) = 0, \quad (5.54)$$

$$i\gamma^\mu \frac{\partial \psi'(x')}{\partial x'^\mu} - m \psi'(x') = 0, \quad (5.55)$$

where $x' = \Lambda x$. There must exist a relation

$$\psi'(x') = S \psi(x). \quad (5.56)$$

If we recall (5.18), it is clear that S is independent of x and acts only on the spinor u . Substituting (5.56) into (5.55) and demanding consistency with (5.54), we obtain

$$S^{-1} \gamma^\mu S = \Lambda^\mu_\nu \gamma^\nu, \quad (5.57)$$

where we have used $\partial/\partial x^\mu = \Lambda^\nu_\mu \partial/\partial x'^\nu$.

EXERCISE 5.12 For a proper infinitesimal Lorentz transformation

$$\Lambda^\nu_\mu = \delta^\nu_\mu + \epsilon^\nu_\mu, \quad (5.58)$$

show that an S which satisfies (5.57) is

$$S_L = 1 - \frac{i}{4} \sigma_{\mu\nu} \epsilon^{\mu\nu}. \quad (5.59)$$

Hence, show that

$$S_L^{-1} = \gamma^0 S_L^\dagger \gamma^0 \quad (5.60)$$

$$\gamma^5 S_L = S_L \gamma^5. \quad (5.61)$$

For space inversion, or the parity operation,

$$\Lambda_\mu^\nu = \begin{pmatrix} 1 & & & \\ & -1 & & \\ & & -1 & \\ & & & -1 \end{pmatrix}.$$

Then, (5.57) becomes

$$\begin{aligned} S_P^{-1} \gamma^0 S_P &= \gamma^0, \\ S_P^{-1} \gamma^k S_P &= -\gamma^k \quad \text{for } k = 1, 2, 3, \end{aligned}$$

which is satisfied by

$$S_P = \gamma^0. \quad (5.62)$$

In the Dirac-Pauli representation of γ^0 , (5.51), the behavior of the four components of ψ under parity is therefore

$$\psi'_{1,2} = \psi_{1,2} \quad \text{and} \quad \psi'_{3,4} = -\psi_{3,4}. \quad (5.63)$$

The “at rest” states, (5.22), are therefore eigenstates of parity, with the positive and negative energy states (that is, the electron and the positron) having opposite intrinsic parities.

Armed with S_L and S_P , we can now check the claimed properties of the bilinear covariants. First, we note that

$$\begin{aligned} \bar{\psi}' &= \psi'^\dagger \gamma^0 = \psi^\dagger S^\dagger \gamma^0 = \psi^\dagger \gamma^0 S^{-1} \\ &= \bar{\psi} S^{-1}, \end{aligned} \quad (5.64)$$

where we have used (5.56) and (5.60). As an example, let us establish the character of $\bar{\psi} \gamma^\mu \psi$. Under Lorentz transformations,

$$\bar{\psi}' \gamma^\mu \psi' = \bar{\psi} S_L^{-1} \gamma^\mu S_L \psi = \Lambda^\mu_\nu (\bar{\psi} \gamma^\nu \psi), \quad (5.65)$$

using (5.64) and (5.57), while under the parity operation,

$$\bar{\psi}' \gamma^\mu \psi' = \bar{\psi} S_P^{-1} \gamma^\mu S_P \psi = \begin{cases} \bar{\psi} \gamma^0 \psi, \\ -\bar{\psi} \gamma^k \psi. \end{cases} \quad (5.66)$$

These transformation properties are precisely what we expect for a Lorentz four-vector.

From (5.56) and (5.64), it follows immediately that $\bar{\psi} \psi$ is a Lorentz scalar. The probability density $\rho = \psi^\dagger \psi$ is not a scalar, but is the time-like component of the

four-vector $\bar{\psi} \gamma^\mu \psi$. Since

$$\gamma^5 S_P = -S_P \gamma^5, \quad (5.67)$$

the presence of γ^5 gives rise to the pseudo-nature of the axial vector and pseudoscalar. For instance, a pseudoscalar is a scalar under proper Lorentz transformations but, unlike a scalar, changes sign under parity.

5.7 Zero-Mass Fermions: The Two-Component Neutrino

We return to the Dirac equation, (5.1),

$$H\psi = (\alpha \cdot \mathbf{p} + \beta m)\psi \quad (5.68)$$

and the discussion at the opening of this chapter. We derived algebraic relations which were demanded of the α_i 's and β , and found that they could be satisfied by 4×4 matrices. However, note that β is not involved in the case of zero-mass particles and that we need only satisfy

$$\alpha_i \alpha_j + \alpha_j \alpha_i = 2\delta_{ij}, \quad \alpha_i = \alpha_i^\dagger \quad (5.69)$$

[see (5.3)]. These relations can be realized by the 2×2 Pauli matrices. We can take $\alpha_i = -\sigma_i$ and $\alpha_i = \sigma_i$, and the massless Dirac equation divides into two *decoupled* equations for two-component spinors $\chi(\mathbf{p})$ and $\phi(\mathbf{p})$:

$$E\chi = -\sigma \cdot \mathbf{p} \chi, \quad (5.70)$$

$$E\phi = +\sigma \cdot \mathbf{p} \phi. \quad (5.71)$$

Each equation is based on the relativistic energy–momentum relation, $E^2 = \mathbf{p}^2$, and so has one positive and one negative energy solution.

Suppose (5.70) is the wave equation for a massless fermion, a neutrino. The positive energy solution has $E = |\mathbf{p}|$ and so satisfies

$$\sigma \cdot \hat{\mathbf{p}} \chi = -\chi. \quad (5.72)$$

That is, χ describes a left-handed neutrino (helicity $\lambda = -\frac{1}{2}$) of energy E and momentum \mathbf{p} . The remaining solution has negative energy. To interpret this, we consider a neutrino solution with energy $-E$ and momentum $-\mathbf{p}$. It satisfies

$$\sigma \cdot (-\hat{\mathbf{p}}) \chi = \chi \quad (5.73)$$

with positive helicity, and hence describes a right-handed antineutrino ($\lambda = +\frac{1}{2}$) of energy E and momentum \mathbf{p} . Symbolically, we say (5.70) describes ν_L and $\bar{\nu}_R$. Such a wave equation was first proposed by Weyl in 1929 but was rejected because of noninvariance under the parity operation P , which takes $\nu_L \rightarrow \nu_R$. For massless neutrinos, this is no longer an objection as weak interactions do not respect parity conservation (see Chapter 12). The second equation, (5.71), describes the other helicity states ν_R and $\bar{\nu}_L$.

Translating these results to four-component form

$$u = \begin{pmatrix} \chi \\ \phi \end{pmatrix}, \quad \alpha = \begin{pmatrix} -\sigma & 0 \\ 0 & \sigma \end{pmatrix}, \quad (5.74)$$

we see that we are here working in the Weyl (or chiral) representation, (5.6). In this representation,

$$\gamma = \begin{pmatrix} 0 & \sigma \\ -\sigma & 0 \end{pmatrix}, \quad \gamma^0 = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}, \quad \gamma^5 = \begin{pmatrix} -I & 0 \\ 0 & I \end{pmatrix}. \quad (5.75)$$

At this stage, it is appropriate to peep ahead at weak interactions, which are discussed in Chapter 12. A vast number of different experiments indicate that leptons enter the “charged-current” weak interactions in a special combination of two of the bilinear covariants. For example, for the electron and its neutrino,

$$J^\mu = \bar{\psi}_e \gamma^{\mu \frac{1}{2}} (1 - \gamma^5) \psi_e. \quad (5.76)$$

We speak of the $V-A$ form of the weak current J^μ , in contrast to the V form of the electromagnetic current (5.17). Our concern here is the presence of the $\frac{1}{2}(1 - \gamma^5)$; this mixture of vector (V) and axial vector (A) ensures that parity is violated, and violated maximally. Indeed, from (5.75),

$$\frac{1}{2}(1 - \gamma^5) u_\nu = \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \chi \\ \phi \end{pmatrix} = \begin{pmatrix} \chi \\ 0 \end{pmatrix} \quad (5.77)$$

and so projects out just ν_L (and $\bar{\nu}_R$). That is, only left-handed neutrinos (and right-handed antineutrinos) are coupled to charged leptons by the weak interactions. As far as we know, neutrinos have only the weak interaction, and hence this is the only way we can observe them. There is thus no empirical evidence for the existence of ν_R (and $\bar{\nu}_L$), and it could well be that they do not exist in nature. Of course, the assertion that only ν_L (and $\bar{\nu}_R$) occur can only be made if the mass is strictly zero. Otherwise, we could perform a Lorentz transformation which would change a ν_L into a ν_R .

EXERCISE 5.13 Show that the operators

$$P_R \equiv \frac{1}{2}(1 + \gamma^5), \quad P_L \equiv \frac{1}{2}(1 - \gamma^5)$$

have the appropriate properties to be (right- and left-hand) projection operators, that is,

$$P_i^2 = P_i, \quad P_L + P_R = 1, \quad P_R P_L = 0.$$

Here, γ^5 is called the chirality operator.

For a massive fermion, we define the projections $\frac{1}{2}(1 \pm \gamma^5)u$ to be right- and left-handed components of u .

EXERCISE 5.14 For a massive fermion, show that handedness is not a good quantum number. That is, show that γ^5 does not commute with the Hamiltonian. However, verify that helicity is conserved but is frame dependent. In particular, show that the helicity is reversed by “overtaking” the particle concerned.

EXERCISE 5.15 Working in the Dirac–Pauli representation of γ -matrices, (5.51), show that at high energies

$$\gamma^5 u^{(s)} \simeq \begin{pmatrix} \sigma \cdot \hat{\mathbf{p}} & 0 \\ 0 & \sigma \cdot \hat{\mathbf{p}} \end{pmatrix} u^{(s)}, \quad (5.78)$$

where $u^{(s)}$ is the electron spinor of (5.27). That is, show that in the extreme relativistic limit, the chirality operator (γ^5) is equal to the helicity operator; and so, for example, $\frac{1}{2}(1 - \gamma^5)u = u_L$ corresponds to an electron of negative helicity.

Of course, the fact that $\frac{1}{2}(1 - \gamma^5)$ projects out negative helicity fermions at high energies does not depend on the choice of representation. We need only choose a representation if we wish to show explicit spinors. The particular advantage of the Dirac–Pauli representation is that it diagonalizes the energy in the nonrelativistic limit (γ^0 is diagonal), whereas the Weyl representation diagonalizes the helicity in the extreme relativistic limit (γ^5 is diagonal).

Finally suppose, for the moment, that a nonzero mass was established for one of the neutrinos (ν_e , ν_μ , ν_τ and so on). Remembering exercise (5.14), how can we have a massive neutrino and still ensure that weak interactions couple only to ν_L and $\bar{\nu}_R$? Majorana neutrinos accomplish this. They are formed by making the neutrino its own antiparticle. Thus, we may identify ν_L and $\bar{\nu}_R$ as two helicity components of a four-component spinor. The other two components, ν_R and $\bar{\nu}_L$ (if these exist), can then be a Majorana fermion of different mass. Clearly this is a different structure to the four-component Dirac spinor for, say, e^\pm . A detailed discussion of Majorana spinors is given by P. Roman in *Theory of Elementary Particles*, North Holland (1961), page 306.

6

Electrodynamics of Spin- $\frac{1}{2}$ Particles

In this chapter, we are going to reach the goal set in Chapter 3: compute cross sections for the electromagnetic interactions of leptons and photons. This is quantum electrodynamics. We proceed by retracing the steps of Chapter 4 for particles now described by the Dirac equation. The result will be Feynman rules for the electromagnetic interactions of spin- $\frac{1}{2}$ leptons and quarks. Not only are the interactions completely physical but so, now, are the particles. They are no longer “pedagogical constructs” of spin zero. We shall finally confront experimental measurements.

6.1 An Electron Interacting with an Electromagnetic Field A^μ

We have seen that a free electron of four-momentum p^μ is described by a four-component wavefunction

$$\psi = u(\mathbf{p}) e^{-ip \cdot x}$$

which satisfies the Dirac equation, (5.19),

$$(\gamma_\mu p^\mu - m)\psi = 0.$$

The equation for an electron in an electromagnetic field A^μ is obtained by the substitution [see (4.1)]

$$p^\mu \rightarrow p^\mu + eA^\mu, \quad (6.1)$$

where we have again taken $-e$ to be the charge of the electron. We find

$$(\gamma_\mu p^\mu - m)\psi = \gamma^0 V\psi, \quad (6.2)$$

where the perturbation is given by

$$\gamma^0 V = -e\gamma_\mu A^\mu. \quad (6.3)$$

The introduction of the γ^0 is to make (6.2) of the form $(E + \dots)\psi = V\psi$, so that the potential energy enters in the same way as in the Schrödinger equation [see, for example, the $-eA^0$ term in the Schrödinger–Pauli equation, (5.31)].

Using first-order perturbation theory, (3.37), the amplitude for the scattering of an electron from state ψ_i to state ψ_f is

$$\begin{aligned} T_{fi} &= -i \int \bar{\psi}_f^\dagger(x) V(x) \psi_i(x) d^4x \\ &= ie \int \bar{\psi}_f \gamma_\mu A^\mu \psi_i d^4x \\ &= -i \int j_\mu^{fi} A^\mu d^4x, \end{aligned} \quad (6.4)$$

where

$$j_\mu^{fi} \equiv -e \bar{\psi}_f \gamma_\mu \psi_i \quad (6.5)$$

$$= -e \bar{u}_f \gamma_\mu u_i e^{i(p_f - p_i) \cdot x}. \quad (6.6)$$

Comparing this result with (5.17), we see that j_μ^{fi} can be regarded as the electromagnetic transition current between electron states i and f .

Recall that the analogous transition current we obtained for a “spinless” electron,

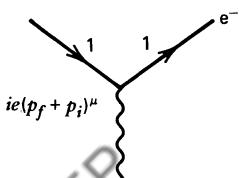
$$j_\mu^{fi} = -e (p_f + p_i)_\mu e^{i(p_f - p_i) \cdot x},$$

resulted in the Feynman rules of Fig. 6.1a. In Fig. 6.1b, we show the corresponding rules that follow for the (physical) spin- $\frac{1}{2}$ electron. The vertex factor is now a 4×4 matrix in spin space. It is sandwiched between column $u^{(s)}(p_i)$ and row $\bar{u}^{(r)}(p_f)$ spinors describing incoming and outgoing electrons of momentum p_i , p_f and in spin states s , r , respectively.

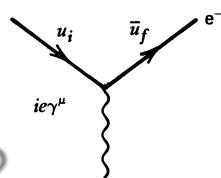
EXERCISE 6.1 A “spinless” electron can interact with A^μ only via its charge; the coupling involves $(p_f + p_i)^\mu$. Show that

$$\bar{u}_f \gamma^\mu u_i = \frac{1}{2m} \bar{u}_f \left((p_f + p_i)^\mu + i \sigma^{\mu\nu} (p_f - p_i)_\nu \right) u_i, \quad (6.7)$$

from which it is possible to establish that the physical spin- $\frac{1}{2}$ electron interacts via both its charge and its magnetic moment; see also Exercise 6.2. Equation (6.7) is known as the Gordon decomposition of the current.



(a) “Spinless” electron



(b) Spin 1/2 electron

Fig. 6.1 Factors for the Feynman rules listed in Section 6.17.

EXERCISE 6.2 Show that in the nonrelativistic limit, the Gordon decomposition, (6.7), of the electron current, (6.6), separates the electron interaction with an electromagnetic field A_μ into a part arising from its charge, $-e$, and a part due to its magnetic moment, $-e/2m$. Assume that A_μ is independent of t , so that (6.4) becomes

$$T_{fi} = -i2\pi\delta(E_f - E_i)\int j_\mu^{(i)} A^\mu d^3x.$$

To identify the magnetic moment interaction ($-\boldsymbol{\mu} \cdot \mathbf{B}$), it suffices to show that

$$\int \left[-\frac{e}{2m} \bar{\psi}_f i\sigma_{\mu\nu} (p_f - p_i)^\nu \psi_i \right] A^\mu d^3x = \int \psi_A^{f\dagger} \left(\frac{e}{2m} \boldsymbol{\sigma} \cdot \mathbf{B} \right) \psi_A^i d^3x,$$

where ψ_A denotes the upper two (or “large”) components of ψ ; compare with eqs. (5.31) and (5.32).

Hint Since $E_i = E_f$, only the space-like components of $(p_f - p_i)$ contribute. Note that $\psi(\sigma_{23}, \sigma_{31}, \sigma_{12})\psi = \bar{\psi}_A \boldsymbol{\sigma} \psi_A$.

6.2 Møller Scattering $e^-e^- \rightarrow e^-e^-$

As an illustration of how to use the QED vertex factor of Fig. 6.1b, we compute the Feynman diagram of Fig. 6.2 for e^-e^- scattering. Repeating the steps in Section 4.2, the transition amplitude is given by [see (4.16)]

$$\begin{aligned} T_{fi} &= -i \int j_\mu^{(1)}(x) \left(-\frac{1}{q^2} \right) j_{(2)}^\mu(x) d^4x \\ &= -i(-e\bar{u}_C \gamma_\mu u_A) \left(-\frac{1}{q^2} \right) (-e\bar{u}_D \gamma^\mu u_B) (2\pi)^4 \delta^{(4)}(p_A + p_B - p_C - p_D), \end{aligned}$$

with $q = p_A - p_C$, where $(2\pi)^4$ times the delta function arises from the integration over the x dependence of the currents [see (6.6)]. Recall that the invariant

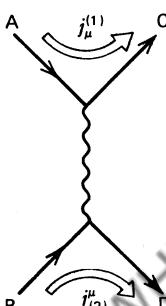
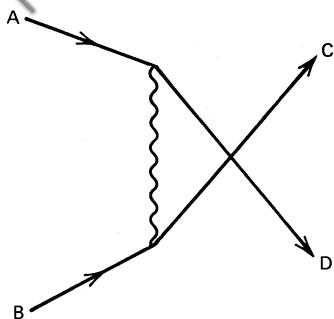


Fig. 6.2 Feynman diagram for $e^-e^- \rightarrow e^-e^-$.

Fig. 6.3 Second diagram for $e^-e^- \rightarrow e^-e^-$.

amplitude \mathcal{M} is defined by

$$T_{fi} = -i(2\pi)^4 \delta^{(4)}(p_A + p_B - p_C - p_D)\mathcal{M},$$

and so we have

$$-i\mathcal{M} = (ie\bar{u}_C\gamma^\mu u_A)\left(\frac{-ig_{\mu\nu}}{q^2}\right)(ie\bar{u}_D\gamma^\nu u_B), \quad (6.8)$$

in agreement with the factors assigned to Fig. 6.1b.

For e^-e^- scattering, there is a second Feynman diagram, Fig. 6.3. The amplitude is obtained from (6.8) with the interchange $C \leftrightarrow D$, but with a relative minus sign on account of the interchange of identical fermions. Thus, the complete (lowest-order) amplitude for Møller scattering is

$$\mathcal{M} = -e^2 \frac{(\bar{u}_C\gamma^\mu u_A)(\bar{u}_D\gamma_\mu u_B)}{(p_A - p_C)^2} + e^2 \frac{(\bar{u}_D\gamma^\mu u_A)(\bar{u}_C\gamma_\mu u_B)}{(p_A - p_D)^2}. \quad (6.9)$$

To calculate the unpolarized cross section, we must amend the cross section formulae of Section 4.3. By unpolarized we mean that no information about the electron spins is recorded in the experiment. To allow for scattering in all possible spin configurations, we therefore have to make the replacement

$$|\mathcal{M}|^2 \rightarrow \overline{|\mathcal{M}|^2} \equiv \frac{1}{(2s_A + 1)(2s_B + 1)} \sum_{\text{all spin states}} |\mathcal{M}|^2, \quad (6.10)$$

where s_A, s_B are the spins of the incoming particles. That is, we average over the spins of the incoming particles and we sum over the spins of the particles in the final state. Clearly, to carry out this calculation is a nontrivial task.

To get some feel for what is involved, we first evaluate the unpolarized cross section for Møller scattering in the *nonrelativistic limit*. In this case, the spin sums become relatively simple. In the limit $|\mathbf{p}| \rightarrow 0$, (5.27) and (5.42) give

$$\begin{aligned} \text{Ingoing } e^- : u^{(s)} &= \sqrt{2m} \begin{pmatrix} \chi^{(s)} \\ 0 \end{pmatrix}, \\ \text{Outgoing } e^- : \bar{u}^{(s)} &= \sqrt{2m} (\chi^{(s)} \quad 0), \end{aligned} \quad (6.11)$$

where $s = 1, 2$ correspond to spin up, down along the z axis [see (5.25)]. Using representation (5.4) in which

$$\gamma^0 = \begin{pmatrix} I & \\ & -I \end{pmatrix}, \quad \gamma = \begin{pmatrix} & \sigma \\ -\sigma & \end{pmatrix}, \quad (6.12)$$

we find

$$\bar{u}^{(s)} \gamma^\mu u^{(s)} = \begin{cases} 2m & \text{if } \mu = 0, \\ 0 & \text{if } \mu \neq 0, \end{cases} \quad (6.13)$$

$$\bar{u}^{(s)} \gamma^\mu u^{(s')} = 0 \quad \text{for all } \mu, \text{ if } s \neq s'.$$

In other words, the spin direction does not change in the scattering of nonrelativistic electrons. This is to be expected, as the electrons interact dominantly via an electric field which cannot change their spin direction. At higher energies (velocities), it is the magnetic field which flips the spins. Inserting (6.13) into (6.9) and labeling the six nonzero amplitudes by the electron spins gives

$$\begin{aligned} \mathcal{M}(\uparrow\uparrow \rightarrow \uparrow\uparrow) &= \mathcal{M}(\downarrow\downarrow \rightarrow \downarrow\downarrow) = -e^2 4m^2 \left(\frac{1}{t} - \frac{1}{u} \right) \\ \mathcal{M}(\uparrow\downarrow \rightarrow \uparrow\downarrow) &= \mathcal{M}(\downarrow\uparrow \rightarrow \downarrow\uparrow) = -e^2 4m^2 \frac{1}{t} \\ \mathcal{M}(\uparrow\downarrow \rightarrow \downarrow\uparrow) &= \mathcal{M}(\downarrow\uparrow \rightarrow \uparrow\downarrow) = e^2 4m^2 \frac{1}{u}, \end{aligned}$$

where $1/t$, $1/u$, defined by (4.43), are the photon propagators. Arrows label the spin-up (down) state of each particle. We can now perform the spin summation (6.10), and find

$$|\mathcal{M}|^2 = \frac{1}{4} (4m^2 e^2)^2 2 \left[\left(\frac{1}{t} - \frac{1}{u} \right)^2 + \frac{1}{t^2} + \frac{1}{u^2} \right]. \quad (6.14)$$

In the center-of-mass frame [see (4.45)],

$$\begin{aligned} t &= -2p^2(1 - \cos \theta) = -4p^2 \sin^2 \frac{\theta}{2}, \\ u &= -2p^2(1 + \cos \theta) = -4p^2 \cos^2 \frac{\theta}{2}, \end{aligned} \quad (6.15)$$

where θ is the scattering angle and $p = |\mathbf{p}_i|$ with $i = A, B, C, D$. Using this, and inserting (6.14) into (4.35), we find that the $e^- e^-$ differential cross section is

$$\frac{d\sigma}{d\Omega} \Big|_{cm} = \frac{m^2 \alpha^2}{16p^4} \left(\frac{1}{\sin^4 \frac{\theta}{2}} + \frac{1}{\cos^4 \frac{\theta}{2}} - \frac{1}{\sin^2 \frac{\theta}{2} \cos^2 \frac{\theta}{2}} \right) \quad (6.16)$$

in the nonrelativistic limit, where $s \approx 4m^2$ and $\alpha \equiv e^2/4\pi$; see (4.5).

6.3 The Process $e^- \mu^- \rightarrow e^- \mu^-$

How are these spin summations done without the benefit of the simplifications introduced by the nonrelativistic approximation? We use the example of $e^- \mu^-$

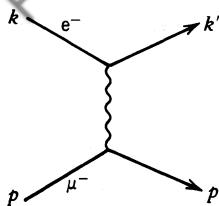


Fig. 6.4 Feynman diagram for electron-muon scattering.

scattering to illustrate the general technique for summing over spins since it has only one (lowest-order) Feynman diagram (Fig. 6.4). The invariant amplitude follows from the Feynman rules [see (6.8)]:

$$\mathcal{M} = -e^2 \bar{u}(k') \gamma^\mu u(k) \frac{1}{q^2} \bar{u}(p') \gamma_\mu u(p). \quad (6.17)$$

The momenta are defined in Fig. 6.4 and $q = k - k'$. To obtain the (unpolarized) cross section, we have to take the square of the modulus of \mathcal{M} and then carry out the spin sums. It is convenient to separate the sums over the electron and muon spins by writing (6.10) as

$$\overline{|\mathcal{M}|^2} = \frac{e^4}{q^4} L_e^{\mu\nu} L_{\mu\nu}^{\text{muon}}, \quad (6.18)$$

where the tensor associated with the electron vertex is

$$L_e^{\mu\nu} \equiv \frac{1}{2} \sum_{(e \text{ spins})} [\bar{u}(k') \gamma^\mu u(k)] [\bar{u}(k') \gamma^\nu u(k)]^*, \quad (6.19)$$

and with a similar expression for $L_{\mu\nu}^{\text{muon}}$.

The spin summations look a forbidding task. Fortunately, well-established trace techniques considerably simplify such calculations. To begin, we note that the second square bracket of (6.19) (a 1×1 matrix for which the complex and hermitian conjugates are the same) is equal to

$$\begin{aligned} [u^\dagger(k') \gamma^0 \gamma^\nu u(k)]^\dagger &= [u^\dagger(k) \gamma^\nu \gamma^0 u(k')] \\ &= [\bar{u}(k) \gamma^\nu u(k)], \end{aligned}$$

where we have used $\gamma^\nu \gamma^0 = \gamma^0 \gamma^\nu$ [see (5.43)]. That is, the complex conjugation in (6.19) simply reverses the order of the matrix product. We now write the complete product in (6.19) explicitly in terms of individual matrix elements (labeled α, β, \dots , with a summation over repeated indices implied)

$$L_e^{\mu\nu} = \frac{1}{2} \underbrace{\sum_{s'} \bar{u}_{\alpha}^{(s')}(k') \gamma_{\alpha\beta}^\mu \sum_s u_{\beta}^{(s)}(k) \bar{u}_\gamma^{(s)}(k) \gamma_{\gamma\delta}^\nu u_\delta^{(s')}(k')}_{(\not{k} + m)_{\alpha\beta}}.$$

The completeness relation (5.47) allows the sums over both the initial and the final electron spins to be performed. The repositioning of u_δ makes this evident; it

can be moved as the matrix character is recorded by the component. Thus, L becomes the trace of the product of four 4×4 matrices,

$$L_e^{\mu\nu} = \frac{1}{2} \text{Tr}((\not{k}' + m)\gamma^\mu(\not{k} + m)\gamma^\nu), \quad (6.20)$$

where m is the mass of the electron. To evaluate L , we use the trace theorems.

6.4 Trace Theorems and Properties of γ -Matrices

Recall that the Dirac γ -matrices satisfy the commutation algebra

$$\gamma^\mu \gamma^\nu + \gamma^\nu \gamma^\mu = 2g^{\mu\nu}. \quad (6.21)$$

As a consequence, it is straightforward to show that the trace of a product of γ -matrices can be evaluated without ever explicitly calculating a matrix product. The trace theorems are (using again the notation $\not{a} = \gamma_\mu a^\mu$):

$$\text{Tr } 1 = 4,$$

Trace of an odd number of γ_μ 's vanishes,

$$\text{Tr}(\not{a}\not{b}) = 4a \cdot b,$$

$$\text{Tr}(\not{a}\not{b}\not{c}\not{d}) = 4[(a \cdot b)(c \cdot d) - (a \cdot c)(b \cdot d) + (a \cdot d)(b \cdot c)], \quad (6.22)$$

$$\text{Tr } \gamma_5 = 0,$$

$$\text{Tr}(\gamma_5 \not{a}\not{b}) = 0,$$

$$\text{Tr}(\gamma_5 \not{a}\not{b}\not{c}\not{d}) = 4i\epsilon_{\mu\nu\lambda\sigma}a^\mu b^\nu c^\lambda d^\sigma,$$

where $\epsilon_{\mu\nu\lambda\sigma} = +1 (-1)$ for $\mu, \nu, \lambda, \sigma$ an even (odd) permutation of 0, 1, 2, 3; and 0 if two indices are the same.

Other useful results for simplifying trace calculations are:

$$\begin{aligned} \gamma_\mu \gamma^\mu &= 4 \\ \gamma_\mu \not{a} \gamma^\mu &= -2\not{a} \\ \gamma_\mu \not{a}\not{b} \gamma^\mu &= 4a \cdot b \\ \gamma_\mu \not{a}\not{b}\not{c}\not{d} \gamma^\mu &= -2\not{c}\not{b}\not{a} \end{aligned} \quad (6.24)$$

EXERCISE 6.3 Making use of (6.21), prove the trace theorems and the identities (6.24).

6.5 $e^- \mu^-$ Scattering and the Process $e^+ e^- \rightarrow \mu^+ \mu^-$

It is straightforward to evaluate the tensor associated with the electron vertex, (6.20), using the trace theorems listed in (6.22). We have

$$\begin{aligned} L_e^{\mu\nu} &= \frac{1}{2} \text{Tr}((\not{k}'\gamma^\mu\not{k}\gamma^\nu) + \frac{1}{2}m^2 \text{Tr}(\gamma^\mu\gamma^\nu), \\ &= 2(k'^\mu k^\nu + k''^\nu k^\mu - (k' \cdot k - m^2)g^{\mu\nu}). \end{aligned} \quad (6.25)$$

The evaluation of $L_{\mu\nu}^{\text{muon}}$ of (6.18) is identical. We find

$$L_{\mu\nu}^{\text{muon}} = 2(p'_\mu p_\nu + p'_\nu p_\mu - (p' \cdot p - M^2)g_{\mu\nu}), \quad (6.26)$$

where M is the mass of the muon. Forming the product of (6.25) and (6.26), we finally arrive at the following “exact” form for the spin-averaged $e^- \mu^- \rightarrow e^- \mu^-$ amplitude, (6.18):

$$\begin{aligned} |\mathcal{M}|^2 = & \frac{8e^4}{q^4} [(k' \cdot p')(k \cdot p) + (k' \cdot p)(k \cdot p')] \\ & - m^2 p' \cdot p - M^2 k' \cdot k + 2m^2 M^2]. \end{aligned} \quad (6.27)$$

In the extreme relativistic limit, we may neglect the terms involving m^2 and M^2 ; therefore,

$$|\mathcal{M}|^2 = \frac{8e^4}{(k - k')^4} [(k' \cdot p')(k \cdot p) + (k' \cdot p)(k \cdot p')]. \quad (6.28)$$

Moreover, in this limit, the (Mandelstam) variables of (4.43) become

$$\begin{aligned} s &\equiv (k + p)^2 \simeq 2k \cdot p \simeq 2k' \cdot p', \\ t &\equiv (k - k')^2 \simeq -2k \cdot k' \simeq -2p \cdot p', \\ u &\equiv (k - p')^2 \simeq -2k \cdot p' \simeq -2k' \cdot p. \end{aligned} \quad (6.29)$$

Thus, at high energies, unpolarized $e^- \mu^-$ scattering, (6.28), is given by

$$|\mathcal{M}|^2 = 2e^4 \frac{s^2 + u^2}{t^2}. \quad (6.30)$$

We may also obtain the amplitude for $e^- e^+ \rightarrow \mu^+ \mu^-$ by “crossing” the above result for $e^- \mu^- \rightarrow e^- \mu^-$ (see Section 4.6). The required interchange is $k' \leftrightarrow -p$, that is, $s \leftrightarrow t$ in (6.30), and we obtain

$$|\mathcal{M}|^2 = 2e^4 \frac{t^2 + u^2}{s^2} \quad (6.31)$$

where now $e^- e^+ \rightarrow \mu^+ \mu^-$ is the s -channel process. The corresponding diagram is drawn in Fig. 6.5. This result for the square of the amplitude can be translated into a differential cross section for $e^+ e^- \rightarrow \mu^+ \mu^-$ scattering using (4.35). In the

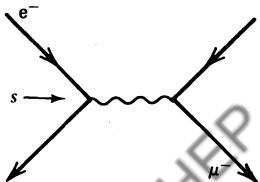


Fig. 6.5 The Feynman diagram for $e^- e^+ \rightarrow \mu^+ \mu^-$. As always, the antiparticles are drawn using only particle (e^- , μ^-) lines.

center-of-mass frame, we have

$$\frac{d\sigma}{d\Omega} \Big|_{cm} = \frac{1}{64\pi^2 s} 2e^4 \left[\frac{1}{2}(1 + \cos^2 \theta) \right],$$

where the quantity in square brackets is $(t^2 + u^2)/s^2$; see (4.45). Using $\alpha = e^2/4\pi$, this becomes

$$\frac{d\sigma}{d\Omega} \Big|_{cm} = \frac{\alpha^2}{4s} (1 + \cos^2 \theta). \quad (6.32)$$

To obtain the reaction cross section, we integrate over θ, ϕ :

$$\sigma(e^+ e^- \rightarrow \mu^+ \mu^-) = \frac{4\pi\alpha^2}{3s}. \quad (6.33)$$

A comparison of this result with PETRA data is shown in Fig. 6.6. The PETRA accelerator consists of a ring of magnets which simultaneously accelerate an electron and positron beam circulating in opposite directions. In selected spots, these beams are crossed, resulting in $e^+ e^-$ interactions with center-of-mass energy $\sqrt{s} = 2E_b$, where E_b is the energy of each beam. Equation (6.33) can be written in

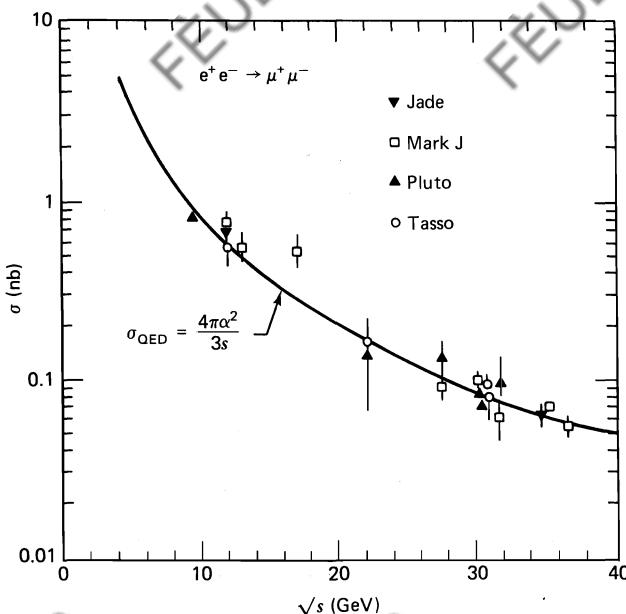
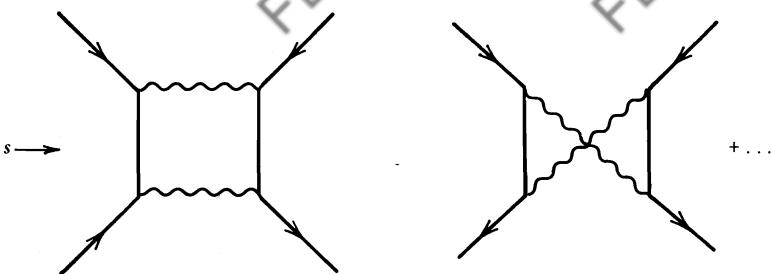


Fig. 6.6 The total cross section for $e^- e^+ \rightarrow \mu^- \mu^+$ measured at PETRA versus the center-of-mass energy.

Fig. 6.7 Some higher-order diagrams for $e^-e^+ \rightarrow \mu^+\mu^-$.

numerical form as

$$\sigma(e^+e^- \rightarrow \mu^+\mu^-) = \frac{20 \text{ (nb)}}{E_b^2 \text{ (in GeV}^2\text{)}}.$$

This quadratic dependence of the annihilation cross section on beam energy predicted by our calculations can be checked by varying the energy of the beams, as displayed in Fig. 6.6. The above results are the contribution from the lowest-order Feynman diagram, Fig. 6.5. There are, of course, corrections of order $\alpha^3, \alpha^4, \dots$, arising due to interference with, or directly from, the amplitudes of higher-order diagrams, such as those of Fig. 6.7.

6.6 Helicity Conservation at High Energies

We can obtain further physical insight into the structure of results like (6.30) and (6.31) by looking at the helicity of the particles. As we will be often interested in highly relativistic interactions, it is useful to study the structure of the electromagnetic current in this limit. We start with (5.78) and note that for a fermion of energy $E \gg m$,

$$\begin{aligned} \frac{1}{2}(1 - \gamma^5)u &= u_L \\ \frac{1}{2}(1 + \gamma^5)u &= u_R. \end{aligned} \tag{6.34}$$

That is, $\frac{1}{2}(1 \pm \gamma^5)$ project out the helicity $\lambda = \pm \frac{1}{2}$ components of a spinor, respectively. We can use this observation to show that

$$\bar{u}\gamma^\mu u \equiv (\bar{u}_L + \bar{u}_R)\gamma^\mu(u_L + u_R) = \bar{u}_L\gamma^\mu u_L + \bar{u}_R\gamma^\mu u_R, \tag{6.35}$$

and so at high energies the electromagnetic interaction [see (6.4) and (6.6)] conserves the helicity of the scattered fermion.

The proof is as follows. First note that

$$\bar{u}_L = u_L^\dagger \gamma^0 = u_L^\dagger \frac{1}{2}(1 - \gamma^5)\gamma^0 = \bar{u}_L \frac{1}{2}(1 + \gamma^5), \tag{6.36}$$

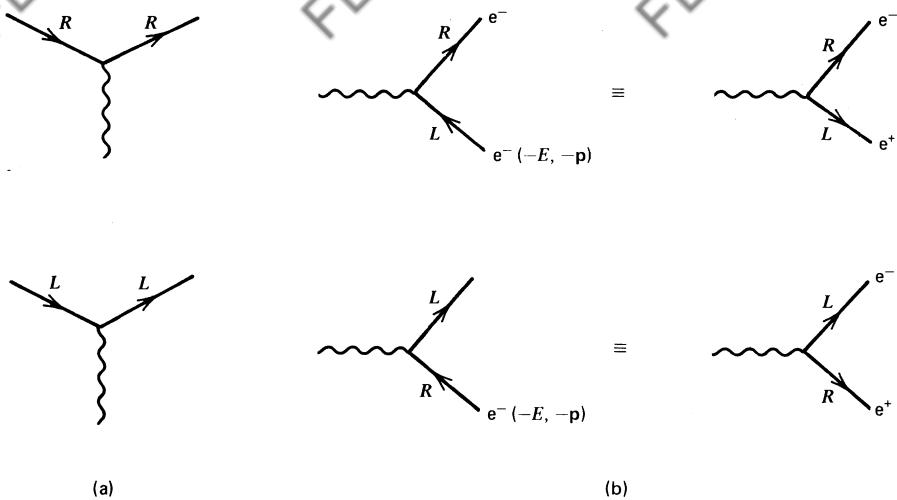


Fig. 6.8 The allowed vertices to $O(m/E)$ for (a) fermion scattering and (b) the crossed or annihilation channel. In all diagrams time increases from left to right.

since $\gamma^5 = \gamma^{5\dagger}$ and $\gamma^5\gamma^0 = -\gamma^0\gamma^5$. Hence,

$$\begin{aligned}\bar{u}_L \gamma^\mu u_R &= \frac{1}{4} \bar{u}(1 + \gamma^5)\gamma^\mu(1 + \gamma^5)u \\ &= \frac{1}{4} \bar{u}\gamma^\mu(1 - \gamma^5)(1 + \gamma^5)u \\ &= 0,\end{aligned}\quad (6.37)$$

where we have used $\gamma^5\gamma^\mu = -\gamma^\mu\gamma^5$ and $(\gamma^5)^2 = 1$. Helicity conservation clearly holds for any vector (or, for that matter, axial vector) interaction at high energies.

The allowed vertices for high-energy fermion scattering are shown in Fig. 6.8a. On the other hand, in an annihilation channel such as e^-e^+ , the vertices of Fig. 6.8b dominate to order m/E , that is, the outgoing fermions have opposite helicity. We shall see throughout the book how the application of these helicity conservation rules provides valuable insights into electromagnetic and weak interactions.

EXERCISE 6.4 Assuming a vector–axial vector form of the weak interaction, explain why the electron emitted in the μ^- -decay process, $\mu^- \rightarrow e^- \bar{\nu}_e \nu_\mu$, must be left-handed. What is the helicity of e^+ from μ^+ decay?

Let us consider e^+e^- annihilation in more detail. Helicity conservation requires that the incoming e^- and e^+ have opposite helicities, see Fig. 6.8b. The same is true for the μ^- and μ^+ in the final state. Thus, in the center-of-mass frame, scattering proceeds from an initial state with $J_z = +1$ or -1 to a final state with $J_{z'} = +1$ or -1 , where the z, z' axes are along the ingoing e^- and outgoing μ^- directions, respectively. One of the four possibilities is sketched in Fig. 6.9. The reaction proceeds via an intermediate photon of spin $j = 1$, and therefore the

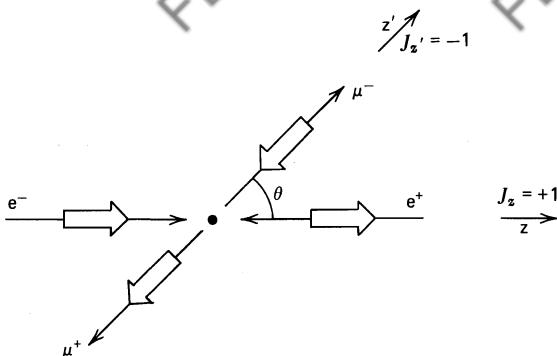


Fig. 6.9 The process $e^- e^+ \rightarrow \mu^- \mu^+$ in a particular helicity configuration: $\mathcal{M}(e_R^- e_L^+ \rightarrow \mu_L^- \mu_R^+) \sim d_{-11}^1(\theta)$.

amplitudes are proportional to the rotation matrices

$$d_{\lambda\lambda'}^j(\theta) \equiv \langle j\lambda' | e^{-i\theta J_y} | j\lambda \rangle, \quad (6.38)$$

where y is perpendicular to the reaction plane, θ is the center-of-mass scattering angle, and λ, λ' are the net helicities along the z, z' axes. We adopt the conventional notation for the rotation matrices. They can be readily worked out from angular momentum theory and are tabulated in many places (e.g., Martin and Spearman, 1970). The four possible helicity amplitudes have the same vertex factors and are thus proportional to [see (4.45) and (2.21)]

$$\begin{aligned} d_{11}^1(\theta) &= d_{-1-1}^1(\theta) = \frac{1}{2}(1 + \cos \theta) \approx -\frac{u}{s} \\ d_{1-1}^1(\theta) &= d_{-11}^1(\theta) = \frac{1}{2}(1 - \cos \theta) \approx -\frac{t}{s}. \end{aligned} \quad (6.39)$$

Spin averaging these amplitudes gives the desired result [see (6.31)]

$$|\overline{\mathcal{M}}|^2 \propto \frac{u^2 + t^2}{s^2}. \quad (6.40)$$

It is a straightforward consequence of angular momentum conservation and embodies, for example, the requirement that the amplitude for the process of Fig. 6.9 must vanish in the forward direction, because here the net helicity is not conserved.

EXERCISE 6.5 Use rotation matrix arguments to show that for “spin-less” electrons and muons

$$\mathcal{M}(e^- e^+ \rightarrow \mu^- \mu^+) \propto \frac{t-u}{s}. \quad (6.41)$$

Compare the s -channel photon contribution of (4.47).

TABLE 6.1

Leading Order Contributions to Representative QED Processes

	Feynman Diagrams		$ \mathcal{M} ^2/2e^4$		
	Forward peak	Backward peak	Forward	Interference	Backward
Möller scattering $e^-e^- \rightarrow e^-e^-$ (Crossing $s \leftrightarrow u$)			$\frac{s^2 + u^2}{t^2} + \frac{2s^2}{tu} + \frac{s^2 + t^2}{u^2}$		$(u \leftrightarrow t \text{ symmetric})$
Bhabha scattering $e^-e^+ \rightarrow e^-e^+$			Forward	“Time-like”	Forward
$e^- \mu^- \rightarrow e^- \mu^-$ (Crossing $s \leftrightarrow t$)			$\frac{s^2 + u^2}{t^2} + \frac{2u^2}{ts} + \frac{u^2 + t^2}{s^2}$	Interference	Time-like
$e^-e^+ \rightarrow \mu^-\mu^+$				$\frac{s^2 + u^2}{t^2}$	

6.7 Survey of $e^-e^+ \rightarrow e^-e^+, \mu^-\mu^+$

We have calculated above some typical QED processes in the extreme relativistic limit. It is instructive to summarize these and related results in Table 6.1. Apart from the interference terms, all the contributions listed in the table follow simply from the last entry, which itself is easily obtained by helicity arguments.

Similar results are found in QCD for the “strong” $qq \rightarrow qq, q\bar{q} \rightarrow q\bar{q}$ interactions via single gluon exchange. In fact, the results are identical except that we must average (sum) over the colors of the initial (final) quarks, in addition to their spins, and make the replacement $\alpha \rightarrow \alpha_s$, where α_s is the quark-gluon coupling introduced in Chapters 1 and 2.

EXERCISE 6.6 Show that the spin-averaged interference term between the two Feynman diagrams for electron-electron scattering is that shown in the table.

Wherever an internal photon line occurs in a Feynman diagram, we shall see in Chapter 13 that we also have the possibility of a contribution from a massive neutral weak boson Z^0 . Although the Feynman rules for the weak bosons are given in detail in Chapter 13, it is worth quickly anticipating the possible effect of γ - Z^0 interference. Processes such as $e^-e^+ \rightarrow e^-e^+$, $\mu^-\mu^+$ with *s*-channel or “time-like” Feynman diagrams should increasingly feel the effects of any such Z bosons as the center-of-mass energy approaches its mass value M_Z . From Table 6.1, it is clear that it is preferable to study $e^-e^+ \rightarrow \mu^-\mu^+$ rather than Bhabha scattering, since in the latter process, Z effects will be swamped by *t*-channel photon exchange. The details are given in Section 13.6.

6.8 $e^- \mu^- \rightarrow e^- \mu^-$ in the Laboratory Frame. Kinematics Relevant to the Parton Model

Before we leave fermion scattering, it is useful to introduce laboratory frame kinematics, that is, the frame where the initial μ is at rest. We can then directly apply these results to electron–quark scattering when probing the structure of hadrons in Chapter 8 and onward.

We return to the “exact” formula (6.27) for $e^-(k) + \mu^-(p) \rightarrow e^-(k') + \mu^-(p')$ and neglect only the terms involving the electron mass m ,

$$\begin{aligned} |\mathcal{D}|^2 &= \frac{8e^4}{q^4} [(k' \cdot p')(k \cdot p) + (k' \cdot p)(k \cdot p') - M^2 k' \cdot k] \\ &= \frac{8e^4}{q^4} \left[-\frac{1}{2} q^2 (k \cdot p - k' \cdot p) + 2(k' \cdot p)(k \cdot p) + \frac{1}{2} M^2 q^2 \right] \quad (6.42) \end{aligned}$$

where $q = k - k'$. To obtain the last line, we have used $p' = k - k' + p$, $k^2 = k'^2 \approx 0$ and $q^2 \approx -2k \cdot k'$.

We wish to evaluate the cross section in the laboratory frame, the frame in which the muon is initially at rest, $p = (M, \mathbf{0})$. The particle momenta in this frame are shown in Fig. 6.10.

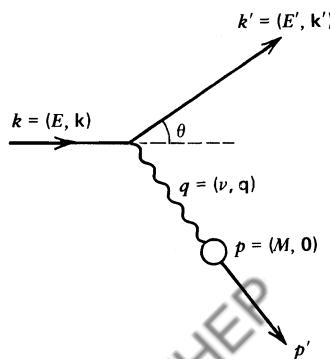


Fig. 6.10 The process $e^- \mu^- \rightarrow e^- \mu^-$ in the laboratory frame.

Evaluating (6.42) in the laboratory frame, we find

$$\begin{aligned} |\mathcal{M}|^2 &= \frac{8e^4}{q^4} \left(-\frac{1}{2} q^2 M(E - E') + 2EE'M^2 + \frac{1}{2} M^2 q^2 \right) \\ &= \frac{8e^4}{q^4} 2M^2 E'E \left\{ 1 + \frac{q^2}{4EE'} - \frac{q^2}{2M^2} \frac{M(E - E')}{2EE'} \right\} \\ &= \frac{8e^4}{q^4} 2M^2 E'E \left\{ \cos^2 \frac{\theta}{2} - \frac{q^2}{2M^2} \sin^2 \frac{\theta}{2} \right\}. \end{aligned} \quad (6.43)$$

To reach the last line, we have used the following kinematic relations:

$$q^2 \approx -2k \cdot k' \approx -2EE'(1 - \cos \theta) = -4EE' \sin^2 \frac{\theta}{2}. \quad (6.44)$$

Also, squaring $q + p = p'$, we obtain

$$q^2 = -2p \cdot q = -2\nu M \quad \text{so} \quad \nu \equiv E - E' = -\frac{q^2}{2M}. \quad (6.45)$$

To calculate the $e^- \mu^- \rightarrow e^- \mu^-$ cross section, we make use of (4.27):

$$\begin{aligned} d\sigma &= \frac{1}{(2E)(2M)} \frac{|\mathcal{M}|^2}{4\pi^2} \frac{d^3 k'}{2E'} \frac{d^3 p'_0}{2p'_0} \delta^{(4)}(p + k - p' - k') \\ &= \frac{1}{4ME} \frac{|\mathcal{M}|^2}{4\pi^2} \frac{1}{2} E' dE' d\Omega \frac{d^3 p'_0}{2p'_0} \delta^{(4)}(p + q - p'). \end{aligned} \quad (6.46)$$

The flux is the product of beam and target densities $(2E)(2M)$ multiplied by the relative velocity which is 1 (i.e., the speed of light) in the limit where the electron mass has been neglected.

EXERCISE 6.7 Justify the following relations:

$$\begin{aligned} \int \frac{d^3 p'}{2p'_0} \delta^{(4)}(p + q - p') &= \int d^3 p' dp'_0 \delta^{(4)}(p + q - p') \theta(p'_0) \delta(p'^2 - M^2) \\ &= \frac{1}{2M} \delta\left(\nu + \frac{q^2}{2M}\right) \end{aligned} \quad (6.47)$$

$$= \frac{1}{2MA} \delta(E' - E/A), \quad (6.48)$$

where $A = 1 + (2E/M)\sin^2 \frac{\theta}{2}$, and the step function $\theta(x)$ is 1 if $x > 0$ and 0 otherwise.

Inserting (6.43) into (6.46) and using (6.47), we obtain

$$\frac{d\sigma}{dE' d\Omega} = \frac{(2\alpha E')^2}{q^4} \left\{ \cos^2 \frac{\theta}{2} - \frac{q^2}{2M^2} \sin^2 \frac{\theta}{2} \right\} \delta\left(\nu + \frac{q^2}{2M}\right). \quad (6.49)$$

Using (6.48), we may perform the dE' integration and, replacing q^2 by (6.44), we finally arrive at the following formula for the differential cross section for $e^- \mu^-$ scattering in the laboratory frame:

$$\left. \frac{d\sigma}{d\Omega} \right|_{\text{lab.}} = \left(\frac{\alpha^2}{4E^2 \sin^4 \frac{\theta}{2}} \right) \frac{E'}{E} \left\{ \cos^2 \frac{\theta}{2} - \frac{q^2}{2M^2} \sin^2 \frac{\theta}{2} \right\}. \quad (6.50)$$

A powerful technique for exploring the internal structure of a target is to bombard it with a beam of high-energy electrons and to observe the angular distribution and energy of the scattered electrons. Such experiments have repeatedly led to major advances in our understanding of the structure of matter. Starting in Chapter 8, we describe how this method has revealed the structure of the proton. Equation (6.50) plays a central role in the story.

EXERCISE 6.8 Show that the cross section for elastic scattering of unpolarized electrons from spinless point-like particles is

$$\left. \frac{d\sigma}{d\Omega} \right|_{\text{lab.}} = \left(\frac{\alpha^2}{4E^2 \sin^4 \frac{\theta}{2}} \right) \frac{E'}{E} \cos^2 \frac{\theta}{2}, \quad (6.51)$$

where as before we neglect the mass of the electron. Justify using (6.18) with $L_{\mu\nu}^{\text{muon}}$ replaced by $(p + p')_\mu (p + p')_\nu$. Comparing the cross section with that for $e^- \mu^- \rightarrow e^- \mu^-$, we see that the $\sin^2(\theta/2)$ in (6.50) is due to scattering from the magnetic moment of the muon.

6.9 Photons. Polarization Vectors

We have already noted that, in the presence of a current density j , the electromagnetic field A^μ satisfies

$$\square^2 A^\mu = j^\mu. \quad (6.52)$$

The following two exercises recall how this equation arises from Maxwell's equations.

EXERCISE 6.9 Maxwell's equations of classical electrodynamics are, *in vacuo*,

$$\begin{aligned}\nabla \cdot \mathbf{E} &= \rho, & \nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} &= 0 \\ \nabla \cdot \mathbf{B} &= 0, & \nabla \times \mathbf{B} - \frac{\partial \mathbf{E}}{\partial t} &= \mathbf{j}\end{aligned}\tag{6.53}$$

(where we are using Heaviside–Lorentz rationalized units, see Appendix C of Aitchison and Hey). Show that these equations are equivalent to the following covariant equation for A^μ :

$$\square^2 A^\mu - \partial^\mu (\partial_\nu A^\nu) = j^\mu,\tag{6.54}$$

with $j^\mu = (\rho, \mathbf{j})$, and where $A^\mu = (\phi, \mathbf{A})$, the four-vector potential, is related to the electric and magnetic fields by

$$\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t} - \nabla \phi, \quad \mathbf{B} = \nabla \times \mathbf{A}.\tag{6.55}$$

Further, show that in terms of the antisymmetric field strength tensor

$$F^{\mu\nu} \equiv \partial^\mu A^\nu - \partial^\nu A^\mu\tag{6.56}$$

Maxwell's equations take the compact form

$$\partial_\mu F^{\mu\nu} = j^\nu,\tag{6.57}$$

and that current conservation, $\partial_\nu j^\nu = 0$, follows as a natural compatibility condition. [Note that $\nabla \times (\nabla \times \mathbf{A}) = -\nabla^2 \mathbf{A} + \nabla(\nabla \cdot \mathbf{A})$.]

EXERCISE 6.10 Verify that \mathbf{E} and \mathbf{B} in (6.55) are unchanged by the gauge transformation

$$A_\mu \rightarrow A'_\mu = A_\mu + \partial_\mu \chi,\tag{6.58}$$

where χ can be any function of x . Use this freedom to write Maxwell's equations in the form

$$\square^2 A^\mu = j^\mu \quad \text{with } \partial_\mu A^\mu = 0.\tag{6.59}$$

The requirement $\partial_\mu A^\mu = 0$ is known as the Lorentz condition. However, even after imposing this, there is still some residual freedom in the choice of the

potential A^μ . We can still make another gauge transformation,

$$A_\mu \rightarrow A'_\mu = A_\mu + \partial_\mu \Lambda \quad (6.60)$$

where Λ is any function that satisfies

$$\square^2 \Lambda = 0. \quad (6.61)$$

This last equation ensures that the Lorentz condition is still satisfied.

Turning now from classical to quantum mechanics, we see that the wavefunction A^μ for a free photon satisfies the equation

$$\square^2 A^\mu = 0, \quad (6.62)$$

which has solutions

$$A^\mu = \epsilon^\mu(\mathbf{q}) e^{-iq \cdot x}. \quad (6.63)$$

The four-vector ϵ^μ is called the polarization vector of the photon. On substituting into the equation, we find that q , the four-momentum of the photon, satisfies

$$q^2 = 0, \quad \text{that is, } m_\gamma = 0. \quad (6.64)$$

The polarization vector has four components and yet it describes a spin-1 particle. How does this come about? First, the Lorentz condition, $\partial_\mu A^\mu = 0$, gives

$$q_\mu \epsilon^\mu = 0, \quad (6.65)$$

and this reduces the number of independent components of ϵ^μ to three. Moreover, we have to explore the consequences of the additional gauge freedom (6.60). Choose a gauge parameter

$$\Lambda = iae^{-iq \cdot x}$$

with a constant so that (6.61) is satisfied. Substituting this, together with (6.63), into (6.60) shows that the physics is unchanged by the replacement

$$\epsilon_\mu \rightarrow \epsilon'_\mu = \epsilon_\mu + aq_\mu. \quad (6.66)$$

In other words, two polarization vectors ($\epsilon_\mu, \epsilon'_\mu$) which differ by a multiple of q_μ describe the same photon. We may use this freedom to ensure that the time component of ϵ^μ vanishes, $\epsilon^0 \equiv 0$; and then the Lorentz condition (6.65) reduces to

$$\boldsymbol{\epsilon} \cdot \mathbf{q} = 0. \quad (6.67)$$

This (noncovariant) choice of gauge is known as the Coulomb gauge.

From (6.67), we see that there are only *two* independent polarization vectors and that they are both transverse to the three-momentum of the photon. For example, for a photon traveling along the z axis, we may take

$$\boldsymbol{\epsilon}_1 = (1, 0, 0), \quad \boldsymbol{\epsilon}_2 = (0, 1, 0). \quad (6.68)$$

A free photon is thus described by its momentum q and a polarization vector $\boldsymbol{\epsilon}_i$. Since $\boldsymbol{\epsilon}_i$ transforms as a vector, we anticipate that it is associated with a particle of spin 1.

EXERCISE 6.11 Determine how the linear combinations

$$\begin{aligned}\boldsymbol{\epsilon}_R &= -\sqrt{\frac{1}{2}}(\boldsymbol{\epsilon}_1 + i\boldsymbol{\epsilon}_2) \\ \boldsymbol{\epsilon}_L &= \sqrt{\frac{1}{2}}(\boldsymbol{\epsilon}_1 - i\boldsymbol{\epsilon}_2)\end{aligned}\quad (6.69)$$

transform under a rotation θ about the z axis. Hence, show that $\boldsymbol{\epsilon}_R$ and $\boldsymbol{\epsilon}_L$ describe a photon of helicity $+1$ and -1 , respectively; $\boldsymbol{\epsilon}_{R,L}$ are called circular polarization vectors.

EXERCISE 6.12 Show that (in the transverse gauge) the completeness relation is

$$\sum_{\lambda=R, L} (\boldsymbol{\epsilon}_\lambda)_i^* (\boldsymbol{\epsilon}_\lambda)_j = \delta_{ij} - \hat{q}_i \hat{q}_j \quad (6.70)$$

If $\boldsymbol{\epsilon}$ were along \mathbf{q} , it would be associated with a helicity-zero photon. This state is missing because of the transversality condition, $\mathbf{q} \cdot \boldsymbol{\epsilon} = 0$. It can only be absent because the photon is massless. We return to a further discussion of photon polarization vectors in Section 6.13.

6.10 More on Propagators. The Electron Propagator

Here, we wish to consolidate our earlier discussion of propagators and also introduce the propagator for the electron.

First, let us recall the nonrelativistic perturbation expansion of the transition amplitude [see (3.44) and (4.48)]

$$T_{fi} = -i2\pi \delta(E_f - E_i) \left(\langle f | V | i \rangle + \sum_{n \neq i} \langle f | V | n \rangle \frac{1}{E_i - E_n} \langle n | V | i \rangle + \dots \right). \quad (6.71)$$

Recall also that we associated factors such as $\langle f | V | n \rangle$ with the vertices and identified $1/(E_i - E_n)$ as the propagator (see Fig. 3.4 and Section 4.8). The state vectors are eigenstates of the Hamiltonian in the absence of V [see (3.29)]

$$H_0 |n\rangle = E_n |n\rangle.$$

Formally, we may therefore rewrite (6.71) as

$$T_{fi} = 2\pi \delta(E_f - E_i) \langle f | (-iV) + (-iV) \frac{i}{E_i - H_0} (-iV) + \dots | i \rangle, \quad (6.72)$$

where we have made use of the completeness relation $\sum |n\rangle \langle n| = 1$. (The prescription for handling the singularity at $E_n = E_i$ is discussed in Section 6.16.) It is natural to take $-iV$, rather than V , as the perturbation parameter. (The $-i$ arises from the i in $i\partial\psi/\partial t = V\psi$, which leads to a time dependence $\exp(-iVt)$ in the

interaction picture.) That is, the vertex factor is $-iV$, and the propagator may thus be regarded as i times the inverse of the Schrödinger operator,

$$-i(E_i - H_0)\psi = -iV\psi, \quad (6.73)$$

acting on the intermediate state.

Let us apply the same technique to the various relativistic wave equations and so deduce the form of the propagators for the corresponding particles.

The Propagator for a Spinless Particle

The form of the Klein–Gordon equation corresponding to (6.73) is

$$i(\square^2 + m^2)\phi = -iV\phi, \quad (6.74)$$

see (4.3). Guided by the relativistic generalization of (6.72), we expect the propagator for a spinless particle to be the inverse of the operator on the left-hand side of (6.74). For an intermediate state of momentum p , this gives

$$\frac{1}{i(-p^2 + m^2)} = \boxed{\frac{i}{p^2 - m^2}}. \quad (6.75)$$

Indeed, we have already discussed how this form arises as the relativistic generalization of the propagator (see Section 4.8).

The Electron Propagator

An electron in an electromagnetic field satisfies

$$(\not{p} - m)\psi = -e\gamma^\mu A_\mu\psi, \quad (6.76)$$

see (6.2) and (6.3). As before, we must multiply by $-i$. Hence, the vertex factor is $ie\gamma^\mu$. The electron propagator is therefore the inverse of $-i$ times the left-hand side of (6.76):

$$\frac{1}{-i(\not{p} - m)} = \frac{i}{\not{p} - m} = \boxed{\frac{i(\not{p} + m)}{p^2 - m^2}} = \frac{i \sum_s u\bar{u}}{p^2 - m^2}, \quad (6.77)$$

where we have used $\not{p}\not{p} = p^2$ and the completeness relation (5.47). The numerator contains the sum over the spin states of the virtual electron; a further discussion is given in Section 6.16.

In summary, the general form of the propagator of a virtual particle is

$$\frac{i \sum_{\text{spins}}}{p^2 - m^2}.$$

The spin sum is the completeness relation; we include all possible spin states of

the propagating particle. We would also integrate over the different momentum states that propagate. For the diagrams we have considered so far, this momentum is fixed by the momenta of the external particles.

We discuss the prescription for handling the singularity at $p^2 = m^2$ in Section 6.16.

6.11 The Photon Propagator

The propagator for a photon is not unique, on account of the freedom in the choice of A^μ . Recall that physics is unchanged by the transformation

$$A^\mu \rightarrow A^\mu + \partial^\mu \chi$$

which, as will be explained in Chapter 14, is associated with the invariance of QED under phase or “gauge” transformations of the wavefunctions of charged particles.

From (6.54), we see that the wave equation for a photon can be written in the form

$$(g^{\nu\lambda} \square^2 - \partial^\nu \partial^\lambda) A_\lambda = j^\nu, \quad (6.78)$$

and, in fact, a photon propagator cannot exist until we remove some of the gauge freedom of A_λ , see Exercise 6.13.

EXERCISE 6.13 Verify that the inverse of the “momentum space operator” of (6.78) does not exist.

Hint Attempt to write the inverse in the most general form satisfying Lorentz covariance

$$A q^2 g_{\mu\nu} + B q_\mu q_\nu \quad (6.79)$$

where A and B are functions of q^2 .

In our discussions so far, we have chosen to work in the Lorentz class of gauges with $\partial_\lambda A^\lambda = 0$. The wave equation, (6.78), then simplifies to

$$g^{\nu\lambda} \square^2 A_\lambda = j^\nu. \quad (6.80)$$

Now, since

$$g_{\mu\nu} g^{\nu\lambda} = \delta_\mu^\lambda, \quad (6.81)$$

where δ_μ^λ equals 1 if $\lambda = \mu$, and 0 otherwise, the propagator (the inverse of the momentum space operator multiplied by $-i$) is

$i \frac{-g_{\mu\nu}}{q^2}$

(6.82)

The discussion in Section 6.10 implies that we should associate $-g_{\mu\nu}$ with the sum over the polarization vectors of the virtual photon. This we tackle in Section 6.13. The gauge condition $\partial_\lambda A^\lambda = 0$ has been imposed covariantly, and the resulting covariant propagator, (6.82), is ideal for QED calculations. It is called the *Feynman propagator*, and we say we are working in the Feynman gauge. Indeed, this is the photon propagator we have been using so far, and we have already entered it into our table of Feynman rules, see Section 6.17.

EXERCISE 6.14 The condition $\partial_\lambda A^\lambda = 0$ does not fully define the propagator. We are at liberty to rewrite wave equation (6.78) as

$$\left[g^{\nu\lambda} \square^2 - \left(1 - \frac{1}{\xi}\right) \partial^\nu \partial^\lambda \right] A_\lambda = j^\nu. \quad (6.83)$$

In this case, use (6.79) to show that the propagator is

$$\frac{i}{q^2} \left(-g_{\mu\nu} + (1 - \xi) \frac{q_\mu q_\nu}{q^2} \right). \quad (6.84)$$

The Feynman gauge takes $\xi = 1$. But in any case, the extra term in the propagator vanishes in QED calculations in which the virtual photon is coupled to conserved currents which satisfy $q_\mu j^\mu = q_\nu j^\nu = 0$.

6.12 Massive Vector Particles

Massive vector (spin 1) particles, denoted W^\pm and Z^0 , play a central role in the theory of weak interactions (see Chapter 12 onward). The wave equation for a spin-1 particle of mass M can be obtained from that for the photon by the replacement $\square^2 \rightarrow \square^2 + M^2$; recall the Klein–Gordon operator (3.19). From (6.78), we see that the wavefunction B_λ for a free particle satisfies

$$(g^{\nu\lambda} (\square^2 + M^2) - \partial^\nu \partial^\lambda) B_\lambda = 0. \quad (6.85)$$

Proceeding exactly as before, we determine the inverse of the momentum space operator by solving

$$(g^{\nu\lambda} (-p^2 + M^2) + p^\nu p^\lambda)^{-1} = \delta_\lambda^\nu (A g_{\mu\nu} + B p_\mu p_\nu) \quad (6.86)$$

for A and B . The propagator, which is the quantity in brackets on the right-hand side of (6.86) multiplied by i , is found to be

$$\frac{i(-g^{\mu\nu} + p^\mu p^\nu / M^2)}{p^2 - M^2}.$$

(6.87)

We can show that the numerator is the sum over the three spin states of the

massive particle when taken on-shell $p^2 = M^2$. We first take the divergence, ∂_ν , of (6.85). Two terms cancel, and we find

$$M^2 \partial^\lambda B_\lambda = 0, \quad \text{that is, } \partial^\lambda B_\lambda = 0. \quad (6.88)$$

Thus, for a massive vector particle, we have no choice but to take $\partial^\lambda B_\lambda = 0$; it is not a gauge condition. As a consequence, the wave equation reduces to

$$(\square^2 + M^2) B_\mu = 0 \quad (6.89)$$

with free-particle solutions

$$B_\mu = \epsilon_\mu e^{-ip \cdot x}. \quad (6.90)$$

The condition (6.88) demands

$$p^\mu \epsilon_\mu = 0 \quad (6.91)$$

and so reduces the number of independent polarization vectors from four to three in a covariant fashion.

EXERCISE 6.15 For a vector particle of mass M , energy E , and momentum \mathbf{p} along the z axis, show that states of helicity λ can be described by polarization vectors

$$\begin{aligned} \epsilon^{(\lambda=\pm 1)} &= \mp(0, 1, \pm i, 0)/\sqrt{2}, \\ \epsilon^{(\lambda=0)} &= (|\mathbf{p}|, 0, 0, E)/M. \end{aligned} \quad (6.92)$$

EXERCISE 6.16 Show that the completeness relation is

$$\sum_{\lambda} \epsilon_{\mu}^{(\lambda)*} \epsilon_{\nu}^{(\lambda)} = -g_{\mu\nu} + \frac{p_{\mu} p_{\nu}}{M^2} \quad (6.93)$$

where the sum is over the three polarization states of the massive vector particle.

6.13 Real and Virtual Photons

For a real photon, we saw that there are only *two* polarization states. Indeed, in Section 6.9 we found we could choose $\epsilon_0^{(\lambda)} \equiv 0$ and $\mathbf{q} \cdot \epsilon^{(\lambda)} = 0$, so that only transverse polarization states remain. Recall that the completeness relation for these polarization vectors is

$$\sum_{T=R, L} \epsilon_i^T * \epsilon_j^T = \delta_{ij} - \hat{q}_i \hat{q}_j, \quad (6.94)$$

where T denotes transverse [see (6.70)].

On the other hand, we have associated with a virtual photon the covariant propagator $i(-g_{\mu\nu})/q^2$, where $-g_{\mu\nu}$ implies we are summing over *four* polarizations.

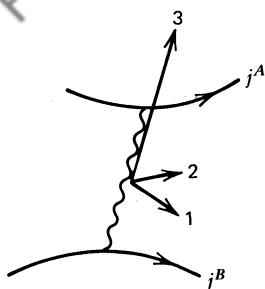


Fig. 6.11 A virtual photon exchanged between two charged particles A and B.

zation states. The completeness relation now reads in an obvious notation

$$\begin{aligned}
 -g_{\mu\nu} &= \sum_{\lambda=1}^4 \epsilon_{\mu}^{(\lambda)*} \epsilon_{\nu}^{(\lambda)} = \sum_T \epsilon_{\mu}^T * \epsilon_{\nu}^T + \epsilon_{\mu}^L * \epsilon_{\nu}^L + \epsilon_{\mu}^S * \epsilon_{\nu}^S \\
 &= \underbrace{(\delta_{ij} - \hat{q}_i \hat{q}_j)}_{\text{transverse}} + \underbrace{\hat{q}_i \hat{q}_j}_{\text{longitudinal}} + (-g_{\mu 0} g_{\nu 0}). \quad (6.95)
 \end{aligned}$$

However, in a sense every photon is virtual, being emitted and then sooner or later being absorbed. How can we reconcile the two descriptions?

Let us look at a typical Feynman diagram, Fig. 6.11, containing a virtual photon exchanged between charged particles. For such diagrams, see, for example, Fig. 6.2, we have found a transition amplitude of the form [see the discussion leading to (6.8)]

$$\begin{aligned}
 T_{fi} &= -i \int j_{\mu}^A(x) \left(\frac{-g^{\mu\nu}}{q^2} \right) j_{\nu}^B(x) d^4x \\
 &= -i \int \underbrace{\left(\frac{j_1^A j_1^B + j_2^A j_2^B}{q^2} \right)}_{\text{transverse}} + \underbrace{\left(\frac{j_3^A j_3^B - j_0^A j_0^B}{q^2} \right)}_{\text{longitudinal/scalar}} d^4x \quad (6.96)
 \end{aligned}$$

where we have taken the photon four-momentum $q^{\mu} = (q^0, 0, 0, |\mathbf{q}|)$. That is, we choose the 3-axis to be along $\hat{\mathbf{q}}$. Recall that charge conservation gives rise to the continuity equation $\partial^{\mu} j_{\mu} = 0$. For both the A, B currents, this implies

$$q^{\mu} j_{\mu} = q^0 j_0 - |\mathbf{q}| j_3 = 0. \quad (6.97)$$

Thus, if the exchanged photon is almost real, $q^0 \approx |\mathbf{q}|$, then $j_3 \approx j_0$ and the longitudinal and scalar contributions cancel each other, leaving only the two transverse contributions. For a real photon, we can therefore make the replacement

$$\sum_T \epsilon_{\mu}^T * \epsilon_{\nu}^T \rightarrow -g_{\mu\nu}. \quad (6.98)$$

On the other hand, for a virtual photon the longitudinal and scalar components cannot be neglected. Indeed, they play an important role. If we use (6.97) to

substitute for j_3^A and j_3^B in (6.96), we find

$$T_{fi} = -i \int \left(\frac{j_1^A j_1^B + j_2^A j_2^B}{q^2} + \frac{j_0^A j_0^B}{|\mathbf{q}|^2} \right) d^4x. \quad (6.99)$$

The first term describes the propagation of virtual photons in transverse polarization states, whereas the second term, with its $|\mathbf{q}|^2$ denominator, is not associated with propagation. Instead, it represents the instantaneous Coulomb interaction between the charges of the two particles, j_0^A and j_0^B . This becomes clear if we rewrite the second term of (6.99) in the form

$$T_{fi}^{\text{Coul}} = -i \int dt \int d^3x_1 \int d^3x_2 \frac{j_0^A(t, \mathbf{x}_1) j_0^B(t, \mathbf{x}_2)}{4\pi |\mathbf{x}_2 - \mathbf{x}_1|}, \quad (6.100)$$

and note that the charges interact without retardation at time t .

EXERCISE 6.17 Verify (6.100) by making use of the Fourier transform

$$\frac{1}{|\mathbf{q}|^2} = \int d^3x e^{i\mathbf{q} \cdot \mathbf{x}} \frac{1}{4\pi |\mathbf{x}|}. \quad (6.101)$$

Finally, by inspection of (6.95), we see that the division of $-g^{\mu\nu}/q^2$ into a transverse propagating contribution and a longitudinal/scalar static contribution is not a Lorentz covariant separation. Only the sum forms a covariant photon propagator.

6.14 Compton Scattering $\gamma e^- \rightarrow \gamma e^-$

Compton scattering is a useful example to work through in detail. Not only do the Feynman diagrams involve both the electron propagator and external photons, but we shall need the form of the amplitude for the analogous process $\gamma^* q \rightarrow g q$ (involving quarks q and gluons g) in the development of QCD in Chapter 10.

The two (lowest-order) Feynman diagrams are shown in Fig. 6.12, and the factors needed to compute the amplitude are shown in detail on the first diagram.

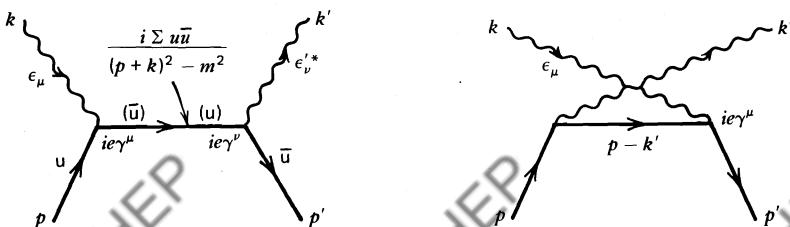


Fig. 6.12 Feynman diagrams for Compton scattering $\gamma e^- \rightarrow \gamma e^-$.

A word of explanation is in order. For the incoming photon we have

$$A_\mu = \epsilon_\mu e^{-ik \cdot x},$$

where ϵ_μ is one of the two transverse polarization vectors. Referring back to (6.4), we see that the $e^{-ik \cdot x}$ is disposed of by the x integration, which results in momentum conservation at the vertex. We are thus led to include only the factor ϵ_μ for an incoming photon line in our table of Feynman rules at the end of the chapter. Similarly, for an outgoing photon, $(\epsilon'_\nu e^{-ik' \cdot x})^*$, we are led to a factor ϵ'^*_ν . Note that, just as in Fig. 6.1b, the structure of the electron-photon vertex is

$$(\bar{u} \gamma^\mu u) \epsilon_\mu, \quad (6.102)$$

but that here \bar{u} is already contained in the electron propagator, whereas for electron scattering vertices it is ϵ_μ that is embodied in the photon propagator.

Using the Feynman rules, we obtain the following amplitudes for the two Feynman diagrams:

$$-i\mathcal{M}_1 = \bar{u}^{(s')}(p') \left[\epsilon'^*(ie\gamma^\nu) \frac{i(\not{p} + \not{k} + m)}{(\not{p} + \not{k})^2 - m^2} (ie\gamma^\mu) \epsilon_\mu \right] u^{(s)}(p) \quad (6.103)$$

$$-i\mathcal{M}_2 = \bar{u}^{(s')}(p') \left[\epsilon_\mu (ie\gamma^\mu) \frac{i(\not{p} - \not{k}' + m)}{(\not{p} - \not{k}')^2 - m^2} (ie\gamma^\nu) \epsilon'^*_\nu \right] u^{(s)}(p), \quad (6.104)$$

where p , s and p' , s' are the momentum, spin state of the ingoing and outgoing electrons, respectively. Similarly, k , ϵ and k' , ϵ' are the momentum, polarization vector of the ingoing and outgoing photons, respectively. Note that the invariant amplitude for Compton scattering ($\mathcal{M}_1 + \mathcal{M}_2$) is symmetric under the interchange (or crossing) of the two photons

$$k, \epsilon \leftrightarrow -k', \epsilon'^*. \quad (6.105)$$

This is another example of crossing symmetry.

What does gauge invariance have to say about Compton scattering? Provided we impose the Lorentz condition $\partial_\mu A^\mu = 0$, we saw that physics is unchanged by the replacement

$$\epsilon_\mu \rightarrow \epsilon_\mu + ak_\mu, \quad (6.106)$$

where ϵ_μ and k_μ are the polarization vector and momentum of the photon, respectively, and a is an arbitrary constant, see (6.66). It therefore follows that if we write the amplitude for Compton scattering in the form

$$\mathcal{M} = \epsilon'^*_\nu \epsilon_\mu T^{\mu\nu}, \quad (6.107)$$

then \mathcal{M} is unchanged by substitution $\epsilon_\mu \rightarrow \epsilon_\mu + ak_\mu$ or $\epsilon'_\nu \rightarrow \epsilon'_\nu + ak'_\nu$. Thus, gauge invariance requires

$$k_\mu T^{\nu\mu} = k'_\nu T^{\nu\mu} = 0. \quad (6.108)$$

EXERCISE 6.18 Show that, individually, the amplitudes \mathcal{M}_1 and \mathcal{M}_2 are not gauge invariant but that their sum indeed satisfies (6.108).

It is instructive to work through the calculation of the Compton scattering amplitude in detail. For simplicity, we neglect the mass of the electron, and so the invariant variables for $\gamma(k) + e(p) \rightarrow \gamma(k') + e(p')$ are

$$\begin{aligned} s &= (k+p)^2 = 2k \cdot p = 2k' \cdot p' \\ t &= (k-k')^2 = -2k \cdot k' = -2p \cdot p' \\ u &= (k-p')^2 = -2k \cdot p' = -2p \cdot k'. \end{aligned} \quad (6.109)$$

The two invariant amplitudes, (6.103) and (6.104), are

$$\begin{aligned} \mathcal{M}_1 &= \epsilon'_\nu * \epsilon_\mu e^2 \bar{u}(p') \gamma^\nu (\not{p} + \not{k}) \gamma^\mu u(p) / s, \\ \mathcal{M}_2 &= \epsilon'_\nu * \epsilon_\mu e^2 \bar{u}(p') \gamma^\mu (\not{p} - \not{k}') \gamma^\nu u(p) / u. \end{aligned} \quad (6.110)$$

To obtain the unpolarized cross section, we must average/sum $|\mathcal{M}_1 + \mathcal{M}_2|^2$ over the initial/final electron and photon spins. Fortunately, this is not as difficult as it first appears. For physical photons, (6.98) applies, and we can make the replacement

$$\sum_T \epsilon_\mu^T * \epsilon_{\mu'}^T \rightarrow -g_{\mu\mu'} \quad (6.111)$$

where T denotes transverse. We have a similar completeness relation for the outgoing photon states, ϵ' . Thus, for example,

$$|\overline{\mathcal{M}_1}|^2 = \frac{e^4}{4s^2} \sum_{s,s'} (\bar{u}^{(s')} \gamma^\nu (\not{p} + \not{k}) \gamma^\mu u^{(s)}) (\bar{u}^{(s)} \gamma_\mu (\not{p} + \not{k}) \gamma_\nu u^{(s')}).$$

The factor $\frac{1}{4}$ is due to averaging over the initial electron and photon spins. The spinor completeness relation, (5.47), allows the sum over $u\bar{u}$ states to be performed (just as we described for e^-e^- scattering in Section 6.3), and we find

$$\begin{aligned} |\overline{\mathcal{M}_1}|^2 &= \frac{e^4}{4s^2} \text{Tr} \left(\underbrace{\not{p}' \gamma^\nu}_{-2\not{p}'} (\not{p} + \not{k}) \gamma^\mu \not{p} \gamma_\mu (\not{p} + \not{k}) \gamma_\nu \right) \\ &= \frac{e^4}{s^2} \text{Tr}(\not{p}' \not{k} \not{p} \not{k}) \\ &= \frac{4e^4}{s^2} 2(\not{p}' \cdot \not{k})(\not{p} \cdot \not{k}) \\ &= 2e^4 \left(-\frac{u}{s} \right), \end{aligned} \quad (6.112)$$

where we have made use of (6.24) and (6.22). Similarly, we obtain

$$\begin{aligned} |\overline{\mathcal{M}_2}|^2 &= 2e^4 \left(-\frac{s}{u} \right), \\ \overline{\mathcal{M}_1 \mathcal{M}_2^*} &= 0. \end{aligned}$$

Thus, the spin-averaged Compton amplitude is

$$\overline{|\mathcal{M}|^2} = \overline{|\mathcal{M}_1 + \mathcal{M}_2|^2} = 2e^4 \left(-\frac{u}{s} - \frac{s}{u} \right). \quad (6.113)$$

EXERCISE 6.19 Repeat the above calculation for an incident virtual photon of mass $k^2 \equiv -Q^2$. Continue to use (6.111). Show that for $\gamma^* e^- \rightarrow \gamma e^-$ (where γ^* denotes a virtual photon),

$$\overline{|\mathcal{M}|^2} = 2e^4 \left(-\frac{u}{s} - \frac{s}{u} + \frac{2Q^2 t}{su} \right). \quad (6.114)$$

We shall make use of this result in Chapter 10.

EXERCISE 6.20 Restore the mass m of the electron and show that at high energy, $s \rightarrow \infty$, the integrated cross section for Compton scattering is

$$\sigma = \frac{1}{64\pi^2 s} \int \overline{|\mathcal{M}|^2} d\Omega \rightarrow \frac{2\pi\alpha^2}{s} \log\left(\frac{s}{m^2}\right). \quad (6.115)$$

Note that at high energy the dominant contribution comes from \mathcal{M}_2 , via a glancing collision in which the u -channel electron is almost on mass shell.

EXERCISE 6.21 Show, by using particle helicities, that high-energy Compton scattering via the first diagram of Fig. 6.12 is, in the center-of-mass frame, given by

$$\begin{aligned} \overline{|\mathcal{M}_1|^2} &\propto |d_{++}^{1/2}(\theta)|^2 + |d_{--}^{1/2}(\theta)|^2 \\ &= (1 + \cos\theta) \simeq -\frac{u}{2s}, \end{aligned} \quad (6.116)$$

in agreement with (6.112). An example of this type of calculation is described in Section 6.6.

6.15 Pair Annihilation $e^+ e^- \rightarrow \gamma\gamma$

EXERCISE 6.22 Draw the lowest-order Feynman diagrams for the pair annihilation process

$$e^+(p_1, s_1) + e^-(p_2, s_2) \rightarrow \gamma(k_1, \epsilon_1) + \gamma(k_2, \epsilon_2).$$



Fig. 6.13 Feynman diagrams for $e^+ e^- \rightarrow \gamma\gamma$.

Check your answer with Fig. 6.13. Use the Feynman rules to show that

$$-i\mathfrak{M} = (ie)^2 \bar{v}^{(s_1)}(p_1) \left(\not{\epsilon}_1^* \frac{i}{p_1 - \not{k}_1 - m} \not{\epsilon}_2^* + \not{\epsilon}_2^* \frac{i}{p_1 - \not{k}_2 - m} \not{\epsilon}_1^* \right) u^{(s_2)}(p_2). \quad (6.117)$$

Verify that, in the high-energy limit, the spin-averaged rate is given by

$$\overline{|\mathfrak{M}|^2} = 2e^4 \left(\frac{u}{t} + \frac{t}{u} \right). \quad (6.118)$$

The $e^+e^- \rightarrow \gamma\gamma$ cross section has both forward and backward peaks, corresponding to the t - and u -channel exchanged electrons being almost on mass shell. Result (6.117) can also be obtained by crossing the amplitude for Compton scattering. To go from $\gamma e^- \rightarrow \gamma e^-$ to $e^+e^- \rightarrow \gamma\gamma$, we simply “cross” the ingoing photon with the outgoing electron,

$$k, \epsilon \rightarrow -k_2, \epsilon_2^* \\ p' \rightarrow -p_1 \quad \text{and} \quad \bar{u}^{(s')}(p') \rightarrow \bar{v}^{(s_1)}(p_1).$$

The initial electron and outgoing photon are unaltered:

$$u^{(s)}(p) \equiv u^{(s_2)}(p_2), \quad k', \epsilon'^* \rightarrow k_1, \epsilon_1^*.$$

Making these substitutions in (6.103) and (6.104) gives the pair annihilation amplitude (6.117).

6.16 The $+ie$ Prescription for Propagators

The heuristic treatment of quantum electrodynamics that we have given so far is based on Feynman’s intuitive space-time approach. Our primary aim has been to motivate the Feynman rules and to calculate physical amplitudes. In so doing, we have avoided a detailed discussion of the underlying propagator theory, although we have made free use of the word “propagator.” Here, we try to rectify this omission, but we urge those interested to read Feynman’s original papers and the chapter on propagator theory in Bjorken and Drell (1964).

Green’s Functions

Propagator theory is based on the Green’s function method of solving inhomogeneous differential equations. We explain the method in terms of a simple example. Suppose we wish to solve Poisson’s equation

$$\nabla^2 \phi(\mathbf{x}) = -\rho(\mathbf{x}) \quad (6.119)$$

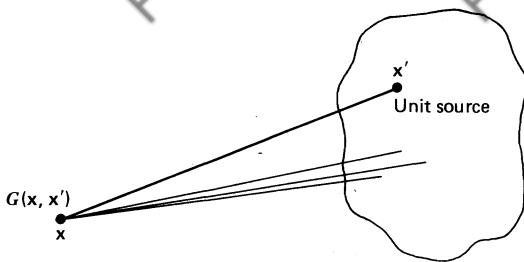


Fig. 6.14 G is the potential at x due to a unit source at x' . We then use the principle of linear superposition to obtain the cumulative potential at x , (6.121), arising from all possible elemental charges $\rho d^3x'$.

for a known charge distribution $\rho(\mathbf{x})$, subject to some boundary condition. It is easier to first solve the “unit source” problem

$$\nabla^2 G = -\delta^{(3)}(\mathbf{x} - \mathbf{x}') \quad (6.120)$$

where $G(\mathbf{x}, \mathbf{x}')$ is the potential at \mathbf{x} due to a unit source at \mathbf{x}' . [For the boundary condition that $G \rightarrow 0$ at large distances, it is easy to show that $G = 1/(4\pi|\mathbf{x} - \mathbf{x}'|)$.] We then move this source over the charge distribution and accumulate the total potential at \mathbf{x} from all possible volume elements d^3x' :

$$\phi(\mathbf{x}) = \int G(\mathbf{x}, \mathbf{x}') \rho(\mathbf{x}') d^3x', \quad (6.121)$$

see Fig. 6.14. We can check directly that ϕ is the desired solution of (6.119) by operating with ∇^2 on (6.121).

The Electron Propagator iS_F

We take the electron propagator as our example and use the Green's function method to solve Dirac's equation, (6.76), for an electron in an electromagnetic field:

$$(i\gamma_\mu \partial^\mu - m)\psi = -e\gamma_\mu A^\mu \psi. \quad (6.122)$$

That is, we first solve the unit source problem

$$(i\gamma_\mu \partial^\mu - m)G_F = \delta^{(4)}(\mathbf{x} - \mathbf{x}'), \quad (6.123)$$

where G_F represents the wave produced at x by a unit source at x' . Once we have found Green's function G_F , we can construct the solution to (6.122):

$$\psi(x) = -e \int d^4x' G_F(x, x') \gamma_\mu A^\mu(x') \psi(x'). \quad (6.124)$$

Note that here ψ appears also on the right-hand side, and so an iterative perturbation series solution in powers of e is obtained.

From translational invariance, $G_F(x, x')$ is a function only of the difference $x - x'$. To solve (6.123), we first Fourier transform to momentum space:

$$G_F(x - x') = \frac{1}{(2\pi)^4} \int S_F(p) e^{-ip \cdot (x - x')} d^4 p. \quad (6.125)$$

Then, on substituting into (6.123), we obtain

$$\frac{1}{(2\pi)^4} \int (\not{p} - m) S_F(p) e^{-ip \cdot (x - x')} d^4 p = \frac{1}{(2\pi)^4} \int e^{-ip \cdot (x - x')} d^4 p,$$

where the right-hand side is the Fourier representation of the delta function. In momentum space, (6.123) therefore becomes simply

$$(\not{p} - m) S_F(p) = 1.$$

That is,

$$S_F(p) = \frac{1}{\not{p} - m} = \frac{\not{p} + m}{p^2 - m^2}. \quad (6.126)$$

So far, this is just a sophisticated version of the derivation of Section 6.10.

To complete the determination of S_F , we need to know how to treat the singularities at

$$p^2 - m^2 = p_0^2 - (\mathbf{p}^2 + m^2) = (p_0 - E)(p_0 + E) = 0.$$

Since the electron is off mass shell, p_0 and $E = (\mathbf{p}^2 + m^2)^{1/2}$ are independent variables. To obtain the correct prescription for integration over the poles at $p_0 = \pm E$, we need to impose the appropriate boundary conditions on $G_F(x - x')$. From (6.125) and (6.126),

$$\begin{aligned} G_F(x - x') &= \frac{1}{(2\pi)^4} \int \frac{\not{p} + m}{(p_0 - E)(p_0 + E)} e^{-ip \cdot (x - x')} d^4 p \\ &= \frac{1}{(2\pi)^4} \int d^3 p e^{i\mathbf{p} \cdot (x - x')} \int_{-\infty}^{\infty} dp_0 \frac{(\gamma_0 p_0 - \boldsymbol{\gamma} \cdot \mathbf{p} + m)}{(p_0 - E)(p_0 + E)} e^{-ip_0(t - t')}. \end{aligned} \quad (6.127)$$

Recall that $G_F(x - x')$ represents the wave produced at x by a unit source at x' . That is, the propagation is from x' to x . Now, we seek an S_F which is associated with the propagation of positive-energy electrons forward in time ($t > t'$) and with negative-energy electrons backward in time ($t < t'$); see Section 3.5. This can be accomplished by performing the p_0 integration along the contour in the complex p_0 plane shown in Fig. 6.15. That is, the required properties of S_F are obtained by choosing the contour along the $\text{Re } p_0$ axis to go below the $p_0 = -E$ pole and above the $p_0 = +E$ pole.

To check that this prescription works, first suppose $t > t'$. Then, from (6.127), we see that to ensure that the contribution from the semicircle vanishes, we must

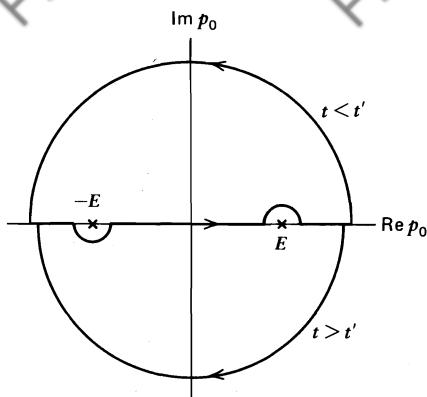


Fig. 6.15 The contours in the complex p_0 plane used to evaluate the dp_0 integral of (6.127).

close the contour in the lower half-plane. We therefore enclose the pole at $p_0 = +E$. Using the Cauchy residue theorem, we obtain

$$\begin{aligned} G_F(x - x') &= \frac{-2\pi i}{(2\pi)^4} \int \frac{d^3 p}{2E} e^{-ip \cdot (x-x')} (\gamma_0 E - \gamma \cdot p + m) \\ &= \frac{-i}{(2\pi)^3} \int \frac{d^3 p}{2E} e^{-ip \cdot (x-x')} (\not{p} + m). \end{aligned} \quad (6.128)$$

Here, $\not{p} + m$ is the operator which projects out the positive-energy electron states, (5.48), and so S_F represents the propagation of $+E$ electrons forward in time.

Now, consider propagation backward in time, $t < t'$. In this case, the semicircle contribution will vanish provided we close the contour in the upper half-plane. We now enclose the pole at $p_0 = -E$, and so

$$G_F(x - x') = \frac{2\pi i}{(2\pi)^4} \int \frac{d^3 p}{(-2E)} e^{ip \cdot (x-x')} e^{-i(-E)(t-t')} (-\gamma_0 E - \gamma \cdot p + m).$$

Since we are integrating over all of three-momentum space, G_F is unchanged by the substitution $p \rightarrow -p$. Therefore,

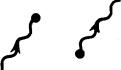
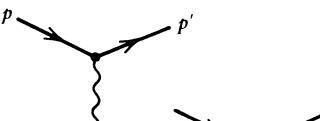
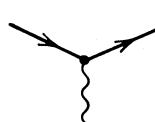
$$G_F(x - x') = \frac{-i}{(2\pi)^3} \int \frac{d^3 p}{2E} e^{ip \cdot (x-x')} (-\not{p} + m), \quad (6.129)$$

where $(-\not{p} + m)$ is the operator which projects out the negative-energy electron states, see (5.48). Thus, S_F represents the propagation of $-E, -p$ electrons backward in time, which is equivalent to the propagation of $+E, +p$ positrons forward in time. We see that the origin of the positron states is the pole at $p_0 = -E$, which was not present in nonrelativistic theory. To be really convinced that propagation in both spin states is included, we can recall the completeness relations (5.47).

The required boundary conditions were imposed by displacing the contour round the poles at $p_0 = \pm E$ as shown in Fig. 6.15. An equivalent prescription is

TABLE 6.2

Feynman Rules for $-i\mathcal{M}$

	Multiplicative Factor
● External Lines	
Spin 0 boson (or antiboson)	 1
Spin $\frac{1}{2}$ fermion (in, out)	 u, \bar{u}
antifermion (in, out)	 \bar{v}, v
Spin 1 photon (in, out)	 $\epsilon_\mu, \epsilon_\mu^*$
● Internal Lines—Propagators (need $+ie$ prescription)	
Spin 0 boson	 $\frac{i}{p^2 - m^2}$
Spin $\frac{1}{2}$ fermion	 $\frac{i(\not{p} + m)}{p^2 - m^2}$
Massive spin 1 boson	 $\frac{-i(g_{\mu\nu} - p_\mu p_\nu/M^2)}{p^2 - M^2}$
Massless spin 1 photon (Feynman gauge)	 $\frac{-ig_{\mu\nu}}{p^2}$
● Vertex Factors	
Photon—spin 0 (charge $-e$)	
Photon—spin $\frac{1}{2}$ (charge $-e$)	
	$ie(p + p')^\mu$
	$ie\gamma^\mu$

Loops: $\int d^4k/(2\pi)^4$ over loop momentum; include -1 if fermion loop and take the trace of associated γ -matrices

Identical Fermions: -1 between diagrams which differ only in $e^- \leftrightarrow e^-$ or initial $e^- \leftrightarrow$ final e^+

to displace the poles slightly off axis and to leave the contour undisturbed. To do this, we write the electron propagator

$$iS_F(p) = i \frac{\not{p} + m}{p^2 - m^2 + i\epsilon}. \quad (6.130)$$

The introduction of $+i\epsilon$, with ϵ infinitesimal and positive, has the effect of displacing the $p_0 = \pm E$ poles slightly below and above the axis, respectively. The same $+i\epsilon$ prescription is required for the other propagators.

An illuminating way to remember the sign of $i\epsilon$ is to regard it as a negative-imaginary contribution to the mass, $m \rightarrow m - i\epsilon/2$, so that the time dependence

$$e^{-iEt} \rightarrow e^{-i(m-i\epsilon/2)t} = e^{-imt}e^{-\epsilon t/2}.$$

Thus, stable particles may be viewed as the limit of unstable particles as the lifetime approaches infinity.

6.17 Summary of the Feynman Rules for QED

The invariant amplitude \mathcal{M} is obtained by drawing all (topologically distinct and connected) Feynman diagrams for the process and assigning multiplicative factors with the various elements of each diagram. The rules are summarized in Table 6.2.

For a photon-spin 0 interaction, there is also a four-particle vertex; see Fig. 6.16. This originates from the $e^2 A^2$ term in (4.4). No corresponding four-particle photon-spin $\frac{1}{2}$ vertex exists, since no A^2 term occurs in the Dirac equation, (6.2), describing an electron in an electromagnetic field.

EXERCISE 6.23 Use the Feynman rules to evaluate $\mathcal{M}(ye^- \rightarrow ye^-)$ corresponding to the two Feynman diagrams of Fig. 6.12 with the electron taken to have spin 0. Show that the result is not invariant under the gauge transformation (6.66). Demonstrate that gauge invariance is restored if diagram 6.16 is included with a vertex factor $2ie^2 g^{\mu\nu}$.

In these chapters, we have considered only the lowest-order Feynman diagrams. The rules generalize to higher-order graphs. However, new features occur. The diagrams contain closed loops of intermediate particles (see, for example, Fig. 6.7). Even after applying four-momentum conservation at each vertex, there still remains an undetermined four-momentum running round a closed loop. We

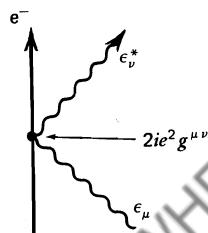


Fig. 6.16 The “sea-gull” diagram for $ye^- \rightarrow ye^-$, with “spinless” electrons.

thus need additional Feynman rules to evaluate such diagrams. First, we must integrate over the loop momentum, $\int d^4k/(2\pi)^4$. We must include a factor -1 for each closed fermion loop, and we have to take the trace of the associated γ -matrices. We discuss this in more detail in Chapter 7.

Unfortunately, loop integrations often lead to divergences. However, all the infinities which occur can be removed by well-established techniques. We say that QED is a renormalizable theory. This is the topic of the next chapter.

7

Loops, Renormalization, Running Coupling Constants, and All That

In this chapter, we attempt to give you a glimpse of the beautiful structure of field theory. Field theory is not the main subject of this book, so you can therefore safely skip this chapter; but a successful reading of it will expose you to such inaccessible concepts as loops, renormalization, and running coupling constants in a concise and physical way, we hope. A consequence is that the discussion is rather incomplete, and a few results are not explicitly derived. But only unrevealing algebra is omitted, which can be found in most field theory books.

7.1 Scattering Electrons Off a Static Charge

We use a simple experiment to demonstrate the concepts introduced in this chapter: the scattering of electrons by a static charge. In lowest order, the process is shown in Fig. 7.1a, in which the static charge is represented by a cross. How are the Feynman rules applied to this particular case? The question is best answered by going back to (6.4) and (6.6), where we find that the amplitude for the process of Fig. 7.1a can be written as

$$T_{fi} = -i \int d^4x j_\mu^{fi}(x) A^\mu(x). \quad (7.1)$$

Here, $j_\mu^{fi}(x)$ is the electron current:

$$j_\mu^{fi} = -e \bar{u}_f \gamma_\mu u_i e^{-iq \cdot x}, \quad (7.2)$$

and $A_\mu(x)$ is the four-vector potential associated with the static charge. As before, $q = p_i - p_f$ in terms of the momenta defined in Fig. 7.1a. Equation (7.1) can be written as

$$T_{fi} = ie \bar{u}_f \gamma_\mu u_i A^\mu(q), \quad (7.3)$$

where $A^\mu(q)$ is the Fourier transform:

$$A^\mu(q) = \int d^4x e^{-iq \cdot x} A^\mu(x). \quad (7.4)$$

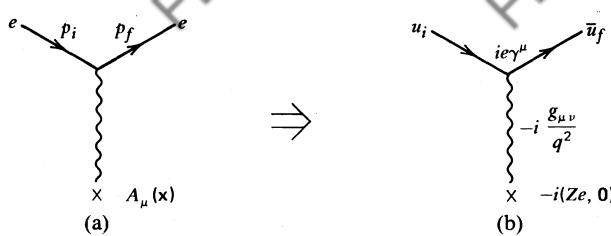


Fig. 7.1 Feynman rules for Rutherford scattering of electrons off a static charge Ze , for example, a nucleus.

For a static source, $A^\mu(x)$ is time independent; therefore

$$\begin{aligned} A^\mu(q) &= \int dt e^{-i(E_i - E_f)t} \int d^3x e^{i\mathbf{q} \cdot \mathbf{x}} A^\mu(\mathbf{x}) \\ &= 2\pi \delta(E_f - E_i) A^\mu(\mathbf{q}). \end{aligned} \quad (7.5)$$

The three-dimensional Fourier transform, $A^\mu(\mathbf{q})$, is best calculated using Maxwell's equations (6.59). For $A^\mu(x)$ independent of t , we have

$$\nabla^2 A^\mu(\mathbf{x}) = -j^\mu(\mathbf{x}), \quad (7.6)$$

and therefore

$$\int d^3x (\nabla^2 A^\mu(\mathbf{x})) e^{i\mathbf{q} \cdot \mathbf{x}} = -j^\mu(\mathbf{q}). \quad (7.7)$$

Now, by partial integration the left-hand side equals

$$\int d^3x A^\mu(\mathbf{x}) (\nabla^2 e^{i\mathbf{q} \cdot \mathbf{x}}) = -|\mathbf{q}|^2 A^\mu(\mathbf{q}), \quad (7.8)$$

and so, combining (7.7) and (7.8), we have

$$A^\mu(\mathbf{q}) = \frac{1}{|\mathbf{q}|^2} j^\mu(\mathbf{q}). \quad (7.9)$$

We substitute this result into (7.5) and find, from (7.3),

$$T_{fi} = i2\pi \delta(E_f - E_i) e\bar{u}_f \gamma_\mu u_i \frac{1}{|\mathbf{q}|^2} j^\mu(\mathbf{q}). \quad (7.10)$$

The covariant amplitude \mathcal{M} is then obtained by removing the δ -function of T_{fi} [see (4.17)]:

$$-i\mathcal{M} = ie\bar{u}_f \gamma_\mu u_i \frac{1}{|\mathbf{q}|^2} j^\mu(\mathbf{q}). \quad (7.11)$$

The electron recoils off the static charge in Fig. 7.1a, and $\mathbf{p}_i \neq \mathbf{p}_f$, but energy conservation in (7.10) implies $E_i = E_f$ or $q_0 = 0$. Therefore,

$$q^2 = -|\mathbf{q}|^2, \quad (7.12)$$

and so (7.11) can be written

$$-i\mathcal{M} = (ie\bar{u}_f \gamma^\mu u_i) \left(\frac{-ig_{\mu\nu}}{q^2} \right) (-ij^\nu(\mathbf{q})). \quad (7.13)$$

We recognize the familiar vertex factor and photon propagator of the Feynman rules for the amplitude ($-i\mathcal{M}$), see Section 6.17. We therefore deduce that the factor $-ij^\nu$ is associated with the source. For a static nucleus of charge Ze ,

$$\begin{aligned} j^0(\mathbf{x}) &= \rho(\mathbf{x}) = Ze \delta(\mathbf{x}) \\ \mathbf{j}(\mathbf{x}) &= 0, \end{aligned} \quad (7.14)$$

and so

$$-i\mathcal{M} = (ie\bar{u}_f \gamma_0 u_i) \left(\frac{-i}{q^2} \right) (-iZe). \quad (7.15)$$

The result is recorded in Fig. 7.1b. For a static nucleus, (7.15) just describes Rutherford scattering. We recognize the familiar result for the angular distribution [see (1.6)]:

$$\frac{d\sigma}{d\Omega} \sim |\mathcal{M}|^2 \sim \frac{1}{\sin^4(\theta/2)}, \quad (7.16)$$

where θ is the deflection angle of the electron, shown in Fig. 7.2. This angular distribution is a result of the q^{-4} behavior of $d\sigma/d\Omega$ obtained by inserting (7.15) into (7.16). Indeed,

$$\begin{aligned} q^2 &= (p_i - p_f)^2 \\ &\simeq -2k^2(1 - \cos\theta) \\ &\simeq -4k^2 \sin^2 \frac{\theta}{2}, \end{aligned} \quad (7.17)$$

where we have neglected the electron mass and used

$$k \equiv |\mathbf{p}_i| = |\mathbf{p}_f|.$$

7.2 Higher-Order Corrections

The previous calculation gives the Rutherford cross section to $O(\alpha^2)$ and is therefore an approximate perturbative result. To $O(\alpha^4)$, other Feynman diagrams have to be included, one of which is shown in Fig. 7.3. When the invariant

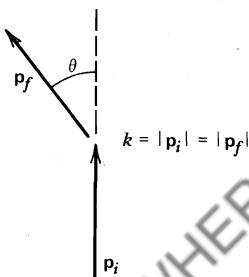


Fig. 7.2 Deflection of an electron by a static charge.

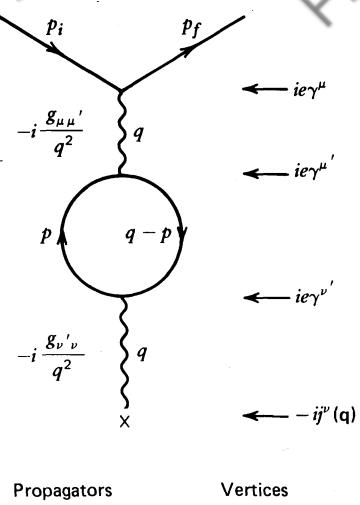


Fig. 7.3 Feynman diagram for Rutherford scattering in which the exchanged photon fluctuates into an $e^- e^+$ pair.

amplitude ($-i\mathcal{M}$) of the $O(\alpha^4)$ diagrams is added to (7.15), a more accurate result is obtained for $d\sigma/d\Omega$. In the specific diagram of Fig. 7.3, the exchanged photon spends some time as a virtual $e^- e^+$ pair; this will lead to a modification of Coulomb's law which results from the lowest-order diagram of Fig. 7.1. We first evaluate the higher-order diagram and then return to discuss this intriguing statement.

By applying the Feynman rules, shown in Fig. 7.3 and Section 6.17, we obtain

$$\begin{aligned}
 -i\mathcal{M} = & (-1)^1 \left(ie\bar{u}_f \gamma^\mu u_i \right) \left(-i \frac{g_{\mu\mu'}}{q^2} \right) \\
 & \times \int \frac{d^4 p}{(2\pi)^4} \left[(ie\gamma^{\mu'})_{\alpha\beta} \frac{i(\not{p} + m)_{\beta\lambda}}{p^2 - m^2} (ie\gamma^{\nu'})_{\lambda\tau} \frac{i(\not{q} - \not{p} + m)_{\tau\alpha}}{(q - p)^2 - m^2} \right] \\
 & \times \left(-i \frac{g_{\nu'\nu}}{q^2} \right) (-ij^\nu(\mathbf{q})). \tag{7.18}
 \end{aligned}$$

The Feynman rules for higher-order diagrams involve some nontrivial extensions of the rules developed in Chapter 6. A factor $(-1)^n$ should be included in an amplitude for a diagram containing n fermion loops [see, for example, the discussion of Mandl (1959) of the Dyson-Wick formalism], hence the factor $(-1)^1$ in (7.18). The only other unfamiliar feature of (7.18) is the $d^4 p/(2\pi)^4$ integration. Its origin is easily uncovered. Although four-momentum is conserved at each vertex, the momentum p , which is circulating around the loop, is unrestricted. The magnitude of the loop four-momentum

$$|p| = (\not{p}^2)^{1/2} = (p_0^2 - |\mathbf{p}|^2)^{1/2}$$

can be zero or infinite or have any value in between. As p is not observable, we have to sum over all possibilities, hence $\int d^4 p$.

The addition of (7.18) to (7.13) can be regarded as a modification to the propagator of the lowest-order result (7.13), namely,

$$\begin{aligned} -i \frac{g_{\mu\nu}}{q^2} &\rightarrow -i \frac{g_{\mu\nu}}{q^2} + \left(-i \frac{g_{\mu\mu'}}{q^2} \right) I^{\mu'\nu'} \left(-i \frac{g_{\nu'\nu}}{q^2} \right) \\ &\rightarrow -i \frac{g_{\mu\nu}}{q^2} + \frac{(-i)}{q^2} I_{\mu\nu} \frac{(-i)}{q^2} \end{aligned} \quad (7.19)$$

where

$$I_{\mu\nu}(q^2) = (-1)^1 \int \frac{d^4 p}{(2\pi)^4} \text{Tr} \left\{ (ie\gamma_\mu) \frac{i(\not{p} + m)}{p^2 - m^2} (ie\gamma_\nu) \frac{i(\not{q} - \not{p} + m)}{(q-p)^2 - m^2} \right\}. \quad (7.20)$$

This $O(\alpha)$ modification to the propagator is shown symbolically in Fig. 7.4. The correction can be calculated once and for all and then substituted into any Feynman diagram.

There is, however, a major problem. $I_{\mu\nu}$ as given by (7.20) apparently has terms of the form $\int |p|^3 d|p|/|p|^2$ for $|p| \rightarrow \infty$, and so the correction diverges. Indeed, a rather lengthy but straightforward calculation shows that $I_{\mu\nu}$ can be written as

$$I_{\mu\nu} = -ig_{\mu\nu} q^2 I(q^2) + \dots \quad (7.21)$$

with

$$I(q^2) = \frac{\alpha}{3\pi} \int_{m^2}^{\infty} \frac{dp^2}{p^2} - \frac{2\alpha}{\pi} \int dz z(1-z) \log \left(1 - \frac{q^2 z(1-z)}{m^2} \right), \quad (7.22)$$

where m is the mass of the electron. The dots in (7.21) represent omitted terms which are proportional to $q_\mu q_\nu$ and vanish when the propagator is coupled to external charges or currents. Equation (7.22) divides $I(q^2)$ into logarithmically divergent and finite contributions. We might have expected $I(q^2)$ to diverge quadratically as $\int |p| d|p|$. However, the divergence turns out to be only logarithmic on account of the “conspiratorial” algebra connected with the rest of the integrand. An explicit derivation of (7.21) and (7.22) is given, for example, in Bjorken and Drell (1964), Jauch and Rohrlich (1976), Scadron (1979), or Sakurai (1967) who, in an appendix, gives a collection of tricks for handling loop integrals.

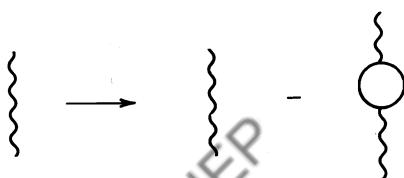


Fig. 7.4 Pictorial representation of (7.19).

Later, we shall study the effects of e^-e^+ loops in the limit of short- or long-range interactions, and so it is useful to evaluate $I(q^2)$ for both large and small values of $(-q^2)$. For $(-q^2)$ small,

$$\log\left(1 - \frac{q^2 z(1-z)}{m^2}\right) \approx -\frac{q^2 z(1-z)}{m^2},$$

and (7.22) becomes

$$I(q^2) \approx \frac{\alpha}{3\pi} \log \frac{M^2}{m^2} + \frac{\alpha}{15\pi} \frac{q^2}{m^2}, \quad (7.23)$$

where for the moment we have introduced a cut-off M^2 to replace ∞ as the upper limit of integration in the first term of (7.22). On the other hand, for $(-q^2)$ large,

$$\log\left(1 - \frac{q^2 z(1-z)}{m^2}\right) \approx \log\left(\frac{-q^2}{m^2}\right)$$

and so, similarly,

$$\begin{aligned} I(q^2) &\approx \frac{\alpha}{3\pi} \log\left(\frac{M^2}{m^2}\right) - \frac{\alpha}{3\pi} \log\left(\frac{-q^2}{m^2}\right) \\ &= \frac{\alpha}{3\pi} \log\left(\frac{M^2}{-q^2}\right). \end{aligned} \quad (7.24)$$

Unless we can dispose of the infinite part of $I(q^2)$ [which appears as $M^2 \rightarrow \infty$ in (7.23) and (7.24)], the result will not be physically meaningful.

The way to proceed is best explained by returning to Rutherford scattering. Including the loop contribution, (7.19), the amplitude (7.15) is

$$-i\mathcal{N} = (ie\bar{u}\gamma_0 u)\left(-\frac{i}{q^2}\right)\left(1 - \frac{\alpha}{3\pi} \log\left(\frac{M^2}{m^2}\right) - \frac{\alpha}{15\pi} \frac{q^2}{m^2} + O(e^4)\right)(-iZe), \quad (7.25)$$

where we have used (7.21) in the small $-q^2$ limit, (7.23), for $I(q^2)$. We may rewrite (7.25) in the form

$$-i\mathcal{N} = (ie_R\bar{u}\gamma_0 u)\left(-\frac{i}{q^2}\right)\left(1 - \frac{e_R^2}{60\pi^2} \frac{q^2}{m^2}\right)(-iZe_R), \quad (7.26)$$

with

$$e_R \equiv e \left(1 - \frac{e^2}{12\pi^2} \log \frac{M^2}{m^2}\right)^{1/2}.$$

(7.27)

To $O(e^4)$, it is easy to verify that (7.25) and (7.26) are mathematically equivalent. In the previous chapters, we were led to believe that e , the charge appearing in the lowest-order Feynman diagrams, is the charge of the electron as measured in

Thomson scattering or any other long-range Coulomb experiment. We never justified this, and it is, in fact, not true! Suppose that e_R in (7.27) is the electric charge listed in the particle data tables, that is, $e_R^2/4\pi = 1/137$. The invariant amplitude (7.26) is now finite. The infinity associated with the cut-off $M \rightarrow \infty$ has been “absorbed” in e_R . This procedure is admittedly bizarre. It is our first contact with “renormalization.” We return to this later, but first we explore the physics content of our new amplitude, which we have contrived to be free of infinities.

7.3 The Lamb Shift

As pointed out in Chapters 3 and 4, T_{fi} (or \mathcal{M}) represents the Fourier transform of the potential. The first term in (7.26), which is proportional to $|\mathbf{q}|^{-2}$, is associated with the Coulomb potential, since

$$V_0(r) = -\frac{Ze_R^2}{(2\pi)^3} \int d^3q e^{i\mathbf{q}\cdot\mathbf{r}} \frac{1}{|\mathbf{q}|^2} = -\frac{Ze_R^2}{4\pi r} \quad (7.28)$$

(see Exercise 6.17). The second term, which represents the quantum effect of the virtual e^-e^+ loop in the propagator of the exchanged photon, contains an extra factor $|\mathbf{q}|^2$ relative to the first. In coordinate space $|\mathbf{q}|^2 \rightarrow \nabla^2$ and, since

$$\frac{1}{(2\pi)^3} \int d^3q e^{i\mathbf{q}\cdot\mathbf{r}} = \delta(\mathbf{r}), \quad (7.29)$$

(7.26) corresponds to an interaction between the electron and the charge Ze_R of the form

$$V(r) = -\left(1 - \frac{e_R^2}{60\pi^2 m^2} \nabla^2\right) \frac{Ze_R^2}{4\pi r}$$

$$V(r) = -\frac{Ze_R^2}{4\pi r} - \frac{Ze_R^4}{60\pi^2 m^2} \delta(r).$$

(7.30)

The extra interaction (including its sign) was anticipated in the discussion of screening in Chapter 1. When $q^2 \rightarrow 0$, the electron probes the static charge Ze_R from a large distance and just interacts via the Coulomb interaction, that is, the first term in (7.30). The charge e_R is by definition the familiar electron charge, the one measured in any long-range electromagnetic interaction, for example, Thomson scattering (see Fig. 1.7). But when the electron comes closer to the nucleus (i.e., $-q^2$ increases), it penetrates the cloud of virtual e^-e^+ pairs which surround it. This leads to an increase in the effective interaction as explained in Fig. 1.6 (note, indeed, that both terms in (7.30) have the same sign), and the second term in (7.30) represents a calculation of this effect to leading order or to “the one-loop level.” The presence of the loop thus leads to an additional attractive force between the electron and the nucleus.

This additional interaction can be detected. Loops are not just some graphical construct. Their presence can be experimentally established. For example, if the source in Fig. 7.1 is a proton ($Z = 1$) and the Feynman graph represents the electron–proton interaction of the hydrogen atom, then (7.30) describes the atomic binding, including the additional attraction when the electron ventures inside the cloud of e^-e^+ pairs screening the charge of the nucleus. This effect, represented to lowest order by the $\delta(\mathbf{r})$ potential, contributes to the energy levels E_{nl} of the hydrogen atom. Its magnitude can be computed using ordinary quantum mechanics. Treating the second term in (7.30) as a perturbation, we obtain a contribution to the “Lamb shift”

$$\begin{aligned}\Delta E_{nl} &= -\frac{e_R^4}{60\pi^2 m^2} |\psi_{nl}(0)|^2 \delta_{l0} \\ &= -\frac{8\alpha_R^3}{15\pi n^3} Ry \delta_{l0}\end{aligned}\quad (7.31)$$

where the ψ_{nl} are the usual hydrogen atom wavefunctions and $Ry = m\alpha_R^2/2$ is the Rydberg constant. The δ_{l0} in (7.31) arises because the $\delta(\mathbf{r})$ potential can only perturb levels described by wavefunctions which are finite at the origin, namely, those with the $l = 0$. This result can be established experimentally by measuring the Lamb shift between the $2s_{1/2}$ and $2p_{1/2}$ levels. These levels are degenerate if loop contributions are not included. Equation (7.31), which for obvious reasons is called the “vacuum polarization” correction, contributes -27 MHz to the total Lamb shift of $+1057$ MHz between the $2s_{1/2}$ and $2p_{1/2}$ levels. As the Lamb shift can be measured to an accuracy of about 0.01%, the shift (7.31) due to the e^-e^+ loop has been verified experimentally. Indeed, this together with other loop contributions exactly reproduces the observed shift (see Section 7.4). We conclude that loop diagrams give real observable effects; but, even more important, our bizarre reinterpretation of the electron charge, (7.27), has received confirmation from experiment.

This eventually leads to the charges depending on q^2 (or, equivalently, on their separation) as in (7.24). For instance, the hydrogen atom is bound by the exchange of photons between the electron and the proton. The Coulomb force leads to a separation of a Bohr radius on the average. In QED, the electron can deviate from its Bohr orbit as a result of, among other things, the fluctuation of the exchanged photon into e^-e^+ pairs. This quantum screening effect reduces the attraction when the electron is far from the proton and increases the force when it approaches the nucleus. The competing effects do not cancel because the Coulomb force falls with r . The net result is an additional attraction over and above the Coulomb potential $-\alpha_R/r$, given by the second term in (7.30).

7.4 More Loops: The Anomalous Magnetic Moment

The vacuum polarization loop only accounts for a fraction of the splitting of the $2s_{1/2}-2p_{1/2}$ levels. Other $O(e^4)$ graphs exist which help to destroy the degener-

acy of the levels obtained when calculating to $O(e^2)$ only. The complete set of $O(e^4)$ graphs is shown in Fig. 7.5. We have calculated the contribution of diagram 7.5a. Each of the other diagrams also contains a loop, which, just as before, will diverge for large loop momentum. These divergences can all be hidden in a redefinition of the charge, mass, or wavefunction of the electron, in the same way that we absorbed the divergent part of the e^-e^+ loop occurring in the photon propagator into the charge e_R of (7.27). The “physics” is contained in the finite terms.

We consider the diagram of Fig. 7.5b next. Just as the loop in the propagator affects the attraction between the charges which it connects, so we anticipate that the loop around the vertex will modify the structure of the electron current $-e\bar{u}_f \gamma_\mu u_i$, see Fig. 7.6. Indeed, a computation of the (finite) piece of the diagram gives in the small ($-q^2$) limit

$$-e\bar{u}_f \gamma_\mu u_i \rightarrow -e\bar{u}_f \left\{ \gamma_\mu \left[1 + \frac{\alpha}{3\pi} \frac{q^2}{m^2} \left(\log \frac{m}{m_\gamma} - \frac{3}{8} \right) \right] - \left[\frac{\alpha}{2\pi} \frac{1}{2m} i\sigma_{\mu\nu} q^\nu \right] \right\} u_i. \quad (7.32)$$

The first square bracket gives an additional contribution to the Lamb shift of similar form to (7.26). A new feature arises here since the loop also diverges for small (infrared) loop momenta $|p|$. In (7.32), we have sidestepped this problem by giving the photon a small fictitious mass m_γ . We shall explain how this takes care of infrared divergences in Chapter 11. The combined effects of (7.26) and (7.32)

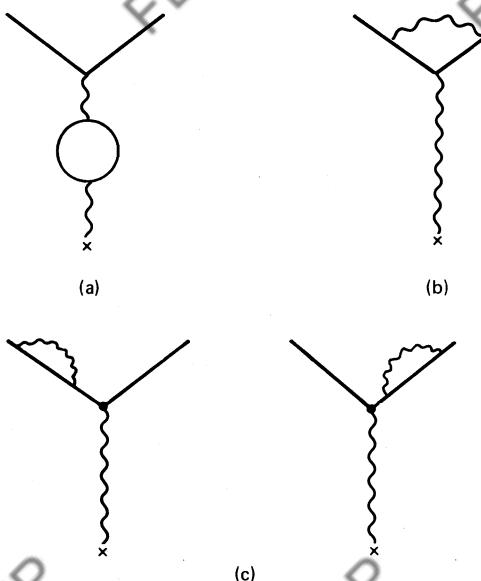


Fig. 7.5 Complete set of $O(\alpha^2)$ Feynman graphs.

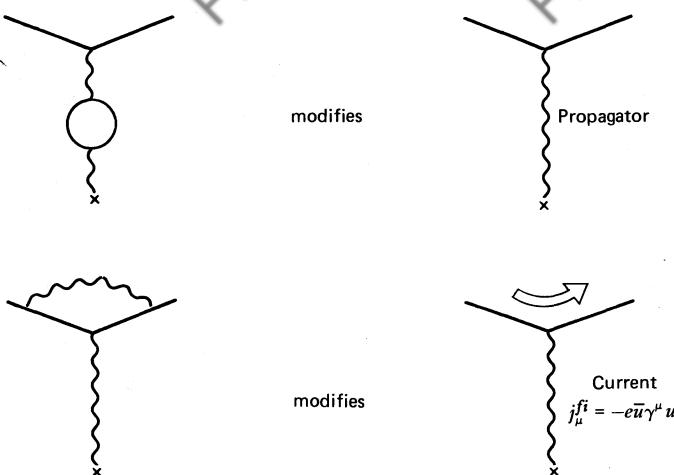


Fig. 7.6

account for the observed value of the Lamb shift. The detailed discussion is rather lengthy and requires a careful treatment of the infrared photons, see Bjorken and Drell (1964) or Jauch and Rohrlich (1976).

An intriguing feature of (7.32) is the second term in square brackets, which modifies the γ_μ Lorentz structure of the electron current. To see the physical implications of this term, we recall the Gordon decomposition of a γ_μ -current, (6.7),

$$-e\bar{u}_f \gamma_\mu u_i = -\frac{e}{2m} \bar{u}_f \left((p_f + p_i)_\mu - i\sigma_{\mu\nu} q^\nu \right) u_i. \quad (7.33)$$

Equation (7.33) exhibits the fact that the electron interacts via both its charge and its magnetic moment. In exercise 6.2, we demonstrated that the $\sigma_{\mu\nu} q^\nu$ term in (7.33) represents a magnetic moment of the electron,

$$\mu = -\frac{e}{2m} \sigma, \quad (7.34)$$

which is often written as

$$\mu = -g \frac{e}{2m} S \quad (7.35)$$

with $S = \frac{1}{2}\sigma$ and the gyromagnetic ratio

$$g = 2. \quad (7.36)$$

The second term in (7.32) is therefore just an extra magnetic moment interaction to that already contained in γ_μ via (7.33). In fact, substituting (7.33) in (7.32), we find, using (7.34), that

$$\mu = -\frac{e}{2m} \left(1 + \frac{\alpha}{2\pi} \right) \sigma, \quad (7.37)$$

or

$$g = 2 + \frac{\alpha}{\pi}. \quad (7.38)$$

The electron thus has an “anomalous” magnetic moment $\alpha/2\pi$ in addition to its Dirac magnetic moment. To be precise, the anomalous part is given by

$$\begin{aligned} \frac{g-2}{2} &= \frac{1}{2}\left(\frac{\alpha}{\pi}\right) - 0.32848\left(\frac{\alpha}{\pi}\right)^2 + (1.49 \pm 0.2)\left(\frac{\alpha}{\pi}\right)^3 + \dots \\ &= (1159655.4 \pm 3.3) \times 10^{-9}, \end{aligned} \quad (7.39)$$

where the first term corresponds to our lowest-order result (7.38), while the second and third terms represent the higher-order contributions. The number of diagrams grows rapidly with the order of α , and the error on the $O(\alpha^3)$ contributions hints at the difficult numerical calculations that are involved. The experimental value of the electron’s anomalous magnetic moment is

$$\left(\frac{g-2}{2}\right)_{\text{exp}} = (1159657.7 \pm 3.5) \times 10^{-9}, \quad (7.40)$$

in excellent agreement with the prediction (7.39). This triumph of QED has been repeated for the muon magnetic moment, providing further evidence that our strange way of handling the infinities is correct.

7.5 Putting the Loops Together: Ward Identities

Equation (7.27) shows how the infinite part of the loop in the photon propagator is hidden by a redefinition of the electron’s charge. When performing the complete $O(e^4)$ calculation, infinite parts of the loops in Fig. 7.5b and 7.5c will also be absorbed into e_R . Suppose we now repeat this calculation for the scattering of a muon, instead of an electron, by a nucleus. Clearly, the first diagram (Fig. 7.5a) contributes an amount to the charge which is independent of the nature of the scattered particle. The result (7.27) is determined by the modified photon propagator and changes the charge of an electron, muon, or any other particle in exactly the same way. But this is not the case for the diagrams of Figs. 7.5b and 7.5c, where the scattered particles are an integral part of the loop. It would appear that we shall get different redefinitions of e for electrons and muons, which would be a serious problem since experimentally the electron and muon charges are equal. It is here that QED displays its full power. A full calculation shows that magically the modification of the charge by the vertex diagram of Fig. 7.5b is exactly canceled by the modification introduced by the diagrams of Fig. 7.5c [see Sakurai (1967), Bjorken and Drell (1964)]. Only the “vacuum polarization” graph of Fig. 7.5a modifies the charge. Equation (7.27) is the full answer, and hence the “renormalized” charge of the electron and muon remain equal.

This cancellation repeats itself in every order of perturbation theory. The electron and muon charges are exactly equal. The conspiracy between the diagrams of Figs. 7.5b and 7.5c reflects a very basic property of (gauge) field theories known as a Ward identity.

7.6 Charge Screening and $e^- \mu^-$ Scattering

Loops in the propagator of the exchanged photon not only modify the interaction of an electron with a static charge but also affect other interactions, for example, electron-muon elastic scattering. The lowest-order $e^- \mu^-$ amplitude is given by (6.50). The $O(e^4)$ vacuum-polarization contribution is readily obtained by replacing the source factor $-ij^\nu = (-iZe, \mathbf{0})$ of (7.18) by the muon current:

$$-ij^\nu = ie \bar{u}(p'_f) \gamma^\nu u(p'_i),$$

see Fig. 7.7. We require the charge to be the renormalized charge, (7.27), and keep the finite piece of $I(q^2)$ of (7.21). We add this higher-order contribution to the lowest-order result. This is an illustration of the fact that propagator corrections are common to all processes and hence can be calculated once and for all. Of course, there are other $O(e^4)$ contributions to $e^- \mu^-$ scattering which we must also include.

7.7 Renormalization

Despite its phenomenological success, the procedure for treating infinities deserves further consideration. How can we justify perturbation theory in α , when in the next order α is accompanied by an infinite coefficient, namely, $\log(M^2/m^2)$, where M is some arbitrary cutoff? We return to (7.27), which drew attention to a very serious shortcoming of our discussion of relativistic quantum mechanics: the quantity that we called the charge, which appears in the lowest-order Feynman amplitudes of Chapter 6, is changed by higher-order interactions. It is therefore

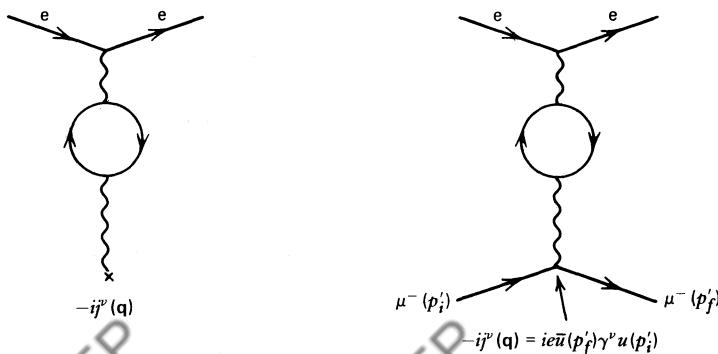
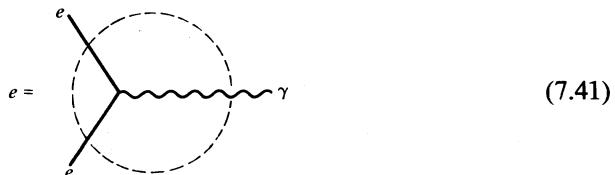
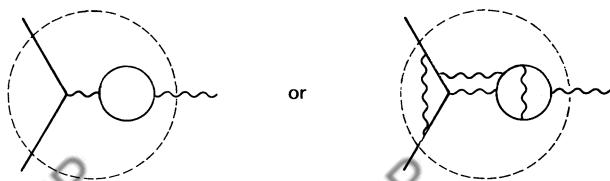


Fig. 7.7 Relation between Rutherford and $e^- \mu^-$ scattering.

not what we thought it was, and it is certainly not the charge the experimentalist measures. We can see this dilemma another way. The charge is associated with the electron–photon coupling, which we symbolically represented as



But this is absurd, because the charge is also



In fact, it is all of these things at once, and that is what the experimentalist measures. Therefore, to label (7.41) as e (the quantity measured in Coulomb's experiment giving $\alpha = 1/137$) is simply wrong. Let us therefore call (7.41) the “bare” charge e_0 . Here, “bare” refers to the fact that the vertex is stripped of all loops. We can now summarize the situation by writing

$$\begin{array}{c} \text{---} \\ | \\ \text{---} \end{array} e = \left[\begin{array}{c} \text{---} \\ | \\ \text{---} \\ e_0 \\ | \\ \text{---} \\ e_0 \\ | \\ \text{---} \end{array} - \begin{array}{c} \text{---} \\ | \\ \text{---} \\ e_0 \\ | \\ \text{---} \\ e_0 \\ | \\ \text{---} \end{array} + \begin{array}{c} \text{---} \\ | \\ \text{---} \\ \text{---} \\ | \\ \text{---} \\ \text{---} \end{array} + \dots \right] \quad (7.42)$$

$\text{at } Q^2 = \mu^2$

where \dots stands for diagrams with all possible propagator modifications. We need only consider modifications to the photon propagator because of the Ward identity of Section 7.5. We also explicitly show the negative sign associated with each loop, see (7.18). The charge e in (7.42) is the charge the experimentalist measures when scattering two low-energy electrons or performing a Coulomb experiment, namely, $e^2/4\pi = 1/137$. Its definition recognizes the fact that e_0 , appearing in the lowest-order Feynman amplitude, is modified by interactions. The relation between e^2 and e_0^2 has to be specified at the particular value of the virtual photon's momentum, say, $q^2 \equiv -Q^2 = -\mu^2$ appropriate to the experiment, as is done in (7.42). It is conventional to introduce Q^2 for $-q^2$, as this quantity is positive.

To $O(e_0^4)$, we can write the relation between e and the bare charge e_0 as

$$e^2 = e_0^2 [1 - I(q^2 = -\mu^2) + O(e_0^4)], \quad (7.43)$$

where $I(q^2)$ is given by (7.19) and (7.21). $I(q^2)$ is $O(e_0^2)$ and represents the result of the one-loop calculation. Indeed, taking the square root of (7.43), we have

$$e = e_0 [1 - \frac{1}{2}I(q^2 = -\mu^2) + O(e_0^4)], \quad (7.44)$$

which is of the form of (7.27) after expansion of the square root. In the diagrammatic notation of Fig. 7.4, this can be written as

$$\begin{array}{c} \text{Diagram A: } e_0 \\ \text{Diagram B: } e \\ \text{Diagram C: } \left[1 + \frac{1}{2} \text{Diagram D} + O(e^4) \right] \\ \text{at } Q^2 = \mu^2 \end{array} \quad (7.44')$$

or, to all orders,

$$e = e_0 (1 + e_0^2 A_1(Q^2) + e_0^4 A_2(Q^2) + \dots)_{\text{at } Q^2 = \mu^2}, \quad (7.45)$$

where $-q^2 \equiv Q^2$. Clearly, $A_1(Q^2)$, which is directly related to $I(-Q^2)$, is an infinite quantity; so are $A_2(Q^2)$ and all subsequent coefficients in (7.45). There is *a priori* nothing wrong with that. It does not matter that a theory is formulated in terms of infinite quantities as long as observable quantities are finite. Extensive use is made of complex quantities in optics, and there is no objection to that as long as the observables are real.

Let us calculate an observable to illustrate this point, for example, $e\mu$ scattering at 90° (see Section 7.6). We fix the angle in order to have an observable $d\sigma/d\Omega(s, t)$ which depends on only one momentum. At 90° , $-t \simeq s/2 \simeq Q^2$. We calculate the invariant amplitude as before, but we now explicitly display the fact that the calculation is in terms of the bare charge e_0 :

$$\begin{aligned} -i \mathfrak{M}(e_0^2) &= \left[\text{Diagram E} - \text{Diagram F} + \dots \right] \\ &\quad \text{at } Q^2 \end{aligned} \quad (7.46)$$

$$= e_0^2 [F_1(Q^2) + e_0^2 F_2(Q^2) + O(e_0^4)]. \quad (7.46')$$

Other diagrams are represented by ... in (7.46). To obtain realistic results, they would have to be considered explicitly. Here we just want to demonstrate the techniques by which perturbation theory can manipulate infinite diagrams, for example, the infinity associated with the loop in the e_0^4 term in (7.46). Indeed, all terms in a perturbative calculation in terms of the bare charge e_0 , like (7.46), are again infinite. Now comes the crucial step: we reparametrize ("renormalize" in

the usual, but unfortunate, terminology) $-i\mathcal{M}(e_0^2)$ in terms of e^2 . To do this, we invert (7.44') [or (7.44)]:

$$\text{Diagram with } e_0 \text{ vertex} = \text{Diagram with } e \text{ vertex} \left[1 + \frac{1}{2} \text{Loop} + O(e^4) \right] \quad (7.47)$$

at $Q^2 = \mu^2$

and use this result to replace the e_0 vertices of (7.46). We obtain

$$\begin{aligned} -i\mathcal{M}(e^2) &= \text{Diagram with } e \text{ vertex at } Q^2 + 2 \left[\text{Diagram with } \frac{1}{2} \text{Loop at } Q^2 = \mu^2 \right] - \text{Diagram with } e \text{ vertex at } Q^2 \\ &\quad + O(e^6) \end{aligned} \quad (7.48)$$

The first two diagrams both come from the first diagram of (7.46); the factor 2 arises because we must replace e_0 by e at each vertex. In the remaining diagram, we can simply write e instead of e_0 , as the difference is $O(e^6)$. Equation (7.48) can be written as

$$\begin{aligned} -i\mathcal{M}(e^2) &= \text{Diagram with } e \text{ vertex at } Q^2 - \left\{ \text{Diagram with } \frac{1}{2} \text{Loop at } Q^2 - \text{Diagram with } e \text{ vertex at } Q^2 \right\} + O(e^6) \\ &\quad (7.49) \end{aligned}$$

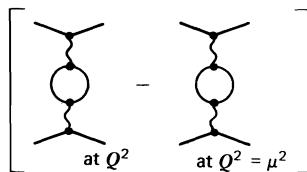
$$= e^2 [F'_1(Q^2) + e^2 F'_2(Q^2) + O(e^4)] \quad (7.49')$$

We have achieved the desired result. Comparing (7.46) and (7.49), we see that we have obtained a new expression for the invariant amplitude in terms of the “experimentalists” charge e as defined by (7.44), that is, as measured in an experiment with $Q^2 = \mu^2$. In doing so, nothing has been added or thrown away; we have just reparametrized the original calculation (7.46). Therefore, clearly

$$\mathcal{M}(e^2) = \mathcal{M}(e_0^2), \quad (7.50)$$

as indeed it must. So, what have we achieved? The e_0^4 term in (7.46) is infinite, the e^4 term in (7.49) is finite! The e^4 term has been split into two terms, one containing a loop at Q^2 and the other a loop at $Q^2 = \mu^2$. The signs of the two terms are opposite. To see in more detail what happens, take, for example, the

result for the loop given by (7.24):



$$\sim \left(\frac{\alpha}{3\pi} \log\left(\frac{M^2}{Q^2}\right) - \frac{\alpha}{3\pi} \log\left(\frac{M^2}{\mu^2}\right) \right) = \frac{\alpha}{3\pi} \log\left(\frac{\mu^2}{Q^2}\right). \quad (7.51)$$

The difference of the two terms is finite; it does not depend on the *ad hoc* cutoff M^2 , which we can now send back off to infinity, where it belongs. We conclude that (7.49), unlike (7.46), defines the observables in terms of finite quantities. The two perturbation expansions are nevertheless equivalent, as we have demonstrated by explicit calculation. The infinite coefficients in the original series, (7.46), arose because e_0 itself is not finite (it is in fact infinitesimal). Once we reorganize the series in terms of the finite quantity e^2 , all the coefficients are finite.

Note that a free parameter μ with the dimensions of mass has slipped into the theory via the reparametrization of the charge. Different choices of μ^2 , the renormalization mass, will lead to different expansions, (7.49'), of the amplitude. We say we are using different renormalization schemes. But $|\mathcal{M}|^2$ is an observable and so must be independent of the value chosen for μ . This requirement can be formulated as follows:

$$\mu \frac{d\mathcal{M}}{d\mu} = \left(\mu \frac{\partial}{\partial \mu} \Big|_e + \mu \frac{\partial e}{\partial \mu} \frac{\partial}{\partial e} \right) \mathcal{M} = 0. \quad (7.52)$$

The dependence of \mathcal{M} on μ , given by the coefficients $F'(Q^2, \mu^2)$ in (7.49'), must be cancelled by the μ -dependence of $e(\mu^2)$. Equation (7.52) is called the “renormalization group equation.” Its importance transcends particle physics; we have done scant justice to it in this passing reference.

7.8 Charge Screening in QED: The Running Coupling Constant

We have seen repeatedly how the charge is modified by the vacuum polarization loop in the photon propagator. We know that the loop will be repeated in higher orders as shown in (7.42). We can rewrite this relation as

$$(7.53)$$

and the geometric series can be summed to give

$$\begin{array}{c} e \\ \diagdown \quad \diagup \\ \text{---} \end{array} = \begin{array}{c} e_0 \\ \diagdown \quad \diagup \\ \text{---} \end{array} \left\{ \frac{1}{1 + \text{---}} \right\}. \quad (7.54)$$

It turns out to be a good idea to redefine the charge including all vacuum polarization loops as given by (7.54).

We showed how the infinities can be removed by working in terms of the physical (renormalized) charge e given by (7.53) at $Q^2 = \mu^2$. In fact, we could have used any value of μ^2 . However, different choices $Q^2 = \mu_1^2, \mu_2^2, \dots$ correspond to perturbation expansions in terms of numerically different values of the physical charge $e(\mu_i^2)$. Indeed, using the notation of (7.43), we have, from (7.54),

$$e^2(Q^2) = e_0^2 \left(\frac{1}{1 + I(q^2)} \right). \quad (7.55)$$

Equation (7.55) explicitly displays the fact that the charge the experimentalist measures depends on the Q^2 of the experiment; $\alpha(Q^2) \equiv e^2(Q^2)/4\pi$ is referred to as the “running coupling constant.”

In the large $Q^2 \equiv -q^2$ limit, $I(q^2)$ is given by (7.24), and (7.55) becomes

$$\alpha(Q^2) = \frac{\alpha_0}{1 - \frac{\alpha_0}{3\pi} \log\left(\frac{Q^2}{M^2}\right)}. \quad (7.56)$$

To eliminate the explicit dependence of $\alpha(Q^2)$ on the cutoff M , we choose a renormalization or reference momentum μ . The renormalization procedure is then to subtract $\alpha(\mu^2)$ from $\alpha(Q^2)$. We find

$$\alpha(Q^2) = \frac{\alpha(\mu^2)}{1 - \frac{\alpha(\mu^2)}{3\pi} \log\left(\frac{Q^2}{\mu^2}\right)} \quad (7.57)$$

for large Q^2 . Equation (7.57) now contains only finite, physically measurable quantities.

The running coupling constant, $\alpha(Q^2)$, describes how the effective charge depends on the separation of the two charged particles. By summing part of all orders of perturbation theory, we have obtained the charge screening of electrodynamics, see Fig. 1.5. As Q^2 increases, the photon sees more and more charge until, at some astronomically large but finite Q^2 , the coupling $\alpha(Q^2)$ is infinite. However, by inserting numerical values, we find that for all practically attainable Q^2 , the variation of α with Q^2 is extremely small; α increases from 1/137 very slowly as Q^2 increases. Of course, as Q^2 increases, other loops (formed, for example, by a $\mu^+ \mu^-$ -pair or a $\bar{u}u$ -quark pair) will also contribute to the variation.