

Figure 16.1. A typical barrier penetration problem. The particle has energy $E > 0$, but is restrained by the barrier since $V_{\max} > E$. But there is an amplitude for it to tunnel through the barrier and escape to infinity as a free particle.

one with negative t corresponding to the extrapolation of the given initial conditions to earlier times. Verify that $U(x, x', z) + U^*(x, x', z)$ is indeed the transform of $U(t)$ for all times by seeing what happens to $\langle x | e^{-(i/\hbar)Ht} | x' \rangle$ under complex conjugation and the exchange $x \leftrightarrow x'$. Likewise, ask what happens to U_{cl} when we transform it for all times. Now you will find that no matter what the sign of $x - x'$ a single right moving trajectory can contribute—it is just that the time of the stationary point, t^* , will change sign with the sign of $x - x'$. (The same goes for a left moving trajectory.)

Tunneling Amplitudes

The WKB formula can also be used to calculate tunneling amplitudes provided $\chi(x) = \{2m[V(x) - E]\}^{1/2}$ varies slowly. As an example, consider a particle trapped in a potential shown in Fig. 16.1. If its energy is positive, there exists some probability for it to penetrate the barrier and escape to infinity as a free particle. In the first approximation, the ratio of ψ at the point of escape, x_e and at the outer wall of the well, x_0 , is

$$\psi(x_e) = \psi(x_0) \exp\left(\frac{i}{\hbar} \int_{x_0}^{x_e} i\{2m[V(x) - E]\}^{1/2} dx\right) \quad (16.2.24)$$

$$\equiv \psi(x_0) e^{-\gamma/2} \quad (16.2.25)$$

The mean lifetime of the particle inside the well may be estimated by the following semiclassical computation. Since the particle inside the well has a kinetic energy $T = E - V = E + V_0$, its velocity is $v = [2m(E + V_0)]^{1/2}/m$ and it bangs against the outer wall at a frequency $f = v/2x_0$. Upon collision, there is a probability of escape $e^{-\gamma}$. Consequently probability of escape in 1 second is

$$R = \frac{[2m(E + V_0)]^{1/2}}{2mx_0} e^{-\gamma} \quad (16.2.26)$$

The mean lifetime is then $\tau = 1/R$.

Note that just as a particle inside can tunnel out a particle with $E < V_{\max}$ can tunnel in from outside and get captured. A standard example of tunneling and capture is provided by α decay in which a nucleus emits an α particle.[‡] At short distances, the force between the α particle and the nucleus is attractive (the nuclear force beats the Coulomb repulsion), while at larger distances, the Coulomb repulsion

[‡] The α particle has two protons and two neutrons in it.

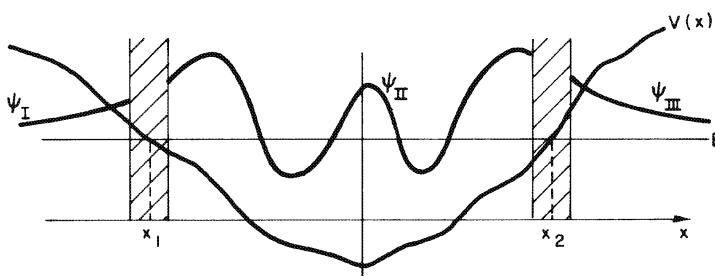


Figure 16.2. A typical bound state problem. The WKB approximation to the wave function works nicely except in the shaded bands near the classical turning points x_1 and x_2 . To join ψ_{II} to ψ_I and ψ_{III} , one solves the Schrödinger equation in these bands after approximating the potential by a linear function within each band.

dominates. The sum of the two potentials is roughly described by a potential of the type shown in Fig. 16.1, with x playing the role of the radial coordinate. (The centrifugal barrier must be added onto the two potentials if the α particle comes out with nonzero orbital angular momentum.) Thus α particles with $E < V_{\max}$ can tunnel out from within, or get captured from without.

Exercise 16.2.4. Alpha particles of kinetic energy 4.2 MeV come tunneling out of a nucleus of charge $Z = 90$ (after emission). Assume that $V_0 = 0$, $x_0 = 10^{-12}$ cm, and $V(x)$ is just Coulombic for $x \geq x_0$. (See Fig. 16.1.) Estimate the mean lifetime. {Hint: Show $\gamma = (8Ze^2/\hbar v)[\cos^{-1} y^{1/2} - y^{1/2}(1-y)^{1/2}]$, where $y = x_0/x_e$. Show that $y \ll 1$ and use $\cos^{-1} y^{1/2} \approx \frac{1}{2}\pi - y^{1/2}$ before calculating numbers.}

The derivation of the tunneling amplitude, Eq. (16.2.24), is not straightforward in the path integral formalism due to the fact that there exists no classical path that can take the particle across the barrier. There is, however, such a path in the “imaginary time” formalism. This will be detailed in Chapter 21.

Bound States

The WKB method can be applied to approximate bound state energies and wave functions. Consider a particle (Fig. 16.2) bound by a potential $V(x)$. In the figure, x_1 and x_2 are the classical turning points for the energy E . Let us see how the quantization of E emerges in this formalism. We know that in the classically forbidden regions I and III, the wave function will be a damped exponential. For instance

$$\psi_{III}(x) \sim \frac{1}{\{2m[V(x)-E]\}^{1/2}} \exp\left(-\frac{1}{\hbar} \int^x \{2m[V(x')-E]\}^{1/2} dx'\right) \quad (16.2.27)$$

In the classically allowed region II, it will be oscillating. We assume it is a real function[‡] with two free parameters A and B :

$$\psi_{\text{II}}(x) = \frac{A}{[p(x)]^{1/2}} \cos \left[\frac{1}{\hbar} \int_{x_1}^x p(x') dx' + B \right] \quad (16.2.28)$$

[The two real parameters A and B replace the one complex parameter $\psi(x_0)$ used previously.] Unfortunately, neither Eq. (16.2.27) from Eq. (16.2.28) is applicable near the turning points. Formally this is because $[p(x)]^{-1/2}$ and $\{2m[V(z)-E]\}^{-1/2}$ blow up there. Physically it is because the wavelength tends to infinity there and the requirement that $V(x)$ varies little over a wavelength becomes impossible to meet. It is therefore impossible to match the approximate ψ_1 , ψ_{II} , and ψ_{III} and thereby determine the allowed energies as one does in the simple examples with piecewise constant potentials. The problem is surmounted as follows. Near each turning point, we define a *transition region* (shaded bands in the figure) inside which we solve the problem by going back to Schrödinger's equation. If $V(x)$ is slowly varying, it may be approximated by a linear function in these regions. For instance near x_1

$$\begin{aligned} V(x) &\simeq V(x_1) + V' \cdot (x - x_1) \\ &= E + V' \cdot (x - x_1) \end{aligned} \quad (16.2.29)$$

The exact solutions with this $V(x)$ are then matched to the WKB solutions outside the shaded region, that is to say, to the damped exponentials on one side and the oscillating cosine on the other.

The analysis[§] near x_1 would yield the following function in region II:

$$\psi_{\text{II}}(x) = \frac{A}{[p(x)]^{1/2}} \cos \left[\frac{1}{\hbar} \int_{x_1}^x p(x') dx' - \frac{\pi}{4} \right] \quad (16.2.30)$$

while the one near x_2 would yield

$$\psi_{\text{II}}(x) = \frac{A'}{[p(x)]^{1/2}} \cos \left[\frac{1}{\hbar} \int_{x_2}^x p(x') dx' + \frac{\pi}{4} \right] \quad (16.2.31)$$

For the two solutions to coincide, A and A' must have the same magnitude and the difference in phase between the two cosines must be a multiple of π :

$$\frac{1}{\hbar} \int_{x_1}^x p(x') dx' - \frac{1}{\hbar} \int_{x_2}^x p(x') dx' - \frac{\pi}{2} = n\pi, \quad n = 0, 1, 2, \dots$$

[‡] Recall Theorem 16, Sec. 5.6.

[§] The details are omitted.

or

$$\int_{x_1}^{x_2} p(x) dx = (n + \frac{1}{2})\pi\hbar \quad (16.2.32)$$

or

$$\oint p(x) dx = (n + \frac{1}{2})2\pi\hbar \quad (16.2.33)$$

where \oint denotes the integral over a full cycle, from x_1 to x_2 and back. If n is even, $A = A'$, if odd $A = -A'$.

Equation (16.2.32) expresses the quantization of energy since the integral and limits on the left-hand side are functions of energy and the other parameters such as particle mass.

As an example, consider a particle in a linear potential $V(x) = k|x|$. The turning points are

$$x_{1,2} = \mp E/k \quad (16.2.34)$$

and the quantization condition is

$$\int_{-E/k}^{E/k} [2m(E - k|x|)]^{1/2} dx = 2 \int_0^{E/k} [2m(E - kx)]^{1/2} dx = (n + \frac{1}{2})\hbar\pi \quad (16.2.35)$$

The n, k, m, \hbar dependence of E can be found by a scaling argument. Let us define a variable y by

$$x = (E/k)y \quad (16.2.36)$$

in terms of which we have

$$2 \int_0^1 (2mE)^{1/2}(1-y)^{1/2} \left(\frac{E}{k}\right) dy = (n + \frac{1}{2})\hbar\pi$$

or

$$E \propto (k)^{2/3}(m)^{-1/3}(n + \frac{1}{2})^{2/3}\hbar^{2/3} \quad (16.2.37)$$

The constant of proportionality may be found by carrying out the y integral. The result is

$$E_n = \left[\frac{3k\hbar\pi}{4(2m)^{1/2}} \left(n + \frac{1}{2}\right) \right]^{2/3} \quad (16.2.38)$$

If the method is applied to the potential $V(x) = \lambda x^4$, we would get, by the scaling argument,

$$E_n = \left[\frac{c\lambda^{1/4}\hbar}{m^{1/2}} \left(n + \frac{1}{2} \right) \right]^{4/3} \quad (16.2.39)$$

where c is a constant that may be found by carrying out a dimensionless integral. If the WKB energy levels are compared to the actual ones (obtained either by analytic solution or by accurate numerical integration of Schrödinger's equation) we would find that the agreement is excellent for all but very small n . In the λx^4 case, for example,

$$\frac{E_0(\text{WKB})}{E_0(\text{numerical})} = 0.716$$

$$\frac{E_1(\text{WKB})}{E_1(\text{numerical})} = 0.992$$

The agreement gets even better as we move up in n . Thus the WKB method complements the variational method, which works best for the lowest levels. The improved accuracy with increasing n is reasonable in view of the fact that as we move up in energy, the transitional region near the turning points (where the approximation breaks down) plays a decreasingly important role.[‡]

What about the WKB wave functions? They too get better at large n , except of course near the turning points, where they blow up due to the $[p(x)]^{-1/2}$ factor. If, however, we actually solve the Schrödinger equation near the turning points after approximating the potential by a linear function, the blowup can be avoided and the agreement with the true eigenfunctions greatly enhanced.

The WKB wave function has another feature that agrees with the exact answer: the wave function has n nodes ($n = 0, 1, 2, \dots$) in the n th level. We see this analytically from Eq. (16.2.30) for $\psi_{ll}(x)$, and Eq. (16.2.32) for the phase integral $(1/\hbar) \int_{x_1}^x p(x') dx'$. As x goes from x_1 to x_2 , the phase ϕ goes from $-\pi/4$ to $n\pi + \pi/4$ and $\cos(\phi)$ vanishes n times. We can in fact understand the quantization rule, Eq. (16.2.32), as follows. If we assume, in the first approximation, that ψ must vanish in the classically forbidden region, it follows that it must undergo an integral number of half-cycles (half-wavelengths) in the interval $x_1 \leq x \leq x_2$. This leads to the *Bohr-Sommerfeld quantization rule*.

$$\int_{x_1}^{x_2} p(x) dx = (n + 1)\hbar\pi, \quad n = 0, 1, 2, \dots \quad (16.2.40)$$

But we know that ψ doesn't vanish at the turning points and has an exponential tail in the classically forbidden region. Consequently the number of half-cycles completed between x_1 and x_2 is somewhat less than $n + 1$. The connection procedure tells us that

[‡] There are some exceptional cases, such as the harmonic oscillator, where the method gives the exact energies for all n .

it is in fact $n + \frac{1}{2}$ and hence the usual quantization rule, Eq. (16.2.32). If, however, ψ actually vanishes at x_1 and x_2 because the potential barrier there is infinite (as in the case of a particle in a box), Eq. (16.2.40) [and not Eq. (16.3.32)] is relevant.[‡] One can also consider an intermediate case where the barrier is infinite at one turning point and not at the other. In this case the quantization rule has an $(n + 3/4)$ factor in it.

The WKB method may also be applied in three dimensions to solve the radial equation in a rotationally invariant problem. In the $l=0$ state, there is no centrifugal barrier, and the WKB wave function has the form

$$U(r) \sim \frac{1}{[p(r)]^{1/2}} \sin \left[\frac{1}{\hbar} \int_0^r p(r') dr' \right], \quad p = \{2m[E - V(r)]\}^{1/2} \quad (16.2.41)$$

where the lower limit in the phase integral is chosen to be 0, so that $U(0)=0$. The quantization condition, bearing in mind that the barrier at $r=0$ is infinite, is

$$\int_0^{r_{\max}} p(r) dr = \left(n + \frac{3}{4} \right) \hbar \pi, \quad n = 0, 1, 2, \dots \quad (16.2.42)$$

where r_{\max} is the turning point. This formula is valid only if $V(r)$ is regular at the origin. If it blows up there, the constant we add to n is not $3/4$ but something else. Also if $l \neq 0$, the centrifugal barrier will alter the behavior near $r=0$ and change both the wave function and this constant.

*Exercise 16.2.5.** In 1974 two new particles called the ψ and ψ' were discovered, with rest energies 3.1 and 3.7 GeV, respectively ($1 \text{ GeV} = 10^9 \text{ eV}$). These are believed to be nonrelativistic bound states of a “charmed” quark of mass $m = 1.5 \text{ GeV}/c^2$ (i.e., $mc^2 = 1.5 \text{ GeV}$) and an antiquark of the same mass, in a linear potential $V(r) = V_0 + kr$. By assuming that these are the $n=0$ and $n=1$ bound states of zero orbital angular momentum, calculate V_0 using the WKB formula. What do you predict for the rest mass of ψ'' , the $n=2$ state? (The measured value is $\approx 4.2 \text{ GeV}/c^2$.) [Hints: (1) Work with GeV instead of eV. (2) There is no need to determine k explicitly.]

Exercise 16.2.6. Obtain Eq. (16.2.39) for the λx^4 potential by the scaling trick.

*Exercise 16.2.7** Find the allowed levels of the harmonic oscillator by the WKB method.

Exercise 16.2.8. Consider the $l=0$ radial equation for the Coulomb problem. Since $V(r)$ is singular at the turning point $r=0$, we can't use $(n+3/4)$.

- (1) Will the additive constant be more or less than $3/4$?
- (2) By analyzing the exact equation near $r=0$, it can be shown that the constant equals 1. Using this constant show that the WKB energy levels agree with the exact results.

[‡] The assumption that $V(x)$ may be linearized near the turning point breaks down and this invalidates Eq. (16.2.29).

Time-Independent Perturbation Theory

17.1. The Formalism

Time-independent perturbation theory is an approximation scheme that applies in the following context: we know the solution to the eigenvalue problem of the Hamiltonian H^0 , and we want the solution to $H = H^0 + H^1$, where H^1 is small compared to H^0 in a sense to be made precise shortly. For instance, H^0 can be the Coulomb Hamiltonian for an electron bound to a proton, and H^1 the addition due to an external electric field that is weak compared to the proton's field at the (average) location of the electron. One refers to H^0 as the *unperturbed Hamiltonian* and H^1 as the *perturbing Hamiltonian* or *perturbation*.

We proceed as follows. We assume that to every eigenket $|E_n^0\rangle \equiv |n^0\rangle$ of H^0 with eigenvalue E_n^0 , there is an eigenket $|n\rangle$ of H with eigenvalue E_n .[†] We then assume that the eigenkets and eigenvalues of H may be expanded in a *perturbation series*:[‡]

$$|n\rangle = |n^0\rangle + |n^1\rangle + |n^2\rangle + \dots \quad (17.1.1)$$

$$E_n = E_n^0 + E_n^1 + E_n^2 + \dots \quad (17.1.2)$$

The superscript k on each term gives the power of (the matrix element of) H^1 that it is expected to come out proportional to. A term with superscript equal to k is called a *k th-order term*. (Clearly a product like $E_n^k |n^k\rangle$ is a term of order $k+k$.) We hope that as the order increases, the terms get systematically smaller; this is when we can say that H^1 is small. When we find explicit formulas for $|n^k\rangle$ and E_n^k , these ideas will be sharpened.

To find the terms in the expansions for $|n\rangle$ and E_n , we start with the eigenvalue equation

$$H|n\rangle = E_n|n\rangle \quad (17.1.3)$$

[†] We are assuming that $|n^0\rangle$ is *nondegenerate*. The degenerate case follows.

[‡] We assume $|n^0\rangle$ is normalized (to unity). The norm of $|n\rangle$ will be discussed shortly.

or

$$(H^0 + H^1)(|n^0\rangle + |n^1\rangle + \dots) = (E_n^0 + E_n^1 + \dots)(|n^0\rangle + |n^1\rangle + \dots) \quad (17.1.4)$$

We approach these equations as we did the differential equation in the WKB approximation. Recall that there we had an equation with terms of order \hbar^{-2} , \hbar^{-1} , etc. We first ignored all but the \hbar^{-2} terms and solved for ϕ_0 . We then fed this into the \hbar^{-1} part to determine ϕ_1 . (We could have gone on this way, though we chose to stop there.) Likewise, in the present case, we first consider the zeroth-order terms of Eq. (17.1.4). We get the equation

$$H^0|n^0\rangle = E_n^0|n^0\rangle \quad (17.1.5)$$

Notice that the zeroth-order quantities $|n^0\rangle$ and E_n^0 are indeed independent of H^1 (or, equivalently, they depend on the zeroth power of H^1). By assumption, this equation may be solved and the eigenvectors $|n^0\rangle$ and eigenvalues E_n^0 determined. So we move on to the first-order terms. We get the equation

$$H^0|n^1\rangle + H^1|n^0\rangle = E_n^0|n^1\rangle + E_n^1|n^0\rangle \quad (17.1.6)$$

Let us dot both sides with $\langle n^0|$. Using $\langle n^0|H^0 = \langle n^0|E_n^0$ and $\langle n^0|n^0\rangle = 1$, we get

$$E_n^1 = \langle n^0|H^1|n^0\rangle \quad (17.1.7)$$

i.e., *the first-order change in energy is the expectation value of H^1 in the unperturbed state.* Notice that E_n^1 is indeed proportional to the first power of H^1 . Let us next dot both sides of Eq. (17.1.6) with $\langle m^0|$, $m \neq n$, to get

$$\langle m^0|H^0|n^1\rangle + \langle m^0|H^1|n^0\rangle = E_n^0\langle m^0|n^1\rangle$$

or

$$\langle m^0|n^1\rangle = \frac{\langle m^0|H^1|n^0\rangle}{E_n^0 - E_m^0} \quad (17.1.8)$$

Since $m \neq n$, this equation determines all the components of $|n^1\rangle$ in the eigenbasis of H^0 , except for the component parallel to $|n^0\rangle$, let's call it $|n_{\parallel}^1\rangle$. We determine it by the requirement that $|n\rangle$ is normalized to this order.[‡] In obvious notation, we have

$$1 = \langle n|n\rangle = (\langle n^0| + \langle n_{\perp}^1| + \langle n_{\parallel}^1|)(|n^0\rangle + |n_{\perp}^1\rangle + |n_{\parallel}^1\rangle) \quad (17.1.9)$$

[‡] Recall that even in eigenvalue problems that can be solved exactly, there is the arbitrariness in the norm of the vector. To this order, only $|n_{\parallel}^1\rangle$ alters the length of $|n^0\rangle$. [See Eq. (17.1.10).]

which leads to

$$1 = \langle n^0 | n^0 \rangle + \langle n_{||}^1 | n^0 \rangle + \langle n^0 | n_{||}^1 \rangle + \text{higher order}$$

or

$$0 = \langle n_{||}^1 | n^0 \rangle + \langle n^0 | n_{||}^1 \rangle + \text{higher order} \quad (17.1.10)$$

This means that

$$\langle n^0 | n_{||}^1 \rangle = i\alpha, \quad \alpha \text{ real} \quad (17.1.11)$$

Using

$$1 + i\alpha = e^{i\alpha} \quad (\text{to this order}) \quad (17.1.12)$$

we get

$$|n\rangle = |n^0\rangle e^{i\alpha} + \sum'_m \frac{|m^0\rangle \langle m^0 | H^1 | n^0 \rangle}{E_n^0 - E_m^0} \quad (17.1.13)$$

where the prime on \sum' means that $m \neq n$. Since $|n\rangle$ has an arbitrariness in its overall phase, even after it is normalized, let us change its phase by the factor $e^{-i\alpha}$ in Eq. (17.1.13). This gets rid of the phase factor multiplying $|n^0\rangle$ and does nothing to the first-order piece, *to this order*. Calling the perturbed eigenket with the new phase also $|n\rangle$, we get the result to first order:

$$|n\rangle = |n^0\rangle + \sum'_m \frac{|m^0\rangle \langle m^0 | H^1 | n^0 \rangle}{E_n^0 - E_m^0} = |n^0\rangle + |n^1\rangle \quad (17.1.14)$$

Notice that $|n^1\rangle$ is orthogonal to $|n^0\rangle$ and proportional to the first power of H^1 (as anticipated). We determine E_n^2 from the second-order part of Eq. (17.1.4):

$$H^0 |n^2\rangle + H^1 |n^1\rangle = E_n^0 |n^2\rangle + E_n^1 |n^1\rangle + E_n^2 |n^0\rangle \quad (17.1.15)$$

Dotting with $\langle n^0 |$ and using the results from lower order ($|n^1\rangle = |n_{||}^1\rangle$) we obtain

$$E_n^2 = \langle n^0 | H^1 | n^1 \rangle \quad (17.1.16)$$

$$\begin{aligned} &= \sum'_m \frac{\langle n^0 | H^1 | m^0 \rangle \langle m^0 | H^1 | n^0 \rangle}{E_n^0 - E_m^0} \\ &= \sum'_m \frac{|\langle n^0 | H^1 | m^0 \rangle|^2}{E_n^0 - E_m^0} \end{aligned} \quad (17.1.17)$$

We can go on to higher orders, but we choose to stop here.

Before we turn to examples, let us consider some general features of our results. First we note that the energy to a given order is determined by the state vector to the next lower order, see Eqs. (17.1.7) and (17.1.16). This is in accord with the remarks made in the study of the variational method. The physics behind this phenomenon will become clear when we consider a few examples. Next we ask under what conditions the perturbation expansion is good, namely, when the correction terms are small compared to the zeroth-order (unperturbed) results. The answer follows from Eq. (17.1.14). A necessary condition for $|n^1\rangle$ to be small compared to $|n^0\rangle$ is that

$$\left| \frac{\langle m^0 | H^1 | n^0 \rangle}{E_n^0 - E_m^0} \right| \ll 1 \quad (17.1.18)$$

Thus we see that the condition depends on (1) the absolute size of H^1 (i.e., if it is due to some external field, the magnitude of the field); (2) the matrix elements of H^1 between unperturbed states; and (3) the energy difference between the levels. If the unperturbed eigenstate is $|n^0\rangle$, the perturbation mixes in orthogonal states $|m^0\rangle$; this mixing is directly proportional to the matrix element $\langle m^0 | H^1 | n^0 \rangle$ and inversely proportional to the energy difference between the two levels, which measures the “rigidity” of the system. If for any reason the above inequality is not fulfilled (say due to degeneracy, $E_n^0 = E_m^0$) we must turn to an alternate formalism called *degenerate perturbation theory* to be described later in this chapter.

17.2. Some Examples

Consider a particle of charge q and mass m in a harmonic oscillator potential $V = \frac{1}{2}m\omega^2x^2$. Suppose we apply an external electric field of magnitude f along the positive x direction. This corresponds to an electrostatic potential $\phi = -fx$ and a potential energy $V = -qfx$. Thus

$$H = H^0 + H^1 = \frac{P^2}{2m} + \frac{1}{2}m\omega^2X^2 - qfX \quad (17.2.1)$$

We wish to handle H^1 by perturbation theory. Let us first calculate the first-order shift in energy, given by

$$E_n^1 = \langle n^0 | H^1 | n^0 \rangle = -qf \langle n^0 | X | n^0 \rangle \quad (17.2.2)$$

where $|n^0\rangle$ is just the n th state of the unperturbed oscillator. We can see that E^1 vanishes in many ways. At a formal level, since

$$X = \left(\frac{\hbar}{2m\omega} \right)^{1/2} (a + a^\dagger) \quad (17.2.3)$$

$$\begin{aligned} E_n^1 &= -qf \int (\psi_n^0)^* x \psi_n^0 dx \\ &= -qf \int |\psi_n^0|^2 x dx \end{aligned} \quad (17.2.4)$$

Now $\psi_n^0(x)$, being the unperturbed eigenfunction, has definite parity $(-1)^n$. Consequently $|\psi_n^0|^2$ is an even function, while the external potential is an odd function. Thus the average interaction with the external field is zero, for the particle is as likely to be found in the region of potential ϕ as in the region of potential $-\phi$. Notice that E_n^1 is the energy of interaction of the *unperturbed configuration* $|n^0\rangle$, with the applied field. Consequently this is not the whole story, for the configuration itself will get modified by the external field to $|n^0\rangle + |n^1\rangle + \dots$, and we should really be considering the energy of interaction of the perturbed configurations and the applied field. But this is a distinction that is at least a second-order effect, for the change $\delta|n\rangle \equiv |n\rangle - |n^0\rangle$ in the configuration is at least a first-order effect and the interaction of $\delta|n\rangle$ with the applied field involves another order of H^1 . So let us calculate the perturbed eigenket to first order and then energy levels to second order. From Eq. (17.1.14).

$$\begin{aligned} |n\rangle &= |n^0\rangle + \sum_m' \frac{|m^0\rangle \langle m^0| - qf(\hbar/2m\omega)^{1/2}(a + a^\dagger)|n^0\rangle}{E_n^0 - E_m^0} \\ &= |n^0\rangle + qf \left(\frac{1}{2m\hbar\omega^3} \right)^{1/2} [(n+1)^{1/2}|(n+1)^0\rangle - n^{1/2}|(n-1)^0\rangle] \end{aligned} \quad (17.2.5)$$

Thus to first order, the perturbation mixes the state $|n^0\rangle$ with the states immediately above and below it. It was stated earlier that $E_n^0 - E_m^0$ measures the “rigidity” of the system. We find in this example that this quantity is proportional to ω , which in the mass-spring case measures the force constant. How does the wave function of the perturbed state look? This is not transparent from the above equation, but we expect that it will represent a probability distribution that is no longer centered at and symmetric about $x=0$, but instead is biased toward positive x (for that is the direction of the external field). We will return to confirm this picture quantitatively.

The second-order energy shift (which reflects the fact that the configuration of the system is not $|n^0\rangle$ but $|n^0\rangle + |n^1\rangle$), is

$$\begin{aligned} E_n^2 &= \langle n^0 | H^1 | n^1 \rangle = \sum_m \frac{|\langle m^0 | H^1 | n^0 \rangle|^2}{E_n^0 - E_m^0} \\ &= q^2 \cdot f^2 \cdot \frac{\hbar}{2m\omega} \left(\frac{n+1}{-\hbar\omega} + \frac{n}{\hbar\omega} \right) = \frac{-q^2 f^2}{2m\omega^2} \end{aligned} \quad (17.2.6)$$

The present problem is a nice testing ground for perturbation theory because it may be exactly solved. This is because H may be written as

$$\begin{aligned} H &= \frac{P^2}{2m} + \frac{1}{2} m\omega^2 X^2 - qfX \\ &= \frac{P^2}{2m} + \frac{1}{2} m\omega^2 \left(X - \frac{qf}{m\omega^2} \right)^2 - \frac{1}{2} \frac{q^2 f^2}{m\omega^2} \end{aligned} \quad (17.2.7)$$

This Hamiltonian also describes an oscillator of frequency ω , but is different in that (i) the oscillator is centered at $x = qf/m\omega^2$; (ii) each state has a constant energy $(-q^2 f^2/2m\omega^2)$ added to it. Thus the eigenfunctions of H are just the eigenfunctions of H^0 shifted by $qf/m\omega^2$ and the eigenvalues are $E_n = E_n^0 - q^2 f^2 / 2m\omega^2$. The classical picture that goes with Eq. (17.2.7) is clear: the effect of a constant force qf on a mass coupled to a spring of force constant $m\omega^2$ is to shift the equilibrium point to $x = qf/m\omega^2$. (Imagine a mass hanging from a spring attached to the ceiling and ask what gravity does to its dynamics.) Let us compare these exact results with the perturbation theory. Consider the energy

$$E_n = E_n^0 - q^2 f^2 / 2m\omega^2 \quad (17.2.8)$$

Since H^1 is proportional to qf , the power of qf gives the order of the term. According to Eq. (17.2.8), there is no first-order shift in energy, and the second-order shift is $-q^2 f^2 / 2m\omega^2$, which agrees with Eq. (17.2.6). Had we tried to go to higher orders, we would have found nothing more.

Now consider the state vectors. The exact result is

$$|n\rangle = T(qf/m\omega^2)|n^0\rangle \quad (17.2.9)$$

where $T(a)$ is the operator that translates the system by an amount a . Since we are working to first order in qf ,

$$\begin{aligned} T(qf/m\omega^2) &= e^{-i(qf/m\omega^2\hbar)P} \simeq I - i\left(\frac{qf}{m\omega^2\hbar}\right)P \\ &= I - i\left(\frac{qf}{m\omega^2\hbar}\right)\left(\frac{\hbar m\omega}{2}\right)^{1/2} \cdot \frac{a - a^\dagger}{i} \end{aligned} \quad (17.2.10)$$

so that

$$\begin{aligned} |n\rangle &= \left[I - \left(\frac{qf}{m\omega^2\hbar}\right)\left(\frac{\hbar m\omega}{2}\right)^{1/2} (a - a^\dagger) \right] |n^0\rangle \\ &= |n^0\rangle + qf\left(\frac{1}{2m\hbar\omega^3}\right)^{1/2} [(n+1)^{1/2}|(n+1)^0\rangle - n^{1/2}|(n-1)^0\rangle] \end{aligned} \quad (17.2.11)$$

which agrees with Eq. (17.2.5). It is clear that computing $|n\rangle$ to higher order in perturbation theory will be equivalent to expanding T to higher orders in qf .

*Exercise 17.2.1.** Consider $H^I = \lambda x^4$ for the oscillator problem.

(1) Show that

$$E_n^I = \frac{3\hbar^2\lambda}{4m^2\omega^2} [1 + 2n + 2n^2]$$

(2) Argue that no matter how small λ is, the perturbation expansion will break down for some large enough n . What is the physical reason?

*Exercise 17.2.2.** Consider a spin-1/2 particle with gyromagnetic ratio γ in a magnetic field $\mathbf{B} = B\hat{\mathbf{i}} + B_0\hat{\mathbf{k}}$. Treating B as a perturbation, calculate the first- and second-order shifts in energy and first-order shift in wave function for the ground state. Then compare the exact answers expanded to the corresponding orders.

Exercise 17.2.3. In our study of the H atom, we assumed that the proton is a point charge e . This leads to the familiar Coulomb interaction ($-e^2/r$) with the electron. (1) Show that if the proton is a uniformly dense charge distribution of radius R , the interaction is

$$\begin{aligned} V(r) &= -\frac{3e^2}{2R} + \frac{e^2r^2}{2R^3}, & r \leq R \\ &= -\frac{e^2}{r}, & r > R \end{aligned}$$

(2) Calculate the first-order shift in the ground-state energy of hydrogen due to this modification. You may assume $e^{-R/a_0} \approx 1$. You should find $E^I = 2e^2R^2/5a_0^3$.

*Exercise 17.2.4.** (1) Prove the Thomas-Reiche-Kuhn sum rule

$$\sum_{n'} (E_{n'} - E_n) |\langle n' | X | n \rangle|^2 = \sum_{n'} (E_{n'} - E_n) \langle n | X | n' \rangle \langle n' | X | n \rangle = \frac{\hbar^2}{2m}$$

where $|n\rangle$ and $|n'\rangle$ are eigenstates of $H = P^2/2m + V(X)$. (Hint: Eliminate the $E_{n'} - E_n$ factor in favor of H .)

(2) Test the sum rule on the n th state of the oscillator.

Exercise 17.2.5 (Hard). We have seen that if we neglect the repulsion e^2/r_{12} between the two electrons in the ground state of He, the energy is $-8 \text{ Ry} = -108.8 \text{ eV}$. Treating e^2/r_{12} as a perturbation, show that

$$\langle 100, 100 | H^I | 100, 100 \rangle = \frac{5}{2} \text{ Ry}$$

so that $E_0^0 + E_0^I = -5.5 \text{ Ry} = -74.8 \text{ eV}$. Recall that the measured value is -78.6 eV and the variational estimate is -77.5 eV . [Hint: $\langle H^I \rangle$ can be viewed as the interaction between two concentric, spherically symmetric exponentially falling charge distributions. Find the potential $\phi(r)$ due to one distribution and calculate the interaction energy between this potential and the other charge distribution.]

Selection Rules

The labor involved in perturbation theory calculations is greatly reduced by the use of selection rules, which allow us to conclude that certain matrix elements of H^1 are zero without explicitly calculating them. They are based on the idea that if

$$[\Omega, H^1] = 0$$

then

$$\langle \alpha_2 \omega_2 | H^1 | \alpha_1 \omega_1 \rangle = 0 \quad \text{unless} \quad \omega_1 = \omega_2 \quad (17.2.12)^\ddagger$$

Proof.

$$0 = \langle \alpha_2 \omega_2 | \Omega H^1 - H^1 \Omega | \alpha_1 \omega_1 \rangle = (\omega_2 - \omega_1) \langle \alpha_2 \omega_2 | H^1 | \alpha_1 \omega_1 \rangle \quad \text{Q.E.D.}$$

Consider for example $H^1 = \lambda Z$, which is invariant under rotations around the z axis. Then $[L_z, H^1] = 0$ and

$$\langle \alpha_2 m_2 | H^1 | \alpha_1 m_1 \rangle = 0 \quad \text{unless} \quad m_2 = m_1 \quad (17.2.13)$$

(This result also follows from the Wigner–Eckart theorem.) Or if H^1 is parity invariant, say $H^1 = \lambda Z^2$, then its matrix element between states of opposite parity is zero.

There is a simple way to understand Eq. (17.2.12). To say that $[\Omega, H^1] = 0$ is to say that H^1 “carries no Ω ”; in other words, when it acts on a state it imparts no Ω to it. We see this as follows. Consider $|\omega_1\rangle$, which carries a definite amount of the variable Ω , namely, ω_1 :

$$\Omega |\omega_1\rangle = \omega_1 |\omega_1\rangle \quad (17.2.14)$$

Let us measure Ω in the state after H^1 acts on it:

$$\Omega(H^1|\omega_1\rangle) = H^1\Omega|\omega_1\rangle = H^1\omega_1|\omega_1\rangle = \omega_1(H^1|\omega_1\rangle) \quad (17.2.15)$$

We find it is the same as before, namely, ω_1 . The selection rule then merely reflects the orthogonality of eigenstates with different ω .

This discussion paves the way for an extension of the selection rule to a case where H^1 carries a definite amount of Ω . For instance, if H^1 is a tensor operator T_k^q , it carries angular momentum (k, q) and we know from the Wigner–Eckart theorem that

$$\langle \alpha_2 j_2 m_2 | T_k^q | \alpha_1 j_1 m_1 \rangle = 0 \quad \text{unless} \quad \begin{cases} j_1 + k \geq j_2 \geq |j_1 - k| \\ m_2 = m_1 + q \end{cases} \quad (17.2.16)$$

[†] α stands for other quantum numbers that label the state.

i.e., that the matrix element vanishes unless $|\alpha_2 j_2 m_2\rangle$ has the angular momentum that obtains when we add to $(j_1 m_1)$ the angular momentum (kq) imparted by the operator. For instance, if $H^l = \lambda Z \sim T_1^0$,

$$\langle \alpha_2 j_2 m_2 | Z | \alpha_1 j_1 m_1 \rangle = 0 \quad \text{unless} \quad \begin{cases} j_2 = j_1 + 1, j_1, j_1 - 1 \\ m_2 = m_1 + 0 \end{cases} \quad (17.2.17)$$

while if $H^l = \lambda X$ or $\lambda Y (\sim T_1^{\pm 1})$, we have

$$\langle \alpha_2 j_2 m_2 | X \text{ or } Y | \alpha_1 j_1 m_1 \rangle = 0 \quad \text{unless} \quad \begin{cases} j_2 = j_1 + 1, j_1, j_1 - 1 \\ m_2 = m_1 \pm 1 \end{cases} \quad (17.2.18)$$

Another example of this type is an operator that is not parity invariant, but parity odd. An example is X , which obeys

$$\Pi^\dagger X \Pi = -X \quad (17.2.19)$$

You can verify that if X acts on a state of definite parity, it changes the parity of the state. Thus the matrix element of X between eigenstates of parity vanishes unless they have opposite parity. More generally, if

$$\Pi^\dagger \Omega \Pi = -\Omega \quad (17.2.20)$$

then the matrix element of Ω between two parity eigenstates vanishes unless they have opposite parity.

We get more selection rules by combining these selection rules. For instance, we can combine the angular momentum and parity selection rules for the vector operators \mathbf{R} to get (in the case of no spin, $\mathbf{J} = \mathbf{L}$),

$$\begin{aligned} \langle \alpha_2 l_2 m_2 | Z | \alpha_1 l_1 m_1 \rangle &= 0 \quad \text{unless} \quad \begin{cases} l_2 = l_1 \pm 1 \\ m_2 = m_1 \end{cases} \\ \langle \alpha_2 l_2 m_2 | X \text{ or } Y | \alpha_1 l_1 m_1 \rangle &= 0 \quad \text{unless} \quad \begin{cases} l_2 = l_1 \pm 1 \\ m_2 = m_1 \pm 1 \end{cases} \end{aligned} \quad (17.2.21)$$

We rule out the possibility $l_2 = l_1$ by the parity selection rule, for states of orbital angular momentum l have definite parity $(-1)^l$. Equation (17.2.21) is called the *dipole selection rule*.

We now consider an example that illustrates the use of these tricks and a few more. The problem is to determine the response of the hydrogen atom in the ground state to a constant external electric field $\mathbf{E} = \mathcal{E}\mathbf{k}$. This is called the *Stark effect*. Let us first calculate H^l . We do this by determining \mathcal{H}^l , its classical counterpart and then making the operator substitution. If \mathbf{r}_1 and \mathbf{r}_2 are the position vectors of the

electron and proton, respectively, and $\phi(\mathbf{r})$ is the electrostatic potential due to \mathbf{E} , then

$$\begin{aligned}\mathcal{H}^1 &= -e\phi(\mathbf{r}_1) + e\phi(\mathbf{r}_2) \\ &= e[\phi(\mathbf{r}_2) - \phi(\mathbf{r}_1)] \\ &= e(\mathbf{r}_1 - \mathbf{r}_2) \cdot \mathbf{E} \quad (\text{recall } \mathbf{E} = -\nabla\phi) \\ &= e\mathbf{r} \cdot \mathbf{E}\end{aligned}\tag{17.2.22}$$

where \mathbf{r} is the relative coordinate or equivalently the position vector of the electron in the CM frame in the limit $m/M=0$. \mathcal{H}^1 is called the *dipole interaction*, for in terms of

$$\boldsymbol{\mu}_e = e(\mathbf{r}_2 - \mathbf{r}_1) = -e\mathbf{r}\tag{17.2.23}$$

the *electric dipole moment* of the system,

$$\mathcal{H}^1 = -\boldsymbol{\mu}_e \cdot \mathbf{E}\tag{17.2.24}$$

(This is the electric analog of $\mathcal{H} = -\boldsymbol{\mu} \cdot \mathbf{B}$).[‡] Thus, for the given electric field

$$H^1 = eZ\mathcal{E}\tag{17.2.25}$$

Let us now calculate the first-order shift in the energy of the ground state $|100\rangle$:[§]

$$E_{100}^1 = \langle 100 | eZ\mathcal{E} | 100 \rangle\tag{17.2.26}$$

We can argue that $E_{100}^1 = 0$ either on the grounds of parity or the Wigner–Eckart theorem. More physically, E_{100}^1 vanishes because in the unperturbed state, the electron probability distribution is spherically symmetric and the electron samples $\phi(\mathbf{r})$ and $\phi(-\mathbf{r}) = -\phi(\mathbf{r})$ equally. Another way to say this is that the unperturbed atom has no mean electric dipole moment $\langle \boldsymbol{\mu} \rangle$ (by parity or the Wigner–Eckart theorem) so that

$$E_{100}^1 = \langle 100 | -\boldsymbol{\mu} \cdot \mathbf{E} | 100 \rangle = -\langle 100 | \boldsymbol{\mu} | 100 \rangle \cdot \mathbf{E} = 0\tag{17.2.27}$$

But we expect the second-order energy shift to be nonzero, for the external field will shift the electron distribution downward and induce a dipole moment which can interact with \mathbf{E} . So let us calculate

$$E_{100}^2 = \sum'_{nlm} \frac{e^2 \mathcal{E}^2 |\langle nlm | Z | 100 \rangle|^2}{E_{100}^0 - E_{nlm}^0}\tag{17.2.28}$$

[‡] In the rest of this chapter we will omit the subscript e on $\boldsymbol{\mu}_e$.

[§] When we discuss hydrogen, we will use the symbol $|nlm\rangle$, rather than $|(nlm)^0\rangle$ to denote the *unperturbed* state.

where

$$E_{100}^0 - E_{nlm}^0 = -Ry \left(1 - \frac{1}{n^2} \right) = Ry \left(\frac{1-n^2}{n^2} \right) \quad (17.2.29)$$

Unlike in the case of the oscillator, the sum now involves an infinite number of terms. Although we can use dipole selection rules to reduce the sum to

$$E_{100}^2 = \sum_{n=2}^{\infty} \frac{e^2 \mathcal{E}^2 |\langle n10|Z|100 \rangle|^2}{E_1^0 - E_n^0} \quad (17.2.30)$$

let us keep the form in Eq. (17.2.28) for a while. There are several ways to proceed.

Method 1. Since the magnitude of the energy denominator grows with n , we have the inequality

$$|E_{100}^2| \leq \frac{e^2 \mathcal{E}^2}{|E_1^0 - E_2^0|} \sum'_{nlm} |\langle nlm|Z|100 \rangle|^2$$

But since

$$\begin{aligned} & \sum'_{nlm} |\langle nlm|Z|100 \rangle|^2 \\ &= \sum'_{nlm} \langle 100|Z|nlm \rangle \langle nlm|Z|100 \rangle \\ &= \sum_{nlm} \langle 100|Z|nlm \rangle \langle nlm|Z|100 \rangle - \langle 100|Z|100 \rangle^2 \\ &= \langle 100|Z^2|100 \rangle - \langle 100|Z|100 \rangle^2 \\ &= a_0^2 - 0 = a_0^2 \end{aligned} \quad (17.2.31)$$

we get

$$\begin{aligned} |E_{100}^2| &\leq \frac{e^2 \mathcal{E}^2}{|(e^2/2a_0)(1-\frac{1}{4})|} a_0^2 \\ &\leq \frac{8a_0^3 \mathcal{E}^2}{3} \end{aligned} \quad (17.2.32)$$

We can also get a lower bound on $|E_{100}^2|$ by keeping just the first term in Eq. (17.2.30) (since all terms have the same sign):

$$|E_{100}^2| \geq \frac{e^2 \mathcal{E}^2}{3e^2/8a_0} |\langle 210|Z|100 \rangle|^2 \quad (17.2.33)$$

Now,

$$|\langle 210|Z|100\rangle|^2 = \frac{2^{15}a_0^2}{3^{10}} \simeq 0.55a_0^2 \quad (17.2.34)$$

so that

$$|E_{100}^2| \geq (0.55)^{\frac{8}{3}}\epsilon^2 a_0^3 \quad (17.2.35)$$

We thus manage to restrict $|E_{100}^2|$ to the interval

$$\frac{8}{3}\epsilon^2 a_0^3 \geq |E_{100}^2| \geq 0.55(\frac{8}{3})\epsilon^2 a_0^3 \quad (17.2.36)$$

Method 2. Consider the general problem of evaluating

$$E_n^2 = \sum_m' \frac{\langle n^0 | H^1 | m^0 \rangle \langle m^0 | H^1 | n^0 \rangle}{E_n^0 - E_m^0} \quad (17.2.37)$$

If it weren't for the energy denominator, we could use the completeness relation to eliminate the sum (after adding and subtracting the $m=n$ term). There exists a way to eliminate the energy denominator.[‡] Suppose we can find an operator Ω such that

$$H^1 = [\Omega, H^0] \quad (17.2.38)$$

then

$$\begin{aligned} E_n^2 &= \sum_m' \frac{\langle n^0 | H^1 | m^0 \rangle \langle m^0 | \Omega H^0 - H^0 \Omega | n^0 \rangle}{E_n^0 - E_m^0} \\ &= \sum_m' \langle n^0 | H^1 | m^0 \rangle \langle m^0 | \Omega | n^0 \rangle \\ &= \langle n^0 | H^1 \Omega | n^0 \rangle - \langle n^0 | H^1 | n^0 \rangle \langle n^0 | \Omega | n^0 \rangle \end{aligned} \quad (17.2.39)$$

which calls for computing just three matrix elements. But it is not an easy problem to find the Ω that satisfies Eq. (17.2.38). (There are, however, exceptions, see Exercise 17.2.7.) A more modest proposal is to find Ω such that

$$H^1 |n^0\rangle = [\Omega, H^0] |n^0\rangle \quad (17.2.40)$$

for a given $|n^0\rangle$. You can verify that this is all it takes to derive Eq. (17.2.39) for this value of n . In the problem we are interested in, we need to solve

$$H^1 |100\rangle = [\Omega, H^0] |100\rangle \quad (17.2.41)$$

[‡] See A. Dalgarno and J. T. Lewis, *Proceedings of the Royal Society*, **A233**, 70 (1955).

By writing this equation in the coordinate basis and assuming Ω is a function of coordinates and not momenta, we can show that

$$\Omega \xrightarrow[\text{coordinate basis}]{} -\frac{ma_0e\mathcal{E}}{\hbar^2} \left(\frac{r^2 \cos \theta}{2} + a_0 r \cos \theta \right) \quad (17.2.42)$$

The exact second-order shift is then

$$\begin{aligned} |E_{100}^2| &= |\langle 100 | H^1 \Omega | 100 \rangle - 0| \\ &= |\langle 100 | eZ\mathcal{E}\Omega | 100 \rangle| \\ &= \frac{9}{4}a_0^3\mathcal{E}^2 = \frac{8}{3}a_0^3\mathcal{E}^2 \cdot \left(\frac{27}{32}\right) \\ &= (0.84)\frac{8}{3}a_0^3\mathcal{E}^2 \end{aligned} \quad (17.2.43)$$

which is roughly in the middle of the interval we restricted it to by Method 1.

Exercise 17.2.6. Verify Eq. (17.2.34).

*Exercise 17.2.7.** For the oscillator, consider $H^1 = -qfX$. Find an Ω that satisfies Eq. (17.2.38). Feed it into Eq. (17.2.39) for E_n^2 and compare with the earlier calculation.

Exercise 17.2.8. Fill in the steps connecting Eqs. (17.2.41) and (17.2.43). Try to use symmetry arguments to reduce the labor involved in evaluating the integrals.

We argued earlier that E_{100}^2 represents the interaction of the induced dipole moment with the applied field. How big is the induced moment μ ? One way to find out is to calculate $\langle \mu \rangle$ in the perturbed ground state. An easier way to extract it from E_{100}^2 . Suppose we take a system that has no intrinsic dipole moment and turn on an external electric field that starts at 0 and grows to the full value of \mathbf{E} . During this time the dipole moment grows from 0 to μ . If you imagine charges $\pm q$ separated by a distance x along \mathbf{E} , you can see that the work done on the system as x changes by dx is

$$\begin{aligned} dW &= -q\mathcal{E} dx \\ &= -\mathcal{E} d\mu \end{aligned} \quad (17.2.44)$$

If we assume that the induced moment is proportional to \mathbf{E} :

$$\mu = \alpha \mathbf{E} \quad (17.2.45)$$

(where α is called the *polarizability*), then

$$dW = -\alpha \mathcal{E} d\mathcal{E}$$

or

$$W = -\frac{1}{2}a\epsilon^2 \quad (17.2.46)$$

We identify W with E_{100}^2 and determine the polarizability

$$\alpha = \frac{18}{4} a_0^3 \approx \frac{18}{4} (0.5 \text{ \AA})^3 \approx 0.56 \text{ \AA}^3 \quad (17.2.47)$$

If we use a more accurate value $a_0 = 0.53 \text{ \AA}$, we get $\alpha = 0.67 \text{ \AA}^3$, which is in excellent agreement with the measured value of 0.68 \AA^3 . For a given \mathbf{E} , we can get μ from Eq. (17.2.45).

Finally note that E_{100}^2 is negative. From Eq. (17.1.17) it is clear that the second-order shift in the ground-state energy is always negative (unless it vanishes). Since E_0^2 measures the energy shift due to the first-order change in the ground-state state vector, we conclude that the system changes its configuration so as to lower its energy of interaction with the external field.

17.3. Degenerate Perturbation Theory

In the face of degeneracy ($E_n^0 = E_m^0$) the condition for the validity of the perturbation expansion,

$$\left| \frac{\langle m^0 | H^1 | n^0 \rangle}{E_n^0 - E_m^0} \right| \ll 1 \quad (17.3.1)$$

is impossible to fulfill. The breakdown of the method may be understood in less formal terms as follows.

Let us consider the case when neither H^0 nor $H^0 + H^1$ is degenerate. For the purposes of this argument imagine that H^1 is due to some external field that can be continuously varied from zero to its full value. As the total Hamiltonian grows from H^0 to $H^0 + H^1$, the corresponding eigenbasis changes continuously from $|n^0\rangle$ to $|n\rangle$. It is this continuous or analytic variation of the eigenbasis with the perturbation that makes it possible to find $|n\rangle$ starting with $|n^0\rangle$, the way one finds the value of some analytic function at the point $x+a$ starting at the point x and using a Taylor series. Consider now the case when H^0 has a degenerate subspace and $H^0 + H^1$ is nondegenerate in this subspace. (More general cases can be handled the same way.) Imagine starting with the basis $|n\rangle$ and slowly turning *off* the perturbation. We will end up with a basis $|\bar{n}^0\rangle$ of H^0 . If we now turn on the perturbation, we can retrace our path back to $|n\rangle$. It is clear that if we start with this basis, $|\bar{n}^0\rangle$, we can evaluate $|n\rangle$ perturbatively. But since H^0 is degenerate, we needn't have started with this basis; we could have started with some other basis $|n^0\rangle$, chosen randomly. But if we start with any basis except $|\bar{n}^0\rangle$, and turn on the external field of infinitesimal size, the change in the basis will not be infinitesimal. It is this nonanalytic behavior that is signaled by the divergence in the first-order matrix element. [This can be compared to the divergence of the first derivative in the Taylor series where $f(x)$ is discontinuous.]

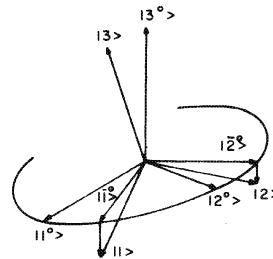


Figure 17.1. An example of the degenerate problem from $V^3(R)$. In the x - y plane, which is the degenerate subspace, we must start not with some arbitrarily chosen pair of basis vectors $|1^0\rangle$ and $|2^0\rangle$, but with the pair $|\bar{1}^0\rangle$ and $|\bar{2}^0\rangle$ which diagonalizes H^1 .

ous.] So, we must start with the right basis in the degenerate space. We have already characterized this basis as one we get if we start with $|n\rangle$ and slowly turn off H^1 . A more useful characterization is the following: it is a basis that diagonalizes H^1 within the degenerate space. Why? Because, if we start with this basis, the first-order perturbation coefficient [Eq. (17.1.8)] does not blow up, for the (off-diagonal) matrix element in the numerator vanishes along with energy denominator whenever $|n^0\rangle$ and $|m^0\rangle$ belong to the degenerate space. Figure 17.1 depicts a simple example from $V^3(R)$, where the x - y plane is the degenerate space and $|1^0\rangle$ and $|2^0\rangle$ are randomly chosen basis vectors in that subspace. The proper starting point is the pair $|\bar{1}^0\rangle$, $|\bar{2}^0\rangle$, which diagonalizes H^1 in the x - y plane.

It is worth noting that to find the proper starting point, we need to find the basis that diagonalizes H^1 only within the degenerate space and not the full Hilbert space. Thus even if we work with infinite-dimensional spaces, the exact diagonalization will usually have to be carried out only in some small, finite-dimensional subspace.

Let us consider, as a concrete example, the Stark effect in the $n=2$ level of hydrogen. (We ignore spin, which is a spectator variable.) Are we to conclude that there is no first-order shift because

$$\langle 2lm|e\mathcal{E}Z|2lm\rangle = 0 \quad (17.3.2)$$

by parity invariance, or equivalently, because the atom in these states has no intrinsic dipole moment? No, because these states need not provide the correct starting points for a perturbative calculation in view of the degeneracy. We must first find the basis in the $n=2$ sector which diagonalizes H^1 . Using the selection rules, which tell us that only two of the 16 matrix elements are nonzero, we get

$$H^1 \rightarrow \begin{array}{c|cccc} nlm & 200 & 210 & 211 & 21-1 \\ \hline 200 & \begin{bmatrix} 0 & \Delta & 0 & 0 \end{bmatrix} \\ 210 & \begin{bmatrix} \Delta & 0 & 0 & 0 \end{bmatrix} \\ 211 & \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix} \\ 21-1 & \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix} \end{array} \quad (17.3.3)$$

where

$$\Delta = \langle 200|e\mathcal{E}Z|210\rangle = -3e\mathcal{E}a_0 \quad (17.3.4)$$

*Exercise 17.3.1.** Use the dipole selection rules to show that H^1 has the above form and carry out the evaluation of Δ .

Since H^1 is just Δ times the Pauli matrix σ_x in the $m=0$ sector, we infer that its eigenvalues are $\pm\Delta$ and that its eigenstates are $[|200\rangle \pm |210\rangle]/2^{1/2}$. In the $|m|=1$ sector the old states $|2, 1, \pm 1\rangle$ diagonalize H^1 . Our calculation tells us the following.

(1) The zeroth-order states stable under the perturbation are $|2, 1, \pm 1\rangle$ and $[|200\rangle \pm |210\rangle]/2^{1/2}$.

(2) The first-order shift E^1 is zero for the first two states and $\pm\Delta$ for the next two. (Note that Δ is negative.)

Notice that the stable eigenstates for which $E^1 \neq 0$ are mixtures of $l=0$ and $l=1$. Thus they have indefinite parity and can have a nonzero intrinsic dipole moment which can interact with E and produce a first-order energy shift. From the energy shift, we infer that the size of the dipole moment is $3ea_0$.

Degenerate perturbation theory is relevant not only when the levels are exactly degenerate but also when they are close, that is to say, when the inequality (17.3.1) is not respected. In that case one must diagonalize $H^0 + H^1$ exactly in the almost degenerate subspace.

Exercise 17.3.2. Consider a spin-1 particle (with no orbital degrees of freedom). Let $H = AS_z^2 + B(S_x^2 - S_y^2)$, where S_i are 3×3 spin matrices, and $A \gg B$. Treating the B term as a perturbation, find the eigenstates of $H^0 = AS_z^2$ that are stable under the perturbation. Calculate the energy shifts to first order in B . How are these related to the exact answers?

Fine Structure

The Coulomb potential ($-e^2/r$) does not give the complete interaction between the electron and the proton, though it does provide an excellent first approximation.[‡] There are “fine-structure” corrections to this basic interaction, which produce energy shifts of the order of α^2 times the binding energy due to the Coulomb potential. Since the electron velocity (in a semiclassical picture) is typically $\beta = v/c \simeq O(\alpha)$, these are corrections of the order $(v/c)^2$ relative to binding energy, which is itself proportional to $(v/c)^2$. Thus these are relativistic in origin. There are two parts to this effect.

The first reflects the fact that to order $(v/c)^4$ the kinetic energy of the electron is not $p^2/2m$ but

$$T = (c^2 p^2 + m^2 c^4)^{1/2} - mc^2 = \frac{p^2}{2m} - \frac{p^4}{8m^3 c^2} + O(p^6 \text{ or } v^6) \quad (17.3.5)$$

We now wish to calculate the effect of this extra term

$$H_T = -P^4/8m^3 c^2 \quad (17.3.6)$$

[‡] We consider here just the fine structure of hydrogen. The analysis may be extended directly to hydrogen-like atoms. We also ignore the difference between the reduced mass and the electron mass.

treating it as a perturbation. Since H_T is rotationally invariant, it is diagonal in the $|nlm\rangle$ basis. (In other words, the $|nlm\rangle$ basis is stable under this perturbation.) So we can forget about the fact that the levels at each n are degenerate and determine E_T^1 simply from

$$E_T^1 = -\frac{1}{8m^3c^2} \langle nlm | P^4 | nlm \rangle \quad (17.3.7)$$

We evaluate the matrix element by noting that

$$P^4 = 4m^2 \left(\frac{P^2}{2m} \right)^2 = 4m^2 \left(H^0 + \frac{e^2}{r} \right)^2 \quad (17.3.8)$$

so that

$$E_T^1 = -\frac{1}{2mc^2} \left[(E_n^0)^2 + 2E_n^0 e^2 \left\langle \frac{1}{r} \right\rangle_{nlm} + e^4 \left\langle \frac{1}{r^2} \right\rangle_{nlm} \right] \quad (17.3.9)$$

From the virial theorem [Eq. (13.1.34)]

$$-\left\langle \frac{e^2}{r} \right\rangle_{nlm} = 2E_n^0 \quad (17.3.10)$$

while from Exercise (17.3.4)

$$\left\langle \frac{e^4}{r^2} \right\rangle_{nlm} = \frac{e^4}{a_0^2 n^3 (l+1/2)} = \frac{4E_0^2 n}{l+1/2} \quad (17.3.11)$$

so that

$$\begin{aligned} E_T^1 &= -\frac{(E_n^0)^2}{2mc^2} \left(-3 + \frac{4n}{l+1/2} \right) \\ &= -\frac{1}{2} (mc^2) \alpha^4 \left[-\frac{3}{4n^4} + \frac{1}{n^3(l+1/2)} \right] \end{aligned} \quad (17.3.12)$$

The other relativistic effect is called the *spin-orbit interaction*. Its origin may be understood as follows. The Coulomb interaction ($-e^2/r$) is the whole story only if the electron is at rest. If it moves at a velocity v , there is an extra term which we find as follows. In the electron rest frame, the proton will be moving at a velocity

($-\mathbf{v}$) and will produce a magnetic field

$$\mathbf{B} = -\frac{e}{c} \frac{\mathbf{v} \times \mathbf{r}}{r^3} \quad (17.3.13)$$

The interaction of the magnetic moment of the electron with this field leads to the *spin-orbit energy*

$$\begin{aligned} \mathcal{H}_{\text{s.o.}} &= -\boldsymbol{\mu} \cdot \mathbf{B} = \frac{e}{mc r^3} \boldsymbol{\mu} \cdot (\mathbf{p} \times \mathbf{r}) \\ &= -\frac{e}{mc} \frac{\boldsymbol{\mu} \cdot \mathbf{l}}{r^3} \end{aligned} \quad (17.3.14)$$

So we expect that in the quantum theory there will be a perturbation

$$\begin{aligned} H_{\text{s.o.}} &= \left(-\frac{e}{mc} \right) \left(-\frac{e}{mc} \right) \frac{\mathbf{S} \cdot \mathbf{L}}{r^3} \\ &= \frac{e^2}{m^2 c^2 r^3} \mathbf{S} \cdot \mathbf{L} \end{aligned} \quad (17.3.15)$$

However, the correct answer is half as big:

$$H_{\text{s.o.}} = \frac{e^2}{2m^2 c^2 r^3} \mathbf{S} \cdot \mathbf{L} \quad (17.3.16)$$

The reason is that the “rest frame of the electron” doesn’t have a fixed velocity relative to the CM of the atom since the motion of the electron is not rectilinear. Thus $\mathcal{H}_{\text{s.o.}}$ deduced in the comoving frame does not directly translate into what must be used in the CM frame. The transformation and the factor of 1/2 were found by Thomas.[‡] In Chapter 20 we will derive Eq. (17.3.16) from the Dirac equation, which has relativistic kinematics built into it. The *Thomas factor* of 1/2 will drop out automatically.

Since $H_{\text{s.o.}}$ involves the spin, we must now reinstate it. Since the states at a given n are degenerate, we must start with a basis that diagonalizes $H_{\text{s.o.}}$. Since we can rewrite $H_{\text{s.o.}}$ as

$$H_{\text{s.o.}} = \frac{e^2}{4m^2 c^2 r^3} [J^2 - L^2 - S^2] \quad (17.3.17)$$

[‡] L. H. Thomas *Nature* **117**, 574 (1926).

$$\langle j', m'; l', 1/2 | H_{\text{s.o.}} | j, m; l, 1/2 \rangle$$

$$= \delta_{j'} \delta_{m'} \delta_{l'} \frac{e^2}{4m^2 c^2} \left\langle \frac{1}{r^3} \right\rangle_{nl} \hbar^2 [j(j+1) - l(l+1) - 3/4] \quad (17.3.18)$$

(Note that two states with the same total jm , but built from different l 's, are orthogonal because of the orthogonality of the spherical harmonics. Thus, for example, at $n=2$, we can build $j=1/2$ either from $l=0$ or $l=1$. The states $|j=1/2, m; 0, 1/2\rangle$ and $|j=1/2, m; 1, 1/2\rangle$ are orthogonal.) Feeding $j=l\pm 1/2$ into Eq. (17.3.18) we get

$$E_{\text{s.o.}}^1 = \frac{\hbar^2 e^2}{4m^2 c^2} \left\langle \frac{1}{r^3} \right\rangle_{nl} \begin{Bmatrix} l \\ -(l+1) \end{Bmatrix} \quad (17.3.19)$$

where the upper and lower values correspond to $j=l\pm 1/2$. Using the result from Exercise 17.3.4

$$\left\langle \frac{1}{r^3} \right\rangle_{nl} = \frac{1}{a_0^3} \frac{1}{n^3 l(l+1/2)(l+1)} \quad (17.3.20)$$

we get

$$E_{\text{s.o.}}^1 = \frac{1}{4} mc^2 \alpha^4 \frac{\begin{Bmatrix} l \\ -(l+1) \end{Bmatrix}}{n^3 l(l+1/2)(l+1)} \quad (17.3.21)$$

This formula has been derived for $l\neq 0$. When $l=0$, $\langle 1/r^3 \rangle$ diverges and $\langle \mathbf{L} \cdot \mathbf{S} \rangle$ vanishes. But if we set $l=0$ in Eq. (17.3.21) we get a finite limit, which in fact happens to give the correct level shift for $l=0$ states. This will be demonstrated when we study the Dirac equation in Chapter 20. The physical origin behind this shift (which is clearly not the spin-orbit interaction) will be discussed then. Since $E_{\text{s.o.}}^1$ and E_T^1 are both α^4 effects, we combine them to get the *total fine-structure energy shift*

$$E_{\text{f.s.}}^1 = E_T^1 + E_{\text{s.o.}}^1 = -\frac{mc^2 \alpha^2}{2n^2} \cdot \frac{\alpha^2}{n} \left(\frac{1}{j+1/2} - \frac{3}{4n} \right) \quad (17.3.22)$$

for both $j=l\pm 1/2$.

The fine-structure formula can be extended to other atoms as well, provided we make the following change in Eq. (17.3.19):

$$\left\langle \frac{e^2}{r^3} \right\rangle \rightarrow \left\langle \frac{1}{r} \frac{dV}{dr} \right\rangle$$

where V is the potential energy of the electron in question. Consider, for example, the $n=4$ states of potassium. We have seen in the prespin treatment that due to penetration and shielding effects the $4s$ level lies below the $4p$ level. If we add spin to this picture, the s state can only become ${}^2S_{1/2}$ while the p state can generate both ${}^2P_{3/2}$ and ${}^2P_{1/2}$. The last two are split by the fine-structure effect[‡] by an amount $(3\hbar^2/4m^2c^2)\langle(1/r)(dV/dr)\rangle$, where V is the potential seen by the $n=4$, $l=1$ electron. In the $4p \rightarrow 4s$ transition, the fine-structure interaction generates two lines in the place of one, with wavelengths 7644.9 Å and 7699.0 Å.

Exercise 17.3.3. Consider the case where H^0 includes the Coulomb plus spin-orbit interaction and H^1 is the effect of a weak magnetic field $\mathbf{B}=B\hat{\mathbf{k}}$. Using the appropriate basis, show that the first-order level shift is related to j_z by

$$E^1 = \left(\frac{eB}{2mc} \right) \left(1 \pm \frac{1}{2l+1} \right) j_z, \quad j = l \pm 1/2$$

Sketch the levels for the $n=2$ level assuming that $E^1 \ll E_{\text{fs}}$.

*Exercise 17.3.4.** We discuss here some tricks for evaluating the expectation values of certain operators in the eigenstates of hydrogen.

(1) Suppose we want $\langle 1/r \rangle_{n\ell m}$. Consider first $\langle \lambda/r \rangle$. We can interpret $\langle \lambda/r \rangle$ as the first-order correction due to a perturbation λ/r . Now this problem can be solved exactly; we just replace e^2 by $e^2 - \lambda$ everywhere. (Why?) So the exact energy, from Eq. (13.1.16) is $E(\lambda) = -(e^2 - \lambda)^2 m / 2n^2 \hbar^2$. The first-order correction is the term linear in λ , that is, $E^1 = me^2 \lambda / n^2 \hbar^2 = \langle \lambda/r \rangle$, from which we get $\langle 1/r \rangle = 1/n^2 a_0$, in agreement with Eq. (13.1.36). For later use, let us observe that as $E(\lambda) = E^0 + E^1 + \dots = E(\lambda=0) + \lambda(dE/d\lambda)_{\lambda=0} + \dots$, one way to extract E^1 from the exact answer is to calculate $\lambda(dE/d\lambda)_{\lambda=0}$.

(2) Consider now $\langle \lambda/r^2 \rangle$. In this case, an exact solution is possible since the perturbation just modifies the centrifugal term as follows:

$$\frac{\hbar^2 l(l+1)}{2mr^2} + \frac{\lambda}{r^2} = \frac{\hbar^2 l'(l'+1)}{2mr^2} \quad (17.3.23)$$

where l' is a function of λ . Now the dependence of E on $l'(\lambda)$ is, from Eq. (13.1.14),

$$E(l') = \frac{-me^4}{2\hbar^2(k+l'+1)^2} = E(\lambda) = E^0 + E^1 + \dots$$

[‡] Actually the splitting at a given l is solely due to the spin-orbit interaction. The kinetic energy correction depends only on l and does not contribute to the splitting between the $P_{3/2}$ and $P_{1/2}$ levels.

Show that

$$\left\langle \frac{\lambda}{r^3} \right\rangle = E^1 = \lambda \frac{dE}{d\lambda} \Big|_{\lambda=0} = \left(\frac{dE}{dl'} \right)_{l'=l} \cdot \left(\frac{dl'}{d\lambda} \right)_{l'=l} \cdot \lambda = \frac{\lambda}{n^3 a_0^3 (l+\frac{1}{2})}$$

Cancelling λ on both sides, we get Eq. (17.3.11).

(3) Consider finally $\langle l/r^3 \rangle$. Since there is no such term in the Coulomb Hamiltonian, we resort to another trick. Consider the radial momentum operator, $p_r = -i\hbar(\partial/\partial r + 1/r)$, in terms of which we may write the radial part of the Hamiltonian

$$\left(\frac{-\hbar^2}{2m} \right) \left(\frac{1}{r^2} \frac{\partial}{\partial r} r^2 \frac{\partial}{\partial r} \right)$$

as $p_r^2/2m$. (Verify this.) Using the fact that $\langle [H, p_r] \rangle = 0$ in the energy eigenstates, and by explicitly evaluating the commutator, show that

$$\left\langle \frac{1}{r^3} \right\rangle = \frac{1}{a_0(l)(l+1)} \left\langle \frac{1}{r^2} \right\rangle$$

combining which with the result from part (2) we get Eq. (17.3.20).

(4) Find the mean kinetic energy using the trick from part (1), this time rescaling the mass. Regain the virial theorem.

Time-Dependent Perturbation Theory

18.1. The Problem

Except for the problem of magnetic resonance, we have avoided studying phenomena governed by a time-dependent Hamiltonian. Whereas in the time-independent case the problem of solving the equation

$$i\hbar|\dot{\psi}\rangle = H|\psi\rangle \quad (18.1.1)$$

reduced to solving the eigenvalue problem of H , in the time-dependent case a frontal attack on the full time-dependent Schrödinger equation becomes inevitable.

In this chapter we consider the perturbative solution to a class of phenomena described by

$$H(t) = H^0 + H^1(t) \quad (18.1.2)$$

where H^0 is a time-independent piece whose eigenvalue problem has been solved and H^1 is a small time-dependent perturbation. For instance, H^0 could be the hydrogen atom Hamiltonian and H^1 the addition due to a weak external electromagnetic field. Whereas in the time-independent case one is interested in the eigenvectors and eigenvalues of H , the typical question one asks here is the following. If at $t=0$ the system is in the eigenstate $|i^0\rangle$ of H^0 , what is the amplitude for it to be in the eigenstate $|f^0\rangle$ ($f \neq i$) at a later time t ? Our goal is to set up a scheme in which the answer may be computed in a perturbation series in powers of H^1 . To zeroth order, the answer to the question raised is clearly zero, for the only effect of H^0 is to multiply $|i^0\rangle$ by a phase factor $\exp(-iE_i^0 t/\hbar)$, which does not alter its orthogonality to $|f^0\rangle$. But as soon as we let H^1 enter the picture, i.e., work to nonzero order, the eigenstates of H^0 cease to be stationary and $|i^0\rangle$ can evolve into a state with a projection along $|f^0\rangle$.

The next section begins with a simple derivation of the first-order transition amplitude for the process $i \rightarrow f$ and is followed by several applications and discussions of special types of perturbations (sudden, adiabatic, periodic, etc.). In Section 3 the expressions for the transition amplitude to any order are derived, following a scheme more abstract than the one used in Section 2. Sections 4 and 5 are concerned with electromagnetic interactions. Section 4 contains a brief summary of relevant concepts from classical electrodynamics, followed by a general discussion of several fine points of the electromagnetic interaction at the classical and quantum levels. It therefore has little to do with perturbation theory. However, it paves the way for the last section, in which first-order perturbation theory is applied to the study of the interaction of atoms with the electromagnetic field. Two illustrative problems are considered, one in which the field is treated classically and the other in which it is treated quantum mechanically.

18.2. First-Order Perturbation Theory

Our problem is to solve Eq. (18.1.1) to first order in H^1 . Since the eigenkets $|n^0\rangle$ of H^0 form a complete basis, we can always expand

$$|\psi(t)\rangle = \sum_n c_n(t) |n^0\rangle \quad (18.2.1)$$

To find $c_n(t)$ given $c_n(0)$ is equivalent to finding $|\psi(t)\rangle$ given $|\psi(0)\rangle$. Now $c_n(t)$ changes with time because of H^0 and H^1 . Had H^1 been absent, we would know

$$c_n(t) = c_n(0) e^{-iE_n^0 t/\hbar} \quad (18.2.2)$$

Let us use this information and write

$$|\psi(t)\rangle = \sum_n d_n(t) e^{-iE_n^0 t/\hbar} |n^0\rangle \quad (18.2.3)$$

If d_n changes with time, it is because of H^1 . So we expect that the time evolution of d_n can be written in a nice power series in H^1 . The equation of motion for $d_f(t)$ is found by operating both sides of Eq. (18.2.3) with $(i\hbar\partial/\partial t - H^0 - H^1)$ to get

$$0 = \sum_n [i\hbar\dot{d}_n - H^1(t)d_n] e^{-iE_n^0 t/\hbar} |n^0\rangle \quad (18.2.4)$$

and then dotting with $\langle f^0 | \exp(iE_f^0 t/\hbar) :$

$$i\hbar\dot{d}_f = \sum_n \langle f^0 | H^1(t) | n^0 \rangle e^{i\omega_n t} d_n(t) \quad (18.2.5a)$$

where

$$\omega_{fi} = \frac{E_f^0 - E_i^0}{\hbar} \quad (18.2.5b)$$

Notice that H^0 has been eliminated in Eq. (18.2.5), which is exact and fully equivalent to Eq. (18.1.1). Let us now consider the case where at $t=0$, the system is in the state $|i^0\rangle$, i.e.,

$$d_n(0) = \delta_{ni} \quad (18.2.6)$$

and ask what $d_f(t)$ is. To *zeroth order*, we ignore the right-hand side of Eq. (18.2.5a) completely, because of the explicit H^1 , and get

$$\dot{d}_f = 0 \quad (18.2.7)$$

in accordance with our expectations. To *first order*, we use the zeroth-order d_n in the right-hand side because H^1 is itself of first order. This gives us the first-order equation

$$\dot{d}_f(t) = \frac{-i}{\hbar} \langle f^0 | H^1(t) | i^0 \rangle e^{i\omega_{fi}t} \quad (18.2.8)$$

the solution to which, with the right initial conditions, is

$$d_f(t) = \delta_{fi} - \frac{i}{\hbar} \int_0^t \langle f^0 | H^1(t') | i^0 \rangle e^{i\omega_{fi}t'} dt' \quad (18.2.9)$$

Since we now know d to first order, we can feed it into the right-hand side of Eq. (18.2.5a) to get an equation for d that is good to second order. Although we can keep going to any desired order in this manner, we stop with the first, since a more compact scheme for calculating transition amplitudes to any desired order will be set up in the next section. At this point we merely note that the first-order calculation is reliable if $|d_f(t)| \ll 1$ ($f \neq i$). If this condition is violated, our calculation becomes internally inconsistent, for we can no longer approximate $d_n(t)$ by δ_{ni} in the right-hand side of Eq. (18.2.5a).

Let us apply our first-order result to a simple problem. Consider a one-dimensional harmonic oscillator in the ground state $|0\rangle$ [‡] of the unperturbed Hamiltonian at $t=-\infty$. Let a perturbation

$$H^1(t) = -e\mathcal{E}X e^{-t^2/\tau^2} \quad (18.2.10)$$

[‡] We shall denote the n th unperturbed state by $|n\rangle$ and not $|n^0\rangle$ in this discussion.

be applied between $t = -\infty$ and $+\infty$. What is the probability that the oscillator is in the state $|n\rangle$ at $t = \infty$? According to Eq. (18.2.9), for $n \neq 0$,

$$d_n(\infty) = \frac{-i}{\hbar} \int_{-\infty}^{\infty} (-e\mathcal{E}) \langle n | X | 0 \rangle e^{-t^2/\tau^2} e^{i\omega t} dt \quad (18.2.11)$$

Since

$$X = \left(\frac{\hbar}{2m\omega} \right)^{1/2} (a + a^\dagger)$$

only $d_1(\infty) \neq 0$. We find that it is (using $a^\dagger |0\rangle = |1\rangle$)

$$\begin{aligned} d_1(\infty) &= \frac{ie\mathcal{E}}{\hbar} \left(\frac{\hbar}{2m\omega} \right)^{1/2} \int_{-\infty}^{\infty} e^{-t^2/\tau^2} e^{i\omega t} dt \\ &= \frac{ie\mathcal{E}}{\hbar} \left(\frac{\hbar}{2m\omega} \right)^{1/2} \cdot (\pi\tau^2)^{1/2} e^{-\omega^2\tau^2/4} \end{aligned} \quad (18.2.12)$$

Thus the probability of the transition $0 \rightarrow 1$ is[‡]

$$P_{0 \rightarrow 1} = |d_1|^2 = \frac{e^2 \mathcal{E}^2 \pi \tau^2}{2m\omega \hbar} e^{-\omega^2\tau^2/2} \quad (18.2.13)$$

This result will be used shortly.

Exercise 18.2.1. Show that if $H^1(t) = -e\mathcal{E}X/[1 + (t/\tau)^2]$, then, to first order,

$$P_{0 \rightarrow 1} = \frac{e^2 \mathcal{E}^2 \pi^2 \tau^2}{2m\omega \hbar} e^{-2\omega\tau}$$

*Exercise 18.2.2.** A hydrogen atom is in the ground state at $t = -\infty$. An electric field $\mathbf{E}(t) = (k\mathcal{E}) e^{-t^2/\tau^2}$ is applied until $t = \infty$. Show that the probability that the atom ends up in any of the $n = 2$ states is, to first order,

$$P(n=2) = \left(\frac{e\mathcal{E}}{\hbar} \right)^2 \left(\frac{2^{15} a_0^2}{3^{10}} \right) \pi \tau^2 e^{-\omega^2\tau^2/2}$$

where $\omega = (E_{2lm} - E_{100})/\hbar$. Does the answer depend on whether or not we incorporate spin in the picture?

We now turn our attention to different types of perturbations.

[‡] Since $d_n(t)$ and $c_n(t)$ differ only by a phase factor, $P(n) = |c_n|^2 = |d_n|^2$.

The Sudden Perturbation

Consider a system whose Hamiltonian changes abruptly over a small time interval ε . What is the change in the state vector as $\varepsilon \rightarrow 0$? We can find the answer without resorting to perturbation theory. Assuming that the change occurred around $t=0$, we get, upon integrating Schrödinger's equation between $t=-\varepsilon/2$ and $\varepsilon/2$,

$$\begin{aligned} |\psi(\varepsilon/2)\rangle - |\psi(-\varepsilon/2)\rangle &= |\psi_{\text{after}}\rangle - |\psi_{\text{before}}\rangle \\ &= \frac{-i}{\hbar} \int_{-\varepsilon/2}^{\varepsilon/2} H(t) |\psi(t)\rangle dt \end{aligned} \quad (18.2.14)$$

Since the integrand on the right-hand side is finite, the integral is of order ε . In the limit $\varepsilon \rightarrow 0$, we get

$$|\psi_{\text{after}}\rangle = |\psi_{\text{before}}\rangle \quad (18.2.15)$$

An instantaneous change in H produces no instantaneous change in $|\psi\rangle$.[‡] Now the limit $\varepsilon \rightarrow 0$ is unphysical. The utility of the above result lies in the fact that it is an excellent approximation if H changes over a time that is very small compared to the natural time scale of the system. The latter may be estimated semiclassically; several examples follow in a moment. For the present, let us consider the case of an oscillator to which is applied the perturbation in Eq. (18.2.10). It is clear that whatever be the time scale of this system, the change in the state vector must vanish as τ , the width of the Gaussian pulse, vanishes. This means in particular that the system initially in the ground state must remain there after the pulse, i.e., the $0 \rightarrow 1$ transition probability must vanish. This being an exact result, we expect that if the transition probability is calculated perturbatively, it must vanish to any given order. (This is like saying that if an analytic function vanishes identically, then so does every term in its Taylor expansion.) Turning to the first-order probability for $0 \rightarrow 1$ in Eq. (18.2.13), we see that indeed it vanishes as τ tends to zero.

A more realistic problem, where ε is fixed, involves a $1s$ electron bound to a nucleus of charge Z which undergoes β decay by emitting a *relativistic* electron and changing its charge to $(Z+1)$. The time the emitted electron takes to get out of the $n=1$ shell is

$$\tau \simeq a_0/Zc \quad (18.2.16)$$

whereas the characteristic time for the $1s$ electron is

$$T \simeq \frac{\text{size of state}}{\text{velocity of } e^-} \simeq \frac{a_0}{Z} / Z\alpha c = \frac{a_0}{Z^2 \alpha c} \quad (18.2.17)$$

so that

$$\tau/T = Z\alpha$$

[‡] We are assuming H is finite in the integral $(-\varepsilon/2, \varepsilon/2)$. If it has a delta function spike, it can produce a change in $|\psi\rangle$, see Exercise 18.2.6.

For Z small, we may apply the sudden approximation and conclude that the state of the atomic electron is the same just before and just after β decay. Of course, this state is not an eigenstate of the charge $(Z+1)$ ion, but rather a superposition of such states (see Exercise 18.2.4).

*Exercise 18.2.3.** Consider a particle in the ground state of a box of length L . Argue on semiclassical grounds that the natural time period associated with it is $T \approx mL^2/\hbar\pi$. If the box expands symmetrically to double its size in time $\tau \ll T$ what is the probability of catching the particle in the ground state of the new box? (See Exercise (5.2.1).)

*Exercise 18.2.4.** In the β decay H^3 (two neutrons + one proton in the nucleus) $\rightarrow (\text{He}^3)^+$ (two protons + one neutron in the nucleus), the emitted electron has a kinetic energy of 16 keV. Argue that the sudden approximation may be used to describe the response of an electron that is initially in the $1s$ state of H^3 . Show that the amplitude for it to be in the ground state of $(\text{He}^3)^+$ is $16(2)^{1/2}/27$. What is the probability for it to be in the state

$$|n=16, l=3, m=0\rangle \text{ of } (\text{He}^3)^+?$$

Exercise 18.2.5. An oscillator is in the ground state of $H = H^0 + H^1$, where the time-independent perturbation H^1 is the linear potential $(-fx)$. If at $t=0$, H^1 is abruptly turned off, show that the probability that the system is in the n th eigenstate of H^0 is given by the Poisson distribution

$$P(n) = \frac{e^{-\lambda}\lambda^n}{n!}, \quad \text{where} \quad \lambda = \frac{f^2}{2m\omega^3\hbar}$$

Hint: Use the formula

$$\exp[A + B] = \exp[A] \exp[B] \exp[-\frac{1}{2}[A, B]]$$

where $[A, B]$ is a c number.

*Exercise 18.2.6.** Consider a system subject to a perturbation $H^1(t) = H^1\delta(t)$. Show that if at $t=0^-$ the system is in the state $|i^0\rangle$, the amplitude to be in a state $|f^0\rangle$ at $t=0^+$ is, to first order,

$$d_f = \frac{-i}{\hbar} \langle f^0 | H^1 | i^0 \rangle \quad (f \neq i)$$

Notice that (1) the state of the system *does* change instantaneously; (2) Even though the perturbation is “infinite” at $t=0$, we can still use first-order perturbation theory if the “area under it” is small enough.

The Adiabatic Perturbation

We now turn to the other extreme and consider a system whose Hamiltonian $H(t)$ changes very slowly from $H(0)$ to $H(\tau)$ in a time τ . If the system starts out at $t=0$ in an eigenstate $|n(0)\rangle$ of $H(0)$, where will it end at time τ ? The *adiabatic theorem* asserts that if the rate of change of H is slow enough, the system will end

up in the corresponding eigenket $|n(\tau)\rangle$ of $H(\tau)$.[†] Rather than derive the theorem and the precise definition of “slow enough” we consider a few illustrative examples.

Consider a particle in a box of length $L(0)$. If the box expands slowly to a length $L(\tau)$, the theorem tells us that a particle that was in the n th state of the box of length $L(0)$ will now be in the n th state of the box of length $L(\tau)$. But how slow is slow enough?

There are two ways to estimate this. The first is a semiclassical method and goes as follows. The momentum of the particle is of the order (dropping factors of order unity like π , n , etc.)

$$p \simeq \frac{\hbar}{L} \quad (18.2.18)$$

and the time it takes to finish one full oscillation is of the order

$$T \simeq \frac{L}{v} = \frac{mL}{p} \simeq \frac{mL^2}{\hbar} \quad (18.2.19)$$

We can say the expansion or contraction is slow if the fractional change in the length of the box per cycle is much smaller than unity:

$$\frac{|\Delta L|_{\text{per cycle}}}{L} \simeq \frac{|dL/dt|mL^2/\hbar}{L} = \frac{mL}{\hbar} \left| \frac{dL}{dt} \right| \ll 1 \quad (18.2.20)$$

This can also be written as

$$\frac{v_{\text{walls}}}{v_{\text{particle}}} \ll 1 \quad (18.2.21)$$

The second approach is less intuitive[§] and it estimates T as

$$T \sim \frac{1}{\omega_{\min}} \quad (18.2.22)$$

[†] This is again a result that is true to any given order in perturbation theory. We shall exploit this fact in a moment.

[§] The logic behind this approach and its superiority over the intuitive one will become apparent shortly in an example where we recover the results of time-independent perturbation theory from the time-dependent one.

where ω_{\min} is the smallest of the transition frequencies between the initial state i and any *accessible* final state f [‡]; it is the smallest of

$$\omega_{fi} = \frac{E_f^0 - E_i^0}{\hbar} \quad (18.2.23)$$

In the present case, since $E_n^0 = (n^2 \hbar^2 \pi^2 / 2mL^2)$, energy differences are of the order \hbar^2 / mL^2 and

$$T \sim \frac{1}{\omega_{\min}} \simeq \frac{mL^2}{\hbar} \quad (18.2.24)$$

which coincides with Eq. (18.2.19). This is not surprising, for we can also write T in Eq. (18.2.19) as

$$T \simeq \frac{mL^2}{\hbar} \simeq \frac{1}{E_i^0/\hbar} \sim \frac{1}{\omega_i} \quad (18.2.25)$$

Thus T in Eq. (18.2.19) is $\sim \hbar/E_i^0$, while T in eq. (18.2.24) is $\sim \hbar/|E_j^0 - E_i^0|_{\min}$. Since the energy levels of a quantum system are all of the same order of magnitude (say a Rydberg or $\hbar\omega$), energies and energy differences are of the same order of magnitude and the two estimates for T are equivalent, unless *the levels are degenerate or nearly so*. In this case, it is $T \sim 1/\omega_{\min}$ that is to be trusted, for it exposes the instability of a degenerate or nearly degenerate system. An explicit example that follows later will illustrate this.

Let us consider one more example of the adiabatic theorem, an oscillator subject to the perturbation

$$H^1(t) = -e\varepsilon X e^{-t^2/\tau^2} \quad (18.2.26)$$

between $-\infty \leq t \leq \infty$. We expect that if τ , which measures the time over which H^1 grows from 0 to its peak, tends to infinity, the change in the system will be adiabatic. Thus, if a system starts in the ground state of $H(-\infty) = H^0$ at $t = -\infty$, it will end up in the ground state of $H(\infty) = H(-\infty) = H^0$. Our first-order formula, Eq. (18.2.13), for $P_{0 \rightarrow 1}$ conforms with this expectation and vanishes exponentially as $\omega\tau \rightarrow \infty$. Our formula also tells us what large τ means: it means

$$\omega\tau \gg 1, \quad \tau \gg 1/\omega \quad (18.2.27)$$

This is what we would expect from the semiclassical estimate or the estimate $T \sim 1/\omega_{\min}$ and the condition $\tau \gg T$.

The adiabatic theorem suggests a way of recovering the results of time-independent perturbation theory from time-dependent theory. Consider a Hamiltonian $H(t)$

[‡] This is a state for which $\langle f^0 | H^1 | i^0 \rangle \neq 0$.

which changes continuously from H^0 at $t=-\infty$ to H^0+H^1 at $t=0$:

$$H(t) = H^0 + e^{t/\tau} H^1, \quad -\infty \leq t \leq 0 \quad (18.2.28)$$

As τ , the rise time of the exponential, goes to infinity, the adiabatic theorem assures us that an eigenstate $|n^0\rangle$ of H^0 at $t=-\infty$ will evolve into the eigenstate $|n\rangle$ of H at $t=0$. If we calculate the state at $t=0$ to a given order in time-dependent theory and let $\tau \rightarrow \infty$, we should get the time-independent formula for the state $|n\rangle$ to that order. To first order, we know that the projection of the state at $t=0$ along $|m^0\rangle$ ($m \neq n$) is

$$\begin{aligned} d_m(0) &= \frac{-i}{\hbar} \int_{-\infty}^0 \langle m^0 | H^1 | n^0 \rangle e^{t/\tau} e^{i\omega_{mn} t} dt \\ &= \frac{(-i/\hbar) \langle m^0 | H^1 | n^0 \rangle}{1/\tau + i\omega_{mn}} \end{aligned} \quad (18.2.29)$$

If we now let $\tau \rightarrow \infty$, we regain the familiar result

$$\langle m^0 | n \rangle = \frac{\langle m^0 | H^1 | n^0 \rangle}{E_n^0 - E_m^0} \quad (18.2.30)$$

In practice, $\tau \rightarrow \infty$ is replaced by some large τ . Equation (18.2.29) tells us what large τ means: it is defined by

$$|1/\tau| \ll |\omega_{\min}|$$

or

$$\tau \gg 1/\omega_{\min} \quad (18.2.31)$$

Thus we see that $T \simeq 1/\omega_{\min}$ is indeed the reliable measure of the natural time scale of the system. In particular, if the system is degenerate (or nearly so), $T \rightarrow \infty$ and it becomes impossible, in practice, to change the state of the system adiabatically.

Let us wind up the discussion on the adiabatic approximation by observing its similarity to the WKB approximation. The former tells us that if the Hamiltonian changes in time from H^0 to H^0+H^1 , the eigenstate $|n^0\rangle$ evolves smoothly into its counterpart $|n\rangle$ in the limit $\tau/T \rightarrow \infty$, where τ is the duration over which the Hamiltonian changes and T is the natural time scale for the system. The latter tells us that if the potential changes in space from V^0 to V^1 , a plane wave of momentum $p^0 = [2m(E - V^0)]^{1/2}$ evolves smoothly into a plane wave of momentum $p^1 = [2m(E - V^1)]^{1/2}$ in the limit $L/\lambda \rightarrow \infty$, where L is the length over which V changes and $\lambda = 2\pi\hbar/p$ is natural length scale for the system.

We shall return to adiabatic evolutions in Chapter 21.

The Periodic Perturbation

Consider a system that is subject to a periodic perturbation, say an atom placed between the plates of a condenser connected to an alternating current (ac) source or in the way of a monochromatic light beam. While in reality these perturbations vary as sines and cosines, we consider here the case

$$H^1(t) = H^1 e^{-i\omega t} \quad (18.2.32)$$

Which is easier to handle mathematically. The sines and cosines can be handled by expressing them in terms of exponentials.

Let us say the system comes into contact with this perturbation at $t=0$. The amplitude for transition from $|i^0\rangle$ to $|f^0\rangle$ in time t ($i \neq f$) is

$$d_f(t) = \left(\frac{-i}{\hbar} \right) \int_0^t \langle f^0 | H^1 | i^0 \rangle e^{i(\omega_{fi} - \omega)t'} dt' \quad (18.2.33)$$

$$= \frac{-i}{\hbar} \langle f^0 | H^1 | i^0 \rangle \frac{e^{i(\omega_{fi} - \omega)t} - 1}{i(\omega_{fi} - \omega)} \quad (18.2.34)$$

The probability for the transition $i \rightarrow f$ is

$$P_{i \rightarrow f} = |d_f|^2 = \frac{1}{\hbar^2} |\langle f^0 | H^1 | i^0 \rangle|^2 \left\{ \frac{\sin[(\omega_{fi} - \omega)t/2]}{(\omega_{fi} - \omega)^{\frac{1}{2}} t} \right\}^2 \quad (18.2.35)$$

Since the function $(\sin^2 x)/x^2$ is peaked at the origin and has a width $\Delta x \approx \pi$, we find that the system likes to go to states f such that

$$|(\omega_{fi} - \omega)t/2| \lesssim \pi$$

or

$$E_f^0 t = (E_i^0 t + \hbar\omega t) \pm 2\hbar\pi$$

or

$$E_f^0 - E_i^0 = \hbar\omega \pm \frac{2\hbar\pi}{t} = \hbar\omega \left(1 \pm \frac{2\pi}{\omega t} \right) \quad (18.2.36)$$

For small t , the system shows no particular preference for the level with $E_f^0 = E_i^0 + \hbar\omega$. Only when $\omega t \gg 2\pi$ does it begin to favor $E_f^0 = E_i^0 + \hbar\omega$. The reason is simple. You and I know the perturbation has a frequency ω , say, because we set the dial on the ac source or tuned our laser to frequency ω . But the system goes by what it knows, starting from the time it made contact with the perturbation. In the beginning, it will not even know it is dealing with a periodic perturbation; it must wait a few cycles to get the message. Thus it can become selective only after a few cycles, i.e., after $\omega t \gg 2\pi$. What does it do meanwhile? It Fourier-analyzes the pulse

into its frequency components and its transition amplitude to a state with $E_f^0 = E_i^0 + \hbar\omega_{fi}$ is proportional to the Fourier component at $\omega = \omega_{fi}$. The t' integral in Eq. (18.2.33) is precisely this Fourier transform.[‡]

What happens if we wait a long time? To find out, we consider the case of a system exposed to the perturbation from $t = -T/2$ to $T/2$ and let $T \rightarrow \infty$. Equation (18.2.33) becomes

$$d_f = \lim_{T \rightarrow \infty} \frac{-i}{\hbar} \int_{-T/2}^{T/2} H_{fi}^1 e^{i(\omega_{fi} - \omega)t'} dt' \quad (18.2.37)$$

$$= \frac{-2\pi i}{\hbar} H_{fi}^1 \delta(\omega_{fi} - \omega) \quad (18.2.38)$$

and

$$P_{i \rightarrow f} = \frac{4\pi^2}{\hbar^2} |H_{fi}^1|^2 \delta(\omega_{fi} - \omega) \delta(\omega_{fi} - \omega) \quad (18.2.39)$$

We handle the product of δ functions as follows:

$$\delta\delta = \lim_{T \rightarrow \infty} \delta(\omega_{fi} - \omega) \frac{1}{2\pi} \int_{-T/2}^{T/2} e^{i(\omega_{fi} - \omega)t} dt \quad (18.2.40)$$

Since the δ function in front of the integral vanishes unless $\omega_{fi} = \omega$, we may set $\omega_{fi} = \omega$ in the integral to obtain

$$\delta\delta = \delta(\omega_{fi} - \omega) \lim_{T \rightarrow \infty} \frac{T}{2\pi} \quad (18.2.41)$$

Feeding this into Eq. (18.2.39) for $P_{i \rightarrow f}$, and dividing by T , we get the *average transition rate*:

$$R_{i \rightarrow f} = \frac{P_{i \rightarrow f}}{T} = \frac{2\pi}{\hbar} |\langle f^0 | H^1 | i^0 \rangle|^2 \delta(E_f^0 - E_i^0 - \hbar\omega) \quad (18.2.42)$$

This is called *Fermi's golden rule* and has numerous applications, some of which will be discussed later in this chapter and in the next chapter. You may be worried about the δ function in $R_{i \rightarrow f}$ and in particular whether first-order perturbation theory is to be trusted when the rate comes out infinite! As we will see, in all practical applications the δ function will get integrated over for one reason or another. The validity of the first-order formula will then depend only on the area under the δ function. (Recall Exercise 18.2.6.)

[‡] The inability of a system to assign a definite frequency to an external perturbation until many cycles have elapsed is a purely classical effect. The quantum mechanics comes in when we relate frequency to energy.

18.3. Higher Orders in Perturbation Theory‡

In Section 18.2 we derived a formula for the transition amplitude from $|i^0\rangle$ to $|f^0\rangle$ to first order in perturbation theory. The procedure for going to higher orders was indicated but not pursued. We address that problem here, using a more abstract formalism, desirable for its compactness and the insight it gives us into the anatomy of the perturbation series.

The basic idea behind the approach is the same as in Section 18.2: we want to isolate the time evolution generated by H^1 , for H^0 by itself causes no transitions between its own eigenstates $|i^0\rangle$ and $|f^0\rangle$. To do this, we must get acquainted with other *equivalent* descriptions of quantum dynamics besides the one we have used so far. The description we are familiar with is called the *Schrödinger picture*. In this picture the state of the particle is described by a vector $|\psi_S(t)\rangle$. (We append a subscript S to all quantities that appear in the Schrödinger picture to distinguish them from their counterparts in other pictures.) The physics is contained in the inner products $\langle\omega_S|\psi_S(t)\rangle$ which give the probabilities

$$P(\omega, t) = |\langle\omega_S|\psi_S(t)\rangle|^2 \quad (18.3.1)$$

for obtaining the result ω when Ω is measured. Here $|\omega_S\rangle$ is the normalized eigenket of the operator $\Omega_S(X_S, P_S)$ with eigenvalue ω . Since X_S and P_S are time independent so are Ω_S and $|\omega_S\rangle$. Thus the physics is contained in the dot product of the moving ket $|\psi_S(t)\rangle$ with the stationary kets $|\omega_S\rangle$.

The time evolution of $|\psi_S(t)\rangle$ is given in general by

$$i\hbar \frac{d}{dt} |\psi_S(t)\rangle = H_S |\psi_S(t)\rangle \quad (18.3.2a)$$

and in our problem by

$$i\hbar \frac{d}{dt} |\psi_S(t)\rangle = [H_S^0 + H_S^1(t)] |\psi_S(t)\rangle \quad (18.3.2b)$$

The expectation values change according to

$$i\hbar \frac{d}{dt} \langle\Omega_S\rangle = \langle[\Omega_S, H_S]\rangle \quad (18.3.3)$$

If we define a propagator $U_S(t, t_0)$ by

$$|\psi_S(t)\rangle = U_S(t, t_0) |\psi_S(t_0)\rangle \quad (18.3.4)$$

‡ This section may be skimmed through by a reader pressured for time.

it follows from Eq. (18.3.2) [because $|\psi_S(t_0)\rangle$ is arbitrary] that

$$i\hbar \frac{dU_S}{dt} = H_S U_S \quad (18.3.5)$$

Here are some formulas (true for all propagators U) that will be useful in what follows (recall Eq. (4.3.16)):

$$\begin{aligned} U^\dagger U &= I \\ U(t_3, t_2)U(t_2, t_1) &= U(t_3, t_1) \\ U(t_1, t_1) &= I \\ U^\dagger(t_1, t_2) &= U(t_2, t_1) \end{aligned} \quad (18.3.6)$$

The Interaction Picture

Since $U_S(t, t_0)$ is a unitary operator, which is the generalization of the rotation operator to complex spaces, we may describe the time evolution of state vectors as “rotations” in Hilbert space.[‡] The rotation is generated by $U_S(t, t_0)$ or equivalently, by $H_S(t) = H_S^0 + H_S^1(t)$. Imagine for a moment that H_S^1 is absent. Then the rotation will be generated by $U_S^0(t)$, which obeys

$$i\hbar \frac{dU_S^0}{dt} = H_S^0 U_S^0 \quad (18.3.7)$$

the formal solution to which is $U_S^0(t, t_0) = e^{-iH_S^0(t-t_0)/\hbar}$. If $H_S^1(t)$ is put back in, both H_S^0 and $H_S^1(t)$ jointly produce the rotation U_S .

These pictorial arguments suggest a way to freeze out the time evolution generated by H_S^0 . Suppose we switch to a frame that rotates at a rate that U_S^0 (or H_S^0) by itself generates. In this frame the state vector moves because $H_S^1 \neq 0$. Let us verify this conjecture. To neutralize the rotation induced by U_S^0 , i.e., to see things from the rotating frame, we multiply $|\psi_S(t)\rangle$ by $(U_S^0)^\dagger$ to get

$$|\psi_I(t)\rangle = [U_S^0(t, t_0)]^\dagger |\psi_S(t)\rangle \quad (18.3.8a)$$

The ket $|\psi_I(t)\rangle$ is the state vector in the rotating frame, or in the *interaction picture*. If we set $t = t_0$ in the above equation, we did

$$|\psi_I(t_0)\rangle = |\psi_S(t_0)\rangle \quad (18.3.8b)$$

[‡] In this section we use the word “rotation” in this generalized sense, and not in the sense of a spatial rotation.

i.e., the interaction and Schrödinger kets coincide at $t = t_0$, which is that instant we switch to the moving frame. The time evolution of $|\psi_I(t)\rangle$ is as follows‡:

$$\begin{aligned} i\hbar \frac{d}{dt} |\psi_I(t)\rangle &= i\hbar \frac{dU_S^{0\dagger}}{dt} |\psi_S\rangle + U_S^{0\dagger} i\hbar \frac{d|\psi_S\rangle}{dt} \\ &= -U_S^{\dagger} H_S^0 |\psi_S\rangle + U_S^{0\dagger} (H_S^0 + H_S^1) |\psi_S\rangle \\ &= U_S^{0\dagger} H_S^1 |\psi_S\rangle \\ &= U_S^{0\dagger} H_S^1 U_S^0 U_S^{0\dagger} |\psi_S\rangle \\ &= U_S^{0\dagger} H_S^1 U_S^0 |\psi_I(t)\rangle \end{aligned}$$

Now

$$(U_S^0)^{\dagger} H_S^1(t) U_S^0 = H_I^1(t) \quad (18.3.9)$$

is the perturbing Hamiltonian as seen in the rotating frame. So we can write

$$i\hbar \frac{d}{dt} |\psi_I(t)\rangle = H_I^1(t) |\psi_I(t)\rangle \quad (18.3.10)$$

So, as we anticipated, the time evolution of the state vector in the interaction picture is determined by the perturbing Hamiltonian, H_I^1 . Despite the fact that the state vector now rotates at a different rate, the physical predictions are the same as in the Schrödinger picture. This is because $P(\omega, t)$ depends only on the inner product between the state vector and the eigenket of Ω with eigenvalue ω , and the inner product between two vectors is unaffected by going to a rotating frame. However, both the state vector and the eigenket appear different in the interaction picture. Just as

$$|\psi_S(t)\rangle \rightarrow U_S^{0\dagger}(t, t_0) |\psi_S(t)\rangle = |\psi_I(t)\rangle$$

so does

$$|\omega_S\rangle \rightarrow U_S^{0\dagger}(t, t_0) |\omega_S\rangle = |\omega_I(t)\rangle \quad (18.3.11)$$

However,

$$\langle \omega_S | \psi_S(t) \rangle = \langle \omega_I(t) | \psi_I(t) \rangle \quad (18.3.12)$$

The time-dependent ket $|\omega_I(t)\rangle$ is just the eigenket of the time-dependent operator

$$\Omega_I(t) = U_S^{0\dagger} \Omega_S U_S^0 \quad (18.3.13)$$

which is just Ω as seen in the rotating frame:

$$\Omega_I(t) |\omega_I(t)\rangle = U_S^{0\dagger} \Omega_S U_S^0 U_S^{0\dagger} |\omega_S\rangle = U_S^{0\dagger} \Omega_S |\omega_S\rangle = \omega |\omega_I(t)\rangle \quad (18.3.14)$$

‡ Whenever the argument of any U is suppressed, it may be assumed to be (t, t_0) .

The time dependence of Ω_I may be calculated by combining Eq. (18.3.13), which defines it, and Eq. (18.3.7), which gives the time evolution of U_S^0 :

$$\begin{aligned} i\hbar \frac{d\Omega_I}{dt} &= i\hbar \frac{dU_S^{0\dagger}}{dt} \Omega_S U_S^0 + U_S^{0\dagger} \Omega_S i\hbar \frac{dU_S^0}{dt} \\ &= U_S^{0\dagger} [\Omega_S, H_S^0] U_S^0 = [\Omega_I, H_I^0] \end{aligned} \quad (18.3.15)$$

In the interaction picture, the operators evolve in response to the unperturbed Hamiltonian H_I^0 .[‡] Whereas in the Schrödinger picture, the entire burden of time evolution lies with the state vectors, in this picture it is shared by the state vectors and the operators (in such a way that the physics is the same).

Let us now address the original problem, of obtaining a perturbation series for the transition amplitude. We define a propagator $U_I(t, t_0)$ in the interaction picture:

$$|\psi_I(t)\rangle = U_I(t, t_0) |\psi_I(t_0)\rangle \quad (18.3.16)$$

which, because of Eq. (18.3.10), obeys

$$i\hbar \frac{dU_I}{dt} = H_I^1 U_I \quad (18.3.17)$$

Once we find $U_I(t)$, we can always go back to $U_S(t)$ by using

$$U_S(t, t_0) = U_S^0(t, t_0) U_I(t, t_0) \quad (18.3.18)$$

which follows from Eqs. (18.3.8) and (18.3.16).

Since H_I^1 depends on time, the solution to Eq. (18.3.17) is not $U_I = \exp(-iH_I^1(t-t_0)/\hbar)$. A formal solution, with the right initial condition, is

$$U_I(t, t_0) = I - \frac{i}{\hbar} \int_{t_0}^t H_I(t') U_I(t', t_0) dt' \quad (18.3.19)$$

as may be readily verified by feeding it into the differential equation. Since U_I occurs on both sides, this is not really a solution, but an *integral equation*, equivalent to the differential equation (18.3.17), with the right initial condition built in. So we have not got anywhere in terms of the exact solution. But the integral equation provides a nice way to carry out the perturbation expansion. Suppose we want U_I to zeroth order. We drop anything with an H_I^1 in Eq. (18.3.19):

$$U_I(t, t_0) = I + O(H_I^1) \quad (18.3.20)$$

[‡] Actually, $H_I^0 = U_S^{0\dagger} H_S^0 U_S^0 = H_S^0$ since $[H_S^0, U_S^0] = 0$ in this problem.

This is to be expected, for if we ignore H_I^1 , the state vectors do not move in the interaction picture.

To first order, we can keep one only power of H_I^1 . So we use the zeroth-order value for U_I in the right-hand side of Eq. (18.3.20) to get

$$U_I(t, t_0) = I - \frac{i}{\hbar} \int_{t_0}^t H_I^1(t') dt' + O(H_I^2) \quad (18.3.21)$$

Before going to the next order, let us compare this with Eq. (18.2.9) for the transition amplitude $d_f(t)$, computed to first order. Recall the definition of $d_f(t)$: it is the projection along $\langle f_S^0 | \exp[iE_f^0(t-t_0)/\hbar] | i_S^0 \rangle$ at time t , of a state that was initially (at $t=t_0$) $| i_S^0 \rangle$ [‡]:

$$d_f(t) = \langle f_S^0 | e^{iE_f^0(t-t_0)/\hbar} U_S(t, t_0) | i_S^0 \rangle \quad (18.3.22)$$

$$\begin{aligned} &= \langle f_S^0 | U_S^\dagger(t, t_0) U_S(t, t_0) | i_S^0 \rangle \\ &= \langle f_S^0 | U_I(t, t_0) | i_S^0 \rangle \end{aligned} \quad (18.3.23)$$

If we feed into this our first-order propagator given in Eq. (18.3.21), we get

$$\begin{aligned} d_f(t) &= \langle f_S^0 | U_I(t, t_0) | i_S^0 \rangle \\ &= \delta_{fi} - \frac{i}{\hbar} \int_{t_0}^t \langle f_S^0 | H_I^1(t') | i_S^0 \rangle dt' \\ &= \delta_{fi} - \frac{i}{\hbar} \int_{t_0}^t \langle f_S^0 | U_S^{0\dagger}(t', t_0) H_S^1 U_S^0(t', t_0) | i_S^0 \rangle dt' \\ &= \delta_{fi} - \frac{i}{\hbar} \int_{t_0}^t (H_S^1)_{fi} e^{i\omega_{fi}(t'-t_0)} dt' \end{aligned} \quad (18.3.24)$$

which agrees with Eq. (18.2.9) if we set $t_0=0$.

Let us now turn to higher orders. By repeatedly feeding into the right-hand side of Eq. (18.3.19) the result for U_I to a known order, we can get U_I to higher orders:

$$\begin{aligned} U_I(t, t_0) &= I - \frac{i}{\hbar} \int_{t_0}^t H_I^1(t') dt' + (-i/\hbar)^2 \int_{t_0}^t \int_{t_0}^{t'} H_I^1(t') H_I^1(t'') dt' dt'' \\ &\quad + (-i/\hbar)^3 \int_{t_0}^t \int_{t_0}^{t'} \int_{t_0}^{t''} H_I^1(t') H_I^1(t'') H_I^1(t''') dt' dt'' dt''' + \dots \end{aligned} \quad (18.3.25)$$

[‡] We have set $t_0=0$ in Section 18.2.

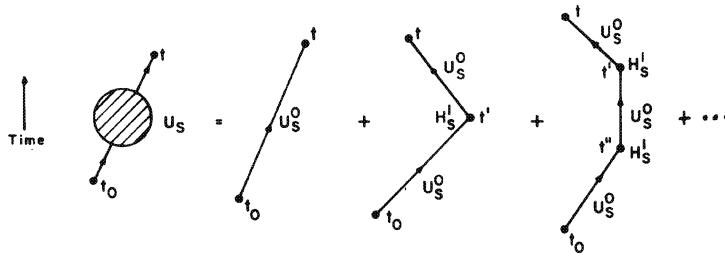


Figure 18.1. A pictorial representation of the perturbation series. The hatched circle represents the full propagator between times t_0 and t . The hatched circle is a sum of many terms, each of which corresponds to a different number of interactions with the perturbation, H_S^1 . Between such interactions, the particle evolves in response to H_S^0 , i.e., is propagated by U_S^0 .

Premultiplying by $U_S^0(t, t_0)$ and expressing H_S^1 in terms of H_S^1 , we get the Schrödinger picture propagator

$$\begin{aligned} U_S(t, t_0) &= U_S^0(t, t_0) - \frac{i}{\hbar} \int_{t_0}^t U_S^0(t, t_0) U_S^{0\dagger}(t', t_0) H_S^1 U_S^0(t', t_0) dt' \\ &\quad + (-i/\hbar)^2 \int_{t_0}^t \int_{t_0}^{t'} U_S^0(t, t_0) U_S^{0\dagger}(t', t_0) H_S^1 U_S^0(t', t_0) U_S^{0\dagger}(t'', t_0) \\ &\quad \times H_S^1 U_S^0(t'', t_0) dt' dt'' + \dots \end{aligned} \quad (18.3.26)$$

$$\begin{aligned} U_S(t, t_0) &= U_S^0(t, t_0) - \frac{i}{\hbar} \int_{t_0}^t U_S^0(t, t') H_S^1 U_S^0(t', t_0) dt' \\ &\quad + (-i/\hbar)^2 \int_{t_0}^t \int_{t_0}^{t'} U_S^0(t, t') H_S^1 U_S^0(t', t'') H_S^1 U_S^0(t'', t_0) dt' dt'' + \dots \end{aligned}$$

The above series could be described by the following words. On the left-hand side we have the complete Schrödinger picture propagator and on the right-hand side a series expansion for it. The first term says the system evolves from t_0 to t in response to just U_S^0 , i.e., in response to H_S^0 . The second term, if we read it from right to left (imagine it acting on some initial state) says the following: the system evolves from t_0 to t' : in response to U_S^0 , there it interacts once with the perturbation and thereafter responds to U_S^0 alone until time t . The integral over t' sums over the possible times at which the single encounter with H_S^1 could have taken place. The meaning of the next and higher terms is obvious. These are represented schematically in Fig. 18.1.

If we consider specifically the transition from the state $|i^0\rangle$ to $|f^0\rangle$ (we drop the subscript S everywhere) we get

$$\begin{aligned} \langle f^0 | U(t, t_0) | i^0 \rangle &= \delta_{fi} e^{-iE_f^0(t-t_0)/\hbar} + \frac{-i}{\hbar} \int_{t_0}^t e^{-iE_f^0(t-t')/\hbar} \langle f^0 | H^1 | i^0 \rangle e^{-iE_i^0(t'-t_0)/\hbar} dt' \\ &\quad + \left(\frac{-i}{\hbar}\right)^2 \int_{t_0}^t \int_{t_0}^{t'} \sum_n e^{-iE_f^0(t-t')/\hbar} \langle f^0 | H^1 | n^0 \rangle \\ &\quad \times e^{-iE_n^0(t'-t'')/\hbar} \langle n^0 | H^1 | i^0 \rangle e^{-iE_i^0(t''-t_0)/\hbar} dt' dt'' + \dots \end{aligned} \quad (18.3.27)$$

upon introducing a complete set of eigenstates of H^0 in the second-order term. The meaning of the first term is obvious. The second (reading right to left) says that between t_0 and t' the eigenstate $|i^0\rangle$ picks up just a phase (i.e., responds to H_S^0 alone). At t' it meets the perturbation, which has an amplitude $\langle f^0|H^1|i^0\rangle$ of converting it to the state $|f^0\rangle$. Thereafter it evolves as the eigenstate $|f^0\rangle$ until time t . The total amplitude to end up in $|f^0\rangle$ is found by integrating over the times at which the conversion could have taken place. Thus the first-order transition corresponds to a one-step process $i \rightarrow f$. At the second order, we see a sum over a complete set of states $|n^0\rangle$. It means the system can go from $|i^0\rangle$ to $|f^0\rangle$ via any *intermediate or virtual* state $|n^0\rangle$ that H^1 can knock $|i^0\rangle$ into. Thus the second-order amplitude describes a two-step process, $i \rightarrow n \rightarrow f$. Higher-order amplitudes have a similar interpretation.

The Heisenberg Picture

It should be evident that there exist not just two, but an infinite number of pictures, for one can go to frames rotating at various speeds. Not all these are worthy of study, however. We conclude this section with one picture that is very important, namely, the *Heisenberg picture*. In this picture, one freezes out the *complete time dependence* of the state vector. The Heisenberg state vector is

$$|\psi_H(t)\rangle = U_S^\dagger(t, t_0)|\psi_S(t)\rangle = |\psi_S(t_0)\rangle \quad (18.3.28)$$

The operators in this picture are

$$\Omega_H(t) = U_S^\dagger \Omega_S U_S \quad (18.3.29)$$

and obey

$$i\hbar \frac{d\Omega_H}{dt} = [\Omega_H, H_H] \quad (18.3.30)$$

*Exercise 18.3.1.** Derive Eq. (18.3.30).

Thus in the Heisenberg picture, the state vectors are fixed and the operators carry the full time dependence. (Since the interaction picture lies between this Heisenberg picture and the Schrödinger picture, in that the operators and the state vectors share the time dependence, it is also called the *intermediate picture*. Another name for it is the *Dirac picture*.)

Notice the similarity between Eq. (18.3.30) and the classical equation

$$\frac{d\omega}{dt} = \{\omega, \mathcal{H}\} \quad (18.3.31)$$

The Heisenberg picture displays the close formal similarity between quantum and classical mechanics: to every classical variable ω there is a quantum operator Ω_H , which obeys similar equations; all we need to do is make the usual substitution $\omega \rightarrow \Omega$, $\{\cdot\} \rightarrow (-i/\hbar)[\cdot, \cdot]$. The similarity between Eqs. (18.3.30) and (18.3.31) is even

more striking if we actually evaluate the commutators and Poisson bracket (PB). Consider, for example, the problem of the oscillator for which

$$H_H = \frac{P_H^2}{2m} + \frac{1}{2} m\omega^2 X_H^2 \quad (18.3.32)$$

Since X_H, P_H are obtained from X_S, P_S by a unitary transformation, they satisfy the same commutation rules

$$[X_H(t), P_H(t)] = U_S^\dagger(t, t_0)[X_S, P_S]U_S(t, t_0) = U_S^\dagger i\hbar I U_S = i\hbar I \quad (18.3.33)$$

Note that the time arguments must be equal in X_H and P_H . Hence Eq. (18.3.33) is called the *equal-time commutation relation*. From Eq. (18.3.30),

$$\dot{X}_H = -\frac{i}{\hbar} [X_H, H_H] = \left(-\frac{i}{\hbar}\right) \frac{i\hbar P_H}{m} = \frac{P_H}{m} \quad (18.3.34a)$$

and likewise

$$\dot{P}_H = -m\omega^2 X_H \quad (18.3.34b)$$

which are identical in form to the classical equations

$$\begin{aligned} \dot{x} &= \frac{\partial \mathcal{H}}{\partial p} = \frac{p}{m} \\ \dot{p} &= -\frac{\partial \mathcal{H}}{\partial x} = -m\omega^2 x \end{aligned} \quad (18.3.35)$$

This is to be expected, because the recipe for quantizing is such that commutators and PB always obey the correspondence [recall Eq. (7.4.40)]

$$\{\omega, \lambda\} = \gamma \rightarrow -\frac{i}{\hbar} [\Omega, \Lambda] = \Gamma \quad (18.3.36)$$

Although the Heisenberg picture is not often used in nonrelativistic quantum mechanics, it is greatly favored in relativistic quantum field theory.

Exercise 18.3.2. In the paramagnetic resonance problem Exercise 14.4.3 we moved to a frame rotating in *real space*. Show that this is also equivalent to a Hilbert space rotation, but that it takes us neither to the interaction nor the Heisenberg picture, except at resonance. What picture is it at resonance? (If $\mathbf{B} = B_0 \mathbf{k} + B \cos \omega t \mathbf{i} - B \sin \omega t \mathbf{j}$, associate B_0 with H_S^0 and B with H_S^1 .)

18.4. A General Discussion of Electromagnetic Interactions

This section contains a summary of several concepts from electro-dynamics that are relevant for the next section. It also deals with certain subtle questions of basic interest, not directly linked to the rest of this chapter.

Classical Electrodynamics

Let us begin with an extremely concise review of this subject.[‡] The response of matter to the electromagnetic field is given by the Lorentz force on a charge q :

$$\mathbf{F} = q \left(\mathbf{E} + \frac{\mathbf{v}}{c} \times \mathbf{B} \right) \quad (18.4.1)$$

The response of the fields to the charges is given by Maxwell's equations:

$$\nabla \cdot \mathbf{E} = 4\pi\rho \quad (18.4.2)$$

$$\nabla \times \mathbf{E} + \frac{1}{c} \frac{\partial \mathbf{B}}{\partial t} = 0 \quad (18.4.3)$$

$$\nabla \cdot \mathbf{B} = 0 \quad (18.4.4)$$

$$\nabla \times \mathbf{B} - \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t} = \frac{4\pi}{c} \mathbf{j} \quad (18.4.5)$$

where ρ and \mathbf{j} are the charge and current densities bound by the continuity equation

$$\nabla \cdot \mathbf{j} + \frac{\partial \rho}{\partial t} = 0 \quad (18.4.6)$$

Exercise 18.4.1. By taking the divergence of Eq. (18.4.5) show that the continuity equation must be obeyed if Maxwell's equations are to be mutually consistent.

The potentials \mathbf{A} and ϕ are now introduced as follows. Equation (18.4.4), combined with the identity $\nabla \cdot \nabla \times \mathbf{A} \equiv 0$, tells us that \mathbf{B} can be written as a curl

$$\mathbf{B} = \nabla \times \mathbf{A} \quad (18.4.7)$$

[‡] For any further information see the classic, *Classical Electrodynamics* by J. D. Jackson, Wiley, New York (1975).

Feeding this into Eq. (18.4.3), we find that

$$\nabla \times \left(\mathbf{E} + \frac{1}{c} \frac{\partial \mathbf{A}}{\partial t} \right) = 0 \quad (18.4.8)$$

Based on the identity $\nabla \times \nabla \phi \equiv 0$, we deduce that $\mathbf{E} + (1/c)\partial \mathbf{A}/\partial t$ can be written as a gradient, or that

$$\mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{A}}{\partial t} - \nabla \phi \quad (18.4.9)$$

If we replace \mathbf{E} and \mathbf{B} by the potentials in the other two Maxwell equations and use the identity $\nabla \times \nabla \times \mathbf{A} \equiv \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A}$ (true in Cartesian coordinates) we get the equations giving the response of \mathbf{A} and ϕ to the charges and currents:

$$\nabla^2 \phi + \frac{1}{c} \frac{\partial}{\partial t} (\nabla \cdot \mathbf{A}) = -4\pi\rho \quad (18.4.10)$$

$$\nabla^2 \mathbf{A} - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} - \nabla \left(\nabla \cdot \mathbf{A} + \frac{1}{c} \frac{\partial \phi}{\partial t} \right) = -\frac{4\pi \mathbf{j}}{c} \quad (18.4.11)$$

Before attacking these equations, let us note that there exists a certain arbitrariness in the potentials \mathbf{A} and ϕ , in that it is possible to change them (in a certain way) without changing anything physical. It may be readily verified that \mathbf{A} and ϕ and

$$\mathbf{A}' = \mathbf{A} - \nabla \Lambda \quad (18.4.12)$$

$$\phi' = \phi + \frac{1}{c} \frac{\partial \Lambda}{\partial t} \quad (18.4.13)$$

where Λ is an arbitrary function, lead to the same fields \mathbf{E} and \mathbf{B} .

*Exercise 18.4.2.** Calculate \mathbf{E} and \mathbf{B} corresponding to (\mathbf{A}, ϕ) and (\mathbf{A}', ϕ') using Eqs. (18.4.7) and (18.4.9) and verify the above claim.

Since the physics, i.e., the force law and Maxwell's equations, is sensitive only to \mathbf{E} and \mathbf{B} , the transformation of the potentials, called a *gauge transformation*, does not affect it. This is known as *gauge invariance*. Λ is called the *gauge parameter*, and (\mathbf{A}, ϕ) and (\mathbf{A}', ϕ') are called *gauge transforms* of each other, or said to be *gauge equivalent*.

Gauge invariance may be exploited to simplify Eqs. (18.4.10) and (18.4.11). We consider the case of the free electromagnetic field ($\rho = \mathbf{j} = 0$), which will be of interest

in the next section. In this case the gauge freedom allows us (see following exercise) to choose \mathbf{A} and ϕ such that

$$\nabla \cdot \mathbf{A} = 0 \quad (18.4.14)$$

$$\phi = 0 \quad (18.4.15)$$

This is called the *Coulomb gauge* and will be used hereafter. There is no residual gauge freedom if we impose the above Coulomb gauge conditions and the requirement that $|\mathbf{A}| \rightarrow 0$ at spatial infinity. The potential in the Coulomb gauge is thus unique and ‘physical’ in the sense that for a given \mathbf{E} and \mathbf{B} there is a unique \mathbf{A} .

*Exercise 18.4.3.** Suppose we are given some \mathbf{A} and ϕ that do not obey the Coulomb gauge conditions. Let us see how they can be transformed to the Coulomb gauge.

- (1) Show that if we choose

$$\Lambda(\mathbf{r}, t) = -c \int_{-\infty}^t \phi(\mathbf{r}, t') dt'$$

and transform to (\mathbf{A}', ϕ') then $\phi' = 0$. \mathbf{A}' is just $\mathbf{A} - \nabla \Lambda$, with $\nabla \cdot \mathbf{A}'$ not necessarily zero.

- (2) Show that if we gauge transform once more to (\mathbf{A}'', ϕ'') via

$$\Lambda' = -\frac{1}{4\pi} \int \frac{\nabla \cdot \mathbf{A}'(\mathbf{r}', t) d^3 r'}{|\mathbf{r} - \mathbf{r}'|}$$

then $\nabla \cdot \mathbf{A}'' = 0$. [Hint: Recall $\nabla^2(1/|\mathbf{r} - \mathbf{r}'|) = -4\pi\delta^3(\mathbf{r} - \mathbf{r}')$.]

- (3) Verify that ϕ'' is also zero by using $\nabla \cdot \mathbf{E} = 0$.

- (4) Show that if we want to make any further gauge transformations *within* the Coulomb gauge, Λ must be time independent and obey $\nabla^2 \Lambda = 0$. If we demand that $|\mathbf{A}| \rightarrow 0$ at spatial infinity, \mathbf{A} becomes unique.

In the Coulomb gauge, the equations of motion for the electromagnetic field (away from charges) simplify to

$$\nabla^2 \mathbf{A} - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} = 0 \quad (18.4.16a)$$

$$\nabla \cdot \dot{\mathbf{A}} = 0 \quad (18.4.16b)$$

$$\nabla \cdot \mathbf{A} = 0 \quad (18.4.16c)$$

The first equation tells us that electromagnetic waves travel at the speed c . Of special interest to us are solutions to these equations of the form[‡]

$$\mathbf{A} = \mathbf{A}_0 \cos(\mathbf{k} \cdot \mathbf{r} - \omega t) \quad (18.4.17)$$

[‡] Here \mathbf{k} denotes the wave vector and not the unit vector along the z axis.

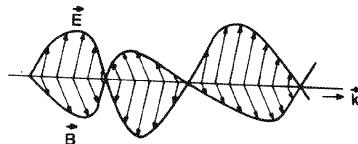


Figure 18.2. The electromagnetic wave at a given time. \mathbf{E} , \mathbf{B} , and \mathbf{k} (the wave vector) are mutually perpendicular.

Feeding this into the wave equation we find

$$\omega^2 = k^2 c^2$$

or

$$\omega = kc \quad (18.4.18)$$

The gauge condition tells us that

$$0 = \nabla \cdot \mathbf{A} = -(\mathbf{k} \cdot \mathbf{A}_0) \sin(\mathbf{k} \cdot \mathbf{r} - \omega t)$$

or

$$\mathbf{k} \cdot \mathbf{A}_0 = 0 \quad (18.4.19)$$

This means that \mathbf{A} must lie in a plane perpendicular to the direction of propagation, i.e., that electromagnetic waves are *transverse*. The electric and magnetic fields corresponding to this solution are

$$\mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{A}}{\partial t} = -\left(\frac{\omega}{c}\right) \mathbf{A}_0 \sin(\mathbf{k} \cdot \mathbf{r} - \omega t) \quad (18.4.20)$$

$$\mathbf{B} = \nabla \times \mathbf{A} = -(\mathbf{k} \times \mathbf{A}_0) \sin(\mathbf{k} \cdot \mathbf{r} - \omega t) \quad (18.4.21)$$

Thus \mathbf{E} and \mathbf{B} are mutually perpendicular and perpendicular to \mathbf{k} (i.e., they are also transverse)—see Fig. 18.2. They have the same magnitude:

$$|\mathbf{E}| = |\mathbf{B}| \quad (18.4.22)$$

The energy flow across unit area (placed normal to \mathbf{k}) per second is (from any standard text)

$$|\mathbf{S}| = \frac{c}{4\pi} |(\mathbf{E} \times \mathbf{B})| = \frac{\omega^2}{4\pi c} |\mathbf{A}_0|^2 \sin^2(\mathbf{k} \cdot \mathbf{r} - \omega t) \quad (18.4.23a)$$

The time average over a cycle is

$$S_{\text{av}} = \frac{\omega^2}{8\pi c} |\mathbf{A}_0|^2 \quad (18.4.23b)$$

The energy per unit volume is

$$u = (1/8\pi) \cdot [|\mathbf{E}|^2 + |\mathbf{B}|^2] \quad (18.4.24)$$

Notice that $|S|$ equals the energy density times the velocity of wave propagation.

The Potentials in Quantum Theory

We now ask if quantum mechanics also is invariant under gauge transformations of the potentials. Let us seek the answer to this question in the path integral approach. Recall that

$$U(\mathbf{r}t, \mathbf{r}'t') = N \sum_{\text{paths}} \exp[iS/\hbar] \quad (18.4.25)$$

where N is a normalization factor and the action

$$S = \int_{t'}^t \mathcal{L} dt'' = \int_{t'}^t \left(\frac{1}{2} m |\dot{\mathbf{r}}|^2 + \frac{q}{c} \mathbf{v} \cdot \mathbf{A} - q\phi \right) dt'' \quad (18.4.26)$$

is to be evaluated along each path P that connects (\mathbf{r}', t') and (\mathbf{r}, t) . Suppose we perform a gauge transformation of the potentials. Then

$$S \rightarrow S_\Lambda = S - \int_{t'}^t \frac{q}{c} \left(\mathbf{v} \cdot \nabla \Lambda + \frac{\partial \Lambda}{\partial t''} \right) dt'' \quad (18.4.27)$$

But

$$\mathbf{v} \cdot \nabla \Lambda + \frac{\partial \Lambda}{\partial t''} = \frac{d\Lambda}{dt''} \quad (18.4.28)$$

is the *total derivative* along the trajectory. Consequently

$$S_\Lambda = S + \frac{q}{c} [\Lambda(\mathbf{r}', t') - \Lambda(\mathbf{r}, t)] \quad (18.4.29)$$

It is clear that S and S_Λ imply the same classical dynamics: varying S and varying S_Λ (to find the path of least actions) are equivalent, since S and S_Λ differ only by

$(q/c)\Lambda$ at the end points, and the latter are held fixed in the variation. Going on to the quantum case, we find from Eqs. (18.4.25) and (18.4.29) that

$$U \rightarrow U_\Lambda = U \cdot \exp\left\{\frac{iq}{\hbar c} [\Lambda(\mathbf{r}', t') - \Lambda(\mathbf{r}, t)]\right\} \quad (18.4.30)$$

Since

$$U(\mathbf{r}, t; \mathbf{r}', t') = \langle \mathbf{r} | U(t, t') | \mathbf{r}' \rangle \quad (18.4.31)$$

we see that effect of the gauge transformation is equivalent to a change in the coordinate basis:

$$|\mathbf{r}\rangle \rightarrow |\mathbf{r}_\Lambda\rangle = e^{(iq\Lambda/\hbar c)} |\mathbf{r}\rangle \quad (18.4.32)$$

which of course cannot change the physics. (Recall, however, the discussion in Section 7.4.) The change in the wave function under the gauge transformation is

$$\psi = \langle \mathbf{r} | \psi \rangle \rightarrow \psi_\Lambda = \langle \mathbf{r}_\Lambda | \psi \rangle = e^{-iq\Lambda(\mathbf{r}, t)/\hbar c} \psi \quad (18.4.33)$$

This result may also be obtained within the Schrödinger approach (see the following exercise).

Exercise 18.4.4 (Proof of Gauge Invariance in the Schrödinger Approach). (1) Write H for a particle in the potentials (\mathbf{A}, ϕ) .

(2) Write down H_Λ , the Hamiltonian obtained by gauge transforming the potentials.

(3) Show that if $\psi(\mathbf{r}, t)$ is a solution to Schrödinger's equation with the Hamiltonian H , then $\psi_\Lambda(\mathbf{r}, t)$ given in Eq. (18.4.33) is the corresponding solution with $H \rightarrow H_\Lambda$.

Although quantum mechanics is similar to classical mechanics in that it is insensitive to gauge transformations of the potentials, it is different in the status it assigns to the potentials. This is dramatically illustrated in the *Aharonov-Bohm effect*, depicted schematically in Fig. 18.3.[‡] The experiment is just the double-slit experiment with one change: there is a small shaded region ($B \neq 0$) where magnetic fluxes comes out of the paper. (You may imagine a tiny solenoid coming out of the paper, inside which are confined the flux lines. These lines must of course return to the other end of the solenoid, but this is arranged not to happen in the experimental region.) The vector potential (in Coulomb gauge) is shown by closed loops surrounding the coil. At a classical level, this variation in the double-slit experiment is expected to make no change in the outcome, for there is no magnetic field along the classical paths P_1 and P_2 . There is, of course, an \mathbf{A} field along P_1 and P_2 , but the potential has no direct significance in classical physics. Its curl, which is significant, vanishes there.

Consider now the quantum case. In the path integral approach, a particle emitted by the source has the following amplitude to end up at a point \mathbf{r} on the screen, before \mathbf{B} is turned on:

$$\psi(\mathbf{r}) \simeq \psi_{P_1}(\mathbf{r}) + \psi_{P_2}(\mathbf{r}) \quad (18.4.34)$$

[‡] For the actual experiment see R. G. Chambers, *Phys. Rev. Lett.*, **5**, 3 (1960).

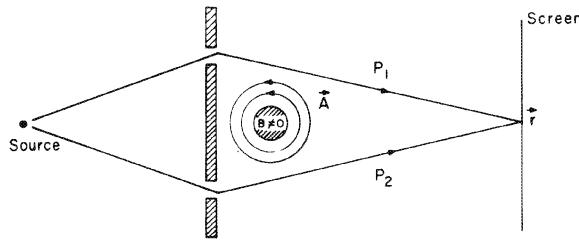


Figure 18.3. An experiment (schematic) that displays the Aharonov-Bohm effect. It is just the double-slit experiment but for the small coil coming out of the paper carrying magnetic flux (indicated by the shaded region marked $B \neq 0$).

where ψ_{P_i} ($i=1, 2$) is the contribution from the classical path P_i and its immediate neighbors. The interference between these two contributions produces the usual interference pattern. Let us turn on \mathbf{B} . Now each path gets an extra factor

$$\exp\left[\frac{iq}{\hbar c} \int_{r'}^r (\mathbf{v} \cdot \mathbf{A}) dt''\right] = \exp\left(\frac{iq}{\hbar c} \int_{\text{source}}^r \mathbf{A} \cdot d\mathbf{r}''\right) \quad (18.4.35)$$

Since $\nabla \times \mathbf{A} = 0$ near P_1 and P_2 , by Stoke's theorem the integral is the same for P_1 and its neighbors and P_2 and its neighbors. But the integral on P_1 is not the same as the integral on P_2 , for these paths surround the coil and

$$\begin{aligned} \int_{P_2} \mathbf{A} \cdot d\mathbf{r} - \int_{P_1} \mathbf{A} \cdot d\mathbf{r} &= \oint \mathbf{A} \cdot d\mathbf{r} = \int_S (\nabla \times \mathbf{A}) \cdot d\mathbf{s} \\ &= \int \mathbf{B} \cdot d\mathbf{s} = \Phi \neq 0 \end{aligned} \quad (18.4.36)$$

where s is any surface bounded by the closed loop $P_1 + P_2$, and Φ is the flux crossing it, i.e., coming out of the paper in Fig. 18.3. Bearing this in mind, we get

$$\psi(\mathbf{r}) = \exp\left(\frac{iq}{\hbar c} \int_{P_1} \mathbf{A} \cdot d\mathbf{r}''\right) \psi_{P_1}(\mathbf{r}) + \exp\left(\frac{iq}{\hbar c} \int_{P_2} \mathbf{A} \cdot d\mathbf{r}''\right) \psi_{P_2}(\mathbf{r}) \quad (18.4.37)$$

Pulling out an overall phase factor, which does not affect the interference pattern, we get

$$\begin{aligned} \psi(\mathbf{r}) &= \left(\begin{array}{c} \text{overall} \\ \text{factor} \end{array} \right) \left[\psi_{P_1}(\mathbf{r}) + \exp\left(\frac{iq}{\hbar c} \oint \mathbf{A} \cdot d\mathbf{r}\right) \psi_{P_2}(\mathbf{r}) \right] \\ &= \left(\begin{array}{c} \text{overall} \\ \text{factor} \end{array} \right) [\psi_{P_1}(\mathbf{r}) + \exp[iq\Phi/\hbar c] \psi_{P_2}(\mathbf{r})] \end{aligned} \quad (18.4.38)$$

By varying \mathbf{B} (and hence Φ) we change the relative phase between the contributions from the two paths and move the interference pattern up and down. Whenever $(q\Phi/\hbar c) = 2n\pi$, the pattern will return to its initial form, as if there were no field.

In other words, an integral multiple of the *flux quantum*

499

TIME-DEPENDENT
PERTURBATION
THEORY

$$\Phi_0 = \frac{2\pi\hbar c}{q} \quad (18.4.39)$$

will not make any observable difference to the quantum mechanics of the particle. This idea is very frequently invoked; we shall do so in Chapter 21.

Let us understand how the particle discerns the magnetic field even though the dominant paths all lie in the $\mathbf{B}=0$ region. Suppose I show you Fig. 18.3 but cover the region where the coil is (the shaded region marked $B \neq 0$); will you know there is magnetic flux coming out of the paper? Yes, because the circulating \mathbf{A} lines will tell you that $\oint \mathbf{A} \cdot d\mathbf{r} = \int \mathbf{B} \cdot d\mathbf{s} \neq 0$.[‡] The classical particle, however, moves along P_1 or P_2 , and can have no knowledge of $\oint \mathbf{A} \cdot d\mathbf{r}$. The best it can do is measure $\nabla \times \mathbf{A}$ locally, and that always equals zero. The quantum particle, on the other hand, “goes along P_1 and P_2 ” (in the path integral sense) and by piecing together what happens along P_1 and P_2 (i.e., by comparing the relative phase of the contributions from the two paths) it can deduce not only the existence of \mathbf{B} , but also the total flux. Notice that although the particle responds to \mathbf{A} and not directly to \mathbf{B} , the response is gauge invariant.

18.5. Interaction of Atoms with Electromagnetic Radiation

We will make no attempt to do justice to this enormous field. We will consider just two illustrative examples. The first is the photoelectric effect in hydrogen (in which the incident radiation knocks the electron out of the atom). The second is the spontaneous decay of hydrogen from an excited state to the ground state (decay in the absence of external fields), which can be understood only if the electromagnetic field is treated as a quantum system.

Photoelectric Effect in Hydrogen

Consider a hydrogen atom in its ground state $|100\rangle$ centered at the origin, and on which is incident the wave

$$\mathbf{A}(\mathbf{r}, t) = \mathbf{A}_0 \cos(\mathbf{k} \cdot \mathbf{r} - \omega t) \quad (18.5.1)$$

For energies $\hbar\omega$ sufficiently large, the bound electron can be liberated and will come flying out. We would like to calculate the rate for this process using Fermi's

[‡] This is like saying that you can infer the existence of a pole in the complex plane and its residue, without actually going near it, by evaluating $1/2\pi i \oint f(z) dz$ on a path that encloses it.

golden rule:

$$R_{i \rightarrow f} = \text{rate of transition } i \rightarrow f = \frac{2\pi}{\hbar} |\langle f^0 | H^1 | i^0 \rangle|^2 \delta(E_f^0 - E_i^0 - \hbar\omega) \quad (18.5.2)$$

Two points need to be explained before the application of this rule:

(1) For the final state, we must use a positive energy eigenstate of the Coulomb Hamiltonian $H^0 = P^2/2m - e^2/r$. Now we argue on intuitive grounds that if the ejected electron is very energetic, we must be able to ignore the pull of the proton on it and describe it by a plane wave $|\mathbf{p}_f\rangle$ in Eq. (18.5.2), with negligible error. While this happens to be the case here, there is a subtle point that is worth noting. If we view the Coulomb attraction of the proton as a perturbation relative to the free-particle Hamiltonian $P^2/2m$, we can write the eigenstate of H^0 as a perturbation series:

$$|f^0\rangle = |\mathbf{p}_f\rangle + \text{higher-order terms}$$

We are certainly right in guessing that $|\mathbf{p}_f\rangle$ dominates the expansion at high energies. But we are assuming more: we are assuming that when we evaluate the matrix element in Eq. (18.5.2) the leading term $|\mathbf{p}_f\rangle$ will continue to dominate the higher-order terms. Clearly, the validity of this assumption depends also on the initial state $|i^0\rangle$ and the operator H^1 . Now it turns out that if the initial state is an s state (as in the present case) the higher-order terms are indeed negligible in computing the matrix element, but not otherwise. For instance if the initial state is a p state, the contribution of the first-order term to the matrix element would be comparable to the contribution from the leading term $|\mathbf{p}_f\rangle$. For more details, you must consult a book that is devoted to the subject.[‡]

(2) The rule applied for potentials of the form $H^1(t) = H^1 e^{-i\omega t}$, whereas here [recall Eq. (14.4.11)],[§]

$$\begin{aligned} H^1(t) &= \frac{-(-e)}{2mc} (\mathbf{A} \cdot \mathbf{P} + \mathbf{P} \cdot \mathbf{A}) \\ &= \frac{e}{mc} \mathbf{A} \cdot \mathbf{P} \quad (\text{because } \nabla \cdot \mathbf{A} = 0) \\ &= \frac{e}{mc} \cos(\mathbf{k} \cdot \mathbf{r} - \omega t) \mathbf{A}_0 \cdot \mathbf{P} \\ &= \frac{e}{2mc} [e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)} + e^{-i(\mathbf{k} \cdot \mathbf{r} - \omega t)}] \mathbf{A}_0 \cdot \mathbf{P} \end{aligned} \quad (18.5.3)$$

[‡] For example, Section 70 of H. Bethe and E. Salpeter, *Quantum Mechanics of One and Two Electron Atoms*, Plenum, New York (1977). This is also a good place to look for other data on this subject. For instance if you want to know what the expectation value of r^{-4} is in the state $|nlm\rangle$ of hydrogen, you will find it here.

[§] We do not include in H^1 the term proportional to $|\mathbf{A}|^2$, which is of second order. The spin interaction $-\gamma \mathbf{S} \cdot \mathbf{B}$ is of the first order, but negligible in the kinematical region we will focus on. This will be demonstrated shortly.

Of the two pieces, only the first has the correct time dependence to induce the transition $i \rightarrow f$ with $E_f > E_i$; the second will be killed by the energy-conserving delta function. Hereafter we ignore the second term and let

$$\begin{aligned} H^1(t) &= \frac{e}{2mc} e^{i\mathbf{k}\cdot\mathbf{r}} \mathbf{A}_0 \cdot \mathbf{P} e^{-i\omega t} \\ &= H^1 e^{-i\omega t} \end{aligned} \quad (18.5.4)$$

With these two points out of the way, we can proceed to evaluate the transition matrix element in the coordinate basis:

$$H_{fi}^1 = \frac{e}{2mc} \frac{1}{(2\pi\hbar)^{3/2}} \left(\frac{1}{\pi a_0^3} \right)^{1/2} \int e^{-i\mathbf{p}_f \cdot \mathbf{r}/\hbar} e^{i\mathbf{k}\cdot\mathbf{r}} \mathbf{A}_0 \cdot (-i\hbar\nabla) e^{-r/a_0} d^3\mathbf{r} \quad (18.5.5)$$

Consider the factor $e^{i\mathbf{k}\cdot\mathbf{r}}$. Recall from Chapter 5 that multiplication of a wave function by $e^{i\mathbf{p}_0 \cdot \mathbf{r}/\hbar}$ adds to the state a momentum \mathbf{p}_0 . Thus the factor $e^{i\mathbf{k}\cdot\mathbf{r}}$ represents the fact that a momentum $\hbar\mathbf{k}$ is imparted by the radiation to the atom.[‡] For any transition between atomic levels, this momentum transferred is negligible compared to the typical momentum p of the electron. We see this as follows. The energy transferred is of the order of a Rydberg:

$$\hbar\omega \sim e^2/a_0 \quad (18.5.6)$$

so that the photon momentum is

$$\hbar k = \frac{\hbar\omega}{c} \approx \frac{e^2}{a_0 c} \quad (18.5.7)$$

On the other hand, the typical momentum of the electron, estimated from the uncertainty principle, is

$$p \sim \frac{\hbar}{a_0} \quad (18.5.8)$$

Thus

$$\frac{\hbar k}{p} \approx \frac{e^2}{\hbar c} \approx \frac{1}{137} \quad (18.5.9)$$

In the present case $\hbar\omega$ is a lot higher because we have a liberated, high-energy electron. But there is still a wide range of ω over which $\hbar k/p \ll 1$. We will work in

[‡] You may be worried that there is the $(-i\hbar\nabla)$ operator between $e^{i\mathbf{k}\cdot\mathbf{r}}$ and the atomic wave function. But since $\nabla \cdot \mathbf{A} = 0$, we can also write $\mathbf{A} \cdot \mathbf{P}$ as $\mathbf{P} \cdot \mathbf{A}$, in which case the $e^{i\mathbf{k}\cdot\mathbf{r}}$ will be right next to the atomic wave function.

this domain. In this domain, the ratio of the spin interaction we neglected, to the orbital interaction we are considering, is roughly

$$\frac{\langle (e/2mc)\mathbf{S} \cdot \mathbf{B} \rangle}{\langle (e/mc)\mathbf{A} \cdot \mathbf{P} \rangle} \simeq \frac{\langle \hbar \boldsymbol{\sigma} \cdot \nabla \times \mathbf{A} \rangle}{\langle \mathbf{A} \cdot \mathbf{P} \rangle} \simeq \frac{\hbar k}{p} \ll 1 \quad (18.5.10)$$

which justifies our neglect.

The domain we are working in may also be described by

$$ka_0 \ll 1 \quad (18.5.11)$$

[Eq. (18.5.9)]. This means that the phase of the wave changes little over the size of the atom. Since the integral in Eq. (18.5.5) is rapidly cut off beyond $r \simeq a_0$ by the wave function e^{-r/a_0} , we may approximate e^{ikr} in the integral as

$$e^{ikr} \simeq 1 \quad (18.5.12)$$

This is called the *electric dipole approximation*.‡ The reason is that in this approximation, the atom sees a *spatially* constant electric field,

$$\begin{aligned} \mathbf{E} &= \frac{-1}{c} \frac{\partial \mathbf{A}}{\partial t} \\ &= \frac{-1}{c} \frac{\partial}{\partial t} \left(\frac{\mathbf{A}_0}{2} e^{-i\omega t} \right) \S \\ &= \frac{i\omega}{2c} \mathbf{A}_0 e^{-i\omega t} \end{aligned} \quad (18.5.13)$$

and couples to it via its electric dipole moment $\mu = -e\mathbf{R}$:

$$H^1(t) = -\mu \cdot \mathbf{E} = \frac{i\omega e}{2c} \mathbf{A}_0 \cdot \mathbf{R} e^{-i\omega t} \quad (18.5.14)$$

This must of course coincide with Eq. (18.5.3) in this approximation:

$$H^1(t) = \frac{e}{2mc} \mathbf{A}_0 \cdot \mathbf{P} e^{-i\omega t} \quad (18.5.15)$$

‡ By keeping higher powers of $\mathbf{k} \cdot \mathbf{r}$ in the expansion, one gets terms known as electric quadrupole, magnetic dipole, electric octupole, magnetic quadrupole, etc. contributions.

§ We ignore the “wrong” frequency part of \mathbf{A} .

The equivalence of Eqs. (18.5.14) and (18.5.15) can be demonstrated in a general situation as follows. Since for any

$$H^0 = \frac{|\mathbf{P}|^2}{2m} + V(\mathbf{R}) \quad (18.5.16a)$$

it is true that

$$[\mathbf{R}, H^0] = \frac{i\hbar}{m} \mathbf{P} \quad (18.5.16b)$$

we find

$$\begin{aligned} \langle f^0 | \mathbf{P} | i^0 \rangle &= \frac{m}{i\hbar} \langle f^0 | \mathbf{R} H^0 - H^0 \mathbf{R} | i^0 \rangle \\ &= \frac{m}{i\hbar} (E_i^0 - E_f^0) \langle f^0 | \mathbf{R} | i^0 \rangle \\ &= im\omega \langle f^0 | \mathbf{R} | i^0 \rangle \end{aligned} \quad (18.5.17)$$

so that

$$\begin{aligned} \langle f^0 | \frac{e}{2mc} \mathbf{A}_0 \cdot \mathbf{P} | i^0 \rangle &= \frac{ie\omega}{2c} \mathbf{A}_0 \cdot \langle f^0 | \mathbf{R} | i^0 \rangle \\ &= \langle f^0 | (-\mathbf{p} \cdot \mathbf{E}) | i^0 \rangle \quad [\text{by Eq. (18.5.14)}] \end{aligned} \quad (18.5.18)$$

Consider now the evaluation of the matrix element H_{fi}^1 in the dipole approximation:

$$H_{fi}^1 = N \int e^{-i\mathbf{p}_f \cdot \mathbf{r}/\hbar} \mathbf{A}_0 \cdot (-i\hbar \nabla) e^{-r/a_0} d^3 \mathbf{r} \quad (18.5.19)$$

where N is a constant given by

$$N = \left(\frac{e}{2mc} \right) \left(\frac{1}{2\pi\hbar} \right)^{3/2} \left(\frac{1}{\pi a_0^3} \right)^{1/2} \quad (18.5.20)$$

If we integrate the ∇ by parts, we get

$$H_{fi}^1 = N \mathbf{A}_0 \cdot \mathbf{p}_f \int e^{-i\mathbf{p}_f \cdot \mathbf{r}/\hbar} e^{-r/a_0} d^3 \mathbf{r} \quad (18.5.21)$$

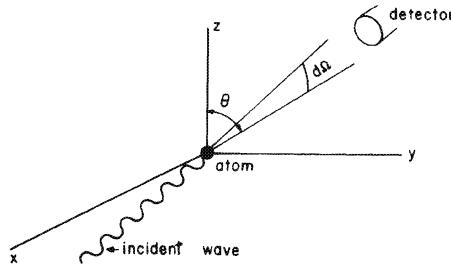


Figure 18.4. The photoelectric effect. In any realistic experiment, the resolution in energy and angle are finite. One asks how many electrons come into the cone of solid angle $d\Omega$ with magnitude of momentum between p and $p + dp$.

(It should now be clear why we prefer the $\mathbf{A}_0 \cdot \mathbf{P}$ form of H^1 to the $\mathbf{A}_0 \cdot \mathbf{R}$ form.) If we choose the z axis along \mathbf{p}_f , the \mathbf{r} integral becomes

$$\begin{aligned}
 & \int_0^\infty \int_{-1}^1 \int_0^{2\pi} e^{-ip_f r \cos \theta / \hbar} e^{-r/a_0 r^2} dr d(\cos \theta) d\phi \\
 &= 2\pi \int_0^\infty \left(\frac{e^{-ip_f r / \hbar} - e^{ip_f r / \hbar}}{-ip_f r / \hbar} \right) e^{-r/a_0 r^2} dr \\
 &= \frac{2\pi \hbar i}{p_f} \left[-\frac{\partial}{\partial (1/a_0)} \right] \int_0^\infty [e^{-(1/a_0 + ip_f / \hbar)r} - e^{-(1/a_0 - ip_f / \hbar)r}] dr \\
 &= \frac{8\pi/a_0}{[(1/a_0)^2 + (p_f/\hbar)^2]^2} \tag{18.5.22}
 \end{aligned}$$

Feeding this into Eq. (18.5.5), and the resulting expression into the golden rule, we get the transition rate

$$\begin{aligned}
 R_{i \rightarrow f} &= \frac{2\pi}{\hbar} \left(\frac{e}{2mc} \right)^2 \frac{1}{8\pi^3 \hbar^3} \frac{1}{\pi a_0^3} \frac{|\mathbf{A}_0 \cdot \mathbf{p}_f|^2 64\pi^2 a_0^6}{[1 + (p_f a_0 / \hbar)^2]^4} \\
 &\times \delta(E_f^0 - E_i^0 - \hbar\omega) \tag{18.5.23}
 \end{aligned}$$

Now the time has come to tackle the δ function. The δ function gives a singular probability distribution for finding a final electron in a state of *mathematically precise momentum* \mathbf{p}_f . This probability is of little interest in practice, where one sets up a detector with a finite opening angle $d\Omega$ and asks how many electrons come into it with magnitude of momentum between p_f and $p_f + dp_f$ (see Fig. 18.4). The δ function tells us that electron momenta are concentrated at

$$\frac{p_f^2}{2m} = E_i^0 + \hbar\omega$$

The contribution from this region is obtained by integrating the δ function over p_f .
Using

505

TIME-DEPENDENT
PERTURBATION
THEORY

$$\delta\left(\frac{p_f^2}{2m} - E_i^0 - \hbar\omega\right) = \frac{m}{p_f} \delta\{p_f - [2m(E_i^0 + \hbar\omega)]^{1/2}\} \quad (18.5.24)$$

we get the rate of transition into the detector to be

$$\begin{aligned} R_{i \rightarrow d\Omega} &= \frac{2\pi}{\hbar} |H_{fi}|^2 m p_f d\Omega \\ &= \frac{4a_0^3 e^2 p_f |\mathbf{A}_0 \cdot \mathbf{p}_f|^2}{m\pi\hbar^4 c^2 [1 + (p_f a_0 / \hbar)^2]^4} d\Omega \end{aligned} \quad (18.5.25)$$

{In this and all following expressions, $p_f = [2m(E_i^0 + \hbar\omega)]^{1/2}$.} Note that the rate depends only on the magnitude of the applied field \mathbf{A}_0 , the angle between the polarization \mathbf{A}_0 and the outgoing momentum, and the magnitude of \mathbf{p}_f or equivalently ω , the frequency of radiation. The formula above tells us that the electron likes to come parallel to \mathbf{A}_0 , that is, to the electric field which rips it out of the atom. The direction of the incident radiation does not appear because we set $e^{i\mathbf{k}\mathbf{r}} = 1$. If we keep the $e^{i\mathbf{k}\mathbf{r}}$ factor it will be seen that the electron momentum is also biased toward \mathbf{k} , reflecting the $\hbar\mathbf{k}$ momentum input.

*Exercise 18.5.1.** (1) By going through the derivation, argue that we can take the $e^{i\mathbf{k}\mathbf{r}}$ factor into account exactly, by replacing \mathbf{p}_f by $\mathbf{p}_f - \hbar\mathbf{k}$ in Eq. (18.5.19).

(2) Verify the claim made above about the electron momentum distribution.

If we integrate $R_{i \rightarrow d\Omega}$ over all angles, we get the total rate for ionization. Choosing \mathbf{A}_0 along the z axis for convenience, we find

$$\begin{aligned} R_{i \rightarrow \text{all}} &= \frac{4a_0^3 e^2 p_f^3 |\mathbf{A}_0|^2}{m\pi\hbar^4 c^2 [1 + (p_f a_0 / \hbar)^2]^4} \iint \cos^2 \theta d(\cos \theta) d\phi \\ &= \frac{16a_0^3 e^2 p_f^3 |\mathbf{A}_0|^2}{3m\hbar^4 c^2 [1 + (p_f a_0 / \hbar)^2]^4} \end{aligned} \quad (18.5.26)$$

Since this is the rate of ionization, and each ionization takes energy $\hbar\omega$ from the beam, the energy absorption rate is

$$\frac{dE_{\text{abs}}}{dt} = \hbar\omega \cdot R_{i \rightarrow \text{all}} \quad (18.5.27)$$

Now the beam brings in energy at the rate $\omega^2 |\mathbf{A}_0|^2 / 8\pi c$ per unit area. Suppose we place, transverse to this beam, a perfectly absorbing disk of area σ . It will absorb

energy at the rate

$$\frac{dE_{\text{abs}}}{dt} = \frac{\sigma |\mathbf{A}_0|^2 \omega^2}{8\pi c} \quad (18.5.28)$$

By comparing Eqs. (18.5.27) and (18.5.28), we see that we can associate with the atom a *photoelectric cross section*

$$\sigma = \frac{8\pi c}{|\mathbf{A}_0|^2 \omega^2} \cdot \hbar\omega \cdot R_{i \rightarrow \text{all}} \quad (18.5.29)$$

$$= \frac{128a_0^3\pi e^2 p_f^3}{3m\hbar^3\omega c[1 + p_f^2 a_0^2/\hbar^2]^4} \quad (18.5.30)$$

in the sense that if an ensemble of N (N large) nonoverlapping (separation $\gg a_0$) hydrogen atoms is placed in the way of the beam, the ensemble will absorb energy like a perfectly absorbent disk of area $N\sigma$. We can also associate a *differential cross section* $d\sigma/d\Omega$, with the energy flowing into a solid angle $d\Omega$:

$$\begin{aligned} \frac{d\sigma}{d\Omega} &= \frac{8\pi c}{|\mathbf{A}_0|^2 \omega^2} \hbar\omega R_{i \rightarrow d\Omega} \\ &= \frac{32a_0^3 e^2 p_f^3 \cos^2 \theta}{mc\omega\hbar^3[1 + p_f^2 a_0^2/\hbar^2]^4} \end{aligned} \quad (18.5.31)$$

In the region where $p_f a_0 / \hbar \gg 1$, the formula simplifies to

$$\frac{d\sigma}{d\Omega} = \frac{32e^2\hbar^5 \cos^2 \theta}{mc\omega p_f^5 a_0^5} \quad (18.5.32)$$

*Exercise 18.5.2.** (1) Estimate the photoelectric cross section when the ejected electron has a kinetic energy of 10 Ry. Compare it to the atom's geometric cross section $\approx \pi a_0^2$.

(2) Show that if we consider photoemission from the $1s$ state of a charge Z atom, $\sigma \propto Z^5$, in the limit $p_f a_0 / Z\hbar \gg 1$.

Field Quantization†

The general formalism, illustrated by the preceding example, may be applied to a host of other phenomena involving the interaction of atoms with radiation. The results are always in splendid agreement with experiment as long the electromagnetic field is of macroscopic strength. The breakdown of the above formalism for weak fields is most dramatically illustrated by the following example. Consider a hydrogen atom in free space (the extreme case of weak field) in the state $|2, l, m\rangle$. What is the rate of decay to the ground state? Our formalism gives an unambiguous answer of

† The treatment of this advanced topic will be somewhat concise. You are urged to work out the missing steps if you want to follow it in depth.

zero, for free space corresponds to $\mathbf{A}=0$ (in the Coulomb gauge), so that $H^1=0$ and the atom should be in the stationary state $|2, l, m\rangle$ forever. But it is found experimentally that the atom decays at a rate $R \simeq 10^9 \text{ second}^{-1}$, or has a mean lifetime $\tau \simeq 10^{-9}$ second. In fact, all excited atoms are found to decay spontaneously in free space to their ground states. This phenomenon cannot be explained within our formalism.

So are we to conclude that our description of free space (which should be the simplest thing to describe) is inadequate? Yes! The description of free space by $\mathbf{A}=\dot{\mathbf{A}}=0$ is *classical*; it is like saying that the ground state of the oscillator is given by $x=p=0$. Now, we know that if the oscillator is treated quantum mechanically, only the *average* quantities $\langle 0|X|0\rangle$ and $\langle 0|P|0\rangle$ vanish in the ground state, and that there are nonzero fluctuations $(\Delta X)^2 = \langle 0|X^2|0\rangle$ and $(\Delta P)^2 = \langle 0|P^2|0\rangle$ about these mean values. In the same way, if the electromagnetic field is treated quantum mechanically, it will be found that free space (which is the ground state of the field) is described by $\langle \mathbf{A} \rangle = \langle \dot{\mathbf{A}} \rangle = 0$ (where \mathbf{A} and $\dot{\mathbf{A}}$ are *operators*)[†] with nonvanishing fluctuations $(\Delta A)^2$, $(\Delta \dot{A})^2$. The free space is dormant only in the average sense; there are always quantum fluctuations of the fields about these mean values. It is these fluctuations that trigger spontaneous decay.

As long as we restrict ourselves to macroscopic fields, the quantum and classical descriptions of the field become indistinguishable. This is why in going from classical to quantum mechanics, i.e., in going from $\mathcal{H}^1 = (e/mc)\mathbf{A} \cdot \mathbf{p}$ to $H^1 = (e/mc)\mathbf{A} \cdot \mathbf{P}$, we merely promoted \mathbf{p} to the operator \mathbf{P} , but let \mathbf{A} continue to be the classical field. For this reason, this treatment is called the *semiclassical treatment*. We now turn to the full quantum mechanical treatment in which \mathbf{A} will become an operator as well.

The basic idea behind quantizing the field is familiar: one finds a complete set of *canonical* coordinates and momenta to describe the classical field, and promotes them to operators obeying canonical commutation relations. One then takes \mathcal{H} , which is just the field energy written in terms of the canonical variables, and obtains H by the usual substitution rule. But there are many obstacles, as we shall see.

Let us start with the coordinaters of the field. If we decide to describe it in terms of the potentials, we have, at each point in space \mathbf{r} , four real coordinates $(\phi(\mathbf{r}), \mathbf{A}(\mathbf{r}))$.[§] Now, we already know that these coordinates are not entirely physical, in that they can be gauge transformed with no observable consequences. For them to be physical, we must constrain them to a point where there is no residual gauge freedom, say by imposing the Coulomb gauge conditions. Although we shall do so eventually, we treat them as genuine coordinates for the present.

What are the momenta conjugate to these coordinates? To find out, we turn to the Lagrangian:

$$\mathcal{L} = \frac{1}{8\pi} \int [|\mathbf{E}|^2 - |\mathbf{B}|^2] d^3\mathbf{r} = \frac{1}{8\pi} \int \left[\left(-\frac{1}{c} \frac{\partial \mathbf{A}}{\partial t} - \nabla \phi \right)^2 - |\nabla \times \mathbf{A}|^2 \right] d^3\mathbf{r} \quad (18.5.33)$$

[†] We depart from our convention here and denote classical and quantum field variables by the *same symbols*, because this is what everyone does in this case.

[§] Now \mathbf{r} is just a label on the field coordinates and not a dynamical variable.

^{||} If you are unfamiliar with this \mathcal{L} : Recall that the field energy is $\int (1/8\pi)[|\mathbf{E}|^2 + |\mathbf{B}|^2] d^3\mathbf{r}$, Eq. (18.4.24). Write this in the gauge $\phi=0$ and change the sign of the term that corresponds to “potential energy.” The above result is just the generalization to the gauge with $\phi \neq 0$.

which, when varied with respect to the potentials, gives Maxwell's equations.[‡] The momentum conjugate to each "coordinate" is the derivative of \mathcal{L} with respect to the corresponding "velocity." It follows that the momentum conjugate to $\phi(\mathbf{r})$ vanishes (at each point \mathbf{r} in space), for $\dot{\phi}(\mathbf{r})$ does not appear in \mathcal{L} . The fact that we are dealing with a coordinate whose conjugate momentum vanishes *identically* tells us that we can not follow the canonical route. But fortunately for us, we have the freedom to work in a gauge where $\phi=0$. So hereafter we can forget all about ϕ and its vanishing conjugate momentum. In particular, we can set $\phi=0$ in Eq. (18.5.33).

Consider now the coordinates $\mathbf{A}(\mathbf{r})$. To find $\Pi_i(\mathbf{r}_0)$, the momentum conjugate to $A_i(\mathbf{r}_0)$, we use the relation

$$\Pi_i(\mathbf{r}_0) = \frac{\partial \mathcal{L}}{\partial \dot{A}_i(\mathbf{r}_0)}$$

In differentiating \mathcal{L} with respect to $\dot{A}_i(\mathbf{r}_0)$, we treat the integral in Eq. (18.5.33) over \mathbf{r} as a sum over the continuous index \mathbf{r} . The partial derivative picks out just the term in the sum carrying the index $\mathbf{r}=\mathbf{r}_0$ (because the velocities at different points are independent variables) and gives[§]

$$\Pi_i(\mathbf{r}_0) = \frac{1}{4\pi c^2} \dot{A}_i(\mathbf{r}_0) = -\frac{E_i(\mathbf{r}_0)}{4\pi c} \quad (18.5.34)$$

or in vector form (dropping the subscript 0 on \mathbf{r})

$$\mathbf{\Pi}(\mathbf{r}) = \frac{1}{4\pi c^2} \dot{\mathbf{A}} = -\frac{\mathbf{E}}{4\pi c} \quad (18.5.35)$$

Note that $\mathbf{\Pi}$ is essentially the electric field.

The natural thing to do at this point would be to promote \mathbf{A} and $\mathbf{\Pi}$ to quantum operators obeying canonical commutation rules, and obtain the quantum Hamiltonian H by the substitution rule. But if we did this, we would not be dealing with

[‡] See for example, H. Goldstein, *Classical Mechanics*, Addison-Wesley, Reading, Massachusetts (1965), page 366.

[§] A more formal treatment is the following. If, say, $\mathcal{L} = \sum_i \dot{q}_i^2$, then we know

$$p_j = \frac{\partial \mathcal{L}}{\partial \dot{q}_j} = \sum_i 2\dot{q}_i \frac{\partial \dot{q}_i}{\partial \dot{q}_j} = \sum_i 2\dot{q}_i \delta_{ij} = 2\dot{q}_j$$

Likewise if

$$\begin{aligned} \mathcal{L} &= \int \sum_i \dot{A}_i^2(\mathbf{r}) d^3\mathbf{r} \\ \frac{\partial \mathcal{L}}{\partial \dot{A}_j(\mathbf{r}_0)} &= \int \sum_i 2\dot{A}_i(\mathbf{r}) \frac{\partial \dot{A}_i(\mathbf{r})}{\partial \dot{A}_j(\mathbf{r}_0)} d^3\mathbf{r} \\ &= \int \sum_i 2\dot{A}_i(\mathbf{r}) \delta_{ij} \delta^3(\mathbf{r} - \mathbf{r}_0) d^3\mathbf{r} = 2\dot{A}_j(\mathbf{r}_0) \end{aligned}$$

electrodynamics. The reason is classical and is as follows. Consider the Lagrangian in Eq. (18.5.33) with ϕ set equal to 0. By varying it with respect to the components of \mathbf{A} we get the vector equation

$$\nabla^2 \mathbf{A} - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} - \nabla(\nabla \cdot \mathbf{A}) = 0 \quad (18.5.36)$$

which is just

$$\nabla \times \mathbf{B} - \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t} = 0$$

in the gauge with $\phi = 0$. Two other equations

$$\begin{aligned} \nabla \cdot \mathbf{B} &= 0 \\ \nabla \times \mathbf{E} + \frac{1}{c} \frac{\partial \mathbf{B}}{\partial t} &= 0 \end{aligned}$$

are identically satisfied if we write \mathbf{E} and \mathbf{B} in terms of \mathbf{A} . (Recall how the potentials were introduced in the first place.) As for the other Maxwell equation, Gauss's law,

$$\nabla \cdot \mathbf{E} = 0$$

it does not follow from anything. (We would get this if we varied \mathcal{L} with respect to ϕ , but we have eliminated ϕ from the picture.) It must therefore be appended as an *equation of constraint* on the momentum Π , which is just \mathbf{E} times a constant. (In contrast to an equation of motion, which has time derivatives in it, an equation of constraint is a relation among the variables at a given time. It signifies that the variables are not independent.) The constraint

$$\nabla \cdot \Pi = 0 \quad (18.5.37)$$

tells us that the components of momenta at nearby points are not independent. (Think of the derivatives in ∇ as differences.) We deduce an important feature of the constraint if we take the divergence of Eq. (18.5.36):

$$0 = \nabla \cdot \nabla^2 \mathbf{A} - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \nabla \cdot \mathbf{A} - \nabla^2 \nabla \cdot \mathbf{A} \rightarrow \frac{\partial}{\partial t} (\nabla \cdot \Pi) = 0 \quad (18.5.38)$$

In other words, the theory without the constraint has a conserved quantity $\nabla \cdot \Pi$, and electrodynamics corresponds to the subset of trajectories in which this constant of motion is zero. Furthermore, if we limit ourselves to these trajectories, we see that $\nabla \cdot \mathbf{A}$ is also a constant of motion. [Write Eq. (18.5.37) as $\nabla \cdot \dot{\mathbf{A}} = 0$.] We shall choose this constant to be zero, i.e., work in Coulomb gauge.

How are we to quantize this theory? One way is to ignore the constraints and to quantize the general theory and then try to pick out the subset of solutions (in

the quantum theory) that correspond to electrodynamics. This can be done but is very hard. Let us therefore tackle the constraints at the classical level. The first problem they pose is that they render the variables \mathbf{A} and $\mathbf{\Pi}$ noncanonical, and we do not have a recipe for quantizing noncanonical variables. Let us verify that the constraints indeed imply that \mathbf{A} and $\mathbf{\Pi}$ are noncanonical. Had they been canonical, they would have obeyed the generalizations of

$$\{q_i, p_j\} = \delta_{ij}$$

(with all other PB zero), namely,

$$\{A_i(\mathbf{r}), \Pi_j(\mathbf{r}')\} = \delta_{ij}\delta^3(\mathbf{r} - \mathbf{r}') \quad (18.5.39)$$

(with all other PB zero.) But if we take the divergence of \mathbf{A} with respect to \mathbf{r} or $\mathbf{\Pi}$ with respect to \mathbf{r}' , we get zero on the left-hand side but not on the right-hand side.

What we would like to do is the following. We would like to trade \mathbf{A} and $\mathbf{\Pi}$ for a new set of variables that are fewer in number but have the constraints built into them. (This would be like trading the variables x , y , and z constrained by $x^2 + y^2 + z^2 = a^2$ for the angles θ and ϕ on a sphere of radius a .) These variables and the corresponding momenta would be canonical and would automatically reproduce electrodynamics if we start with \mathcal{H} written in terms of these. To quantize, we promote these variables to operators obeying canonical commutation rules. The Hamiltonian and other operators would then be obtained by the substitution rule.

Now the problem with the constraints

$$\nabla \cdot \mathbf{A} = 0, \quad \nabla \cdot \mathbf{\Pi} = 0 \quad (18.5.40)$$

called *transversality constraints* (for a reason that will follow) is that they are not algebraic, but differential equations. To render them algebraic, we will trade \mathbf{A} and $\mathbf{\Pi}$ for their Fourier transforms, since differential equations in coordinate space become algebraic when Fourier transformed. It is our hope that the algebraic constraints among the Fourier coefficients will be easier to implement. We will find that this is indeed the case. We will also find a bonus when we are done: the Fourier coefficients are normal coordinates; i.e., when we express the Hamiltonian

$$\mathcal{H} = \frac{1}{8\pi} \int [16\pi^2 c^2 |\mathbf{\Pi}|^2 + |\nabla \times \mathbf{A}|^2] d^3\mathbf{r} \quad (18.5.41)$$

(which is obtained from \mathcal{L} by changing the sign of the potential energy term and eliminating \mathbf{A} in favor of $\mathbf{\Pi}$) in terms of these, it becomes a sum over oscillator Hamiltonians of decoupled oscillators. This result could have been anticipated for the following reason. If we use the relation $|\nabla \times \mathbf{A}|^2 = -\mathbf{A} \cdot \nabla^2 \mathbf{A}$, valid when $\nabla \cdot \mathbf{A} = 0$, we get

$$\mathcal{H} = \frac{1}{8\pi} \int \sum_i \sum_j [16\pi^2 c^2 \Pi_i(\mathbf{r}) \delta_{ij} \Pi_j(\mathbf{r}) - A_i(\mathbf{r}) \nabla^2 \delta_{ij} A_j(\mathbf{r})] d^3\mathbf{r} \quad (18.5.42)$$

which is of the same form as Eq. (7.1.10). [Remember that when we sandwich the derivative operator or the identity operator between two elements of function space, there will be only one (explicit) sum over the continuous index \mathbf{r} , the other one being eaten up by the delta functions in the matrix elements.] As the normal modes are the eigenvectors of ∇^2 (which we know are plane waves) the passage to the Fourier coefficients is the passage to normal coordinates.

With all these preliminaries out of the way, let us turn to the Fourier transform of the *unconstrained A*:

$$\mathbf{A}(\mathbf{r}) = \int [\mathbf{a}(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{r}} + \mathbf{a}^*(\mathbf{k}) e^{-i\mathbf{k}\cdot\mathbf{r}}] d^3k \quad (18.5.43)$$

This expansion deserves a few comments.

(1) Since we are Fourier transforming a vector \mathbf{A} , the Fourier coefficients are vectors $\mathbf{a}(\mathbf{k})$. [You may view Eq. (18.5.43) as giving three Fourier expansions, one for each component of \mathbf{A} .]

(2) Since $\mathbf{A}(\mathbf{r})$ is a *real* function, the Fourier coefficient at \mathbf{k} and $-\mathbf{k}$ must be complex conjugates. Our expansion makes this apparent. Stated differently, one *real* vector function \mathbf{A} in coordinate space cannot specify one *complex* vector function $\mathbf{a}(\mathbf{k})$ in \mathbf{k} space: if we multiply both sides with $e^{-i\mathbf{k}_0\cdot\mathbf{r}}$ and integrate over \mathbf{r} , we find that this is indeed the case:

$$\int e^{-i\mathbf{k}_0\cdot\mathbf{r}} \mathbf{A}(\mathbf{r}) d^3r = (2\pi)^3 [\mathbf{a}(\mathbf{k}_0) + \mathbf{a}^*(-\mathbf{k}_0)] \quad (18.5.44)$$

i.e., $\mathbf{A}(\mathbf{r})$ is seen to determine only the combination $\mathbf{a}(\mathbf{k}) + \mathbf{a}^*(-\mathbf{k})$. We shall exploit this point shortly.

(3) There is no time argument shown in Eq. (18.5.43) because we view it as linear relations between two sets of coordinates, such as the relations

$$x_1 = \frac{x_I + x_{II}}{2^{1/2}}$$

$$x_2 = \frac{x_I - x_{II}}{2^{1/2}}$$

which are understood to be true at all times. The discrete labels 1, 2, I, and II are replaced here by the continuous labels \mathbf{r} and \mathbf{k} .

We similarly expand Π (before the transversality constraint is imposed) as

$$\Pi(\mathbf{r}) = \frac{1}{4\pi i c} \int k [\mathbf{a}(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{r}} - \mathbf{a}^*(\mathbf{k}) e^{-i\mathbf{k}\cdot\mathbf{r}}] d^3k \quad (18.5.45)$$

The factor $(k/4\pi i c)$ is pulled out to simplify future manipulations. Note that the same function $\mathbf{a}(\mathbf{k})$ appears here. There is no conflict, since $\Pi(\mathbf{r})$ determines a different

combination:

$$\int \Pi(\mathbf{r}) e^{-i\mathbf{k}_0 \cdot \mathbf{r}} d^3 \mathbf{r} = \frac{(2\pi)^3}{4\pi c} k_0 [\mathbf{a}(\mathbf{k}_0) - \mathbf{a}^*(-\mathbf{k}_0)] \quad (18.5.46)$$

It is clear that Eqs. (18.5.44) and (18.5.46) may be solved for $\mathbf{a}(\mathbf{k})$ in terms of \mathbf{A} and Π : the two real vector functions $\mathbf{A}(\mathbf{r})$ and $\Pi(\mathbf{r})$ determine one complex vector function $\mathbf{a}(\mathbf{k})$. Consider now the vector $\mathbf{a}(\mathbf{k})$ at a given \mathbf{k} . We can expand it in terms of any three orthonormal vectors. Rather than choose them to be the unit vectors along the x , y , and z directions, let us choose them (with an eye on the constraints) as a function of \mathbf{k} , in the following way:

$$\begin{aligned} \left. \begin{aligned} \mathbf{\epsilon}(\mathbf{k}1) \\ \mathbf{\epsilon}(\mathbf{k}2) \end{aligned} \right\} & \text{ orthonormal vectors in the plane perpendicular to } \mathbf{k} \\ \mathbf{\epsilon}(\mathbf{k}3) & \text{ a unit vector parallel to } \mathbf{k} \end{aligned} \quad (18.5.47)$$

If we now expand $\mathbf{a}(\mathbf{k})$ (at each \mathbf{k}) as

$$\mathbf{a}(\mathbf{k}) = \sum_{\lambda=1}^3 (c^2/4\pi^2\omega)^{1/2} a(\mathbf{k}\lambda) \mathbf{\epsilon}(\mathbf{k}\lambda) \quad (18.5.48)$$

(where $\omega = kc$) and feed this into the expansions for \mathbf{A} and Π , we get

$$\mathbf{A}(\mathbf{r}) = \sum_{\lambda} \int \left(\frac{c^2}{4\pi^2\omega} \right)^{1/2} [a(\mathbf{k}\lambda) \mathbf{\epsilon}(\mathbf{k}\lambda) e^{i\mathbf{k}\cdot\mathbf{r}} + a^*(\mathbf{k}\lambda) \mathbf{\epsilon}(\mathbf{k}\lambda) e^{-i\mathbf{k}\cdot\mathbf{r}}] d^3 \mathbf{k} \quad (18.5.49a)$$

$$\Pi(\mathbf{r}) = \sum_{\lambda} \int \frac{1}{i} \left(\frac{\omega}{64\pi^4 c^2} \right)^{1/2} [a(\mathbf{k}\lambda) \mathbf{\epsilon}(\mathbf{k}\lambda) e^{i\mathbf{k}\cdot\mathbf{r}} - a^*(\mathbf{k}\lambda) \mathbf{\epsilon}(\mathbf{k}\lambda) e^{-i\mathbf{k}\cdot\mathbf{r}}] d^3 \mathbf{k} \quad (18.5.49b)$$

These equations relate the old coordinates—three real components of \mathbf{A} and three real components of Π at each point in \mathbf{r} space—to three complex components of \mathbf{a} at each point in \mathbf{k} space. Since \mathbf{A} and Π are canonical variables before we impose transversality, their PB are

$$\begin{aligned} \{A_i(\mathbf{r}), A_j(\mathbf{r}')\} &= 0 \\ \{\Pi_i(\mathbf{r}), \Pi_j(\mathbf{r}')\} &= 0 \\ \{A_i(\mathbf{r}), \Pi_j(\mathbf{r}')\} &= \delta_{ij} \delta^3(\mathbf{r} - \mathbf{r}') \end{aligned} \quad (18.5.50)$$

From these we may deduce (after some hard work) that

$$\begin{aligned} \{a(\mathbf{k}\lambda), a(\mathbf{k}'\lambda')\} &= 0 = \{a^*(\mathbf{k}\lambda), a^*(\mathbf{k}'\lambda')\} \\ \{a(\mathbf{k}\lambda), a^*(\mathbf{k}'\lambda')\} &= -i\delta_{\lambda\lambda'}\delta^3(\mathbf{k} - \mathbf{k}') \end{aligned} \quad (18.5.51)$$

We now address the problem of imposing the constraints, i.e., of regaining electrodynamics. The conditions $\nabla \cdot \mathbf{A} = 0$ and $\nabla \cdot \mathbf{\Pi} = 0$ tell us [when we apply them to Eqs. (18.5.43) and (18.5.45) and project both sides onto some given \mathbf{k}],

$$\begin{aligned}\mathbf{k} \cdot [\mathbf{a}(\mathbf{k}) + \mathbf{a}^*(-\mathbf{k})] &= 0 \\ \mathbf{k} \cdot [\mathbf{a}(\mathbf{k}) - \mathbf{a}^*(-\mathbf{k})] &= 0\end{aligned}$$

from which we deduce that

$$\mathbf{k} \cdot \mathbf{a}(\mathbf{k}) = 0 \quad (18.5.52)$$

The two differential equations of constraint have reduced, as anticipated, to (a complex) algebraic constraint. Imposing it on Eq. (18.5.48), we find [using $\mathbf{k} \cdot \boldsymbol{\varepsilon}(\mathbf{k}, 1 \text{ or } 2) = 0$],

$$a(\mathbf{k}3) = 0 \quad (18.5.53)$$

Thus the constraint tells us something very simple: every $a(\mathbf{k}3)$ is zero. (Since it forces $\mathbf{a}(\mathbf{k})$ to lie in a plane transverse to \mathbf{k} , we call it the transversality constraint). Implementation of the transversality constraint is very simple in momentum space: hereafter we let λ take on only the values 1 and 2. Also, setting $a(\mathbf{k}3) = 0$ does not change the PB between the remaining a 's. Equation (18.5.49) for \mathbf{A} and $\mathbf{\Pi}$ continues to hold, with λ so restricted. However, these fields are now guaranteed to meet the transversality conditions.

Now for the other nice feature of these conditions. If we express \mathcal{H} in terms of these, we get

$$\mathcal{H} = \sum_{\lambda=1}^2 \int \omega [a^*(\mathbf{k}\lambda) a(\mathbf{k}\lambda)] d^3\mathbf{k} \quad (18.5.54)$$

Thus $a(\mathbf{k}\lambda)$ are normal coordinates in the sense that \mathcal{H} contains no cross terms between a 's carrying different labels. If we want to get the familiar oscillators, we define real variables

$$\begin{aligned}q(\mathbf{k}\lambda) &= \frac{1}{(2\omega)^{1/2}} [a(\mathbf{k}\lambda) + a^*(\mathbf{k}\lambda)] \\ p(\mathbf{k}\lambda) &= \frac{1}{i} \left(\frac{\omega}{2} \right)^{1/2} [a(\mathbf{k}\lambda) - a^*(\mathbf{k}\lambda)]\end{aligned} \quad (18.5.55)$$

which satisfy the canonical PB relations [as you may verify by combining Eqs. (18.5.51) and (18.5.55)]. In terms of these variables

$$\mathcal{H} = \sum_{\lambda} \int \left[\frac{1}{2} p^2(\mathbf{k}\lambda) + \frac{\omega^2}{2} q^2(\mathbf{k}\lambda) \right] d^3\mathbf{k} \quad (18.5.56)$$

Thus we find that the radiation field is equivalent to a collection of decoupled oscillators: there is an oscillator at each \mathbf{k} and λ ($=1$ or 2) with frequency $\omega = kc$. The quantization of the radiation field then reduces to the quantization of the oscillator, which has already been accomplished in Chapter 7.

Since $q(\mathbf{k}\lambda)$ and $p(\mathbf{k}\lambda)$ are independent canonical coordinates describing the field, we can quantize the field by promoting these to operators Q and P obeying canonical commutation rules:

$$[Q(\mathbf{k}\lambda), P(\mathbf{k}'\lambda')] = i\hbar\{q, p\} = i\hbar\delta_{\lambda\lambda'}\delta^3(\mathbf{k} - \mathbf{k}')$$

with all other commutators vanishing. As in the case of a single oscillator, it proves useful to work with the combination

$$a(\mathbf{k}\lambda) = \left(\frac{\omega}{2\hbar}\right)^{1/2} Q + i\left(\frac{1}{2\omega\hbar}\right)^{1/2} P$$

and its adjoint

$$a^\dagger(\mathbf{k}\lambda) = \left(\frac{\omega}{2\hbar}\right)^{1/2} Q - i\left(\frac{1}{2\omega\hbar}\right)^{1/2} P \quad (18.5.57)\ddagger$$

which obey

$$[a(\mathbf{k}\lambda), a^\dagger(\mathbf{k}'\lambda')] = \delta_{\lambda\lambda'}\delta^3(\mathbf{k} - \mathbf{k}') \quad (18.5.58)$$

and in terms of which \mathbf{A} and Π ,[§] which are now Hermitian operators, are given by

$$\mathbf{A} = \sum_\lambda \int \left(\frac{\hbar c^2}{4\pi^2\omega}\right)^{1/2} [a(\mathbf{k}\lambda)\epsilon(\mathbf{k}\lambda)e^{i\mathbf{k}\cdot\mathbf{r}} + a^\dagger(\mathbf{k}\lambda)\epsilon(\mathbf{k}\lambda)e^{-i\mathbf{k}\cdot\mathbf{r}}] d^3\mathbf{k} \quad (18.5.59a)$$

$$\Pi = \sum_\lambda \int \frac{1}{i} \left(\frac{\hbar\omega}{64\pi^4c^2}\right)^{1/2} [a(\mathbf{k}\lambda)\epsilon(\mathbf{k}\lambda)e^{i\mathbf{k}\cdot\mathbf{r}} - a^\dagger(\mathbf{k}\lambda)\epsilon(\mathbf{k}\lambda)e^{-i\mathbf{k}\cdot\mathbf{r}}] d^3\mathbf{k} \quad (18.5.59b)$$

To find H , we first symmetrize \mathcal{H} , i.e., $a^*a \rightarrow \frac{1}{2}(a^*a + aa^*)$, make the operator substitution, and use Eq. (18.5.58), to get

$$H = \sum_\lambda \int [a^\dagger(\mathbf{k}\lambda)a(\mathbf{k}\lambda) + \frac{1}{2}]\hbar\omega d^3\mathbf{k} \quad (18.5.60)$$

[†] A small point, in case you are following all the details: a and a^\dagger above are the operators corresponding to the classical variables $a/\hbar^{1/2}$ and $a^*/\hbar^{1/2}$. To see this, invert Eq. (18.5.55). All we need hereafter are Eqs. (18.5.57)–(18.5.59).

[§] We use the same symbols for the classical and quantum variables in order to follow a widely used convention in this case. It should be clear from the context which is which.

Let us now consider the eigenstates of H . In the field ground state $|0\rangle$, all the oscillators are in their respective ground states. Thus any lowering operator will annihilate $|0\rangle$:

$$a(\mathbf{k}\lambda)|0\rangle = 0 \quad \text{for all } \mathbf{k}, \lambda \quad (18.5.61)$$

The energy of this state, called the vacuum state or simply *vacuum*, is

$$E_0 = \sum_{\lambda} \int \frac{\hbar\omega}{2} d^3\mathbf{k} \quad (18.5.62)$$

which is the sum over the zero point energies of the oscillators. This constant energy E_0 has no physical consequences.

We now verify the results claimed earlier. In this ground state

$$\begin{aligned} \langle 0 | \mathbf{A} | 0 \rangle &\sim \langle 0 | (a + a^\dagger) | 0 \rangle = 0 \\ \langle 0 | \mathbf{\Pi} | 0 \rangle &\sim \langle 0 | (a - a^\dagger) | 0 \rangle = 0 \end{aligned} \quad (18.5.63)$$

In the above equation we have omitted a lot of irrelevant factors; only the central idea—that \mathbf{A} and $\mathbf{\Pi}$ are linear combinations of creation and destruction operators and hence have no diagonal matrix elements in $|0\rangle$ —is emphasized. On the other hand,

$$\begin{aligned} \langle 0 | |\mathbf{A}|^2 | 0 \rangle &\neq 0 \\ \langle 0 | |\mathbf{\Pi}|^2 | 0 \rangle &\neq 0 \end{aligned} \quad (18.5.64)$$

for the same reason that $\langle X^2 \rangle \neq 0$, $\langle P^2 \rangle \neq 0$ for a single oscillator.

If we act on $|0\rangle$ with one of the raising operators, we get

$$a^\dagger(\mathbf{k}\lambda)|0\rangle = |\mathbf{k}\lambda\rangle \quad (18.5.65)$$

where the labels \mathbf{k} and λ tell us that the oscillator bearing that label has gone to its first excited level. This state has energy $\hbar\omega = \hbar k c$ above E_0 as may be verified by letting H act on it and using Eqs. (18.5.58) and (18.5.61). What about the momentum content? Any standard textbook on electrodynamics will tell us that the momentum of the field is given, in classical physics, by

$$\mathcal{P} = \frac{1}{4\pi c} \int (\mathbf{E} \times \mathbf{B}) d^3\mathbf{r} \quad (18.5.66)$$

If we calculate the corresponding quantum operator we will find that it is given by

$$\mathbf{P} = \sum_{\lambda} \int [a^\dagger(\mathbf{k}\lambda)a(\mathbf{k}\lambda)] \hbar \mathbf{k} d^3\mathbf{k} \quad (18.5.67)$$

It is clear on inspection or explicit operation that

$$\mathbf{P}|\mathbf{k}\lambda\rangle = \hbar\mathbf{k}|\mathbf{k}\lambda\rangle \quad (18.5.68)$$

Thus the state $|\mathbf{k}\lambda\rangle$ has momentum $\hbar\mathbf{k}$.

If we apply $a^\dagger(\mathbf{k}\lambda)$ on the vacuum n times, we will create a state with energy $n\hbar\omega$ and momentum $n\hbar\mathbf{k}$. This allows us to view the action of $a^\dagger(\mathbf{k}\lambda)$ as the creation of particles of momenta $\hbar\mathbf{k}$ and energy $\hbar\omega$. These particles, called photons, are massless since

$$m^2c^4 = E^2 - c^2p^2 = (\hbar\omega)^2 - (\hbar kc)^2 = 0 \quad (18.5.69)$$

In terms of photons, we have the correspondence

$$\left\{ \begin{array}{l} \text{quantum state} \\ \text{of field} \end{array} \right\} \leftrightarrow \left\{ \begin{array}{l} \text{quantum state of} \\ \text{each oscillator} \end{array} \right\} \leftrightarrow \left\{ \begin{array}{l} \text{number of photons} \\ \text{at each } \mathbf{k} \text{ and } \lambda \end{array} \right\}$$

For future use, let us obtain the wave function of the photon in the state (\mathbf{k}, λ) . We begin by deducing the normalization of the states. Combining Eqs. (18.5.65) and (18.5.58) we get

$$\begin{aligned} \langle \mathbf{k}'\lambda' | \mathbf{k}\lambda \rangle &= \langle 0 | a(\mathbf{k}'\lambda') a^\dagger(\mathbf{k}\lambda) | 0 \rangle \\ &= \langle 0 | a^\dagger a + \delta_{\lambda\lambda'} \delta^3(\mathbf{k} - \mathbf{k}') | 0 \rangle \\ &= \delta_{\lambda\lambda'} \delta^3(\mathbf{k} - \mathbf{k}') \end{aligned} \quad (18.5.70)$$

(assuming $\langle 0 | 0 \rangle = 1$). The $\delta^3(\mathbf{k} - \mathbf{k}')$ factor and the fact that $\hbar\mathbf{k}$ is the momentum of the state tell us that the wave function corresponding to $|\mathbf{k}, \lambda\rangle$ is

$$\psi \sim \frac{1}{(2\pi)^{3/2}} e^{i\mathbf{k}\cdot\mathbf{r}} \quad (18.5.71)$$

We use the \sim sign instead of the \rightarrow sign because λ has not entered the wave function yet. From the $\delta_{\lambda\lambda'}$ factor and the way λ entered the picture in the first place, we conclude that λ represents the polarization vector:

$$|\mathbf{k}\lambda\rangle \rightarrow \frac{\boldsymbol{\epsilon}(\mathbf{k}\lambda) e^{i\mathbf{k}\cdot\mathbf{r}}}{(2\pi)^{3/2}} \quad (18.5.72)$$

You may be unhappy over the fact that unlike the $e^{i\mathbf{k}\cdot\mathbf{r}}/(2\pi)^{3/2}$ factor, which followed from analyzing the momentum content of the state [i.e., from the analysis of Eq. (18.5.68)], the $\boldsymbol{\epsilon}$ was pulled out of a hat. It too may be deduced, starting with angular momentum considerations. We do not do so here.

Since the wave function of the photon is not a scalar, it has spin. Furthermore, since $\boldsymbol{\epsilon}$ is a three-component object, the spin is unity. However, the requirement that $\mathbf{k} \cdot \boldsymbol{\epsilon} = 0$ imposes a constraint on the possible orientations of photon spin. Consider, for example, a photon moving along the z axis. The condition $\mathbf{k} \cdot \boldsymbol{\epsilon} = 0$ tells us that

ϵ cannot have a component along the z axis. What does this mean? The component of ϵ parallel to the z axis is characterized by the fact that it remains invariant under rotations around the z axis, i.e., transforms like an $s_z=0$ state. So we conclude that the photon can have only $s_z=\pm\hbar$, but not $s_z=0$. More generally, the spin of the photon can only take values $\pm\hbar$ parallel to its momentum. The component of spin parallel to momentum is called *helicity*. The transversality condition restricts the helicity to be $\pm\hbar$ —it precludes helicity zero.[‡]

We consider one last feature of photons before turning to the problem that started this inquiry, namely, spontaneous decay. Consider a state with one photon in $(\mathbf{k}\lambda)$ and another in $(\mathbf{k}'\lambda')$:

$$|\mathbf{k}\lambda, \mathbf{k}'\lambda'\rangle = a^\dagger(\mathbf{k}\lambda)a^\dagger(\mathbf{k}'\lambda')|0\rangle \quad (18.5.73)$$

If we exchange the photon states we get the state

$$|\mathbf{k}'\lambda', \mathbf{k}\lambda\rangle = a^\dagger(\mathbf{k}'\lambda')a^\dagger(\mathbf{k}\lambda)|0\rangle \quad (18.5.74)$$

But since $[a^\dagger, a^\dagger]=0$, the two state vectors coincide, as they should for identical bosons.

Spontaneous Decay

Consider the spontaneous decay of the hydrogen atom from $|2lm\rangle$ to $|100\rangle$. The perturbing Hamiltonian is still given by the substitution rule

$$\mathcal{H}^1 = \frac{e}{mc} \mathbf{A} \cdot \mathbf{p} \rightarrow H^1 = \frac{e}{mc} \mathbf{A} \cdot \mathbf{P} \quad (18.5.75)$$

but the \mathbf{A} in H^1 is now the operator in Eq. (18.5.59a).

The initial state of the system (atom + field) is

$$|i^0\rangle = |2lm\rangle \otimes |0\rangle \quad (18.5.76)$$

The final state is

$$|f^0\rangle = |100\rangle \otimes |\mathbf{k}\lambda\rangle \quad (18.5.77)$$

The perturbation H^1 is time independent (\mathbf{A} is the operator in the Schrödinger picture) and

$$E_f^0 - E_i^0 = E_{100} + \hbar\omega - E_{2lm} \quad (18.5.78)$$

[‡] The graviton, which is massless and has spin 2, also has only two helicity states, $\pm 2\hbar$. This is a general feature of massless bosons with spin.

From Fermi's golden rule, we get[‡]

$$R_{i \rightarrow f} = \frac{2\pi}{\hbar} \left| \langle f^0 | \frac{e}{mc} \mathbf{A} \cdot \mathbf{P} | i^0 \rangle \right|^2 \delta(E_{100} + \hbar\omega - E_{2lm}) \quad (18.5.79)$$

Consider

$$\langle f^0 | \mathbf{A} \cdot \mathbf{P} | i^0 \rangle = \langle 100 | \langle \mathbf{k}\lambda | \mathbf{A} | 0 \rangle \cdot \mathbf{P} | 2lm \rangle \quad (18.5.80)$$

Now, \mathbf{A} is a sum over a 's and a^\dagger 's with different labels. The only relevant one is $a^\dagger(\mathbf{k}\lambda)$, which raises $|0\rangle$ to $|\mathbf{k}\lambda\rangle$. Thus, including the factors that accompany $a^\dagger(\mathbf{k}\lambda)$,

$$\langle \mathbf{k}\lambda | \mathbf{A} | 0 \rangle = \left(\frac{\hbar c^2}{4\pi^2 \omega} \right)^{1/2} \epsilon(\mathbf{k}\lambda) e^{i\mathbf{k}\cdot\mathbf{r}} \quad (18.5.81)$$

so that

$$\langle f^0 | \mathbf{A} \cdot \mathbf{P} | i^0 \rangle = \left(\frac{\hbar c^2}{4\pi^2 \omega} \right)^{1/2} \int \psi_{100}^* e^{i\mathbf{k}\cdot\mathbf{r}} \epsilon \cdot (-i\hbar\nabla) \psi_{2lm} d^3\mathbf{r}$$

In the dipole approximation, this becomes, upon using Eq. (18.5.17),[§]

$$\langle f^0 | \mathbf{A} \cdot \mathbf{P} | i^0 \rangle = \left(\frac{\hbar c^2}{4\pi^2 \omega} \right)^{1/2} (im\omega) \int \psi_{100}^* \epsilon \cdot \mathbf{r} \psi_{2lm} d^3\mathbf{r} \quad (18.5.82)$$

From parity considerations, it is clear that only $l=1$ is relevant. Writing $\epsilon \cdot \mathbf{r}$ in the spherical basis (recall Exercise 15.3.2),

$$\begin{aligned} \epsilon \cdot \mathbf{r} &= \sum_{-1}^{+1} (-1)^q \epsilon_1^q r_1^{-q} \\ &= -\epsilon_1^1 r_1^{-1} + \epsilon_1^0 r_1^0 - \epsilon_1^{-1} r_1^{+1} \end{aligned} \quad (18.5.83)$$

where

$$\epsilon_1^{\pm 1} = \mp \frac{\epsilon_x \pm i\epsilon_y}{2^{1/2}}, \quad \epsilon_1^0 = \epsilon_z \quad (18.5.84)$$

[‡] In the photoelectric effect, the field is treated as an external time-dependent perturbation that acts on the atom, and the $\hbar\omega$ in the delta function reflects this time dependence. In the present case, the field is part of the system and the $\hbar\omega$ stands for the change in its energy.

[§] We are unfortunately forced to use the symbol m for the mass as well as the z component of angular momentum. It should be clear from the context what m stands for.

and from Eq. (12.5.42),

$$r_1^{\pm 1} = \left(\frac{4\pi}{3}\right)^{1/2} r Y_1^{\pm 1}, \quad r_1^0 = \left(\frac{4\pi}{3}\right)^{1/2} r Y_1^0 \quad (18.5.85)$$

we get

$$\begin{aligned} \int \psi_{100}^* \boldsymbol{\epsilon} \cdot \mathbf{r} \psi_{21m} d^3 \mathbf{r} &= \left(\frac{4\pi}{3}\right)^{1/2} \int R_{10} r R_{21} r^2 dr \\ &\times \left[\int Y_0^{0*} (-\varepsilon_1^1 Y_1^{-1} + \varepsilon_1^0 Y_1^0 - \varepsilon_1^{-1} Y_1^{+1}) Y_1^m d\Omega \right] \\ &= \left(\frac{3}{2}\right)^{1/2} \frac{2^8}{3^5} \frac{a_0}{3^{1/2}} (+\varepsilon_1^1 \delta_{m,+1} + \varepsilon_1^0 \delta_{m,0} + \varepsilon_1^{-1} \delta_{m,-1}) \quad (18.5.86) \end{aligned}$$

The evaluation of the integrals (like so many other steps in this high-speed treatment) is left as an exercise. The modulus squared of the above quantity is

$$\frac{3}{2} \frac{2^{16}}{3^{10}} \frac{a_0^2}{3} [|\varepsilon_1^1|^2 \delta_{m,-1} + |\varepsilon_1^0|^2 \delta_{m,0} + |\varepsilon_1^{-1}|^2 \delta_{m,1}]$$

If we average over the three initial m 's (i.e., over an ensemble of such atoms randomly distributed with respect to m), this reduces to

$$\frac{3}{2} \frac{2^{16}}{3^{10}} \frac{a_0^2}{3} \frac{1}{3} (\varepsilon_x^2 + \varepsilon_y^2 + \varepsilon_z^2) = \frac{2^{15} a_0^2}{3^{11}} \quad (18.5.87)$$

Notice that the result is independent of the direction of $\boldsymbol{\epsilon}$. This is to be expected since the atom has no sense of direction after the angular (m) averaging. The transition rate is

$$R_{i \rightarrow f} = \frac{2\pi}{\hbar} \left(\frac{e}{mc}\right)^2 \frac{\hbar c^2}{4\pi^2 \omega} m^2 \omega^2 \frac{2^{15} a_0^2}{3^{11}} \delta(E_{100} + \hbar\omega - E_{2lm}) \quad (18.5.88)$$

where \bar{i} means the initial state is averaged over all orientations.

If we sum over all possible photon momenta and two possible polarizations at each momentum, we get, using

$$\int \delta(E_{100} + \hbar\omega - E_{2lm}) k^2 dk d\Omega = \frac{4\pi k^2}{\hbar c}$$

where

$$k = \frac{\omega}{c} = \frac{E_{2lm} - E_{100}}{\hbar c} = \frac{e^2}{2a_0\hbar c} \left(1 - \frac{1}{4}\right) = \frac{3e^2}{8a_0\hbar c}$$

the total decay rate

$$R_{i \rightarrow \text{all}} = \left(\frac{2}{3}\right)^8 \alpha^5 \frac{mc^2}{\hbar} \quad (18.5.89)$$

Recall that

$$\begin{aligned} \frac{mc^2}{\hbar} &= \frac{mc^2 c}{\hbar c} \simeq \frac{0.5 \times 10^6 \text{ eV } c}{2000 \text{ eV } \text{\AA}} \\ &\simeq 0.25 \times 10^3 \text{ \AA}^{-1} c \end{aligned}$$

Now $c = 3 \times 10^{10} \text{ cm/sec} = 3 \times 10^{18} \text{ \AA/sec}$. So

$$\frac{mc^2}{\hbar} \simeq 10^{21} \text{ sec}^{-1}$$

and

$$\begin{aligned} R_{i \rightarrow \text{all}} &\simeq (0.67)^8 \left(\frac{1}{137}\right)^5 10^{21} \text{ seconds}^{-1} \\ &\simeq 0.6 \times 10^9 \text{ seconds}^{-1} \end{aligned}$$

The corresponding mean lifetime is

$$\tau = 1/R \simeq 1.6 \times 10^{-9} \text{ seconds} \quad (18.5.90)$$

in excellent agreement with experiment.

Even if the fields are macroscopic, we can use the full quantum theory, though the semiclassical treatment will give virtually identical results. The relation of the two approaches may be described as follows. Consider a process in which an atom goes from the state i_a to the state f_a and the field goes from the state with n photons in (\mathbf{k}, λ) to $n+1$ photons in (\mathbf{k}, λ) .[†] The result we get in the quantum mechanical treatment of this process, which involves the *emission* of a photon, will agree with the semiclassical calculation if we use a classical field \mathbf{A} whose energy density[§] is the same as that of $(n+1)$ photons in (\mathbf{k}, λ) . The 1 in $n+1$ is all important at small n , and contains the key to spontaneous decay. If we consider a process where a photon is *absorbed*, so that $n \rightarrow n-1$, the semiclassical method gives the correct answer if we

[†] We do not concern ourselves with other modes, which are spectators.

[§] The wavelength and polarization are of course the same as that of the photons.

use a classical field \mathbf{A} such that the energy density is that of the n photons. The appearance of the $(n+1)$ and n factors is easy to understand in the oscillator language. When a photon is created, the amplitude goes as

$$\langle n+1|a^\dagger|n\rangle = (n+1)^{1/2} \langle n+1|n+1\rangle \quad (18.5.91)$$

which gives the factor $(n+1)$ in the probability, while if it is destroyed,

$$\langle n-1|a|n\rangle = n^{1/2} \langle n-1|n-1\rangle \quad (18.5.92)$$

which gives a factor n in the probability.

It is conventional to separate the emission probability proportional to $n+1$ into the *probability for induced emission*, proportional to n , and the *probability for spontaneous emission*, proportional to 1. The induced emission is induced by the preexisting photons, and the spontaneous emission is—well, spontaneous.

The $(n+1)$ factor in the emission probability is a feature of bosons in general: the probability of a system emitting a boson into a quantum state already occupied by n bosons (of the same kind), is $(n+1)$ times larger than the probability of emission into that state if it is initially unoccupied. This principle is exploited in a laser, which contains a cavity full of atoms in an excited state, ready to emit photons of a fixed frequency but arbitrary directions for \mathbf{k} and λ . The geometry of the cavity is such that photons of a certain \mathbf{k} and λ get trapped in it. Consequently, these trapped photons stay back to influence more and more atoms to emit into the mode $(\mathbf{k}\lambda)$. This is why we call it *light amplification by stimulated emission of radiation*. (This general principle, in modified form, is exploited in television also: this is the whole idea behind canned laughter.)

Scattering Theory

19.1. Introduction

One of the best ways to understand the structure of particles and the forces between them is to scatter them off each other. This is particularly true at the quantum level where the systems cannot be seen in the literal sense and must be probed by indirect means. The scattering process gives us information about the projectile, the target, and the forces between them. A natural way to proceed (when possible) is to consider cases where two of these are known and learn about the third. Consider, for example, experiments at the Stanford Linear Accelerator Center in which high-energy photons were used to bombard static neutrons. The structure of the photon and its coupling to matter are well understood—the photon is a point particle to an excellent approximation and couples to electric charge in a way we have studied in some detail. It therefore serves as an excellent probe of the neutron. For instance, the very fact that the neutron, which is electrically neutral, interacts with the photon tells us that the neutron is built out of charged constituents (whose total charge add up to zero). These scattering experiments also revealed that the neutron's constituents have spin $\frac{1}{2}$, and fractional charges $(\frac{2}{3}e, -\frac{1}{3}e)$, a picture that had been arrived at from another independent line of reasoning. Furthermore they also indicated that the interaction between these constituents (called quarks) gets very weak as they get close. This information has allowed us to choose, from innumerable possible models of the interquark force, one that is now considered most likely to succeed, and goes by the name of quantum chromodynamics (QCD), a subject that is being vigorously investigated by many particle physicists today.

Scattering theory is a very extensive subject and this chapter aims at giving you just the flavor of the basic ideas. For more information, you must consult books devoted to this subject.[‡]

A general scattering event is of the form

$$a(\alpha) + b(\beta) + \dots \rightarrow f(\gamma) + g(\delta) + \dots$$

where $\{a, b, \dots\}$ are particle names and $\{\alpha, \beta, \gamma, \dots\}$ are the kinematical variables

[‡] See, for example, the excellent book by J. R. Taylor, *Scattering Theory*, Wiley, New York (1971). Any details omitted here due to lack of space may be found there.

specifying their states, such as momentum, spin, etc. We are concerned only with nonrelativistic, elastic scattering of structureless spinless particles.

In the next three sections, we deal with a formalism that describes a single particle scattering from a potential $V(\mathbf{r})$. As it stands, the formalism describes a particle colliding with an immobile target whose only role is to provide the potential. (This picture provides a good approximation to processes where a light particle collides with a very heavy one, say an α particle colliding with a heavy nucleus.) In Section 19.6 we see how, upon proper interpretation, the same formalism describes two-body collisions in the CM frame. In that section we will also see how the description of the scattering process in the CM frame can be translated to another frame, called the lab frame, where the target is initially at rest. It is important to know how to pass from one frame to the other, since theoretical calculations are most easily done in the CM frame, whereas most experiments are done in the lab frame.

19.2. Recapitulation of One-Dimensional Scattering and Overview

Although we are concerned here with scattering in three dimensions, we begin by recalling one-dimensional scattering, for it shares many common features with its three-dimensional counterpart. The practical question one asks is the following: If a beam of nearly monoenergetic particles with mean momenta $\langle P \rangle = \hbar k_0$ are incident from the far left ($x \rightarrow -\infty$) on a potential $V(x)$ which tends to zero as $|x| \rightarrow \infty$, what fraction T will get transmitted and what fraction R will get reflected?[‡] It is not *a priori* obvious that the above question can be answered, since the mean momentum does not specify the quantum states of the incoming particles. But it turns out that if the individual momentum space wave functions are sharply peaked at $\hbar k_0$, the reflection and transmission probabilities depend only on k_0 and not on the detailed shapes of the wave functions. Thus it is possible to calculate $R(k_0)$ and $T(k_0)$ that apply to every particle in the beam. Let us recall some of the details.

(1) We start with some wave packet, say a Gaussian, with $\langle P \rangle = \hbar k_0$ and $\langle X \rangle \rightarrow -\infty$.

(2) We expand this packet in terms of the eigenfunctions ψ_k of $H = T + V$ with coefficients $a(k)$. The functions ψ_k have the following property:

$$\begin{aligned} \psi_k &\xrightarrow[x \rightarrow -\infty]{} A e^{-ikx} + B e^{ikx} \\ &\xrightarrow[x \rightarrow \infty]{} C e^{ikx} \end{aligned} \quad (19.2.1)$$

In other words, the asymptotic form of ψ_k contains an incident wave $A e^{ikx}$ and a reflected wave $B e^{-ikx}$ as $x \rightarrow -\infty$, and just a transmitted wave $C e^{ikx}$ as $x \rightarrow \infty$. Although the most general solution also contains a $D e^{-ikx}$ piece as $x \rightarrow \infty$, we set

[‡] In general, the particle can come in from the far right as well. Also $V(x)$ need not tend to zero at both ends, but to constants V_+ and V_- as $x \rightarrow \pm\infty$. We assume $V_+ = V_- = 0$ for simplicity. We also assume $|xV(x)| \rightarrow 0$ as $|x| \rightarrow \infty$, so that the particle is asymptotically free ($\psi \sim e^{\pm ikx}$).

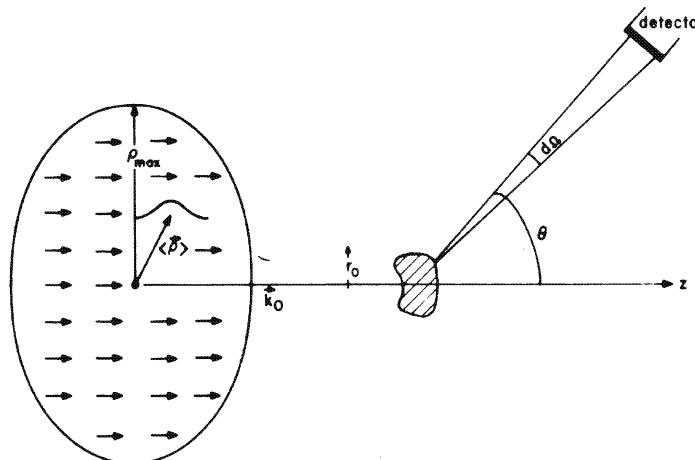


Figure 19.1. A schematic description of scattering. The incident particles, shown by arrows, are really described by wave packets (only one is shown) with mean momentum $\langle \mathbf{P} \rangle = \langle \hbar \mathbf{k}_0 \rangle$ and mean impact parameter $\langle \mathbf{p} \rangle$ uniformly distributed in the \mathbf{p} -plane out to $p_{\max} \gg r_0$, the range of the potential. The shaded region near the origin stands for the domain where the potential is effective. The detector catches all particles that emerge in the cone of opening angle $d\Omega$. The beam is assumed to be coming in along the z axis.

$D=0$ on physical grounds: the incident wave $A e^{ikx}$ can only produce a right-going wave as $x \rightarrow \infty$.

(3) We propagate the wave packet in time by attaching to the expansion coefficients $a(k)$ the time dependence $e^{-iEt/\hbar}$, where $E = \hbar^2 k^2 / 2\mu$. We examine the resulting solution as $t \rightarrow \infty$ and identify the reflected and transmitted packets. From the norms of these we get R and T respectively.

(4) We find at this stage that if the incident packet is sharply peaked in momentum space at $\hbar k_0$, R and T depend only on k_0 and not on the detailed shape of the wave function. Thus the answer to the question raised at the outset is that a fraction $R(k_0)$ of the incident particles will get reflected and a fraction $T(k_0)$ will get transmitted.

(5) Having done all this hard work, we find at the end that the same result could have been obtained by considering just one eigenfunction ψ_{k_0} and taking the ratios of the transmitted and reflected current densities to the incident current density.

The scattering problem in three dimensions has many similarities with its one-dimensional counterpart and also several differences that inevitably accompany the increase in dimensionality. First of all, the incident particles (coming out of the accelerator) are characterized, not by just the mean momentum $\langle \mathbf{P} \rangle = \hbar \mathbf{k}_0$, but also by the fact that they are uniformly distributed in the *impact parameter* \mathbf{p} , which is the coordinate in the plane perpendicular to \mathbf{k}_0 (Fig. 19.1). The distribution is of course not uniform out to $\rho \rightarrow \infty$, but only up to $\rho_{\max} \gg r_0$, where r_0 , the *range of the potential*, is the distance scale beyond which the potential is negligible. [For instance, if $V(r) = e^{-r^2/a^2}$, the range $r_0 \approx a$.] The problem is to calculate the rate at which particles get scattered into a far away detector that subtends a solid angle $d\Omega$ in the direction (θ, ϕ) measured relative to the beam direction (Fig. 19.1). To be

precise, one wants the *differential cross section* $d\sigma/d\Omega$ defined as follows:

$$\frac{d\sigma(\theta, \phi)}{d\Omega} = \frac{\text{number of particles scattered into } d\Omega/\text{sec}}{\text{number incident/sec/area in the } \mathbf{p} \text{ plane}} \quad (19.2.2)$$

The calculation of $d\sigma/d\Omega$ proceeds as follows.‡

- (1) One takes some initial wave packet with mean momentum $\langle \mathbf{P} \rangle = \hbar \mathbf{k}_0$ and mean impact parameter $\langle \mathbf{p} \rangle$. The mean coordinate in the beam direction is not relevant, as long as it is far away from the origin.
- (2) One expands the wave packet in terms of the eigenfunctions $\psi_{\mathbf{k}}$ of $H = T + V$ which are of the form

$$\psi_{\mathbf{k}} = \psi_{\text{inc}} + \psi_{\text{sc}} \quad (19.2.3)$$

where ψ_{inc} is the incident wave $e^{i\mathbf{k}\cdot\mathbf{r}}$ and ψ_{sc} is the scattered wave. One takes only those solutions in which ψ_{sc} is purely outgoing. We shall have more to say about ψ_{sc} in a moment.

- (3) One propagates the wave packet by attaching the time-dependence factor $e^{-iEt/\hbar}$ ($E = \hbar^2 k^2 / 2\mu$) to each coefficient $a(\mathbf{k})$ in the expansion.
- (4) One identifies the scattered wave as $t \rightarrow \infty$, and calculates the probability current density associated with it. One integrates the total flow of probability into the cone $d\Omega$ at (θ, ϕ) . This gives the probability that the incident particle goes into the detector at (θ, ϕ) . One finds that if the momentum space wave function of the incident wave packet is sharply peaked at $\langle \mathbf{P} \rangle = \hbar \mathbf{k}_0$, the probability of going into $d\Omega$ depends only on $\hbar \mathbf{k}_0$ and $\langle \mathbf{p} \rangle$. Call this probability $P(\mathbf{p}, \mathbf{k}_0 \rightarrow d\Omega)$.
- (5) One considers next a beam of particle with $\eta(\mathbf{p})$ particles per second per unit area in the \mathbf{p} plane. The number scattering into $d\Omega$ per second is

$$\eta(d\Omega) = \int P(\mathbf{p}, \mathbf{k}_0 \rightarrow d\Omega) \eta(\mathbf{p}) d^2 \mathbf{p} \quad (19.2.4)$$

Since in the experiment $\eta(\mathbf{p}) = \eta$, a constant, we have from Eq. (19.2.2)

$$\frac{d\sigma}{d\Omega} d\Omega = \frac{\eta(d\Omega)}{\eta} = \int P(\mathbf{p}, \mathbf{k}_0 \rightarrow d\Omega) d^2 \mathbf{p} \quad (19.2.5)$$

- (6) After all this work is done one finds that $d\sigma/d\Omega$ could have been calculated from considering just the static solution $\psi_{\mathbf{k}_0}$ and computing in the limit $r \rightarrow \infty$, the ratio of the probability flow per second into $d\Omega$ associated with ψ_{sc} , to the incident probability current density associated with $e^{i\mathbf{k}_0 \cdot \mathbf{r}}$. The reason the time-dependent picture reduces to the time-independent picture is the same as in one dimension: as we broaden the incident wave packet more and more in coordinate space, the incident and scattered waves begin to coexist in a steady-state configuration, $\psi_{\mathbf{k}_0}$. What about

‡ We do not consider the details here, for they are quite similar to the one-dimensional case. The few differences alone are discussed. See Taylor's book for the details.

the average over $\langle \rho \rangle$? This is no longer necessary, since the incident packet is now a plane wave $e^{ik_0 r}$ which is already uniform in ρ .†

Let us consider some details of extracting $d\sigma/d\Omega$ from ψ_{k_0} . Choosing the z axis parallel to \mathbf{k}_0 and dropping the subscript 0, we obtain

$$\psi_{\mathbf{k}} = e^{ikz} + \psi_{\text{sc}}(r, \theta, \phi) \quad (19.2.6)$$

where θ and ϕ are defined in Fig. 19.1. Although the detailed form of ψ_{sc} depends on the potential, we know that far from the origin it satisfies the free-particle equation [assuming $rV(r) \rightarrow 0$ as $r \rightarrow \infty$].

$$(\nabla^2 + k^2)\psi_{\text{sc}} = 0 \quad (r \rightarrow \infty) \quad (19.2.7)$$

and is purely outgoing.

Recalling the general solution to the free-particle equation (in a region that excludes the origin) we get

$$\psi_{\text{sc}} \xrightarrow[r \rightarrow \infty]{} \sum_l \sum_m (A_l j_l(kr) + B_l n_l(kr)) Y_l^m(\theta, \phi) \quad (19.2.8)$$

Notice that we do not exclude the Neumann functions because they are perfectly well behaved as $r \rightarrow \infty$. Since

$$\begin{aligned} j_l(kr) &\xrightarrow[r \rightarrow \infty]{} \sin(kr - l\pi/2)/(kr) \\ n_l(kr) &\xrightarrow[r \rightarrow \infty]{} -\cos(kr - l\pi/2)/(kr) \end{aligned} \quad (19.2.9)$$

it must be that $A_l/B_l = -i$, so that we get a purely outgoing wave e^{ikr}/kr . With this condition, the asymptotic form of the scattered wave is

$$\psi_{\text{sc}} \xrightarrow[r \rightarrow \infty]{} \frac{e^{ikr}}{kr} \sum_l \sum_m (-i)^l (-B_l) Y_l^m(\theta, \phi) \quad (19.2.10)$$

or

$$\psi_{\text{sc}} \xrightarrow[r \rightarrow \infty]{} \frac{e^{ikr}}{r} f(\theta, \phi) \S \quad (19.2.11)$$

and

$$\psi_{\mathbf{k}} \xrightarrow[r \rightarrow \infty]{} e^{ikz} + f(\theta, \phi) \frac{e^{ikr}}{r} \quad (19.2.12)$$

where f is called the *scattering amplitude*.

† Let us note, as we did in one dimension, that a wave packet does not simply become a plane wave as we broaden it, for the former has norm unity and the latter has norm $\delta^3(0)$. So it is assumed that as the packet is broadened, its norm is steadily increased in such a way that we end up with a plane wave. In any case, the overall norm has no significance.

§ Actually f also depends on k ; this dependence is not shown explicitly.

To get the differential cross section, we need the ratio of the probability flowing into $d\Omega$ per second to the incident current density. So what are \mathbf{j}_{sc} and \mathbf{j}_{inc} , the incident and scattered current densities? Though we have repeatedly spoken of these quantities, they are not well defined unless we invoke further physical ideas. This is because there is only one current density \mathbf{j} associated with $\psi_{\mathbf{k}}$ and it is *quadratic* in $\psi_{\mathbf{k}}$. So \mathbf{j} is not just a sum of two pieces, one due to e^{ikz} and one due to ψ_{sc} ; there are cross terms.[‡] We get around this problem as follows. We note that as $r \rightarrow \infty$, ψ_{sc} is negligible compared to e^{ikz} because of the $1/r$ factor. So we calculate the incident current due to e^{ikz} to be

$$\begin{aligned} |j_{inc}| &= \left| \frac{\hbar}{2\mu i} (e^{-ikz} \nabla e^{ikz} - e^{ikz} \nabla e^{-ikz}) \right| \\ &= \frac{\hbar k}{\mu} \end{aligned} \quad (19.2.13)$$

We cannot use this trick to calculate \mathbf{j}_{sc} into $d\Omega$ because ψ_{sc} never dominates over e^{ikz} . So we use another trick. We say that e^{ikz} is really an abstraction for a wave that is limited in the transverse direction by some $\rho_{max} (\gg r_0)$. Thus in any realistic description, only ψ_{sc} will survive as $r \rightarrow \infty$ for $\theta \neq 0$.[§] (For a given ρ_{max} , the incident wave is present only for $\delta\theta \lesssim \rho_{max}/r$. We can make $\delta\theta$ arbitrarily small by increasing the r at which the detector is located.) With this in mind we calculate (for $\theta \neq 0$)

$$\mathbf{j}_{sc} = \frac{\hbar}{2\mu i} (\psi_{sc}^* \nabla \psi_{sc} - \psi_{sc} \nabla \psi_{sc}^*) \quad (19.2.14)$$

Now

$$\nabla = \mathbf{e}_r \frac{\partial}{\partial r} + \mathbf{e}_\theta \frac{\partial}{r \partial \theta} + \mathbf{e}_\phi \frac{1}{r \sin \theta} \frac{\partial}{\partial \phi} \quad (19.2.15)$$

The last two pieces in ∇ are irrelevant as $r \rightarrow \infty$. When the first acts on the asymptotic ψ_{sc} ,

$$\frac{\partial}{\partial r} f(\theta, \phi) \frac{e^{ikr}}{r} = f(\theta, \phi) ik \frac{e^{ikr}}{r} + O\left(\frac{1}{r^2}\right)$$

so that

$$\mathbf{j}_{sc} = \frac{\mathbf{e}_r}{r^2} |f|^2 \frac{\hbar k}{\mu} \quad (19.2.16)$$

[‡] We did not have to worry about this in one dimension because j due to $A e^{ikx} + B e^{-ikx}$ is $(\hbar k/\mu)(|A|^2 - |B|^2) = j_{inc} + j_{ref}$ with no cross terms.

[§] In fact, only in this more realistic picture is it sensible to say that the particles entering the detectors at $\theta \neq 0$ are scattered (and not unscattered incident) particles. At $\theta = 0$, there is no way (operationally) to separate the incident and scattered particles. To compare theory with experiment, one extracts $f(\theta = 0)$ by extrapolating $f(\theta)$ from $\theta \neq 0$.

Probability flows into $d\Omega$ at the rate

$$\begin{aligned} R(d\Omega) &= \mathbf{j}_{\text{sc}} \cdot \mathbf{e}_r r^2 d\Omega \\ &= |f|^2 \frac{\hbar k}{\mu} d\Omega \end{aligned} \quad (19.2.17)$$

Since it arrives at the rate

$$\begin{aligned} j_{\text{inc}} &= \hbar k / \mu \text{ sec}^{-1} \text{ area}^{-1} \\ \frac{d\sigma}{d\Omega} d\Omega &= \frac{R(d\Omega)}{j_{\text{inc}}} = |f|^2 d\Omega \end{aligned}$$

so that finally

$$\frac{d\sigma}{d\Omega} = |f(\theta, \phi)|^2 \quad (19.2.18)$$

Thus, in the time-independent picture, the calculation of $d\sigma/d\Omega$ reduces to the calculation of $f(\theta, \phi)$.

After this general discussion, we turn to specific calculations. In the next section the calculation of $d\sigma/d\Omega$ is carried out in the time-dependent picture *to first order*. In Section 4, we calculate $d\sigma/d\Omega$ to first order in the time-independent picture. (The two results agree, of course.) In Section 5, we go beyond perturbation theory and discuss some general features of f for spherically symmetric potentials. Two-particle scattering is discussed in Section 6.

19.3. The Born Approximation (Time-Dependent Description)

Consider an initial wave packet that is so broad that it can be approximated by a plane wave $|\mathbf{p}_i\rangle$. Its fate after scattering is determined by the propagator $U(t_f \rightarrow \infty, t_i \rightarrow -\infty)$, that is, by the operator

$$S = \lim_{\substack{t_f \rightarrow \infty \\ t_i \rightarrow -\infty}} U(t_f, t_i)$$

which is called the *S matrix*. The probability of the particle entering the detector in the direction (θ, ϕ) with opening angle $d\Omega$ is the probability that the final momentum \mathbf{p}_f lies in a cone of opening angle $d\Omega$ in the direction (θ, ϕ) :

$$P(\mathbf{p}_i \rightarrow d\Omega) = \sum_{\mathbf{p}_f \text{ in } d\Omega} |\langle \mathbf{p}_f | S | \mathbf{p}_i \rangle|^2$$

If we evaluate S or U to first order, treating V as a perturbation, the problem reduces to the use of Fermi's Golden Rule, which tells us that the transition rate is

$$R_{i \rightarrow d\Omega} = \frac{dP(\mathbf{p}_i \rightarrow d\Omega)}{dt} \quad (19.3.1)$$

$$= \frac{2\pi}{\hbar} \left[\int_0^\infty |\langle \mathbf{p}_f | V | \mathbf{p}_i \rangle|^2 \delta\left(\frac{p_f^2 - p_i^2}{2\mu}\right) p_f^2 dp_f \right] d\Omega \quad (19.3.1)$$

$$= \frac{2\pi}{\hbar} |\langle \mathbf{p}_f | V | \mathbf{p}_i \rangle|^2 \mu \mathbf{p}_i d\Omega \quad (19.3.2)$$

(Hereafter $p_f = p_i = p = \hbar k$ is understood.) This transition rate is just the rate of the flow of probability into $d\Omega$. Since the probability comes in at a rate [recall $j = \rho v |\mathbf{p}_i\rangle \rightarrow (2\pi\hbar)^{-3/2} e^{i\mathbf{p}_i \cdot \mathbf{r}/\hbar}$]

$$j_{\text{inc}} = \frac{\hbar k}{\mu} \left(\frac{1}{2\pi\hbar} \right)^3 \quad (19.3.3)$$

in the direction \mathbf{p}_i , the differential cross section, which measures the rate at which probability is intercepted (and channeled off to $d\Omega$), is

$$\begin{aligned} \frac{d\sigma}{d\Omega} d\Omega &= \frac{R_{i \rightarrow d\Omega}}{j_{\text{inc}}} = (2\pi)^4 \mu^2 \hbar^2 |\langle \mathbf{p}_f | V | \mathbf{p}_i \rangle|^2 d\Omega \\ \frac{d\sigma}{d\Omega} &= \left| \frac{\mu}{2\pi\hbar^2} \int e^{-i\mathbf{q} \cdot \mathbf{r}'} V(\mathbf{r}') d^3 \mathbf{r}' \right|^2 \end{aligned} \quad (19.3.4)$$

where

$$\hbar \mathbf{q} = \mathbf{p}_f - \mathbf{p}_i \quad (19.3.5)$$

is the *momentum transferred* to the particle. For later reference note that

$$|\mathbf{q}|^2 = |\mathbf{k}_f - \mathbf{k}_i|^2 = 2k^2(1 - \cos \theta) = 4k^2 \sin^2(\theta/2) \quad (19.3.6)$$

Thus the dependence of $d\sigma/d\Omega$ on the incident energy and the scattering angle is through the combination $|\mathbf{q}| \equiv q = 2k \sin(\theta/2)$.

By comparing Eqs. (19.3.4) and (19.2.18) we can get $f(\theta)$, up to a phase factor of unit modulus (relative to the incident wave). We shall see later that this factor is -1 . So,

$$f(\theta, \phi) = \frac{-\mu}{2\pi\hbar^2} \int e^{-i\mathbf{q} \cdot \mathbf{r}'} V(\mathbf{r}') d^3 \mathbf{r}' \quad (19.3.7)$$

Thus, in this *Born approximation*, $f(\theta, \phi) = f(\mathbf{q})$ is just the Fourier transform of the potential with respect to momentum transfer (up to a constant factor).

Hereafter we focus on potentials that are spherically symmetric: $V(\mathbf{r}) = V(r)$. In this case, we can choose the z' direction parallel to \mathbf{q} in the $d^3\mathbf{r}'$ integration, so that

$$\begin{aligned} f(\theta, \phi) &= \frac{-\mu}{2\pi\hbar^2} \int e^{-iqr' \cos\theta'} V(r') d(\cos\theta') d\phi' r'^2 dr' \\ &= \frac{-2\mu}{\hbar^2} \int \frac{\sin qr'}{q} V(r') r' dr' \\ &= f(\theta) \end{aligned} \quad (19.3.8)$$

That f should be independent of ϕ in this case could have been anticipated. The incident wave e^{ikz} is insensitive to a change in ϕ , i.e., to rotations around the z axis. The potential, being spherically symmetric, also knows nothing about ϕ . It follows that f cannot pick up any dependence on ϕ . In the language of angular momentum, the incident wave has no l_z and this feature is preserved in the scattering. Consequently the scattered wave must also have no l_z , i.e., be independent of ϕ .

Let us calculate $f(\theta)$ for the *Yukawa potential*

$$V(r) = \frac{g e^{-\mu_0 r}}{r} \quad (19.3.9)$$

From Eq. (19.3.8),

$$\begin{aligned} f(\theta) &= -\frac{2\mu g}{\hbar^2 q} \int_0^\infty \frac{e^{iqr'} - e^{-iqr'}}{2i} e^{-\mu_0 r'} dr' \\ &= \frac{-2\mu g}{\hbar^2 (\mu_0^2 + q^2)} \end{aligned} \quad (19.3.10)$$

$$\frac{d\sigma}{d\Omega} = \frac{4\mu^2 g^2}{\hbar^4 [\mu_0^2 + 4k^2 \sin^2(\theta/2)]^2} \quad (19.3.11)$$

If we now set $g = Ze^2$, $\mu_0 = 0$, we get the cross section for *Coulomb scattering* of a particle of charge e on a potential $\phi = Ze/r$ (or $V = Ze^2/r$):

$$\begin{aligned} \left. \frac{d\sigma}{d\Omega} \right|_{\text{Coulomb}} &= \frac{\mu^2 (Ze^2)^2}{4p^4 \sin^4(\theta/2)} \\ &= \frac{(Ze^2)^2}{16E^2 \sin^4(\theta/2)} \end{aligned} \quad (19.3.12)$$

where $E = p^2/2\mu$ is the kinetic energy of the incident particle. This answer happens to be exact quantum mechanically as well as classically. (It was calculated classically by Rutherford and is called the *Rutherford cross section*.) Although we managed to get the right $d\sigma/d\Omega$ by taking the $\mu_0 \rightarrow 0$ limit of the Yukawa potential calculation,

there are some fine points to note. First of all, the Coulomb potential cannot be handled by the formulation we have developed, since the potential does not vanish faster than r^{-1} . In other words, the asymptotic form

$$\psi \xrightarrow[r \rightarrow \infty]{} e^{ikz} + f(\theta, \phi) \frac{e^{ikr}}{r}$$

is not applicable here since the particle is never free from the influence of the potential. (This manifests itself in the fact that the total cross section is infinite: if we try to integrate $d\sigma/d\Omega$ over θ , the integral diverges as $\int d\theta/\theta^3$ as $\theta \rightarrow 0$.) It is, however, possible to define a scattering amplitude $f_c(\theta)$ in the following sense. One finds that as $r \rightarrow \infty$, there are positive energy eigensolutions to the Coulomb Hamiltonian of the form‡

$$\psi \xrightarrow[r \rightarrow \infty]{} \tilde{e}^{ikz} + f_c(\theta) \left(\frac{\tilde{e}^{ikr}}{r} \right) \quad (19.3.13)$$

where the tilde tells us that these are not actually plane or spherical waves, but rather these objects modified by the long-range Coulomb force. For example

$$\frac{\tilde{e}^{ikr}}{r} = \frac{e^{i(kr - \gamma \ln kr)}}{r} \quad (19.3.14)$$

$$\gamma = \frac{Ze^2\mu}{\hbar^2 k} \quad (19.3.15)$$

is the distorted spherical wave, familiar to us from Section 12.6. By comparing the ratio of flux into $d\Omega$ to flux coming in (due to these distorted waves) one finds that

$$\frac{d\sigma}{d\Omega} = |f_c|^2$$

where

$$f_c(\theta) = -\frac{\gamma}{2k(\sin \theta/2)^2} \exp(-i\gamma \ln \sin^2 \theta/2 + \text{const}) \quad (19.3.16)$$

and where the constant is *purely imaginary*. Comparing this to the Yukawa amplitude, Eq. (19.3.10), after setting $\mu_0 = 0$, $g = Ze^2$, we find agreement up to the exponential phase factor. This difference does not show up in $d\sigma/d\Omega$, but will show up when we consider identical-particle scattering later in this chapter.

‡ See A. Messiah, *Quantum Mechanics*, Wiley, New York (1966), page 422.

*Exercise 19.3.1.** Show that

$$\sigma_{\text{Yukawa}} = 16\pi r_0^2 \left(\frac{g\mu r_0}{\hbar^2} \right)^2 \frac{1}{1 + 4k^2 r_0^2}$$

where $r_0 = 1/\mu_0$ is the range. Compare σ to the geometrical cross section associated with this range.

*Exercise 19.3.2** (1) Show that if $V(r) = -V_0 \theta(r_0 - r)$,

$$\frac{d\sigma}{d\Omega} = 4r_0^2 \left(\frac{\mu V_0 r_0^2}{\hbar^2} \right)^2 \frac{(\sin qr_0 - qr_0 \cos qr_0)^2}{(qr_0)^6}$$

(2) Show that as $kr_0 \rightarrow 0$, the scattering becomes isotropic and

$$\sigma \approx \frac{16\pi r_0^2}{9} \left(\frac{\mu V_0 r_0^2}{\hbar^2} \right)^2$$

*Exercise 19.3.3.** Show that for the Gaussian potential, $V(r) = V_0 e^{-r^2/r_0^2}$,

$$\begin{aligned} \frac{d\sigma}{d\Omega} &= \frac{\pi r_0^2}{4} \left(\frac{\mu V_0 r_0^2}{\hbar^2} \right)^2 e^{-q^2 r_0^2/2} \\ \sigma &= \frac{\pi^2}{2k^2} \left(\frac{\mu V_0 r_0^2}{\hbar^2} \right)^2 (1 - e^{-2k^2 r_0^2}) \end{aligned}$$

[Hint: Since $q^2 = 2k^2(1 - \cos \theta)$, $d(\cos \theta) = -d(q^2)/2k^2$.]

Let us end this section by examining some general properties of $f(\theta)$. We see from Eq. (19.3.7) that at *low energies* ($k \rightarrow 0$), $q = 2k \sin(\theta/2) \rightarrow 0$ and

$$\begin{aligned} f(\theta) &\sim -\frac{\mu}{2\pi\hbar^2} \int V(\mathbf{r}') d^3\mathbf{r}' \\ &\cong -\frac{\mu V_0 r_0^3}{\hbar^2} \end{aligned} \tag{19.3.17}$$

where V_0 is some effective height of V , and r_0 is some effective range. At *high energies*, the exponential factor $e^{-iq'\cos\theta'}$ oscillates rapidly. This means that the scattered waves coming from different points \mathbf{r}' add with essentially random phases, except in the small range where the phase is stationary:

$$\begin{aligned} qr' \cos \theta' &\lesssim \pi \\ 2k \sin(\theta/2)r_0 &\lesssim \pi \quad (\text{since } r' \cos \theta' \cong r_0) \\ k\theta r_0 &\lesssim \pi \quad (\sin \theta/2 \simeq \theta/2) \end{aligned}$$

Thus the scattering amplitude is appreciable only in a small forward cone of angle (dropping constants of order unity)

$$\theta \lesssim \frac{1}{kr_0} \quad (19.3.18)$$

These arguments assume $V(r')$ is regular near $r' = 0$. But in some singular cases [$V \propto (r')^{-3}$, say] the r' integral is dominated by small r' and $kr' \cos \theta'$ is not necessarily a large phase. Both the Yukawa and Gaussian potential (Exercise 19.3.3) are free of such pathologies and exhibit this forward peak at high energies.

Exercise 19.3.4. Verify the above claim for the Gaussian potential.

When can we trust the Born approximation? Since we treated the potential as a perturbation, our guess would be that it is reliable at high energies. We shall see in the next section that this is indeed correct, but that the Born approximation can also work at low energies provided a more stringent condition is satisfied.

19.4. Born Again (The Time-Independent Description)

In this approach, the central problem is to find solutions to the full Schrödinger equation

$$(\nabla^2 + k^2)\psi_{\mathbf{k}} = \frac{2\mu}{\hbar^2} V \psi_{\mathbf{k}} \quad (19.4.1)$$

of the form

$$\psi_{\mathbf{k}} = e^{i\mathbf{k}\cdot\mathbf{r}} + \psi_{\text{sc}} \quad (19.4.2a)$$

where

$$\psi_{\text{sc}} \xrightarrow[r \rightarrow \infty]{} f(\theta, \phi) \frac{e^{ikr}}{r} \quad (19.4.2b)$$

In the above, θ and ϕ are measured relative to \mathbf{k} , chosen along the z axis (Fig. 19.1). One approaches the problem as follows. One finds a *Green's function* $G^0(\mathbf{r}, \mathbf{r}')$ which satisfies

$$(\nabla^2 + k^2)G^0(\mathbf{r}, \mathbf{r}') = \delta^3(\mathbf{r} - \mathbf{r}') \quad (19.4.3)$$

in terms of which the *formal* general solution to Eq. (19.4.1) is

$$\psi_k(\mathbf{r}) = \psi^0(\mathbf{r}) + \frac{2\mu}{\hbar^2} \int G^0(\mathbf{r}, \mathbf{r}') V(\mathbf{r}') \psi_k(\mathbf{r}') d^3 r' \quad (19.4.4)$$

where $\psi^0(\mathbf{r})$ is an arbitrary free-particle solution of energy $\hbar^2 k^2 / 2\mu$:

$$(\nabla^2 + k^2) \psi^0 = 0 \quad (19.4.5)$$

We will soon nail down ψ^0 using the boundary conditions.

Applying $\nabla^2 + k^2$ to both sides of Eq. (19.4.4) one may easily verify that ψ_k indeed is a solution to Eq. (19.4.1). The idea here is quite similar to what is employed in solving Poisson's equation for the electrostatic potential in terms of the charge density ρ :

$$\nabla^2 \phi = -4\pi\rho$$

One first finds G , the response to a point charge at \mathbf{r}' :

$$\nabla^2 G = -4\pi\delta^3(\mathbf{r} - \mathbf{r}')$$

Exercise 12.6.4 tells us that

$$G(\mathbf{r}, \mathbf{r}') = G(\mathbf{r} - \mathbf{r}') = \frac{1}{|\mathbf{r} - \mathbf{r}'|}$$

One then views ρ as a superposition of point charges and, since Poisson's equation is linear, obtains ϕ as the sum of ϕ 's produced by these charges:

$$\phi(\mathbf{r}) = \int G(\mathbf{r} - \mathbf{r}') \rho(\mathbf{r}') d^3 r' = \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d^3 r'$$

(By acting on both sides with ∇^2 and using $\nabla^2 G = -4\pi\delta^3$, you may verify that ϕ satisfies Poisson's equation.)

One can add to this $\phi(\mathbf{r})$ any ϕ^0 that satisfies $\nabla^2 \phi^0 = 0$. Using the boundary condition $\phi = 0$ when $\rho = 0$, we get rid of ϕ^0 .

In the scattering problem we pretend that the right-hand side of Eq. (19.4.1) is some given source and write Eq. (19.4.4) for ψ_k in terms of the Green's function. The only catch is that the source for ψ_k is ψ_k itself. Thus Eq. (19.4.4) is really not a solution, but an integral equation for ψ_k . The motivation for converting the differential equation to an integral equation is similar to that in the case of $U_1(t, t_0)$: to obtain a perturbative expansion for ψ_k in powers of V . *To zeroth order in V* , Eq. (19.4.2a) tells us that ψ_k is $e^{i\mathbf{k}\cdot\mathbf{r}}$, since there is no scattered wave if V is neglected; whereas Eq. (19.4.4) tells us that $\psi_k = \psi^0$, since the integral over \mathbf{r}' has an explicit power of V in it while ψ^0 has no dependence on V [since it is the solution to Eq. (19.4.5)]. We are thus able to nail down the arbitrary function ψ^0 in Eq. (19.4.4):

$$\psi^0 = e^{i\mathbf{k}\cdot\mathbf{r}} \quad (19.4.6)$$

and conclude that in the present scattering problem

$$\psi_k = e^{i\mathbf{k}\cdot\mathbf{r}} + \frac{2\mu}{\hbar^2} \int G^0(\mathbf{r}, \mathbf{r}') V(\mathbf{r}') \psi_k(\mathbf{r}') d^3\mathbf{r}' \quad (19.4.7)$$

Upon comparing this to Eq. (19.4.2a) we see that we are associating the second piece with the scattered wave. For consistency of interpretation, it must contain purely outgoing waves at spatial infinity. Since $G^0(\mathbf{r}, \mathbf{r}')$ is the scattered wave produced by a point source at \mathbf{r}' , it is necessary that $G^0(\mathbf{r}, \mathbf{r}')$ be purely outgoing asymptotically. This is an additional physical constraint on G^0 over and above Eq. (19.4.3). As we shall see, this constraint, together with Eq. (19.4.3), will determine G^0 for us uniquely.

Imagine that we have found this G^0 . We are now in a position to obtain a perturbative solution for ψ_k starting with Eq. (19.4.7). To zeroth order we have seen that $\psi_k = e^{i\mathbf{k}\cdot\mathbf{r}}$. To go to first order, we feed the zeroth-order ψ_k into the right-hand side and obtain

$$\psi_k = e^{i\mathbf{k}\cdot\mathbf{r}} + \frac{2\mu}{\hbar^2} \int G^0(\mathbf{r}, \mathbf{r}') V(\mathbf{r}') e^{i\mathbf{k}\cdot\mathbf{r}'} d^3\mathbf{r}' + O(V^2) \quad (19.4.8)$$

If we feed this first-order result back into the right-hand side of Eq. (19.4.7), we get (in symbolic form) the result good to second order:

$$\psi_k = \psi^0 + \frac{2\mu}{\hbar^2} G^0 V \psi^0 + \left(\frac{2\mu}{\hbar^2} \right)^2 G^0 V G^0 V \psi^0 + O(V^3)$$

and so on.

Let us now turn to the determination of G^0 , starting with Eq. (19.4.3):

$$(\nabla^2 + k^2) G^0(\mathbf{r}, \mathbf{r}') = \delta^3(\mathbf{r} - \mathbf{r}')$$

We note that this equation does not have a unique solution, since, given any solution, we can get another by adding to it a function η^0 that obeys the homogeneous equation

$$(\nabla^2 + k^2) \eta^0 = 0$$

Conversely, any two columns G^0 and $G^{0'}$ can differ only by some η^0 . So we will first find the simplest G^0 we can, and then add whatever η^0 it takes to make the sum purely outgoing.

Since $(\nabla^2 + k^2)$ and $\delta^3(\mathbf{r} - \mathbf{r}')$ are invariant under the overall translation of \mathbf{r} and \mathbf{r}' , we know the equation admits translationally invariant solutions[‡]:

$$G^0(\mathbf{r}, \mathbf{r}') = G^0(\mathbf{r} - \mathbf{r}')$$

[‡] Note that if an equation has some symmetry, like rotational invariance, it means only that rotationally invariant solutions exist, and not that all solutions are rotationally invariant. For example, the hydrogen atom Hamiltonian is rotationally invariant, but the eigenfunctions are not in general. But there are some (with $l=m=0$) which are.

Replace $\mathbf{r} - \mathbf{r}'$ by \mathbf{r} for convenience. [Once we find $G^0(\mathbf{r})$, we can replace \mathbf{r} by $\mathbf{r} - \mathbf{r}'$.] So we want to solve

$$(\nabla^2 + k^2)G^0(\mathbf{r}) = \delta^3(\mathbf{r}) \quad (19.4.9)$$

For similar reasons as above, we look for a rotationally invariant solution

$$G^0(\mathbf{r}) = G^0(r)$$

Writing

$$G^0(r) = \frac{U(r)}{r}$$

we find that for $r \neq 0$, $U(r)$ satisfies

$$\frac{d^2 U}{dr^2} + k^2 U = 0$$

the general solution to which is

$$U(r) = A e^{ikr} + B e^{-ikr}$$

or

$$G^0(r) = \frac{A e^{ikr}}{r} + \frac{B e^{-ikr}}{r} \quad (19.4.10)$$

where A and B are arbitrary constants at this point. Since we want G^0 to be purely outgoing we set $B=0$:

$$G^0(r) = \frac{A e^{ikr}}{r} \quad (19.4.11)$$

We find A by calculating $(\nabla^2 + k^2)G^0(r)$ as $r \rightarrow 0$ [‡]

$$(\nabla^2 + k^2)G^0(r) \xrightarrow{r \rightarrow 0} -4\pi A \delta^3(\mathbf{r}) \quad (19.4.12)$$

[‡] We use $\nabla^2(\psi\chi) = \psi\nabla^2\chi + \chi\nabla^2\psi + 2\nabla\psi \cdot \nabla\chi$ and $\nabla^2 = r^{-2}(\partial/\partial r)r^2 \partial/\partial r$ on a function of r alone.

which gives us

$$G^0(r) = -\frac{e^{ikr}}{4\pi r} \quad (19.4.13)$$

We cannot add any η^0 to this solution, without destroying its purely outgoing nature, since the general form of the free-particle solution, regular in all space, is

$$\eta^0(\mathbf{r}) = \sum_{l=0}^{\infty} \sum_{m=-l}^l C_{lm} j_l(kr) Y_l^m(\theta, \phi) \quad (19.4.14)$$

and since, as $r \rightarrow \infty$, the spherical Bessel functions are made up of incoming and outgoing waves of equal amplitude

$$j_l(kr) \xrightarrow[r \rightarrow \infty]{} \frac{\sin(kr - l\pi/2)}{kr} = \frac{e^{i(kr - l\pi/2)} - e^{-i(kr - l\pi/2)}}{2ikr} \quad (19.4.15)$$

Let us now feed

$$G^0(\mathbf{r}, \mathbf{r}') = G^0(\mathbf{r} - \mathbf{r}') = -\frac{e^{ik|\mathbf{r} - \mathbf{r}'|}}{4\pi|\mathbf{r} - \mathbf{r}'|} \quad (19.4.16)$$

into Eq. (19.4.7) to obtain

$$\begin{aligned} \psi_{\mathbf{k}} &= e^{i\mathbf{k}\cdot\mathbf{r}} - \frac{2\mu}{4\pi\hbar^2} \int \frac{e^{ik|\mathbf{r} - \mathbf{r}'|}}{|\mathbf{r} - \mathbf{r}'|} V(\mathbf{r}') \psi_{\mathbf{k}}(\mathbf{r}') d^3\mathbf{r}' \\ &= e^{i\mathbf{k}\cdot\mathbf{r}} + \psi_{sc} \end{aligned} \quad (19.4.17)$$

Let us now verify that as $r \rightarrow \infty$, ψ_{sc} has the desired form $f(\theta, \phi) e^{ikr}/r$. Our first instinct may be to approximate as follows:

$$\frac{e^{ik|\mathbf{r} - \mathbf{r}'|}}{|\mathbf{r} - \mathbf{r}'|} \underset{|r'| \lesssim r_0}{\approx} \frac{e^{ikr}}{r}$$

in the \mathbf{r}' integral since \mathbf{r}' is confined to $|\mathbf{r}'| \lesssim r_0$ (the range), whereas $r \rightarrow \infty$. That this is wrong is clear from the fact that if we do so, the corresponding f has no θ or ϕ dependence. Let us be more careful. We first approximate

$$\begin{aligned} |\mathbf{r} - \mathbf{r}'| &= (r^2 + r'^2 - 2\mathbf{r} \cdot \mathbf{r}')^{1/2} \\ &= r \left[1 + \left(\frac{r'}{r} \right)^2 - 2 \frac{\mathbf{r} \cdot \mathbf{r}'}{r^2} \right]^{1/2} = \end{aligned}$$

$$\begin{aligned}
 &= r \left(1 - 2 \frac{\mathbf{r} \cdot \mathbf{r}'}{r^2} \right)^{1/2} + O\left[\left(\frac{r'}{r}\right)^2\right] r \\
 &\approx r \left(1 - \frac{\mathbf{r} \cdot \mathbf{r}'}{r^2} \right)
 \end{aligned} \tag{19.4.18}$$

We have thrown away the term quadratic in (r'/r) and used the approximation $(1+x)^n \approx 1+nx$ for small x . So

$$\frac{1}{|\mathbf{r} - \mathbf{r}'|} = \frac{1}{r[1 - (\mathbf{r} \cdot \mathbf{r}')/r^2]} \approx \frac{1}{r} \left(1 + \frac{\mathbf{r} \cdot \mathbf{r}'}{r^2} \right) \tag{19.4.19}$$

Whereas replacing $|\mathbf{r} - \mathbf{r}'|^{-1}$ in the integral leads to errors which vanish as $r \rightarrow \infty$, this is not so for the factor $e^{ik|\mathbf{r} - \mathbf{r}'|}$. We have

$$\begin{aligned}
 k|\mathbf{r} - \mathbf{r}'| &= kr \left(1 - \frac{\mathbf{r} \cdot \mathbf{r}'}{r^2} \right) \\
 &= kr - k\hat{r} \cdot \mathbf{r}' \\
 &= kr - \mathbf{k}_f \cdot \mathbf{r}' \tag{19.4.20}
 \end{aligned}$$

where \mathbf{k}_f is the wave vector of the detected particle: it has the same magnitude (k) as the incident particle and points in the direction (\hat{r}) of observation (Fig. 19.2). Consequently

$$\frac{e^{ik|\mathbf{r} - \mathbf{r}'|}}{|\mathbf{r} - \mathbf{r}'|} \approx \frac{e^{ikr}}{r} e^{-i\mathbf{k}_f \cdot \mathbf{r}'} \tag{19.4.21}$$

and

$$\psi_{\mathbf{k}} \xrightarrow[r \rightarrow \infty]{e^{i\mathbf{k} \cdot \mathbf{r}}} \frac{e^{ikr}}{r} \frac{2\mu}{4\pi\hbar^2} \int e^{-i\mathbf{k}_f \cdot \mathbf{r}'} V(\mathbf{r}') \psi_{\mathbf{k}}(\mathbf{r}') d^3 \mathbf{r}' \tag{19.4.22}$$

Thus the solution we have found has the desired form as $r \rightarrow \infty$. Equation (19.4.22) of course does not determine $f(\theta, \phi)$ since $\psi_{\mathbf{k}}$ is present in the \mathbf{r}' integration.

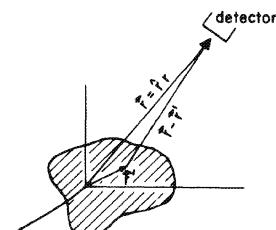


Figure 19.2. The particle is observed at the point \mathbf{r} . The \mathbf{r}' integration is restricted to the shaded region which symbolizes the range of the potential.

However, to any desired order this $\psi_{\mathbf{k}}$ can be replaced by the *calculable* lower-order approximation. In particular, to first order,

$$f(\theta, \phi) = -\frac{2\mu}{4\pi\hbar^2} \int e^{-i\mathbf{k}_i \cdot \mathbf{r}'} V(\mathbf{r}') e^{i\mathbf{k}_i \cdot \mathbf{r}'} d^3 \mathbf{r}' \quad (19.4.23)$$

where we have added a subscript i to \mathbf{k} to remind us that it is the initial or incident wave vector. We recognize $f(\theta, \phi)$ to be just the Born approximation calculated in the last section [Eq. (19.3.7)]. The phase factor -1 , relative to the incident wave was simply assumed there. The agreement between the time-dependent and time-independent calculations of f persists to all orders in the perturbation expansion.

There is another way (involving Cauchy's theorem) to solve

$$(\nabla^2 + k^2) G^0(\mathbf{r}) = \delta^3(\mathbf{r}) \quad (19.4.24)$$

Fourier transforming both sides, we get

$$\left(\frac{1}{2\pi}\right)^{3/2} \int e^{-i\mathbf{q} \cdot \mathbf{r}} (\nabla^2 + k^2) G^0(\mathbf{r}) d^3 \mathbf{r} = \left(\frac{1}{2\pi}\right)^{3/2} \quad (19.4.25)$$

If we let ∇^2 act to the left (remember it is Hermitian) we get

$$(k^2 - q^2) \left(\frac{1}{2\pi}\right)^{3/2} \int e^{-i\mathbf{q} \cdot \mathbf{r}} G^0(\mathbf{r}) d^3 \mathbf{r} = \left(\frac{1}{2\pi}\right)^{3/2} \quad (19.4.26)$$

$$(k^2 - q^2) G^0(\mathbf{q}) = \left(\frac{1}{2\pi}\right)^{3/2} \quad (19.4.27)$$

As always, going to momentum space has reduced the differential equation to an algebraic equation. The solution is

$$G^0(\mathbf{q}) = \frac{1}{(2\pi)^{3/2}(k^2 - q^2)} \quad (19.4.28)$$

except at the point $q=k$ where $G^0(\mathbf{q})$ diverges. The reason for this divergence is the following. Equation (19.4.24) is the coordinate space version of the abstract equation

$$(D^2 + k^2) G^0 = I \quad (19.4.29)$$

where

$$D^2 = D_x^2 + D_y^2 + D_z^2 \quad (19.4.30)$$

(D_x is just the x derivative operator D introduced in Section 1.10, and D_y and D_z are y and z derivative operators.) Thus G^0 is the inverse of $(D^2 + k^2)$:

$$G^0 = (D^2 + k^2)^{-1} \quad (19.4.31)$$

Now, we know that we cannot invert an operator that has a vanishing determinant or equivalently (for a Hermitian operator, since it can be diagonalized) a zero eigenvalue. The operator $(D^2 + k^2)$ has a zero eigenvalue since

$$(\nabla^2 + k^2)\psi = 0 \quad (19.4.32)$$

has nontrivial (plane wave) solutions. We therefore consider a slightly different operator, $D^2 + k^2 + i\varepsilon$, where ε is positive and infinitesimal. This too has a zero eigenvalue, but the corresponding eigenfunctions are plane waves of *complex wave number*. Such functions are not part of the space we are restricting ourselves to, namely, the space of functions normalized to unity or the Dirac delta function.[‡] Thus $D^2 + k^2 + i\varepsilon$ may be inverted *within the physical Hilbert space*. Let us call the corresponding Green's function G_ε^0 . At the end of the calculation we will send ε to zero.[§]

Clearly

$$G_\varepsilon^0(\mathbf{q}) = \frac{1}{(2\pi)^{3/2}} \frac{1}{k^2 + i\varepsilon - q^2} \quad (19.4.33)$$

The coordinate space function is given by the inverse transform:

$$G_\varepsilon^0(\mathbf{r}) = \frac{1}{(2\pi)^3} \int \frac{e^{i\mathbf{qr}}}{k^2 + i\varepsilon - q^2} d^3\mathbf{q} \quad (19.4.34)$$

We choose the q_z axis parallel to \mathbf{r} . If θ and ϕ are the angles in \mathbf{q} space,

$$\begin{aligned} G_\varepsilon^0(\mathbf{r}) &= \frac{1}{8\pi^3} \int \frac{e^{iqr \cos \theta}}{k^2 + i\varepsilon - q^2} d(\cos \theta) d\phi q^2 dq \\ &= \frac{1}{4\pi^2} \int_0^\infty \frac{e^{iqr} - e^{-iqr}}{iqr} \frac{q^2 dq}{k^2 + i\varepsilon - q^2} \end{aligned} \quad (19.4.35a)$$

$$= \frac{1}{4\pi^2} \int_{-\infty}^\infty \frac{e^{iqr}}{iqr} \frac{q^2 dq}{k^2 + i\varepsilon - q^2} \quad (19.4.35b)$$

$$= \frac{-i}{4\pi^2 r} \int_{-\infty}^\infty \frac{e^{iqr} q dq}{k^2 + i\varepsilon - q^2} \quad (19.4.35c)$$

[‡] Recall from Section 1.10 that if k is complex, the norm diverges exponentially.

[§] This is called the “ $i\varepsilon$ prescription.” Throughout the analysis ε will be considered only to first order.

[In going from (19.4.35a) to (19.4.35b) above, we changed q to $-q$ in the e^{-iqr} piece.]

We proceed to evaluate the above integral by means of Cauchy's residue theorem, which states that for any analytic function $f(z)$ of a complex variable z ,

$$\oint f(z) dz = 2\pi i \sum_j R(z_j) \quad (19.4.36)$$

where \oint denotes integration around a closed contour in the complex z plane and $R(z_j)$ is the residue of the pole at the point z_j lying inside the contour.[‡]

Let us view q as a complex variable which happens to be taking only real values ($-\infty$ to $+\infty$) in Eq. (19.4.35).

We are trying to evaluate the integral of the function

$$w(q) = \frac{-i}{4\pi^2 r} \frac{e^{iqr} q}{k^2 + i\varepsilon - q^2} \quad (19.4.37)$$

along the real axis from $-\infty$ to $+\infty$.

This function has poles where

$$k^2 + i\varepsilon - q^2 = 0$$

or (to first order in ε),

$$(k + q + i\eta)(k - q + i\eta) = 0 \quad (\eta \cong \varepsilon/2k) \quad (19.4.38)$$

These poles are shown in Fig. 19.3.

We are not yet ready to use Cauchy's theorem because we do not have a closed contour. Let us now close the contour via a large semicircle C_ρ whose radius $\rho \rightarrow \infty$. Now we can use Cauchy's theorem, but haven't we changed the quantity we wanted to calculate? No, because C_ρ does not contribute to the integral as $\rho \rightarrow \infty$. To see this, let us write $q = \rho e^{i\theta}$ on C_ρ . Then

$$w(q) \xrightarrow[\rho \rightarrow \infty]{} (\text{const}) \frac{e^{iqr}}{q} \quad (19.4.39)$$

and

[‡] Recall that if

$$f(z) \xrightarrow[z \rightarrow z_j]{} \frac{R(z_j)}{z - z_j}$$

then

$$R(z_j) = \lim_{z \rightarrow z_j} f(z)(z - z_j)$$

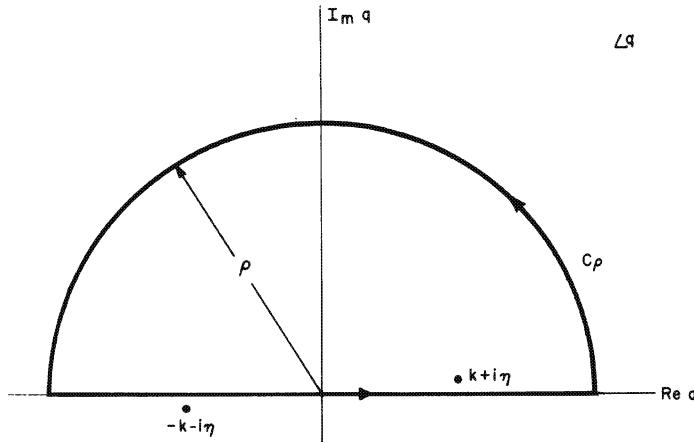


Figure 19.3. The poles of the function $w(q)$ in the complex q plane. We want the integral along the real axis from $-\infty$ to $+\infty$. We add to it the contribution along C_ρ (which vanishes as ρ tends to ∞) in order to close the contour of integration and to use Cauchy's theorem.

$$\int_{C_\rho} w(q) dq \sim \int_{C_\rho} e^{iqr} \frac{dq}{q} = \int_0^\pi e^{i\rho r(\cos \theta + i \sin \theta)} i d\theta \quad (19.4.40)$$

Except for the tiny region near $\theta=0$ (which contributes only an infinitesimal amount) the integral vanishes since $e^{-\rho r \sin \theta} \rightarrow 0$ as $\rho \rightarrow \infty$. We now use Cauchy's theorem. The only pole enclosed is at $q=k+i\eta$. The residue here is

$$R(k+i\eta) = \lim_{q \rightarrow k+i\eta} (q-k-i\eta)w(q) = \frac{i}{8\pi^2 r} e^{i(k+i\eta)r} \quad (19.4.41)$$

and

$$G^0(\mathbf{r}) = \lim_{\eta \rightarrow 0} 2\pi i R = -\frac{e^{ikr}}{4\pi r} \quad (19.4.42)$$

Notice that although the $i\varepsilon$ ($\varepsilon > 0$) prescription happens to give the right answer here, there are other ways to evaluate the integral, which may be appropriate in other contexts. For example if we choose $\varepsilon < 0$, we get a purely incoming wave, since η changes sign and the pole near $q \cong -k$ gets into the contour.

Validity of the Born Approximation

Since in the Born approximation one replaces $\psi_{\mathbf{k}} = e^{i\mathbf{k}\cdot\mathbf{r}'} + \psi_{\text{sc}}$ by just $e^{i\mathbf{k}\cdot\mathbf{r}'}$ in the right-hand side of the integral Eq. (19.4.17), it is a good approximation only if $|\psi_{\text{sc}}| \ll |e^{i\mathbf{k}\cdot\mathbf{r}'}|$ in the region $|\mathbf{r}'| \lesssim r_0$. Since we expect ψ_{sc} to be largest near the origin,

let us perform a comparison there, using Eq. (19.4.17) itself to evaluate $\psi_{\text{sc}}(0)$:

$$\frac{|\psi_{\text{sc}}(0)|}{|e^{ikz}(0)|} = |\psi_{\text{sc}}(0)| = \left| \frac{2\mu}{4\pi\hbar^2} \int \frac{e^{ikr'}}{r'} V(\mathbf{r}') e^{-i\mathbf{k}_r \cdot \mathbf{r}'} d^3 \mathbf{r}' \right| \quad (19.4.43)$$

Let us assume $V(\mathbf{r}) = V(r)$. In this case a rough criterion for the validity of the Born approximation is

$$\frac{2\mu}{\hbar^2 k} \left| \int e^{ikr'} \sin kr' V(r') dr' \right| \ll 1 \quad (19.4.44)$$

Exercise 19.4.1. Derive the inequality (19.4.44).

At low energies, $kr' \rightarrow 0$, $e^{ikr'} \rightarrow 1$, $\sin kr' \rightarrow kr'$ and we get the condition

$$\frac{2\mu}{\hbar^2} \left| \int r' V(r') dr' \right| \ll 1 \quad (19.4.45)$$

If $V(r)$ has an effective depth (or height) V_0 and range r_0 , the condition becomes (dropping constants of order unity)

$$\frac{\mu V_0 r_0^2}{\hbar^2} \ll 1 \quad (19.4.46)$$

The low energy condition may be written as

$$V_0 \ll \frac{\hbar^2}{\mu r_0^2}$$

Now a particle confined to a well of dimension r_0 must have a momentum of order \hbar/r_0 and a kinetic energy of order $\hbar^2/\mu r_0^2$. The above inequality says that if the Born approximation is to work at low energies, the potential must be too shallow to bind a particle confined to a region of size r_0 .

At high energies when $kr_0 \gg 1$ let us write inside the integral in Eq. (19.4.44)

$$e^{ikr'} \sin kr' = \frac{e^{2ikr'} - 1}{2i}$$

and drop the exponential which will be oscillating too rapidly within the range of the potential and keep just the -1 part to get the following condition:

$$\frac{\mu}{\hbar^2 k} \left| \int V(r') dr' \right| \ll 1 \quad (19.4.47)$$

which could be rewritten as

$$\frac{\mu V_0 r_0^2}{\hbar^2} \ll kr_0 \quad (19.4.48)$$

We found that the Born approximation can be good even at low energies if the inequality (19.4.46) is satisfied. In fact, if it is, the Born approximation is good at all energies, i.e., Eq. (19.4.48) is automatically satisfied.

19.5. The Partial Wave Expansion

We have noted that if $V(\mathbf{r}) = V(r)$, $f(\theta, \phi) = f(\theta)$. Actually f is also a function of the energy $E = \hbar^2 k^2 / 2\mu$, though this dependence was never displayed explicitly. Since any function of θ can be expanded in terms of the Legendre polynomials

$$P_l(\cos \theta) = \left(\frac{4\pi}{2l+1} \right)^{1/2} Y_l^0 \quad (19.5.1)$$

we can expand $f(\theta, k)$ in terms of $P_l(\cos \theta)$ with k -dependent coefficients:

$$f(\theta, k) = \sum_{l=0}^{\infty} (2l+1)a_l(k)P_l(\cos \theta) \quad (19.5.2)$$

One calls $a_l(k)$ the *lth partial wave amplitude*. It has the following significance. The incident plane wave e^{ikz} is composed of states of all angular momenta [from Eq. (12.6.41)]:

$$e^{ikz} = e^{ikr \cos \theta} = \sum_{l=0}^{\infty} i^l (2l+1)j_l(kr)P_l(\cos \theta) \quad (19.5.3)$$

Since the potential conserves angular momentum, each angular momentum component scatters independently. The amplitude a_l is a measure of the scattering in the angular momentum l sector.

As it stands, the expansion in Eq. (19.5.2) has not done anything for us: we have traded one function of the two variables (θ and k) for an infinite number of functions $a_l(k)$ of one variable k . What makes the expansion useful is that at low energies, only the first few $a_l(k)$ are appreciably different from zero. In this case, one manages to describe the scattering in terms of just a few functions a_0, a_1, \dots of one variable. The following heuristic argument (corroborated by explicit calculations) is usually given to explain why the scattering is restricted to a few low l values at low k . Coming out of the accelerator is a uniform beam of particles moving along the z axis. All particles in a cylinder of radius ρ and thickness $d\rho$ (ρ is the impact parameter) have angular momentum

$$\hbar l \approx \hbar k \rho \quad (19.5.4)$$

If the potential has range r_0 , particles with $\rho > r_0$ will “miss” the target. Thus there will be scattering only up to

$$l_{\max} = k\rho_{\max} \approx kr_0 \quad (19.5.5)$$

[Conversely, by measuring l_{\max} (from the angular dependence of f) we can deduce the range of the potential.]

*Exercise 19.5.1.** Show that for a 100-MeV (kinetic energy) neutron incident on a fixed nucleus, $l_{\max} \approx 2$. (Hint: The range of the nuclear force is roughly a Fermi = 10^{-5} Å. Also $\hbar c \approx 200$ MeV F is a more useful mnemonic for nuclear physics.)

Given a potential $V(r)$, how does one calculate $a_l(k)$ in terms of it? In other words, how is a_l related to the solution to the Schrödinger equation for angular momentum l ? We begin by considering a free particle. Using

$$j_l(kr) \xrightarrow[r \rightarrow \infty]{} \frac{\sin(kr - l\pi/2)}{kr} \quad (19.5.6)$$

we get, from Eq. (19.5.3),

$$e^{ikz} \xrightarrow[r \rightarrow \infty]{} \frac{1}{2ik} \sum_{l=0}^{\infty} i^l (2l+1) \left(\frac{e^{i(kr - l\pi/2)}}{r} - \frac{e^{-i(kr - l\pi/2)}}{r} \right) P_l(\cos \theta) \quad (19.5.7)$$

$$= \frac{1}{2ik} \sum_{l=0}^{\infty} (2l+1) \left(\frac{e^{ikr}}{r} - \frac{e^{-ikr}}{r} \right) P_l(\cos \theta) \quad (19.5.8)$$

upon using $i = e^{i\pi/2}$. Thus at each angular momentum we have incoming and outgoing waves of the same amplitude. (Their phases differ by $l\pi$ because the repulsive centrifugal barrier potential is present at $l \neq 0$ even for a free particle.) The probability currents associated with the two waves are equal and opposite.[‡] This equality is expected since in this steady state there should be no net probability flux flowing into the origin or coming out of it. (This balance should occur separately for each l , since scattering in each l is independent due to angular momentum conservation.)

What happens if we turn on a potential? As $r \rightarrow \infty$, the radial wave functions must reduce to the free-particle wave function, although there can be a *phase shift* $\delta_l(k)$ due to the potential:

$$R_l(r) = \frac{U_l(r)}{r} \xrightarrow[r \rightarrow \infty]{} \frac{A_l \sin[kr - l\pi/2 + \delta_l(k)]}{r} \quad (19.5.9)$$

[‡] Once again, can we speak of the current associated with a given l and also with the incoming and outgoing waves at a given l ? Yes. If we calculate the total \mathbf{j} (which will have only a radial part as $r \rightarrow \infty$) and integrate over all angles, the orthogonality of P_l 's will eliminate all interference terms between different l 's. There will also be no interference between the incoming and outgoing waves. [See footnote related to Eq. (19.2.13).]

where A_l is some constant. So

$$\psi_{\mathbf{k}}(\mathbf{r}) \xrightarrow[r \rightarrow \infty]{} \sum_{l=0}^{\infty} A_l \frac{(e^{i(kr - l\pi/2 + \delta_l)} - e^{-i(kr - l\pi/2 + \delta_l)}) P_l(\cos \theta)}{r} \quad (19.5.10)$$

To find A_l , we note that since $V(r)$ produces only an outgoing wave, the *incoming* waves must be the same for $\psi_{\mathbf{k}}$ and the plane wave $e^{i\mathbf{k}\cdot\mathbf{r}} = e^{ikz}$. Comparing the coefficients of e^{-ikr}/r in Eqs. (19.5.8) and (19.5.10), we get

$$A_l = \frac{2l+1}{2ik} e^{i(l\pi/2 + \delta_l)} \quad (19.5.11)$$

Feeding this into Eq. (19.5.10) we get

$$\psi_{\mathbf{k}}(\mathbf{r}) \xrightarrow[r \rightarrow \infty]{} \frac{1}{2ikr} \sum_{l=0}^{\infty} (2l+1)[e^{ikr} e^{2i\delta_l} - e^{-i(kr - l\pi)}] P_l(\cos \theta) \quad (19.5.12)$$

$$= e^{ikz} + \left[\sum_{l=0}^{\infty} (2l+1) \left(\frac{e^{2i\delta_l} - 1}{2ik} \right) P_l(\cos \theta) \right] \frac{e^{ikr}}{r} \quad (19.5.13)$$

Comparing this to Eq. (19.5.2) we get

$$a_l(k) = \frac{e^{2i\delta_l} - 1}{2ik} \quad (19.5.14)$$

Thus, to calculate $a_l(k)$, one must calculate the phase shift δ_l in the asymptotic wave function.

A comparison of Eqs. (19.5.12) and (19.5.8) tells us that the effect of the potential is to attach a phase factor $e^{2i\delta_l}$ to the outgoing wave. This factor does not change the probability current associated with it and the balance between the total incoming and outgoing currents is preserved. This does not mean there is no scattering, since the angular distribution is altered by this phase shift.

One calls

$$S_l(k) = e^{2i\delta_l(k)} \quad (19.5.15)$$

the *partial wave S matrix element* or the *S* matrix for angular momentum l . Recall that the *S* matrix is just the $t \rightarrow \infty$ limit of $U(t, -t)$. It is therefore a function of the Hamiltonian. Since in this problem \mathbf{L} is conserved, *S* (like *H*) will be diagonal in the common eigenbasis of energy ($E = \hbar^2 k^2 / 2\mu$), angular momentum (l), and z component of angular momentum ($m = 0$). Since *S* is unitary (for *U* is), its eigenvalues $S_l(k)$ must be of the form $e^{i\theta}$ and here $\theta = 2\delta_l$. If we go to some other basis, say the $|\mathbf{p}\rangle$ basis, $\langle \mathbf{p}' | S | \mathbf{p} \rangle$ will still be elements of a unitary matrix, but no longer diagonal, for \mathbf{p} is not conserved in the scattering process.

If we rewrite $a_l(k)$ as

$$a_l(k) = \frac{e^{2i\delta_l} - 1}{2ik} = \frac{e^{i\delta_l} \sin \delta_l}{k} \quad (19.5.16)$$

we get

$$f(\theta) = \frac{1}{k} \sum_{l=0}^{\infty} (2l+1) e^{i\delta_l} \sin \delta_l P_l(\cos \theta) \quad (19.5.17)$$

The total cross section

$$\sigma = \int |f|^2 d\Omega$$

is given by

$$\sigma = \frac{4\pi}{k^2} \sum_{l=0}^{\infty} (2l+1) \sin^2 \delta_l \quad (19.5.18)$$

upon using the orthogonality relations for the Legendre polynomials

$$\int P_l(\cos \theta) P_m(\cos \theta) d(\cos \theta) = \frac{2}{2l+1} \delta_{lm}$$

Note that σ is a sum of partial cross sections at each l :

$$\sigma = \sum_{l=0}^{\infty} \sigma_l, \quad \sigma_l = \frac{4\pi}{k^2} (2l+1) \sin^2 \delta_l \quad (19.5.19)$$

Each σ_l has an upper bound σ_l^{\max} , called the *unitarity bound*

$$\sigma_l < \sigma_l^{\max} = \frac{4\pi}{k^2} (2l+1) \quad (19.5.20)$$

The bound is saturated when $\delta_l = n\pi/2$, n odd.

Comparing Eqs. (19.5.17) and (19.5.18) and using $P_l(\cos \theta) = 1$ at $\theta = 0$, we get

$$\sigma = \frac{4\pi}{k} \operatorname{Im} f(0) \quad (19.5.21)$$

This is called the *optical theorem*. It is not too surprising that there exists a relation between the total cross section and the forward amplitude, for the following reason. The incident plane wave brings in some current density in the z direction. Some of it gets scattered into the various directions. This must reflect itself in the form of a

decrease in current density behind the target, i.e., in the $\theta = 0$ direction. The decrease can only occur because the incident plane wave and the scattered wave in the forward direction interfere destructively. It is of course not obvious why just the imaginary part of $f(0)$ is relevant or where the factor $4\pi/k$ comes from. To find out, you must do Exercise 19.5.6.

A Model Calculation of δ_l : The Hard Sphere

Consider a hard sphere, which is represented by

$$\begin{aligned} V(r) &= \infty, & r < r_0 \\ &= 0, & r > r_0 \end{aligned} \quad (19.5.22)$$

We now proceed to solve the radial Schrödinger equation, look at the solution as $r \rightarrow \infty$, and identify the phase shift. Clearly the (unnormalized) radial function $R_l(r)$ vanishes inside $r \leq r_0$. Outside, it is given by the free-particle function:

$$R_l(r) = A_l j_l(kr) + B_l n_l(kr) \quad (19.5.23)$$

(We keep the n_l function since it is regular for $r > 0$.) The coefficients A_l and B_l must be chosen such that

$$R_l(r_0) = 0 \quad (19.5.24)$$

to ensure the continuity of the wave function at $r = r_0$. Thus

$$\frac{B_l}{A_l} = -\frac{j_l(kr_0)}{n_l(kr_0)} \quad (19.5.25)$$

From Eq. (12.6.32), which gives the asymptotic form of j_l and n_l ,

$$\begin{aligned} R_l(r) &\xrightarrow[r \rightarrow \infty]{} \frac{1}{kr} [A_l \sin(kr - l\pi/2) - B_l \cos(kr - l\pi/2)] \\ &= \frac{(A_l^2 + B_l^2)^{1/2}}{kr} \left[\sin\left(kr - \frac{l\pi}{2} + \delta_l\right) \right] \end{aligned} \quad (19.5.26)$$

where

$$\delta_l = \tan^{-1}\left(\frac{-B_l}{A_l}\right) = \tan^{-1}\left[\frac{j_l(kr_0)}{n_l(kr_0)}\right] \quad (19.5.27)$$

For instance [from Eq. (12.6.31)]

$$\begin{aligned}\delta_0 &= \tan^{-1} \left[\frac{\sin(kr_0)/kr_0}{-\cos(kr_0)/kr_0} \right] \\ &= -\tan^{-1} \tan(kr_0) \\ &= -kr_0\end{aligned}\quad (19.5.28)$$

It is easy to understand the result: the hard sphere has pushed out the wave function, forcing it to start its sinusoidal oscillations at $r=r_0$ instead of $r=0$. In general, repulsive potentials give negative phase shifts (since they slow down the particle and reduce the phase shift per unit length) while attractive potentials give positive phase shifts (for the opposite reason). This correspondence is of course true only if δ is small, since δ is defined only modulo π . For instance, if the phase shift $kr_0=\pi$, a_0 vanishes and s -wave scattering does not expose the hard sphere centered at the origin.

Consider the hard sphere phase shift as $k \rightarrow 0$. Using

$$\begin{aligned}j_l(x) &\xrightarrow{x \rightarrow 0} x^l / (2l+1)!! \\ n_l(x) &\xrightarrow{x \rightarrow 0} -x^{-(l+1)} (2l-1)!!\end{aligned}$$

we get

$$\tan \delta_l \underset{k \rightarrow 0}{\cong} \delta_l \propto (kr_0)^{2l+1} \quad (19.5.29)$$

This agrees with the intuitive expectation that at low energies there should be negligible scattering in the high angular momentum states. The above $(kr_0)^{2l+1}$ dependence of δ_l at low energies is true for any reasonable potential, with r_0 being some length scale characterizing the range. [Since there is no hard and fast definition of range, we can *define* the range of any potential to be the r_0 that appears in Eq. (19.5.29).] Notice that although $\delta_0 \propto k^1$, the partial cross section does not vanish because $\sigma_0 \propto k^{-2} \sin^2 \delta_0 \sim k^{-2} \delta_0^2 \neq 0$, as $k \rightarrow 0$.

Resonances

The partial cross section σ_l is generally very small at low energies since $\delta_l \propto (k)^{2l+1}$ as $k \rightarrow 0$. But it sometimes happens that δ_l rises very rapidly from 0 to π [or more generally, from $n\pi$ to $(n+1)\pi$] in a very small range of k or E . In this region, near $k=k_0$ or $E=E_0$, we may describe δ_l by

$$\delta_l = \delta_h + \tan^{-1} \left(\frac{\Gamma/2}{E_0 - E} \right) \quad (19.5.30)$$

where δ_b is some *background phase* ($\cong n\pi$) that varies very little. The corresponding cross section, neglecting δ_b , is

$$\begin{aligned}\sigma_l &= \frac{4\pi}{k^2} (2l+1) \sin^2 \delta_l \\ &= \frac{4\pi}{k^2} (2l+1) \frac{(\Gamma/2)^2}{(E_0 - E)^2 + (\Gamma/2)^2}\end{aligned}\quad (19.5.31)$$

σ_l is described by a bell-shaped curve, called the *Breit–Wigner form*, with a maximum height σ_l^{\max} (the unitarity bound) and a half-width $\Gamma/2$. This phenomenon is called a *resonance*.

In Eq. (19.5.31) for σ_l , valid only near E_0 , we have treated Γ as a constant. Its k dependence may be deduced by noting that as $k \rightarrow 0$, we have [from Eq. (19.5.29)],

$$\sigma_l \sim \frac{1}{k^2} \sin^2 \delta_l \cong \frac{1}{k^2} \delta_l^2 \cong \frac{(kr_0)^{4l+2}}{k^2}$$

which implies

$$\Gamma/2 = (kr_0)^{2l+1} \gamma \quad (19.5.32)$$

where γ is some constant with dimensions of energy. Thus the expression for σ_l that is valid over a wider range is

$$\sigma_l = \frac{4\pi}{k^2} (2l+1) \frac{[\gamma(kr_0)^{2l+1}]^2}{(E - E_0)^2 + [\gamma(kr_0)^{2l+1}]^2} \quad (19.5.33)$$

For any $l \neq 0$, σ_l is damped in the entire low-energy region by the net k^{4l} factor, except near E_0 , where a similar factor in the denominator neutralizes it. Clearly, as l goes up, the resonances get sharper. The situation at $l=0$ (where σ_0 starts out nonzero at $k=0$) depends on the potential. More on this later.

We would now like to gain some insight into the dynamics of resonances. We ask what exactly is going on at a resonance, in terms of the underlying Schrödinger equation. We choose to analyze the problem through the S matrix. Near a resonance we have

$$S_l(k) = e^{2i\delta_l} = \frac{e^{i\delta_l}}{e^{-i\delta_l}} = \frac{1 + i \tan \delta_l}{1 - i \tan \delta_l} = \frac{E - E_0 - i\Gamma/2}{E - E_0 + i\Gamma/2} \quad (19.5.34)$$

Although k and E are real in any experiment (and in our analysis so far), let us think of $S_l(k)$ as a function of complex E or k . Then we find that the resonance corresponds to a pole in S_l at a complex point,

$$E = E_0 - i\Gamma/2 \quad (19.5.35)$$

or

$$k = k_0 - i\eta/2 \quad (19.5.36)$$

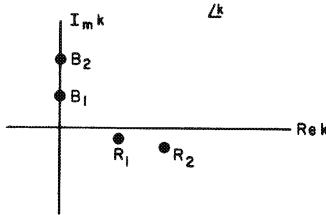


Figure 19.4. Some of the singularities of $S_l(k)$ in the complex k plane. The dots on the positive imaginary axis stand for bound state poles and the dots below the real axis stand for resonance poles. The physical or experimentally accessible region is along the real axis, where S_l has the form $e^{2i\delta_l}$.

where $E_0 = \hbar^2 k_0^2 / 2\mu$ and $\Gamma = \eta \hbar^2 k_0 / \mu$ (for small η and Γ). Since Γ and η are small, the pole is very close to the real axis, which is why we trust the form of S_l that is valid near the point $E = E_0$ on the real axis.

What is the implication of the statement that the resonance corresponds to a (nearby) pole in $S_l(k)$? To find out, we take a new look at bound states in terms of the S matrix. Recall that for k real and positive, if

$$R_{kl}(r) \xrightarrow[r \rightarrow \infty]{} \frac{A e^{ikr}}{r} + \frac{B e^{-ikr}}{r} \quad (19.5.37)$$

then [from Eqs. (19.5.9) and (19.5.10) or Eq. (19.5.12)],

$$e^{2i\delta_l} = S_l(k) = \frac{A}{B} = \frac{\text{outgoing wave amplitude}}{\text{incoming wave amplitude}} \quad (19.5.38)$$

(up to a constant factor $i^2 l$). We now define $S_l(k)$ for complex k as follows: solve the radial equation with k set equal to a complex number, find $R(r \rightarrow \infty)$, and take the ratio A/B . Consider now the case $k = i\kappa$ ($\kappa > 0$), which corresponds to E real and negative. Here we will find

$$R_{kl}(r) \xrightarrow[r \rightarrow \infty]{} \frac{A e^{-\kappa r}}{r} + \frac{B e^{\kappa r}}{r} \quad (19.5.39)$$

Whereas $S_l(k = i\kappa)$ is well defined, the corresponding R_{kl} does not interest us, since it is not normalizable. But recall that for some special values of k , R_{kl} is exponentially damped and describes the wave function of a bound state. These bound states correspond to k such that $B = 0$, or $S_l(k) = \infty$. Thus poles of $S_l(k)$ at $k = i\kappa$ correspond to bound states.

So a resonance, which is a pole at $k = k_0 - i\eta$ must also be some kind of bound state. (See Fig. 19.4 for poles of the S matrix.) We next argue heuristically as follows.[‡] Since the bound state at $E = E_B$ (a negative number) has the time dependence

$$e^{-iE_B t/\hbar}$$

[‡] This result may be established rigorously.

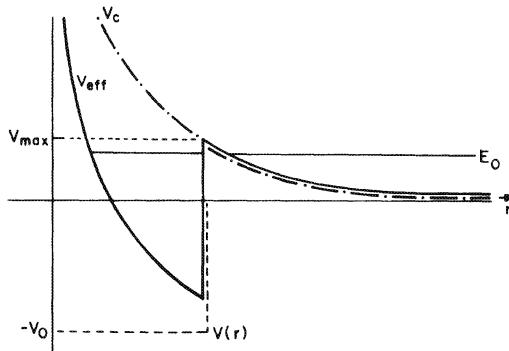


Figure 19.5. A typical potential that can sustain resonances. The centrifugal repulsion V_c (dot-dash line) plus the actual attractive potential (dotted line) gives the effective potential V_{eff} (solid line). The figure shows an example where there would have been a bound state at E_0 but for tunneling. But because of tunneling the particle can leak out, and by the same token, a particle can come from outside with positive energy E_0 , form a metastable bound state (with a lifetime inversely proportional to the tunneling probability), and then escape. This is called resonance.

the resonance must have a time dependence

$$e^{-i(E_0 - i\Gamma/2)t/\hbar} = e^{-iE_0 t/\hbar} e^{-\Gamma t/2\hbar}$$

This describes a state of positive energy E_0 , but whose norm falls exponentially with a half-life $t \sim \hbar/\Gamma$. Thus, a resonance, corresponding to a pole at $E = E_0 - i\Gamma/2$, describes a *metastable* bound state of energy E_0 and lifetime $t = \hbar/\Gamma$.‡

So we must next understand how a positive-energy particle manages to form a metastable bound state. Consider the case where $V(r)$ is attractive, say a square well of depth V_0 and range r_0 . The potential appearing in the radial equation is $V_{\text{eff}} = V + V_c$, where V_c is the centrifugal repulsion (Fig. 19.5). The main point is that V_{eff} is attractive at short distances and repulsive at long distances. Consider now a particle with energy $E_0 < V_{\text{max}}$, such that if tunneling is ignored, the particle can form a bound state inside the attractive region, i.e., we can fit in an integral number of half-wavelengths. But tunneling is of course present and the particle can escape to infinity as a free particle of energy E_0 . Conversely, a free particle of energy E_0 shot into the potential can penetrate the barrier and form a metastable bound state and leak out again. This is when we say resonance is formed. This picture also explains why the resonances get narrower as l increases: as l increases, V_c grows, tunneling is suppressed more, and the lifetime of the metastable state grows. We can also see why $l=0$ is different: there is no repulsive barrier due to V_c . If $V = V_{\text{eff}}$ is purely attractive, only genuine (negative energy) bound states are possible. The closest thing to a resonance is the formation of a bound state near zero energy (Exercise 19.5.4). If, however, V itself has the form of V_{eff} in Fig. 19.5, resonances are possible.

*Exercise 19.5.2.** Derive Eq. (19.5.18) and provide the missing steps leading to the optical theorem, Eq. (19.5.21).

‡ The energy is not strictly E_0 because the uncertainty principle does not allow us to define a precise energy for a state of finite lifetime. E_0 is the mean energy.

Exercise 19.5.3. (1) Show that $\sigma_0 \rightarrow 4\pi r_0^2$ for a hard sphere as $k \rightarrow 0$.

(2) Consider the other extreme of kr_0 very large. From Eq. (19.5.27) and the asymptotic forms of j_l and n_l show that

$$\sin^2 \delta_l \xrightarrow[kr_0 \rightarrow \infty]{} \sin^2(kr_0 - l\pi/2)$$

so that

$$\begin{aligned} \sigma = \sum_{l=0}^{l_{\max}=kr_0} \sigma_l &\cong \frac{4\pi}{k^2} \int_0^{kr_0} (2l) \sin^2 \delta_l dl \\ &\cong 2\pi r_0^2 \end{aligned}$$

if we approximate the sum over l by an integral, $2l+1$ by $2l$, and the oscillating function $\sin^2 \delta$ by its mean value of $1/2$.

*Exercise 19.5.4.** Show that the s -wave phase shift for a square well of depth V_0 and range r_0 is

$$\delta_0 = -kr_0 + \tan^{-1} \left(\frac{k}{k'} \tan k'r_0 \right)$$

where k' and k are the wave numbers inside and outside the well. For k small, kr_0 is some small number and we ignore it. Let us see what happens to δ_0 as we vary the depth of the well, i.e., change k' . Show that whenever $k' \simeq k'_n = (2n+1)\pi/2r_0$, δ_0 takes on the resonant form Eq. (19.5.30) with $\Gamma/2 = \hbar^2 k_n / \mu r_0$, where k_n is the value of k when $k' = k'_n$. Starting with a well that is too shallow to have any bound state, show k'_1 corresponds to the well developing its first bound state, at zero energy. (See Exercise 12.6.9.) (Note: A zero-energy bound state corresponds to $k=0$.) As the well is deepened further, this level moves down, and soon, at k'_2 , another zero-energy bound state is formed, and so on.

Exercise 19.5.5. Show that even if a potential absorbs particles, we can describe it by

$$S_l(k) = \eta_l(k) e^{2i\delta_l}$$

where $\eta_l (< 1)$, is called the *inelasticity factor*.

(1) By considering probability currents, show that

$$\begin{aligned} \sigma_{\text{inel}} &= \frac{\pi}{k^2} \sum_{l=0}^{\infty} (2l+1)[1 - \eta_l^2] \\ \sigma_{\text{el}} &= \frac{\pi}{k^2} \sum_{l=0}^{\infty} (2l+1)(1 + \eta_l^2 - 2\eta_l \cos 2\delta_l) \end{aligned}$$

and that once again

$$\sigma_{\text{tot}} = \frac{4\pi}{k} \operatorname{Im} f(0)$$

- (2) Consider a “black disk” which absorbs everything for $r \leq r_0$ and is ineffective beyond. Idealize it by $\eta = 0$ for $l \leq kr_0$; $\eta = 1$, $\delta = 0$ for $l > kr_0$. Show that $\sigma_{\text{el}} = \sigma_{\text{inel}} \approx \pi r_0^2$. Replace the sum by an integral and assume $kr_0 \gg 1$. (See Exercise 19.5.3.) Why is σ_{inel} always accompanied by σ_{el} ?

Exercise 19.5.6. (The Optical Theorem). (1) Show that the radial component of the current density due to interference between the incident and scattered waves is

$$j_r^{\text{int}} \underset{r \rightarrow \infty}{\sim} \left(\frac{\hbar k}{\mu} \right) \frac{1}{r} \operatorname{Im} [i e^{ikr(\cos \theta - 1)} f^*(\theta) \cos \theta + i e^{ikr(1 - \cos \theta)} f(\theta)]$$

(2) Argue that as long as $\theta \neq 0$, the average of j_r^{int} over any small solid angle is zero because $r \rightarrow \infty$. [Assume $f(\theta)$ is a smooth function.]

(3) Integrate j_r^{int} over a tiny cone in the forward direction and show that (see hint)

$$\int_{\text{forward cone}} j_r^{\text{int}} r^2 d\Omega = - \left(\frac{\hbar k}{\mu} \right) \frac{4\pi}{k} \operatorname{Im} f(0)$$

Thus, if we integrate the total current in the region behind the target, we find that the interference term (important only in the near-forward direction, behind the target) produces a depletion of particles, casting a “shadow.” The total number of particles (per second) missing in the shadow region is given by the above expression for the integrated flux. Equating this loss to the product of the incident flux $\hbar k / \mu$ and the cross section σ , we regain the optical theorem. (Hint: Since θ is small, set $\sin \theta \approx \theta$, $\cos \theta = 1 - \theta^2/2$ using the judgment. In evaluating the upper limit in the θ integration, use the idea introduced in Chapter 1, namely, that the limit of a function that oscillates as its argument approaches infinity is equal to its average value.)

19.6. Two-Particle Scattering

In this section we will see how the differential cross section for two-body scattering may be extracted from the solution of the Schrödinger equation for the relative coordinate with a potential $V(\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2)$. Let us begin by considering the total and differential cross sections for two-body scattering. Let σ be the total cross section for the scattering of the two particles. Imagine a beam of projectiles with density ρ_1 and *magnitude* of velocity v_1 colliding *head on* with the beam of targets with parameters ρ_2 and v_2 . How many collisions will there be per second? We know that if there is only one target and it is at rest,

$$\begin{aligned} \text{No. of collisions/sec} &= \sigma \times \text{incident projectiles/sec/area} \\ &= \sigma \rho_1 v_1 \end{aligned} \tag{19.6.1}$$

Here we modify this result to take into account that (1) there are ρ_2 targets per *unit volume* (ρ_2 is assumed so small that the targets scatter independently of each other),

and (2) the targets are moving toward the projectiles at a relative velocity $v_{\text{rel}} = v_1 + v_2$. Consequently we have

$$\begin{aligned}\text{No. of collisions/sec/volume of interaction} &= \sigma \rho_1 (v_1 + v_2) \rho_2 \\ &= \sigma \rho_1 \rho_2 v_{\text{rel}}\end{aligned}\quad (19.6.2)$$

Note that σ is the same for all observers moving along the beam-target axis.

What about the differential cross section? It *will* depend on the frame. In the lab frame, where the target is initially at rest, we define, in analogy with Eq. (19.6.2),

No. of projectiles scattered into $d(\cos \theta_L) d\phi_L / \text{sec/vol}$

$$= \frac{d\sigma}{d\Omega_L} d\Omega_L \rho_1 \rho_2 v_{\text{rel}} \quad (19.6.3)$$

Here v_{rel} is just the projectile velocity and θ_L and ϕ_L are angles in the lab frame measured relative to the projectile direction. (We can also define a $d\sigma/d\Omega_L$ in terms of how many *target* particles are scattered into $d\Omega_L$, but it would not be an independent quantity since momentum conservation will fix the fate of the target, given the fate of the projectile.) The only other frame we consider is the CM frame, where $(d\sigma/d\Omega) d\Omega$ is defined as in Eq. (19.6.3).‡ We relate $d\sigma/d\Omega$ to $d\sigma/d\Omega_L$ by the following argument. Imagine a detector in the lab frame at (θ_L, ϕ_L) which subtends an angle $d\Omega_L$. The number of counts it registers is an absolute, frame-independent quantity, although its orientation and acceptance angle $d\Omega$ may vary from frame to frame. (For example, a particle coming at right angles to the beam axis in the lab frame will be tilted forward in a frame moving backward.) So we deduce the following equality from Eq. (19.6.2) after noting the frame invariance of $\rho_1 \rho_2 v_{\text{rel}}$:

$$\frac{d\sigma}{d\Omega_L} d\Omega_L = \frac{d\sigma}{d\Omega} d\Omega \quad (19.6.4)$$

or

$$\frac{d\sigma}{d\Omega_L} = \frac{d\sigma}{d\Omega} \frac{d\Omega}{d\Omega_L} \quad (19.6.5)$$

We will consider first the calculation of $d\sigma/d\Omega$, and then $d\Omega/d\Omega_L$.

Let us represent the state of the two colliding particles, long before they begin to interact, by the product wave function (in some general frame):

$$\psi_{\text{inc}} = e^{i\mathbf{k}_1 \cdot \mathbf{r}_1} e^{i\mathbf{k}_2 \cdot \mathbf{r}_2} \quad (19.6.6)$$

‡ The CM variables will carry no subscripts.

We should remember that these plane waves are idealized forms of broad wave packets. Assuming both are moving along the z axis,

$$\begin{aligned}\psi_{\text{inc}} &= e^{ik_1 z_1} e^{ik_2 z_2} \\ &= \exp\left[i(k_1 + k_2)\left(\frac{z_1 + z_2}{2}\right)\right] \exp\left[i\left(\frac{k_1 - k_2}{2}\right)(z_1 - z_2)\right] \\ &= \psi_{\text{inc}}^{\text{CM}}(z_{\text{CM}}) \psi_{\text{inc}}^{\text{rel}}(z)\end{aligned}\quad (19.6.7)$$

Since the potential affects only the relative coordinate, the plane wave describes the CM completely; there is no scattering for the CM as a whole. On the other hand, $\psi_{\text{inc}}^{\text{rel}}(z)$ will develop a scattered wave and become

$$\begin{aligned}\psi(z) &= e^{ikz} + \psi_{\text{sc}}(\mathbf{r}) \\ &\xrightarrow[r \rightarrow \infty]{} e^{ikz} + f(\theta, \phi) e^{ikr}/r\end{aligned}\quad (19.6.8)$$

where we have dropped the superscript “rel,” since the argument z makes it obvious, and set $(k_1 - k_2)/2$ equal to k . Thus the static solution for the entire system is

$$\begin{aligned}\psi_{\text{system}}(\mathbf{r}_1, \mathbf{r}_2) &= \psi^{\text{CM}}(z_{\text{CM}})[e^{ikz} + \psi_{\text{sc}}(\mathbf{r})] \\ &\xrightarrow[r \rightarrow \infty]{} \psi^{\text{CM}}(z_{\text{CM}})[e^{ikz} + f(\theta, \phi) e^{ikr}/r]\end{aligned}\quad (19.6.9)$$

If we go to the CM frame, $\psi^{\text{CM}}(z_{\text{CM}}) = e^{i(k_1 + k_2)z_{\text{CM}}} = 1$, since $k_1 + k_2 = 0$ defines this frame. So we can forget all about the CM coordinate. The scattering in the CM frame is depicted in Fig. 19.6. The classical trajectories are not to be taken literally;

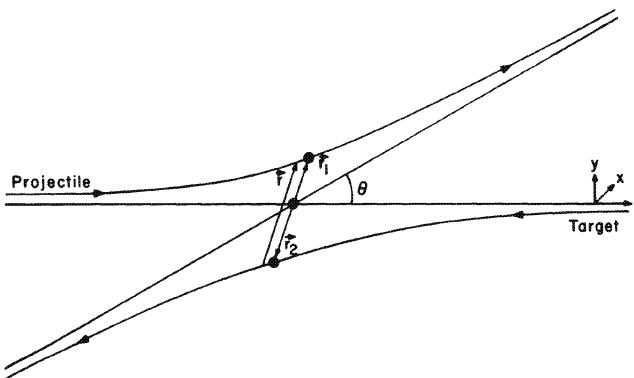


Figure 19.6. Two-body scattering in the CM frame. The projectile and target coordinates are \mathbf{r}_1 and \mathbf{r}_2 , respectively. The relative coordinate $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$ is slightly displaced in the figure for clarity. Since \mathbf{r}_1 and \mathbf{r} are always parallel, the probability that the projectile scatters into $d\Omega$ is the same as the probability that the fictitious particle described by \mathbf{r} scatters into $d\Omega$. To find the latter, we must solve the Schrödinger equation for \mathbf{r} .

they merely define the relative coordinate \mathbf{r} and the individual coordinates \mathbf{r}_1 (projectile) and \mathbf{r}_2 (target).

What we want is the rate at which the projectile scatters into $d\Omega$. But since \mathbf{r}_1 is parallel to \mathbf{r} , this equals the rate at which the fictitious particle described by \mathbf{r} scatters into solid angle $d\Omega$. We find this rate by solving the Schrödinger equation for the relative coordinate. Having done so, and having found

$$\psi(\mathbf{r}) \xrightarrow[r \rightarrow \infty]{} e^{ikz} + f(\theta, \phi) \frac{e^{ikr}}{r} \quad (19.6.10)$$

we recall from Eq. (19.2.17) the rate of scattering into $d\Omega$:

$$R_{i \rightarrow d\Omega} = |f(\theta, \phi)|^2 \frac{\hbar k}{\mu} d\Omega \quad (19.6.11)$$

Note that is the rate *per unit volume* of target-beam interaction, since the probability density for the CM is unity. To extract $d\sigma/d\Omega$ from $R_{i \rightarrow d\Omega}$ above we turn to Eq. (19.6.3) which defines $d\sigma/d\Omega$ (upon dropping the subscript L). Since the definition makes sense only for a flux of wave packets and since we are dealing with plane waves here, we replace the number scattered into $d\Omega$ per second by the probability flowing into $d\Omega$ per second, and the particle densities ρ_1 and ρ_2 by probability densities of the colliding beams. Since the colliding beams ($e^{ikz} = e^{ik(z_1 - z_2)} = e^{ikz_1} \cdot e^{-ikz_2}$) are plane waves of unit modulus, $\rho_1 = \rho_2 = 1$. How about v_{rel} ? Remember that in the CM frame

$$m_1 v_1 = m_2 v_2$$

so

$$\begin{aligned} v_{\text{rel}} &= v_1 + v_2 = v_1 \left(1 + \frac{m_1}{m_2} \right) = v_1 \left(\frac{m_2 + m_1}{m_2} \right) = m_1 v_1 \left(\frac{m_2 + m_1}{m_1 m_2} \right) \\ &= \hbar k \left(\frac{m_2 + m_1}{m_1 m_2} \right) = \frac{\hbar k}{\mu} \end{aligned} \quad (19.6.12)$$

So

$$\frac{d\sigma}{d\Omega} d\Omega = \frac{R_{i \rightarrow d\Omega}}{\rho_1 \rho_2 (v_1 + v_2)} = \frac{|f|^2 (\hbar k / \mu)}{\hbar k / \mu} d\Omega$$

or

$$\frac{d\sigma}{d\Omega} = |f|^2 \quad (19.6.13)$$

Thus the $d\sigma/d\Omega$ we calculated in the previous sections for a single particle scattering off a potential $V(\mathbf{r})$ can also be interpreted as the CM cross section for two bodies interacting via a potential $V(\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2)$.

Passage to the Lab Frame

We now consider the passage to the lab frame, i.e., the calculation of $d\Omega/d\Omega_L$. We discuss the equal mass case, leaving the unequal mass case as an exercise. Figure 19.7a shows the particles coming in with momenta p and $-p$ along the z axis in the CM frame. If \mathbf{p}' is the final momentum of the projectile,

$$\tan \theta = \frac{(p_x'^2 + p_y'^2)^{1/2}}{p_z'} \equiv \frac{p'_\perp}{p_z'} \quad (19.6.14a)$$

$$\tan \phi = p_y'/p_x' \quad (19.6.14b)$$

To go to the lab frame we must move leftward at a speed p/m . In this frame, all momenta get an increment in the z direction (only) equal to p . (Thus T , the target, will be at rest before collision.) The scattering angles in the lab frame are given by

$$\tan \theta_L = p'_\perp / (p_z' + p) \quad (19.6.15a)$$

$$\tan \phi_L = p_y'/p_x' \quad (19.6.15b)$$

Comparing Eqs. (19.6.14) and (19.6.15) we get

$$\phi_L = \phi \quad (19.6.16)$$

$$\begin{aligned} \tan \theta_L &= \frac{p'_\perp}{p_z' + p} = \frac{p'_\perp/p}{p_z'/p + 1} = \frac{\sin \theta}{\cos \theta + 1} \\ &= \tan(\theta/2) \quad (\text{using } |\mathbf{p}'| = p) \end{aligned}$$

So

$$\theta_L = \theta/2 \quad (19.6.17)$$

One consequence of the result is that $\theta_L \leq \pi/2$. Given Eqs. (19.6.16) and (19.6.17) it is a simple matter to relate $d\sigma/d\Omega$ to $d\sigma/d\Omega_L$.

*Exercise 19.6.1.** (1) Starting with Eqs. (19.6.16) and (19.6.17), show that the relation between $d\sigma/d\Omega$ and $d\sigma/d\Omega_L$ is

$$\left. \frac{d\sigma}{d\Omega_L} \right|_{\theta_0} = \left. \frac{d\sigma}{d\Omega} \right|_{2\theta_0} 4 \cos \theta_0$$

(2) Show that $\theta_L \leq \pi/2$ by using just energy and momentum conservation.

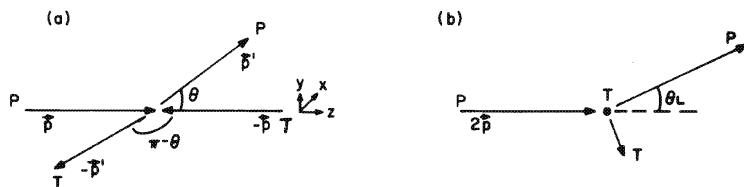


Figure 19.7. (a) Collision of two equal masses in the CM frame. The labels P and T refer to projectile and target. The angle ϕ equals $\frac{1}{2}\pi$ in the figure. (b) The same collision in the lab frame (where T is initially at rest).

(3) For unequal mass scattering, show that

$$\tan \theta_L = \frac{\sin \theta}{\cos \theta + (m_1/m_2)}$$

where m_2 is the target mass.

Scattering of Identical Particles

Consider the scattering of two identical spin-zero bosons in their CM frame. We must describe them by a symmetrized wave function. Under the exchange $\mathbf{r}_1 \leftrightarrow \mathbf{r}_2$; $\mathbf{r}_{CM} = (\mathbf{r}_1 + \mathbf{r}_2)/2$ is invariant while $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$ changes sign. So $\psi^{CM}(\mathbf{r}_{CM})$ is automatically symmetric. We must symmetrize $\psi(\mathbf{r})$ by hand:

$$\psi_{sym}(\mathbf{r}) \xrightarrow[r \rightarrow \infty]{} (e^{ikz} + e^{-ikz}) + [f(\theta, \phi) + f(\pi - \theta, \phi + \pi)] e^{ikr}/r \quad (19.6.18)$$

We have used the fact that under $\mathbf{r} \rightarrow -\mathbf{r}$, $\theta \rightarrow \pi - \theta$ and $\phi \rightarrow \phi + \pi$. The scattering amplitude is thus

$$f_{sym}(\theta, \phi) = f(\theta, \phi) + f(\pi - \theta, \phi + \pi) \quad (19.6.19)$$

Note that f_{sym} is consistent with the fact that since the particles are identical, one cannot say which one scattered into (θ, ϕ) and which one into $(\pi - \theta, \phi + \pi)$ (Fig. 19.7). The differential cross section is

$$\begin{aligned} \frac{d\sigma}{d\Omega} &= |f(\theta, \phi) + f(\pi - \theta, \phi + \pi)|^2 \\ &= |f(\theta, \phi)|^2 + |f(\pi - \theta, \phi + \pi)|^2 + 2 \operatorname{Re}[f(\theta, \phi)f^*(\pi - \theta, \phi + \pi)] \end{aligned} \quad (19.6.20)$$

The first two terms are what we would get if we had two distinguishable particles and asked for the rate at which one or the other comes into $d\Omega$. The third term gives the usual quantum mechanical interference that accompanies identical particles. There are two features worth noting about Eq. (19.6.20):

- (1) To find σ , we must integrate over only 2π radians and not 4π radians (if not, we will count each *distinguishable* event twice).

(2) Recall that when we obtained the Rutherford cross section by taking the $\mu_0 \rightarrow 0$ limit of the Yukawa cross section, we got the right answer although $f(\theta)$ was not right: it did not contain the exponential phase factor that comes from a careful treatment of the Coulomb potential [see Eq. (19.3.16) and the sentences following it.] When we consider the Coulomb scattering of identical bosons (of charge e , say) the interference terms expose the inadequacy of the $\mu_0 \rightarrow 0$ approach. The correct cross section is[‡]

$$\frac{d\sigma}{d\Omega} = \left(\frac{e^2}{4E} \right)^2 \left[\frac{1}{\sin^4 \theta/2} + \frac{1}{\cos^4 \theta/2} + \frac{2 \cos(\gamma \ln \tan^2 \theta/2)}{\sin^2 \theta/2 \cos^2 \theta/2} \right] \quad (19.6.21)$$

whereas the $\mu_0 \rightarrow 0$ trick would not have given the $\cos(\gamma \ln \tan^2 \theta/2)$ factor. (The classical Rutherford treatment would not give the third term at all. Notice, however, that as $\hbar \rightarrow 0$, it oscillates wildly and averages to zero over any realistic detector.)

Consider now the scattering of two identical spin-1/2 fermions, say electrons. Let us assume that the spin variables are spectators, except for their role in the statistics: in the triplet state the spatial function is antisymmetric, while in the singlet it is symmetric. If the electrons are assumed to come in with random values of s_z , the triplet is three times as likely as the singlet and the average cross section will be

$$\begin{aligned} \frac{d\sigma}{d\Omega} = & \frac{3}{4} |f(\theta, \phi) - f(\pi - \theta, \phi + \pi)|^2 \\ & + \frac{1}{4} |f(\theta, \phi) + f(\pi - \theta, \phi + \pi)|^2 \end{aligned} \quad (19.6.22)$$

For Coulomb scattering of electrons this becomes

$$\frac{d\sigma}{d\Omega} = \left(\frac{e^2}{4E} \right)^2 \left[\frac{1}{\sin^4 \theta/2} + \frac{1}{\cos^4 \theta/2} - \frac{\cos(\gamma \ln \tan^2 \theta/2)}{\sin^2 \theta/2 \cos^2 \theta/2} \right] \quad (19.6.23)$$

Exercise 19.6.2. Derive Eq. (19.6.21) using Eq. (19.3.16) for $f_c(\theta)$.

Exercise 19.6.3. Assuming $f = f(\theta)$ show that $(d\sigma/d\Omega)_{\pi/2} = 0$ for fermions in the triplet state.

[‡] $\gamma = e^2 \mu / \hbar^2 k$ here.

The Dirac Equation

Nonrelativistic quantum mechanics, which was developed in the previous chapters, is very successful when applied to problems like the hydrogen atom, where the typical velocity (speaking semiclassically) is small compared to c . (Recall $v/c = \beta = \alpha \approx 1/137$ in the ground state.) But even in this case, there are measurable (fine-structure) corrections of the order of $(v/c)^4$ which have to be put in by hand. If these corrections are to emerge naturally and if relativistic systems (high- Z atoms, for example) are to be described well, it is clear that we need an equation for the electron that has relativity built into it from the start. Such an equation was discovered by Dirac. We study it here with the *main goal* of seeing the coherent emergence of several concepts that were introduced disjointly at various stages—the spin of the electron, its magnetic moment ($g=2$), the spin-orbit, and other fine-structure corrections.

In the last section we address some general questions that accompany Dirac's formulation and indicate the need for quantum field theory.

20.1. The Free-Particle Dirac Equation

Let us consider the simplest case, of a free particle. We start by stating the relation between classical mechanics and the free-particle Schrödinger equation in a way that facilitates generalization. If we start with the nonrelativistic relation

$$\mathcal{H} = \frac{|\mathbf{p}|^2}{2m} = \frac{p^2}{2m} \quad (20.1.1)$$

and make the substitution

$$\begin{aligned} \mathbf{p} &\rightarrow \mathbf{P} \\ \mathcal{H} &\rightarrow i\hbar \frac{\partial}{\partial t} \end{aligned} \quad (20.1.2)$$

and let both sides act on a state vector $|\psi\rangle$, we get Schrödinger's equation

$$i\hbar \frac{\partial |\psi\rangle}{\partial t} = \frac{P^2}{2m} |\psi\rangle \quad (20.1.3)$$

A natural starting point for the relativistic equation is the corresponding relation due to Einstein

$$\mathcal{H} = (c^2 p^2 + m^2 c^4)^{1/2} \quad (20.1.4)$$

If we make the substitution mentioned above, we get

$$i\hbar \frac{\partial |\psi\rangle}{\partial t} = (c^2 P^2 + m^2 c^4)^{1/2} |\psi\rangle \quad (20.1.5)$$

This equation is undesirable because it treats space and time asymmetrically. To see this, we first go to the momentum basis, where \mathbf{P} is just \mathbf{p} and the square root may be expanded in a series:

$$i\hbar \frac{\partial \psi(\mathbf{p}, t)}{\partial t} = mc^2 \left(1 + \frac{p^2}{2m^2 c^2} - \frac{p^4}{8m^4 c^4} + \dots \right) \psi(\mathbf{p}, t) \quad (20.1.6)$$

If we now transform to the coordinate basis, each p^2 becomes $(-\hbar^2 \nabla^2)$ and the asymmetry between space and time is manifest. What we want is an equation that is of the same order in both space and time.

There are two ways out. One is to replace Eq. (20.1.4) by

$$\mathcal{H}^2 = c^2 p^2 + m^2 c^4 \quad (20.1.7)$$

and obtain, upon making the operator substitution,

$$\frac{\partial^2 |\psi\rangle}{\partial t^2} = \left(-\frac{c^2 P^2}{\hbar^2} - \frac{m^2 c^4}{\hbar^2} \right) |\psi\rangle \quad (20.1.8a)$$

In the coordinate basis this becomes

$$\left[\frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \nabla^2 + \left(\frac{mc^2}{\hbar} \right)^2 \right] \psi = 0 \quad (20.1.8b)$$

This is called the *Klein-Gordon equation* and has the desired symmetry between space and time. But we move along, since ψ here is a scalar and cannot describe the electron. It is, however, a good candidate for pions, kaons, etc., which are spinless.

The second alternative, due to Dirac, is the following. Let us suppose that the quantity in the square root in Eq. (20.1.5) can be written as a perfect square of a quantity that is linear in \mathbf{P} . We can then take the square root (which will give us

our Hamiltonian) and obtain an equation that is of the first order in time and space. So let us write

$$\begin{aligned} c^2 P^2 + m^2 c^4 &= (c\alpha_x P_x + c\alpha_y P_y + c\alpha_z P_z + \beta mc^2)^2 \\ &= (\mathbf{c}\alpha \cdot \mathbf{P} + \beta mc^2)^2 \end{aligned} \quad (20.1.9)$$

where α and β are to be determined by matching both sides of

$$\begin{aligned} c^2(P_x^2 + P_y^2 + P_z^2) + m^2 c^4 &= [c^2(\alpha_x^2 P_x^2 + \alpha_y^2 P_y^2 + \alpha_z^2 P_z^2) + \beta^2 m^2 c^4] \\ &\quad + [c^2 P_x P_y (\alpha_x \alpha_y + \alpha_y \alpha_x) + \text{and cyclic permutations}] \\ &\quad + [mc^3 P_x (\alpha_x \beta + \beta \alpha_x) + x \rightarrow y \rightarrow z] \end{aligned} \quad (20.1.10)$$

(We have assumed that α and β are space independent, which is a reasonable assumption for a free particle.) These equations tell us that

$$\begin{aligned} \alpha_i^2 &= \beta^2 = 1 \quad (i = x, y, z) \\ \alpha_i \alpha_j + \alpha_j \alpha_i &= [\alpha_i, \alpha_j]_+ = 0 \quad (i \neq j) \\ \alpha_i \beta + \beta \alpha_i &= [\alpha_i, \beta]_+ = 0 \end{aligned} \quad (20.1.11)$$

It is evident that α and β are not c numbers. They are matrices and furthermore Hermitian (so that the Hamiltonian $H = c\alpha \cdot \mathbf{P} + \beta mc^2$ is Hermitian), traceless, and have eigenvalues ± 1 . (Recall the results of Exercise 1.8.8). They must also be even dimensional if the last two properties are to be compatible. They cannot be 2×2 matrices, since, as we saw in Exercise 14.3.8, the set of three Pauli matrices with these properties cannot be enlarged to include a fourth. So they must be 4×4 matrices. They are not unique (since $\alpha \rightarrow S^\dagger \alpha S$, $\beta \rightarrow S^\dagger \beta S$ preserves the desired properties if S is unitary.) The following four are frequently used and will be used by us:

$$\alpha = \begin{bmatrix} 0 & \sigma \\ \sigma & 0 \end{bmatrix}, \quad \beta = \begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix} \quad (20.1.12)$$

In the above, σ and I are 2×2 matrices.[‡] We now have the *Dirac equation*:

$$i\hbar \frac{\partial |\psi\rangle}{\partial t} = (c\alpha \cdot \mathbf{P} + \beta mc^2) |\psi\rangle \quad (20.1.13)$$

with α and β known. Hereafter we work exclusively in the coordinate basis. However, we depart from our convention and use the symbol \mathbf{P} , reserved for the momentum operator in the abstract, to represent it in the coordinate basis (instead of using $-i\hbar \nabla$). This is done to simplify the notation in what follows.

[‡] For example, β is a 4×4 diagonal matrix with the first two entries $+1$ and the next two entries -1 .

The fact that α and β in

$$i\hbar \frac{\partial \psi}{\partial t} = (c\alpha \cdot \mathbf{P} + \beta mc^2)\psi \quad (20.1.14)$$

are 4×4 matrices implies that ψ is a four-component object. It is called a *Lorentz spinor*. Our reaction is mixed. We are happy that relativity, plus the requirement that the equation be first order in time and space, have led naturally to a multicomponent wave function. But we are distressed that ψ has four components instead of two. In the next two sections we will see how, despite this apparent problem, the Dirac equation describes electrons.

For later use, let us note that since the Hamiltonian is Hermitian, the norm of the state is conserved. In the coordinate basis this means

$$\int \psi^\dagger \psi d^3r = \text{const} \quad (20.1.15)$$

Just as in the nonrelativistic case, this global conservation law has a local version also. (See exercise below.)

*Exercise 20.1.1.** Derive the continuity equation

$$\frac{\partial P}{\partial t} + \nabla \cdot \mathbf{j} = 0$$

where $P = \psi^\dagger \psi$ and $\mathbf{j} = c\psi^\dagger \alpha \psi$.

20.2. Electromagnetic Interaction of the Dirac Particle

In this central section, we see how several properties of the electron emerge naturally from the Dirac equation. As a first step, we couple the particle to the potential (\mathbf{A}, ϕ) . We then consider the equation to order $(v/c)^2$ and show that the particle can be described by a two-component wave function and that it has $g=2$. Finally we consider the equation to order $(v/c)^4$ and see the fine-structure emerge.

The coupling of the electromagnetic potentials is suggested by the classical Hamiltonian for a particle of charge q :

$$\mathcal{H} = [(\mathbf{p} - q\mathbf{A}/c)^2 c^2 + m^2 c^4]^{1/2} + q\phi \quad (20.2.1)$$

which leads us to

$$i\hbar \frac{\partial \psi}{\partial t} = [c\alpha \cdot (\mathbf{P} - q\mathbf{A}/c) + \beta mc^2 + q\phi]\psi \quad (20.2.2)$$

The Electron Spin and Magnetic Moment

To see just these two features emerge, we can set $\phi = 0$ and work to order $(v/c)^2$. If we look for energy eigenstates

$$\psi(t) = \psi e^{-iEt/\hbar}$$

of Eq. (20.2.2), we get

$$E\psi = (c\alpha \cdot \pi + \beta mc^2)\psi \quad (20.2.3)$$

where

$$\pi = \mathbf{P} - q\mathbf{A}/c \quad (20.2.4)$$

is the kinetic ($m\mathbf{v}$) momentum operator. We now write ψ as

$$\psi = \begin{bmatrix} \chi \\ \Phi \end{bmatrix} \quad (20.2.5)$$

where χ and Φ are two-component spinors. Equation (20.2.3), with α and β explicitly written, becomes

$$\begin{bmatrix} E - mc^2 & -c\sigma \cdot \pi \\ -c\sigma \cdot \pi & E + mc^2 \end{bmatrix} \begin{bmatrix} \chi \\ \Phi \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (20.2.6)$$

which means

$$(E - mc^2)\chi - c\sigma \cdot \pi\Phi = 0 \quad (20.2.7)$$

and

$$(E + mc^2)\Phi - c\sigma \cdot \pi\chi = 0 \quad (20.2.8)$$

The second equation tells us that

$$\Phi = \left(\frac{c\sigma \cdot \pi}{E + mc^2} \right) \chi \quad (20.2.9)$$

Let us examine the term in brackets at low velocities. The denominator is

$$E + mc^2 = E_S + 2mc^2 \quad (20.2.10)$$

where $E_S = E - mc^2$ is the energy that appears in Schrödinger's equation. At low velocities, since $E_S \ll mc^2$ ‡

$$E + mc^2 \cong 2mc^2 \quad (20.2.11)$$

The numerator is of the order mvc , where mv is the typical momentum of the state. So

$$\left| \frac{\Phi}{\chi} \right| \cong \frac{1}{2} \left(\frac{v}{c} \right) \ll 1 \quad (20.2.12)$$

For this reason χ and Φ are called the *large and small components*, respectively. The terminology is of course appropriate only in the nonrelativistic domain. In this domain

$$\Phi \cong \frac{\sigma \cdot \pi}{2mc} \chi \quad (20.2.13)$$

and Eq. (20.2.7) becomes

$$E_S \chi = c \sigma \cdot \pi \Phi = \frac{(\sigma \cdot \pi)(\sigma \cdot \pi)}{2m} \chi \quad (20.2.14)$$

This is called the *Pauli equation*.§ If we use the identity

$$\sigma \cdot A \sigma \cdot B = A \cdot B + i \sigma \cdot A \times B \quad (20.2.15)$$

and

$$\pi \times \pi = \frac{iq\hbar}{c} B \quad (20.2.16)$$

we get

$$\left[\frac{(\mathbf{P} - q\mathbf{A}/c)^2}{2m} - \frac{q\hbar}{2mc} \sigma \cdot \mathbf{B} \right] \chi = E_S \chi \quad (20.2.17)$$

It is evident that this equation describes a spin- $\frac{1}{2}$ particle with $g = 2$. It is therefore appropriate to electrons. {Although $g = 2$ emerges so naturally from Dirac theory,

‡ $E_S = T + V = \frac{\pi^2}{2m} + V \underset{\text{(virial theorem)}}{\cong} O\left(\frac{\pi^2}{m}\right) = mv^2, \quad \frac{E_S}{mc^2} \cong \left(\frac{v}{c}\right)^2 \ll 1$

§ Actually the Pauli equation is the time-dependent version, with $i\hbar\dot{\chi}$ on the left-hand side.

it is incorrect to say that we need relativity to get this result. If we write the free-particle Schrödinger equation as

$$\frac{(\boldsymbol{\sigma} \cdot \mathbf{P})^2}{2m} \chi = E_S \chi$$

[since $(\boldsymbol{\sigma} \cdot \mathbf{P})^2 = P^2$] and then couple the vector potential \mathbf{A} as prescribed by *nonrelativistic* mechanics ($\mathbf{P} \rightarrow \mathbf{P} - q\mathbf{A}/c$), we get $g = 2$. Of course spin is introduced artificially here, but $g = 2$ is not.]

*Exercise 20.2.1.** Derive Eq. (20.2.16).

*Exercise 20.2.2.** Solve for the exact levels of the Dirac particle in a uniform magnetic field $\mathbf{B} = B_0 \mathbf{k}$. Assume $\mathbf{A} = (B_0/2)(-y\mathbf{i} + x\mathbf{j})$. Consult Exercise 12.3.8. (Write the equation for χ .)

Hydrogen Fine Structure

We now apply the Dirac equation to the case

$$V = e\phi = -e^2/r \quad (20.2.18)$$

that is to say, the electron in the hydrogen atom. (The proton is assumed to be fixed, i.e., infinitely massive.) The small and big components obey the following coupled equations:

$$(E - V - mc^2)\chi - c\boldsymbol{\sigma} \cdot \mathbf{P}\Phi = 0 \quad (20.2.19)$$

$$(E - V + mc^2)\Phi - c\boldsymbol{\sigma} \cdot \mathbf{P}\chi = 0 \quad (20.2.20)$$

The second one tells us that

$$\Phi = (E - V + mc^2)^{-1} c\boldsymbol{\sigma} \cdot \mathbf{P}\chi \quad (20.2.21)$$

(Since \mathbf{P} can differentiate V , the order of the factors is important.) If we feed this into the first, we get

$$(E - V - mc^2)\chi = c\boldsymbol{\sigma} \cdot \mathbf{P} \left[\frac{1}{E - V + mc^2} \right] c\boldsymbol{\sigma} \cdot \mathbf{P}\chi \quad (20.2.22)$$

If we approximate $E - V + mc^2$ on the right-hand side as $2mc^2$, we get

$$\begin{aligned} E_S \chi &= \left[\frac{(\boldsymbol{\sigma} \cdot \mathbf{P})^2}{2m} + V \right] \chi \\ &= \left[\frac{P^2}{2m} + V \right] \chi \end{aligned} \quad (20.2.23)$$

This is just the nonrelativistic Schrödinger equation we solved in Chapter 13. Notice that the Hamiltonian is order $(v/c)^2$ since it is quadratic in the momentum. To see the fine structure, we must go to order $(v/c)^4$. We do this by expanding $(E - V + mc^2)^{-1}$ on the right-hand side to one more order in v^2/c^2 :

$$\begin{aligned} \frac{1}{E - V + mc^2} &= \frac{1}{2mc^2 + E_S - V} = \frac{1}{2mc^2} \left(1 + \frac{E_S - V}{2mc^2} \right)^{-1} \\ &\cong \frac{1}{2mc^2} \left(1 - \frac{E_S - V}{2mc^2} \right) = \frac{1}{2mc^2} - \frac{E_S - V}{4m^2c^4} \end{aligned} \quad (20.2.24)$$

Equation (20.2.22) now becomes

$$E_S \chi = \left[\frac{P^2}{2m} + V - \frac{\boldsymbol{\sigma} \cdot \mathbf{P}(E_S - V)\boldsymbol{\sigma} \cdot \mathbf{P}}{4m^2c^2} \right] \chi \quad (20.2.25)$$

We cannot view this as the time-independent Schrödinger equation (i.e., as $E_S \chi = H\chi$) since E_S appears on *both sides*. By now even our spinal column knows how to respond to such a crisis. The right-hand side is a power series in v^2/c^2 . The first two terms are of the order v^2/c^2 , and the third is expected to be of order v^4/c^4 . Now the two $\boldsymbol{\sigma} \cdot \mathbf{P}$ factors in the third term use up a factor v^2/c^2 . So we need $E_S - V$ only to order v^2/c^2 . This we get from the same equation truncated to this order:

$$(E_S - V)\chi = \frac{P^2}{2m}\chi \quad (20.2.26)$$

We cannot use this result directly in Eq. (20.2.25) since $E_S - V$ there does not act on χ directly; there is $\boldsymbol{\sigma} \cdot \mathbf{P}$ in the way. So we do the following:

$$\begin{aligned} (E_S - V)\boldsymbol{\sigma} \cdot \mathbf{P}\chi &= \boldsymbol{\sigma} \cdot \mathbf{P}(E_S - V)\chi + \boldsymbol{\sigma} \cdot [E_S - V, \mathbf{P}]\chi \\ &= (\boldsymbol{\sigma} \cdot \mathbf{P}) \frac{P^2}{2m}\chi + \boldsymbol{\sigma} \cdot [\mathbf{P}, V]\chi \end{aligned} \quad (20.2.27)$$

Feeding this into Eq. (20.2.25) we get

$$\begin{aligned} E_S \chi &= \left\{ \frac{P^2}{2m} + V - \frac{P^4}{8m^3c^2} - \frac{(\boldsymbol{\sigma} \cdot \mathbf{P})(\boldsymbol{\sigma} \cdot [\mathbf{P}, V])}{4m^2c^2} \right\} \chi \\ &= \left\{ \frac{P^2}{2m} + V - \frac{P^4}{8m^3c^2} - \frac{i\boldsymbol{\sigma} \cdot \mathbf{P} \times [\mathbf{P}, V]}{4m^2c^2} - \frac{\mathbf{P} \cdot [\mathbf{P}, V]}{4m^2c^2} \right\} \chi \\ &= H\chi \end{aligned} \quad (20.2.28)$$

using once again the identity (20.2.15). We recognize the third term to be just the relativistic correction to the kinetic energy. It is just [recall Eq. (17.3.6)]

$$H_T = -\frac{P^4}{8m^3c^2} \quad (20.2.29)$$

The fourth term is the spin-orbit interaction, $H_{s.o.}$ of Eq. (17.3.16):

$$\begin{aligned} & \frac{-i\sigma \cdot \mathbf{P} \times [\mathbf{P}, V]}{4m^2c^2} \\ &= \frac{-i\sigma \cdot \mathbf{P} \times [-i\hbar\nabla(-e^2/r)]}{4m^2c^2} \quad \left\{ \text{using } [P, f(x)] = -i\hbar \frac{df}{dx} \right\} \\ &= \frac{-\hbar e^2 \sigma \cdot \mathbf{P} \times \mathbf{r}}{4m^2c^2 r^3} = \frac{\hbar e^2}{4m^2c^2 r^3} \sigma \cdot \mathbf{r} \times \mathbf{P}_+^\dagger \\ &= \frac{e^2}{2m^2c^2 r^3} \mathbf{S} \cdot \mathbf{L} = H_{s.o.} \end{aligned} \quad (20.2.30)$$

Notice that the Thomas factor is built in.

Consider now the fifth and last term. It upsets the whole interpretation because it is not Hermitian (check this). So if the quantity in brackets in Eq. (20.2.28) is used as a Hamiltonian we will find

$$\int |\chi|^2 d^3\mathbf{r} \neq \text{const in time}$$

But this is not surprising, since the conservation law that comes from the Dirac equation is

$$\int \psi^\dagger \psi d^3\mathbf{r} = \int [|\chi|^2 + |\Phi|^2] d^3\mathbf{r} = \text{const} \quad (20.2.31)$$

It follows that χ is not a good candidate for the Schrödinger wave function to this order in v/c . [It was all right when we worked to order $(v/c)^2$.] We find the right one as follows. Note that

$$\Phi = \frac{c\sigma \cdot \mathbf{P}}{E - V + mc^2} \chi = \frac{c\sigma \cdot \mathbf{P}}{2mc^2 + E_S - V} \chi \cong \frac{\sigma \cdot \mathbf{P}}{2mc} \chi \quad (20.2.32)$$

[‡] Although \mathbf{P} is a differential operator, $\mathbf{P} \times \mathbf{r} = -\mathbf{r} \times \mathbf{P}$, just as if \mathbf{P} and \mathbf{r} were c numbers, because the cross product never involves the product of a given coordinate and its conjugate momentum. This point was made earlier in the book when it was stated that there was no ordering ambiguity in passing from $\mathbf{l} = \mathbf{r} \times \mathbf{p}$ to \mathbf{L} .

(The neglected terms makes corrections of order v^6/c^6 in the end.) Consequently

$$|\Phi|^2 = \frac{\chi^\dagger (\boldsymbol{\sigma} \cdot \mathbf{P})(\boldsymbol{\sigma} \cdot \mathbf{P})\chi}{(2mc)^2} = \chi^\dagger \frac{P^2}{4m^2c^2} \chi$$

and so, from Eq. (20.2.31),

$$\begin{aligned} \int \chi^\dagger \left(1 + \frac{P^2}{4m^2c^2} \right) \chi d^3r &= \int \left[\left(1 + \frac{P^2}{8m^2c^2} \right) \chi \right]^\dagger \cdot \left(1 + \frac{P^2}{8m^2c^2} \right) \chi d^3r \\ &= \text{const} \end{aligned} \quad (20.2.33)$$

using $(1+x) = (1+x/2)(1+x/2) + O(x^2)$ and the Hermiticity of P^2 . Consequently, the candidate for the Schrödinger wave function is

$$\chi_S = \left(1 + \frac{P^2}{8m^2c^2} \right) \chi \quad (20.2.34)$$

for it will have a time-independent norm. (To the present accuracy, that is. If we go to higher and higher orders in v^2/c^2 , Φ will creep in more and more.)

The equation for χ_S is obtained by eliminating χ in Eq. (20.2.28):

$$\begin{aligned} E_S \left(1 + \frac{P^2}{8m^2c^2} \right)^{-1} \chi_S &= H \left(1 + \frac{P^2}{8m^2c^2} \right)^{-1} \chi_S \\ E_S \chi_S &= \left(1 + \frac{P^2}{8m^2c^2} \right) H \left(1 - \frac{P^2}{8m^2c^2} \right) \chi_S \\ &= \left(H + \left[\frac{P^2}{8m^2c^2}, H \right] \right) \chi_S \quad (\text{to this order in } v/c) \\ &= H_S \chi_S \end{aligned} \quad (20.2.35)$$

In evaluating the commutator, we need consider just the v^2/c^2 part of H , since $P^2/8m^2c^2$ is $O(v^2/c^2)$ and we are working to order v^4/c^4 . So

$$H_S = H + \left[\frac{\mathbf{P} \cdot \mathbf{P}}{8m^2c^2}, V \right]$$

is the desired Schrödinger Hamiltonian. The extra piece the above analysis yields combines with the non-Hermitian piece in Eq. (20.2.28) to form the *Darwin*

term H_D :

$$\begin{aligned}
 H_D &= \frac{1}{8m^2c^2} (-2\mathbf{P}\cdot[\mathbf{P}, V] + [\mathbf{P}\cdot\mathbf{P}, V]) \\
 &= \frac{-1}{8m^2c^2} [\mathbf{P}\cdot[\mathbf{P}, V]] \quad (\text{using the chain rule for commutators of products}) \\
 &= \frac{\hbar^2}{8m^2c^2} \nabla^2 V \quad \{\text{using } [P, f(x)] = -i\hbar df/dx \text{ twice}\} \\
 &= \frac{e^2\hbar^2\pi}{2m^2c^2} \delta^3(\mathbf{r})
 \end{aligned} \tag{20.2.37}$$

Thus the Darwin term affects only the s states.[‡] In the ground state, for example,

$$\langle 100 | H_D | 100 \rangle = \frac{e^2\hbar^2\pi}{2m^2c^2} \frac{1}{\pi a_0^3} = \frac{1}{2} mc^2 \alpha^4$$

and in general

$$\langle n00 | H_D | n00 \rangle = \frac{1}{2} \frac{mc^2 \alpha^4}{n^3} \tag{20.2.38}$$

Recall that in our previous treatment of fine structure we obtained a spin-orbit shift valid only for $l \neq 0$ and then applied it to $l=0$ as well, without any real justification. The result we got for $l=0$ is just what H_D generated above, which was our reason for doing what we did then. Thus $H_{\text{s.o.}}$ (relevant for $l \neq 0$) and H_D (relevant only for $l=0$) *together* conspire to produce a fine-structure shift that is smooth in l . The physics behind the Darwin term has nothing to do with spin-orbit coupling (for there is no such thing for $l=0$). Rather, it reflects the fact that in a relativistic theory, the particle cannot be localized to better than its Compton wavelength \hbar/mc . Thus the potential that is relevant is not $V(\mathbf{r})$ but some smeared average around the point \mathbf{r} :

$$\begin{aligned}
 \overline{V(\mathbf{r})} &= V(\mathbf{r}) + \sum_i \overline{\frac{\partial V}{\partial r_i}} \delta r_i + \frac{1}{2!} \sum_i \sum_j \overline{\frac{\partial^2 V}{\partial r_i \partial r_j}} \delta r_i \delta r_j + O(\delta r^3) \\
 &= V(\mathbf{r}) + \frac{1}{6} (\delta r)^2 \nabla^2 V + O(\delta r^3)
 \end{aligned} \tag{20.2.39}$$

where, in the averaging, we have assumed that fluctuations in the various directions are uncorrelated and spherically symmetric. If we now feed in $\delta r \simeq \hbar/mc$, we get the right sign and almost the right magnitude for the Darwin term [see Eq. (20.2.37)].

[‡] Recall that only in these states is ψ nonzero at the origin.

Although we chose to work to order v^4/c^4 , the Dirac equation can be solved exactly in the Coulomb case, $V = -e^2/r$. The resulting energy spectrum is

$$E_{nj} = mc^2 \left[1 + \left(\frac{\alpha}{n - (j + \frac{1}{2}) + [(j + \frac{1}{2})^2 - \alpha^2]^{1/2}} \right)^2 \right]^{-1/2} \quad (20.2.40)$$

If we expand this in powers of α , we get the rest energy, the Schrödinger energy, the fine-structure energy, and so on. Notice that the states of a given n and j are degenerate to all orders in α .

Whereas the above formula is in fantastic agreement with experiment,[‡] it is not the last word. For example, very precise measurements show that the $2S_{1/2}$ level is above the $2P_{1/2}$ level. This phenomenon, called the *Lamb shift*, can be understood only if the electromagnetic field is treated quantum mechanically.

20.3. More on Relativistic Quantum Mechanics

With the principal goal of this chapter achieved in the last section, we direct our attention to certain phenomena that come out of Dirac theory but were not apparent in the last few pages. Let us first note that the union of relativity and quantum mechanics produces the following problem: relativity allows particle production given enough energy, and quantum mechanics allows arbitrarily large energy violations over short times. Consequently the degrees of freedom of a relativistic system are neither fixed nor finite; a system that initially has one particle can evolve into a state with 15 of them. Why doesn't this problem appear in the Dirac theory, which seems like a single-particle theory? The answer is that it does appear, but in the guise of *negative-energy solutions*. Let us see what these are and how they lead to proliferation of the degrees of freedom.

Consider the free-particle Dirac equation (with $\hbar = c = 1$)

$$i \frac{\partial \psi}{\partial t} = (\mathbf{a} \cdot \mathbf{P} + \beta m) \psi \quad (20.3.1)$$

Let us look for plane wave solutions

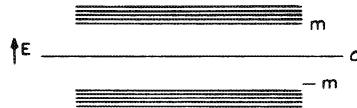
$$\psi = w(\mathbf{p}) e^{i(\mathbf{p} \cdot \mathbf{r} - Et)} \quad (20.3.2)$$

where $w(\mathbf{p})$ is a spinor that has no space-time dependence. It satisfies

$$Ew = (\mathbf{a} \cdot \mathbf{p} + \beta m)w \quad (20.3.3)$$

[‡] After hyperfine interactions are taken into account.

Figure 20.1. In the Dirac theory there are two continuous bands of energy available to the free particle; one goes from $+m$ up to ∞ and the other goes from $-m$ down to $-\infty$.



or in terms of χ and Φ ,

$$\begin{bmatrix} E-m & -\sigma \cdot \mathbf{p} \\ -\sigma \cdot \mathbf{p} & E+m \end{bmatrix} \begin{bmatrix} \chi \\ \Phi \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (20.3.4)$$

If $\mathbf{p}=0$, χ and Φ decouple. The equation for χ is

$$(E-m)\chi=0 \rightarrow E=m \quad (20.3.5)$$

which is fine. It says a particle at rest has energy $E=m$ and is described by an arbitrary two-component spinor which we identify as the spin degree of freedom.

The equation for Φ is

$$(E+m)\Phi=0 \rightarrow E=-m \quad (20.3.6)$$

Now even a layperson will tell you that E is supposed to be mc^2 not $-mc^2$. The significance of the four components of ψ are evident in the rest frame: there are two possible spin orientations and two signs of the energy. The problem persists for $\mathbf{p} \neq 0$ as well. Here we find

$$\chi = \frac{\sigma \cdot \mathbf{p}}{E-m} \Phi \quad (20.3.37)$$

$$\Phi = \frac{\sigma \cdot \mathbf{p}}{E+m} \chi \quad (20.3.8)$$

These are consistent only if

$$\frac{p^2}{E^2 - m^2} = 1$$

or

$$E^2 = p^2 + m^2$$

or

$$E = \pm (p^2 + m^2)^{1/2} \quad (20.3.9)$$

The energy levels corresponding to these two options are shown in Fig. 20.1. What do we do with the negative-energy solutions? If there are no interactions, positive-energy electrons will stay where they are and we can postulate that there

are no negative-energy electrons. But there are always some perturbations acting on all electrons and these can induce all positive-energy electrons to cascade down to the negative-energy states. How do we understand the stability of positive-energy electrons?

There are two ways out, one due to Dirac and one due to Feynman.[‡] Dirac postulated that the negative-energy states are all occupied—that what we call the vacuum is really the occupied (but *unobservable*) sea of negative-energy electrons. If we accept this, the stability problem is solved by the exclusion principle, which prevents the positive-energy electrons from decaying to the occupied negative-energy states. This picture has some profound consequences. Suppose we give a negative-energy electron enough energy (at least $2m$) for it to come to a positive-energy state. Now we have a positive-energy, charge $-e$ object. But we also have created a hole in the “Dirac sea.” Since the filled Dirac sea was postulated to be unobservable, the hole is observable; it represents an increase in charge by $+e$ (the disappearance of $-e$ = appearance of $+e$), and an increase in energy by $|E|$, if $-|E|$ was the energy of the electron ejected from the sea.[§] Thus the hole, which has charge $+e$ and *positive* energy, is created along with the electron. It is called a *positron*. Its mass can easily be shown to be m . Positrons were observed a few years after Dirac’s theory of holes was published.

When an electron meets a positron, i.e., a hole in the sea, it jumps in and we lose both particles, though some energy (at least equal to $2m$) will be liberated in the form of photons. (Hereafter we will occasionally refer to these particles as e^- , e^+ , and γ , respectively.)

The trouble with Dirac’s solution is that it doesn’t apply to spinless particles, which don’t obey the Pauli principle but which do have the same problem of negative-energy solutions, as one can see by plugging a plane wave solution into the Klein-Gordon equation. (In fact this was the reason the Klein-Gordon equation was rejected in the early days and Dirac sought a first-order equation.) So let us turn to Feynman’s resolution, which applies to bosons *and* fermions.

Feynman’s idea is the following: *negative-energy particles can only travel backward in time*. Let us see first how this resolves the problem and then how the statement is actually implemented in quantum theory. Consider a negative-energy particle that is created at the space-time point c and travels backward to d , where it is destroyed (Fig. 20.2a). To us, who move forward in time and see space-time in equal-time slices, this is what will seem to be happening:

- (1) $t < t_d$ Nothing anywhere.
- (2) $t = t_d$ Negative energy $-|E|$ and charge $-e$ are destroyed, i.e., world energy goes up by $|E|$ and charge goes up by e relative to the past. A positron is born.
- (3) $t = t_c$ Negative energy is created, charge $-e$ is created. This wipes out the positron.
- (4) $t > t_c$ Nothing anywhere.

[‡] In its basic form, the idea exploited by Feynman was pointed out by Stueckelberg.

[§] Recall the story of the fellow who got so used to the midnight express going past his house that one day when it failed to show up, we woke up screaming “What’s that noise?”

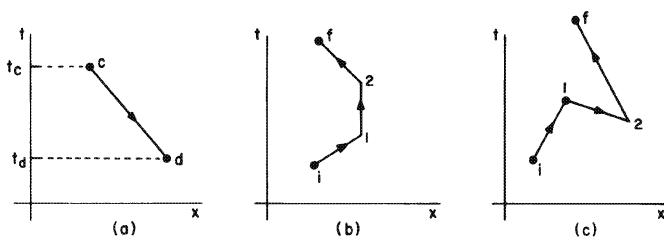


Figure 20.2. (a) A negative-energy particle is created at c , travels back in time to d , where it is destroyed. To us, who move forward in time, it will seem as though an antiparticle of positive energy is created at d and destroyed at c . (b) A normal second order scattering process. (c) A second-order process that involves back-scattering in time. Between times 2 and 1 we will see a particle-antiparticle pair in addition to the original particle.

Thus the process makes perfect sense and represents a positron *created at d and destroyed at c*.

How does Feynman ensure that negative-energy states propagate backward? Here is a sketchy description. Recall that the Schrödinger propagator we have used so far is (in the coordinate basis)

$$U_S(\mathbf{r}, t; \mathbf{r}', t') = \sum_n \psi_n(\mathbf{r}) \psi_n^*(\mathbf{r}') e^{-iE_n(t-t')} \quad (20.3.10)$$

where ψ_n is an energy eigenfunction labeled by a generic quantum number n .

Since every term in the sum satisfies the Schrödinger equation, it is clear that

$$\left(i \frac{\partial}{\partial t} - H \right) U_S = 0 \quad (20.3.11)$$

given this U_S and $\psi(t')$ at some initial time, we can get $\psi(t)$ at a later time ($t > t'$):

$$\psi(t) = U_S \psi(t') \quad (\text{schematic}) \quad (20.3.12)$$

Now note that although we use U_S to propagate ψ forward in time, it can also propagate it backward, since $U_S \neq 0$ for $t < t'$. To avoid this possibility explicitly, let us work with

$$G_S(\mathbf{r}t, \mathbf{r}'t') = \theta(t - t') U_S(\mathbf{r}t, \mathbf{r}'t') \quad (20.3.13)$$

which simply cannot propagate ψ backward. The equation satisfied by G_S is

$$\begin{aligned} \left(i \frac{\partial}{\partial t} - H \right) G_S &= \left[i \frac{\partial}{\partial t} \theta(t - t') \right] \sum_n \psi_n(\mathbf{r}) \psi_n^*(\mathbf{r}') e^{-iE_n(t-t')} \\ &= i\delta(t - t') \delta^3(\mathbf{r} - \mathbf{r}') \\ &= i\delta^4(x - x') \quad [x = (t, \mathbf{r})] \end{aligned} \quad (20.3.14)$$

[We have used the completeness of the eigenfunctions, and $\dot{\theta}(t-t') = \delta(t-t')$.]

The propagator in Dirac theory, G_D , obeys a similar equation. Consider the *free-particle case*. Here

$$\left(i \frac{\partial}{\partial t} - H^0 \right) G_D^0 = i\delta^4(x-x') \quad (20.3.15)$$

with H^0 the free-particle Dirac Hamiltonian. The solution is

$$G_D^0(x, x') = \theta(t-t') \left(\sum_{n+} + \sum_{n-} \right) \quad (20.3.16)$$

where $\sum_{n\pm}$ denote sums over *positive- and negative-energy eigenfunctions*, respectively. [If we throw away \sum_{n-} we lose completeness and won't get $i\delta^4$ on the right-hand side of Eq. (20.3.15).] Although G_D^0 satisfies the requisite equation, it has the negative-energy solutions propagating forward in time. Now here is the trick. G_D^0 is not a unique solution to Eq. (20.3.15); we can add or subtract any solution to the free-Dirac equation, provided we subtract it for all times. (If we subtract it only for $t > 0$, say, we are subtracting a θ function times the solution, which doesn't obey the homogeneous equation.) Let us subtract all negative-energy solutions for all times. This gives us *Feynman's propagator*

$$G_F^0(x, x') = \theta(t-t') \sum_{n+} - \theta(t'-t) \sum_{n-} \quad (20.3.17)$$

Consider now some initial state $\psi_i(t')$ which is composed of just positive-energy solutions. G_F^0 will propagate it forward in time, since $\psi_i(t')$ is orthogonal to every term in \sum_{n-} . Thus $G_F^0 \psi_i(t') = \psi_f(t)$ contains only positive-energy components and keeps moving forward. On the other hand if $\psi_i(t')$ is built out of negative-energy components only, it is orthogonal to every term in \sum_{n+} and gets propagated backwards from t' to t . We will see it as a positron propagating from t to t' .

Consider now the electron in some external potential V . The exact propagation of the electron can be described by a perturbation series based on G_F^0 ; and in schematic form,

$$\psi_f(t) = G_F^0(t, t') \psi_i(t') + \sum_{t''} G_F^0(t, t'') V(t'') G_F^0(t'', t') \psi_i(t') + \dots$$

We can represent these multiple scattering events by diagrams very much like the ones in Section 18.3. There is just one difference. Consider a second-order process. There is of course the usual double scattering in which the electron just gets scattered forward in time (Fig. 20.2b). But now there is also the possibility that the potential scatters it backward in time at 1 and then forward at 2 (Fig. 20.2c). As we move forward in time, we first see the electron, then an e^+e^- pair created at 2, then the annihilation of the e^+ with the original e^- at 1 and finally the arrival of the created e^- at f . Since the electron can wiggle and jiggle any number of times (as we go to higher orders in the expansion) the intermediate stages can contain any number of

e^+e^- pairs. This is how the degrees of freedom proliferate in a relativistic theory. Even though we started with a one-particle equation, particle production creeps in through the negative-energy solutions—either because the latter imply an infinite sea of sleeping particles which can be awakened or because they allow a single electron to go back and forth in time, thereby becoming many particles at a given time. Although particle production (at least pair production) can be handled in the present formulation, it is time to learn quantum field theory, which provides a natural framework for handling the creation and destruction of particles. We have already seen one example, namely, the quantized electromagnetic field, whose quanta, the photons, can be created and destroyed by operators a^\dagger and a . We need a theory in which particles like electrons and positrons can also be created and destroyed. You are ready for that subject.[‡]

[‡] See for example J. D. Bjorken and S. D. Drell, *Relativistic Quantum Mechanics and Relativistic Quantum Fields*, McGraw-Hill, New York (1964), or C. Itzykson and J. B. Zuber, *Quantum Field Theory*, McGraw-Hill, New York (1980).

Path Integrals: Part II

In this chapter we return to path integrals for a more detailed and advanced treatment. The tools described here are so widely used in so many branches of physics, that it makes sense to include them in a book such as this. This chapter will be different from the earlier ones in that it will try to introduce you to a variety of new topics without giving all the derivations in the same detail as before. It also has a list of references to help you pursue any topic that attracts you. The list is not exhaustive and consists mostly of pedagogical reviews or books. From the references these references contain, you can pursue any given topic in greater depth. All this will facilitate the transition from course work to research.

In Chapter 8 the path integral formula for the propagator was simply postulated and shown to lead to the same results as the operator methods either by direct evaluation of the propagator (in the free particle case) or by showing once and for all that the Schrödinger equation followed from the path integral prescription for computing the time evolution.

We begin this chapter by doing the reverse: we start with the operator Hamiltonian $H = P^2/2m + V$ and derive the propagator for it as a path integral. We shall see that there are many types of path integrals one can derive. We will discuss

- The configuration space path integral, discussed in Chapter 8.
- The phase space path integral.
- The coherent state path integral.

You will see that the existence of many path integrals is tied to the existence of many resolutions of the identity, i.e., to the existence of many bases.

Following this we will discuss two applications: to the Quantum Hall Effect (QHE) and a recent development called the Berry Phase.

We then turn to imaginary time quantum mechanics and its relation to statistical mechanics (classical and quantum) as well the calculation of tunneling amplitudes by a semiclassical approximation. You will learn about instantons, the transfer matrix formulation, and so on.

Finally, we discuss path integrals for two problems with no classical limit: a spin Hamiltonian and a fermionic oscillator.

21.1. Derivation of the Path Integral

Let us assume that the Hamiltonian is time-independent and has the form

$$H = \frac{P^2}{2m} + V(X) \quad (21.1.1)$$

The propagator is defined by

$$U(xt; x'0) \equiv U(x, x', t) = \langle x | \exp\left(-\frac{i}{\hbar} H t\right) | x' \rangle \quad (21.1.2)$$

It was stated in Chapter 8 that U may be written as a sum over paths going from $(x'0)$ to (xt) . We will now see how this comes about.

First, it is evident that we may write

$$\exp\left(-\frac{i}{\hbar} H t\right) = \left[\exp\left(-\frac{i}{\hbar} H \frac{t}{N}\right) \right]^N \quad (21.1.3)$$

for any N . This merely states that $U(t)$, the propagator for a time t , is the product of N propagators $U(t/N)$. Let us define

$$\varepsilon = \frac{t}{N} \quad (21.1.4)$$

and consider the limit $N \rightarrow \infty$. Now we can write

$$\exp\left(-\frac{i\varepsilon}{\hbar} (P^2/2m + V(X))\right) \simeq \exp\left(-\frac{i\varepsilon}{2m\hbar} P^2\right) \cdot \exp\left(-\frac{i\varepsilon}{\hbar} V(X)\right) \quad (21.1.5)$$

because of the fact that

$$e^A e^B = e^{A+B+1/2[A,B]+\dots} \quad (21.1.6)$$

which allows us to drop the commutator shown (and other higher-order nested commutators not shown) on the grounds that they are proportional to higher powers of ε which is going to 0. While all this is fine if A and B are finite dimensional matrices with finite matrix elements, it is clearly more delicate for operators in Hilbert space which could have large or even singular matrix elements. We will simply assume that in the limit $\varepsilon \rightarrow 0$ the \simeq sign in Eq. (21.1.5) will become the equality sign for the purpose of computing any reasonable physical quantity.

So we have to compute

$$\langle x| \underbrace{\exp\left(-\frac{i\varepsilon}{2m\hbar} P^2\right) \cdot \exp\left(-\frac{i\varepsilon}{\hbar} V(X)\right)}_{N \text{ times}} \cdot \exp\left(-\frac{i\varepsilon}{2m\hbar} P^2\right) \cdot \exp\left(-\frac{i\varepsilon}{\hbar} V(X)\right) \dots |x'\rangle \quad (21.1.7)$$

The next step is to introduce the resolution of the identity:

$$I = \int_{-\infty}^{\infty} dx |x\rangle \langle x| \quad (21.1.8)$$

between every two adjacent factors of $U(t/N)$. Let us illustrate the outcome by considering $N=3$. We find (upon renaming x, x' as x_3, x_0 for reasons that will be clear soon)

$$\begin{aligned} U(x_3, x_0, t) &= \int \prod_{n=1}^2 dx_n \langle x_3 | \exp\left(-\frac{i\varepsilon}{2m\hbar} P^2\right) \exp\left(-\frac{i\varepsilon}{\hbar} V(X)\right) |x_2\rangle \\ &\quad \times \langle x_2 | \exp\left(-\frac{i\varepsilon}{2m\hbar} P^2\right) \exp\left(-\frac{i\varepsilon}{\hbar} V(X)\right) |x_1\rangle \\ &\quad \times \langle x_1 | \exp\left(-\frac{i\varepsilon}{2m\hbar} P^2\right) \exp\left(-\frac{i\varepsilon}{\hbar} V(X)\right) |x_0\rangle \end{aligned} \quad (21.1.9)$$

Consider now the evaluation of the matrix element

$$\langle x_n | \exp\left(-\frac{i\varepsilon}{2m\hbar} P^2\right) \cdot \exp\left(-\frac{i\varepsilon}{\hbar} V(X)\right) |x_{n-1}\rangle \quad (21.1.10)$$

When the rightmost exponential operates on the ket to its right, the operator X gets replaced by the eigenvalue x_{n-1} . Thus,

$$\begin{aligned} \langle x_n | \exp\left(-\frac{i\varepsilon}{2m\hbar} P^2\right) \cdot \exp\left(-\frac{i\varepsilon}{\hbar} V(X)\right) |x_{n-1}\rangle \\ = \langle x_n | \exp\left(-\frac{i\varepsilon}{2m\hbar} P^2\right) |x_{n-1}\rangle \exp\left(-\frac{i\varepsilon}{\hbar} V(x_{n-1})\right) \end{aligned} \quad (21.1.11)$$

Consider now the remaining matrix element. It is simply the free particle propagator from x_{n-1} to x_n in time ε . We know what it is [say from Eq. (5.1.10)] or the following exercise

$$\langle x_n | \exp\left(-\frac{i\varepsilon}{2m\hbar} P^2\right) |x_{n-1}\rangle = \left[\frac{m}{2\pi i \hbar \varepsilon} \right]^{1/2} \exp\left[\frac{im(x_n - x_{n-1})^2}{2\hbar \varepsilon} \right] \quad (21.1.12)$$

Exercise 21.1.1. Derive the above result independently of Eq. (5.1.10) by introducing a resolution of the identity in terms of momentum states between the exponential operator and the position eigenket in the left-hand side of Eq. (21.1.12). That is, use

$$I = \int_{-\infty}^{\infty} \frac{dp}{2\pi\hbar} |p\rangle\langle p| \quad (21.1.13)$$

where the plane wave states have a wave function given by

$$\langle x|p\rangle = e^{ipx/\hbar} \quad (21.1.14)$$

which explains the measure for the p integration.

Resuming our derivation, we now have

$$\begin{aligned} \langle x_n | \exp\left(-\frac{i\varepsilon}{2m\hbar} P^2\right) \cdot \exp\left(-\frac{i\varepsilon}{\hbar} V(X)\right) | x_{n-1} \rangle \\ = \left[\frac{m}{2\pi i\hbar\varepsilon} \right]^{1/2} \exp\left[\frac{im(x_n - x_{n-1})^2}{2\hbar\varepsilon} \right] \exp\left(-\frac{i\varepsilon}{\hbar} V(x_{n-1}) \right) \end{aligned} \quad (21.1.15)$$

Collecting all such factors (there are just three in this case with $N=3$), we can readily see that for general N

$$\begin{aligned} U(x_N, x_0, t) &= \left(\frac{m}{2\pi i\hbar\varepsilon} \right)^{1/2} \left[\prod_{n=1}^{N-1} \left(\frac{m}{2\pi i\hbar\varepsilon} \right)^{1/2} dx_n \right] \\ &\times \exp\left[\sum_{n=1}^N \frac{im(x_n - x_{n-1})^2}{2\hbar\varepsilon} - \frac{i\varepsilon}{\hbar} V(x_{n-1}) \right] \end{aligned} \quad (21.1.16)$$

If we drop the V terms we see that this is in exact agreement with the free particle path integral of Chapter 8. For example, the measure for integration has exactly N factors of B^{-1} as per Eq. (8.4.8), of which $N-1$ accompany the x -integrals. With the V term, the integrand is just the discretized version of $\exp(iS/\hbar)$:

$$\begin{aligned} &\exp\left[\sum_{n=1}^N \frac{im(x_n - x_{n-1})^2}{2\hbar\varepsilon} - \frac{i\varepsilon}{\hbar} V(x_{n-1}) \right] \\ &= \exp \frac{i}{\hbar} \varepsilon \sum_{n=1}^N \left[\frac{m(x_n - x_{n-1})^2}{2\varepsilon^2} - V(x_{n-1}) \right] \end{aligned} \quad (21.1.17)$$

We can go back to the continuum notation and write all this as follows:

$$U(x, x', t) = \int [\mathcal{D}x] \exp \left[\frac{i}{\hbar} \int_0^t \mathcal{L}(x, \dot{x}) dt \right] \quad (21.1.18)$$

where

$$\int [\mathcal{D}x] = \lim_{N \rightarrow \infty} \left(\frac{m}{2\pi i \hbar \varepsilon} \right)^{1/2} \int \left[\prod_{n=1}^{N-1} \left(\frac{m}{2\pi i \hbar \varepsilon} \right)^{1/2} dx_n \right] \quad (21.1.19)$$

The continuum notation is really a schematic for the discretized version that preceded it, and we need the latter to define what one means by the path integral. It is easy to make many mistakes if one forgets this. In particular, there is no reason to believe that replacing differences by derivatives is always legitimate. For example, in this problem, in a time ε , the variable being integrated over typically changes by $\mathcal{O}(\varepsilon^{1/2})$ and not $\mathcal{O}(\varepsilon)$, as explained in the discussion before Eq. (8.5.6). The works in the Bibliography at the end of this chapter discuss some of the subtleties. The continuum version is, however, very useful to bear in mind since it exposes some aspects of the theory that would not be so transparent otherwise. It is also very useful for getting the picture at the semiclassical level and for finding whatever connection there is between the macroscopic world of smooth paths and the quantum world. We will take up some examples later.

The path integral derived above is called the *Configuration Space* path integral or simply the path integral. We now consider another one. Let us go back to

$$\langle x_N | \underbrace{\exp \left(-\frac{i\varepsilon}{2m\hbar} P^2 \right) \cdot \exp \left(-\frac{i\varepsilon}{\hbar} V(X) \right) \cdot \exp \left(-\frac{i\varepsilon}{2m\hbar} P^2 \right) \cdot \exp \left(-\frac{i\varepsilon}{\hbar} V(X) \right) \dots}_{N \text{ times}} | x_0 \rangle \quad (21.1.20)$$

Let us now introduce resolutions of the identity between *every exponential* and the next. We need two versions

$$I = \int_{-\infty}^{\infty} dx |x\rangle \langle x| \quad (21.1.21)$$

$$I = \int_{-\infty}^{\infty} \frac{dp}{2\pi\hbar} |p\rangle \langle p| \quad (21.1.22)$$

where the plane wave states have a wave function given by

$$\langle x | p \rangle = e^{ipx/\hbar} \quad (21.1.23)$$

Let us first set $N=3$ and insert three resolutions of the identity in terms of p -states and two in terms of x -states with x and p resolutions alternating. This gives us

$$\begin{aligned} U(x_3, x_0, t) = & \int [\mathcal{D}p \mathcal{D}x] \langle x_3 | \exp\left(-\frac{i\varepsilon}{2m\hbar} P^2\right) | p_3 \rangle \\ & \times \langle p_3 | \exp\left(-\frac{i\varepsilon}{\hbar} V(X)\right) | x_2 \rangle \langle x_2 | \exp\left(-\frac{i\varepsilon}{2m\hbar} P^2\right) | p_2 \rangle \\ & \times \langle p_2 | \exp\left(-\frac{i\varepsilon}{\hbar} V(X)\right) | x_1 \rangle \langle x_1 | \exp\left(-\frac{i\varepsilon}{2m\hbar} P^2\right) | p_1 \rangle \\ & \times \langle p_1 | \exp\left(-\frac{i\varepsilon}{\hbar} V(X)\right) | x_0 \rangle \end{aligned} \quad (21.1.24)$$

where

$$\int [\mathcal{D}p \mathcal{D}x] = \underbrace{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{2N-1 \text{ times}} \prod_{n=1}^N \frac{dp_n}{2\pi\hbar} \prod_{n=1}^{N-1} dx_n \quad (21.1.25)$$

Evaluating all the matrix elements of the exponential operators is trivial since each operator can act on the eigenstate to its right and get replaced by the eigenvalue. Collecting all the factors (a strongly recommended exercise for you) we obtain

$$U(x, x', t) = \int [\mathcal{D}p \mathcal{D}x] \exp\left[\sum_{i=1}^N \left(\frac{-i\varepsilon}{2m\hbar} p_n^2 + \frac{i}{\hbar} p_n(x_n - x_{n-1}) - \frac{i\varepsilon}{\hbar} V(x_{n-1}) \right) \right] \quad (21.1.26)$$

This formula derived for $N=3$ is obviously true for any N . In the limit $N \rightarrow \infty$, i.e., $\varepsilon \rightarrow 0$, we write schematically in continuous time (upon multiplying and dividing the middle term by ε), the following continuum version:

$$U(x, x', t) = \int [\mathcal{D}p \mathcal{D}x] \exp\left[\frac{i}{\hbar} \int_0' [p \dot{x} - \mathcal{H}(x, p)] dt \right] \quad (21.1.27)$$

where $\mathcal{H} = p^2/2m + V(x)$ and $(x(t), p(t))$ are now written as functions of a continuous variable t . This is the *Phase Space Path Integral* for the propagator. The continuum version is very pretty [with the Lagrangian in the exponent, but expressed in terms of (x, p)] but is only a schematic for the discretized version preceding it.

In our problem, since p enters the Hamiltonian quadratically, it is possible to integrate out all the N variables p_n . Going back to the discretized form, we isolate

$$\begin{aligned} & \prod_1^N \int_{-\infty}^{\infty} \frac{dp_n}{2\pi\hbar} \exp \left[\left(\frac{-i\varepsilon}{2m\hbar} p_n^2 + \frac{i}{\hbar} p_n (x_n - x_{n-1}) \right) \right] \\ &= \prod_1^N \left(\frac{m}{2\pi i \hbar \varepsilon} \right)^{1/2} \exp \left[\frac{i m (x_n - x_{n-1})^2}{2\hbar \varepsilon} \right] \end{aligned} \quad (21.1.28)$$

If we now bring in the x -integrals we find that this gives us exactly the configuration space path integral, as it should.

Note that if p does not enter the Hamiltonian in a separable quadratic way, it will not be possible to integrate it out and get a path integral over just x , in that we do not know how to do non-Gaussian integrals. In that case we can only write down the phase space path integral.

We now turn to two applications that deal with the path integrals just discussed.

The Landau Levels

We now discuss a problem that is of great theoretical interest in the study of QHE (see Girvin and Prange). We now explore some aspects of it, not all having to do with functional integrals. Consider a particle of mass μ and charge q in the x - y plane with a uniform magnetic field B along the z -axis. This is a problem we discussed in Exercise (12.3.8). Using a vector potential

$$\mathbf{A} = \frac{B}{2} (-y\mathbf{i} + x\mathbf{j}) \quad (21.1.29)$$

we obtained a Hamiltonian

$$H = \frac{[P_x + qYB/2c]^2}{2\mu} + \frac{[P_y - qXB/2c]^2}{2\mu} \quad (21.1.30)$$

You were asked to verify that

$$Q = \frac{(cP_x + qYB/2)}{qB} \quad P = (P_y - qBX/2c) \quad (21.1.31)$$

were canonical variables with $[Q, P] = i\hbar$. It followed that H was given by the formula

$$H = \frac{P^2}{2\mu} + \frac{1}{2} \mu \omega_0^2 Q^2 \quad (21.1.32)$$

and had a harmonic oscillator spectrum with spacing $\hbar\omega_0$, where

$$\omega_0 = qB/\mu c \quad (21.1.33)$$

is the *cyclotron frequency*. In terms of

$$a = \left(\frac{\mu \omega_0}{2\hbar} \right)^{1/2} Q + i \left(\frac{1}{2\mu \omega_0 \hbar} \right)^{1/2} P \quad (21.1.34)$$

and its adjoint, we can write

$$H = [a^\dagger a + \frac{1}{2}] \hbar \omega_0 \quad (21.1.35)$$

We seem to have gone from a problem in two dimensions to a one-dimensional oscillator problem. How can that be? The point is that there is another canonical pair

$$P' = \frac{(cP_x - qYB/2)}{qB} \quad Q' = (P_y + qBX/2c) \quad (21.1.36)$$

which commutes with Q, P and does not enter H .

Exercise 21.1.2. If you do not recall the details of Exercise (12.3.8), provide all the missing steps in the derivation starting at Eq. (21.1.29) and ending with Eq. (21.1.35). Check the advertised commutation rules for (Q', P') .

The cyclic character of (Q', P') is reflected in the fact that the levels of the oscillator, called *Landau Levels*, are infinitely degenerate. To see this degeneracy consider the *Lowest Landau Level*, abbreviated LLL. The states in this level obey the equation

$$a|0\rangle = 0 \quad (21.1.37)$$

which becomes in the coordinate representation

$$\left[\frac{\partial}{\partial z^*} + \frac{qB}{4\hbar c} z \right] \psi_0(z, z^*) = 0 \quad (21.1.38)$$

wherein we have switched to complex coordinates

$$z = x + iy \quad z^* = x - iy \quad (21.1.39)$$

If we make the ansatz

$$\psi_0(z, z^*) = \exp \left[-\frac{qB}{4\hbar c} zz^* \right] u(z, z^*) \quad (21.1.40)$$

we find the beautiful result

$$\frac{\partial}{\partial z^*} u(z, z^*) = 0 \quad (21.1.41)$$

as the defining rule for the LLL. Thus u is any *analytic function*, i.e., function of the combination $x+iy$. The family of such functions is clearly infinitely large, with the monomials $[z^m | m=0, 1, 2, \dots]$ serving as a linearly independent basis. Thus the ground state function ψ_0 is not a unique function as in the case of the truly one dimensional oscillator but a superposition of functions of the form

$$\psi_{0,m} = z^m \exp\left[-\frac{qB}{4\hbar c} zz^*\right] \quad (21.1.42)$$

I now make the following assertions:

- For large m the probability density for the particle is concentrated at some radius $r_m = \sqrt{2mr_0}$ where $r_0 = \sqrt{c\hbar/qB}$ is called the *magnetic length*.
- If the system is not infinite in size, but is a disc of radius R , the biggest value of m that can fit in, and hence N , the number of LLL states that fit into the disc, is given by

$$N = \frac{\pi R^2 B}{\Phi_0} \quad (21.1.43)$$

where the numerator is the flux through the sample and the denominator is the flux quantum of Eq. (18.4.39):

$$\Phi_0 = \frac{2\pi\hbar c}{q} \quad (21.1.44)$$

*Exercise 21.1.3.** (Mandatory if you wish to follow the discussion of the QHE.) Derive the equation for the LLL in the coordinate representation by providing the missing steps in the derivation. Prove the above assertions. Note that N , the number of states in the LLL, is given by the flux through the sample in units of the flux quantum.

In the following discussion we will hold N , i.e., the field and sample dimensions, fixed.

In the study of the QHE one is interested in the problems of an electron gas designed to live in two dimensions. (Since the electron charge is $q=-e$, our formulas will hold with $q=e$ if the sign of the vector potential and field are reversed at the outset. Henceforth imagine this has been done and that q stands for the magnitude of the electron charge.) The electron spin is frozen along the applied field and has no interesting dynamics. In particular it is the burden of the orbital wave function to ensure antisymmetry. In a real-life problem one is also required to consider the interaction between the electrons as well as interaction between the electrons and any external scalar potential $V(x, y)$ due to the background medium. It is assumed that both these interactions have a scale much smaller than the gap $\hbar qB/\mu c$ between Landau levels. Thus at low temperatures, one would like a simplified problem with the Hilbert space restricted to the LLL. What does this problem look like?

The path integral can tell us that. We will work out the answer for the case where electron-electron interaction is zero. (Only then do the electrons propagate

independently and we can write out a functional integral for just one electron.) The action is

$$S = \int \left[\frac{\mu}{2} (\dot{x}^2 + \dot{y}^2) + \frac{qB}{2c} (-y\dot{x} + x\dot{y}) - V(x, y) \right] dt \quad (21.1.45)$$

where the terms linear in velocity represent the $(q/c)\mathbf{v} \cdot \mathbf{A}$ in the Lagrangian in the gauge we are using. To get the low-energy physics we must banish the higher Landau levels. *Since the gap to the higher levels is $\hbar qB/\mu c$ this is readily done by taking the limit $\mu \rightarrow 0$.* (In this limit the zero point energy of the oscillator, which gives the energy of the LLL, diverges. It is assumed this constant is subtracted out of the Hamiltonian.) This gives us the low-energy action

$$S_{LLL} = \int \left[\frac{qB}{c} x\dot{y} - V(x, y) \right] dt \quad (21.1.46)$$

where we have done an integration by parts to combine the two terms linear in velocity. (The surface term will not affect the equations of motion.)

Notice the interesting result that the action is that of a *phase space path integral with y and $\partial \mathcal{L}/\partial \dot{y} = (qB/c)x \equiv \bar{x}$ as canonically conjugate variables.* $V(y, \bar{x})$ now plays the role of the Hamiltonian for this problem. Since we have just one coordinate and one momentum, the problem of the LLL is essentially one-dimensional.

In the semiclassical picture, the orbits will obey Hamilton's equations:

$$\dot{y} = \frac{\partial V}{\partial \bar{x}} \quad \dot{\bar{x}} = -\frac{\partial V}{\partial y} \quad (21.1.47)$$

and one can try to do Bohr-Sommerfeld quantization. At the quantum level, V can become a complicated differential operator since \bar{x} will turn into the y -derivative. I leave the details and applications of the semiclassical picture to the references.

Now you might object that if we did not have the operator solution telling us that the levels of the problem go as μ^{-1} it might not occur to us to consider the limit $\mu \rightarrow 0$ in order to isolate the low energy physics. This is not so. We will simply argue that in the limit of low energies, i.e., low frequencies, terms in the action with higher time derivatives can be neglected compared to those with fewer ones. This would allow us to throw out the same kinetic energy term. (Now you can do this even in a problem without the magnetic field, but this would leave you with very little interesting dynamics. Here we have some linear derivatives left over, i.e., here the low-energy physics is the physics of the entire infinitely degenerate LLL.) In problems where such nontrivial dynamics is left, one usually finds that variables that used to commute become canonically conjugate.

Exercise 21.1.4. Study the semiclassical orbits and show that the motion is on contours of constant V . (Hint: Consider the gradient of V .)

How can X and Y suddenly become noncommuting when by postulate they are commuting? The answer is simply that if two matrices commute in a given space (the full Hilbert space), their truncations to a subspace (here the states of the LLL) need not. What is nice is that the commutator of X and Y , instead of being something ugly is a constant, making the pair canonically conjugate (upon trivial rescaling).

Exercise 21.1.5. Consider the commuting 3×3 matrices Ω and Λ from Exercise (1.8.10). If you truncate them by dropping the third row and column, show that the 2×2 truncations do not commute.

Consider a finite system with N electrons, i.e., a system with a fully filled LLL, *with one electron per state in the LLL*. Ignore all interactions between the electrons or with the medium. What is its ground state? Since their spins are polarized along the field, the spatial wave function must be antisymmetric in the electron spatial coordinates and be analytic. An unnormalized product wave function for the N particles is

$$\Psi_P = z_1^0 z_2^1 z_3^2 \cdots z_N^{N-1} \exp\left(-\frac{qB}{4\hbar c} \sum_i z_i^* z_i\right) \equiv u_P \exp\left(-\frac{qB}{4\hbar c} \sum_i z_i^* z_i\right) \quad (21.1.48)$$

When antisymmetrized, this leads to

$$u_A = \prod_{i=1}^N \prod_{j=1}^{i-1} (z_i - z_j) \quad (21.1.49)$$

Exercise 21.1.6. Verify the above equation for the three particle case. Show this also by writing out the (3×3) determinant as in Eq. (10.3.36). (In all these manipulations, the exponential factor in the wave function, which is totally symmetric in the coordinates, plays no part.)

This wave function is unique since there is just one way to place N (spin polarized) electrons in N states. So we know the unique ground state for the fully filled LLL in the noninteracting limit. But even if we consider the interactions between electrons, this is the only antisymmetric wave function we can write for this problem where the number of states equals the number of electrons, *if we do not want to go above the LLL*.

Now, the really interesting problem is one where in the same field and sample, we have a smaller number vN of electrons where $1/v$ is an odd integer. (This is one of the cases where the experiments show surprising results.) We say the system has a *filling factor* v meaning it has v times the maximum allowed number of particles in the LLL. The fully filled LLL is thus given by $v=1$. Whereas previously we put an electron in each LLL state (and there was just one way to do it and hence one antisymmetric wave function), now there is more than one way and hence many possible superpositions of LLL wave functions that can be candidates for the ground

state of a system of electrons interacting via the Coulomb potential. Laughlin proposed the following wave function:

$$u_v = \prod_{i=1}^{vN} \prod_{j=1}^{i-1} (z_i - z_j)^{1/v} \quad (21.1.50)$$

Let us verify that this wave function fits the description. Pick any one particle coordinate, say z_1 (since the particles are identical) and observe that the highest power that occurs is (for large N)

$$z_1^{1/v \cdot vN} = z_1^N$$

Thus the size of the biggest wave function precisely matches the sample size. Next note that the function is antisymmetric under exchange of any two coordinates since $1/v$ is odd. Lastly note that the electrons nicely avoid each other (due to the high-order zero when any two coordinates approach each other) thereby minimizing their repulsive interaction. Not surprisingly, this happens to be an excellent ground state wave function at these filling factors, (for small $1/v$).

The Berry Phase

The problem in question has to do with the adiabatic approximation. Recall the example of the particle in a box of size L . Let us say it is in the ground state. Suppose the box slowly expands with time as some function $L(t)$. The adiabatic principle states that if the expansion is slow enough, the particle will be in the ground state of the box of size $L(T)$ at time T . Likewise the particle that starts out in the state $|n(L(0))\rangle$ will find itself in the instantaneous eigenstate $|n(L(t))\rangle$ at time t .

More generally, if the particle Hamiltonian is given by $H(R(t))$ where R is some external coordinate which changes slowly and appears parametrically in H , the adiabatic principle tells us that the particle will sit in the n th instantaneous eigenket of $H(R(t))$ at a time t if it started out in the n th eigenstate of $H(R(0))$.

What is the solution to the Schrödinger equation in this approximation? Here is a reasonable guess:

$$|\psi(t)\rangle = \exp\left(-\frac{i}{\hbar} \int_0^t E_n(t') dt'\right) |n(t)\rangle \quad (21.1.51)$$

where

$$H(t)|n(t)\rangle = E_n(t)|n(t)\rangle \quad (21.1.52)$$

First note that if H does not vary with time, the above answer is clearly correct, with the phase factor appropriate to energy E_n . The above formula recognizes that the instantaneous energy varies with time and gives the accumulated phase shift, just as the WKB wave function gives the phase as the spatial integral of a position-dependent momentum for a particle moving in a nonconstant $V(x)$.

Over the years, many people, notably Herzberg and Longuet-Higgins and Mead and Truhlar, recognized various problems with this formula and found ways to fix them. The whole problem was brought into sharp focus and synthesized by Berry. You are urged to read his very lucid writings and the collection of related papers (with helpful commentary) edited by Shapere and Wilczek, referred to in the Bibliography at the end of the chapter.

To see what is missing in the above ansatz, let us modify it as follows:

$$|\psi(t)\rangle = c(t) \exp\left(-\frac{i}{\hbar} \int_0^t E_n(t') dt'\right) |n(t)\rangle \quad (21.1.53)$$

where the extra factor $c(t)$ must be equal to unity if the old ansatz is right. Let us apply the Schrödinger equation to this state:

$$\left(i\hbar \frac{\partial}{\partial t} - H(t)\right) |\psi(t)\rangle = 0 \quad (21.1.54)$$

When the time derivative acts, it generates three terms: one from the derivative of the accumulated phase factor (which neutralizes the action of H on the eigenket), one from the derivative of $c(t)$ and *one from the derivative of the instantaneous eigenket*. The last two terms lead to the following equation (on dotting both sides with the instantaneous bra):

$$\dot{c}(t) = -c(t) \langle n(t) | \frac{d}{dt} |n(t)\rangle \quad (21.1.55)$$

with a solution

$$c(t) = c(0) \exp\left(-\int_0^t \langle n(t') | \frac{d}{dt'} |n(t')\rangle dt'\right) = c(0) e^{i\gamma} \quad (21.1.56)$$

$$\gamma = i \int_0^t \langle n(t') | \frac{d}{dt'} |n(t')\rangle dt' \quad (21.1.57)$$

The impressive thing is not to find this extra phase, called the *Berry phase* or the *geometric phase*, but to recognize that it can have measurable consequences. After all, we have been learning all along that the phase of a ket makes no difference to any measurable quantity. Since the instantaneous kets themselves are defined only

up to a phase factor, we can choose a new set and modify the extra phase. If we choose

$$|n'(t)\rangle = e^{i\chi(t)} |n(t)\rangle \quad (21.1.58)$$

then we find

$$i\langle n'(t)| \frac{d}{dt} |n'(t)\rangle = i\langle n(t)| \frac{d}{dt} |n(t)\rangle - \frac{d\chi(t)}{dt} \quad (21.1.59)$$

suggesting that perhaps we could choose $\chi(t)$ so as to completely neutralize the extra phase. It had been generally assumed that such a choice could always be made and the extra phase forgotten.

Suppose now that the parameter that changes with time and causes the Hamiltonian to change returns to its starting value after time T so that:

$$H(T) = H(0) \quad (21.1.60)$$

Now it is no longer obvious that we can get rid of the extra phase. We find

$$i \oint \langle n'(t)| \frac{d}{dt} |n'(t)\rangle dt = i \oint \langle n(t)| \frac{d}{dt} |n(t)\rangle dt - (\chi(T) - \chi(0)) \quad (21.1.61)$$

Now the choice of phase factors is quite arbitrary, but it must meet the requirement that the assignment is single-valued, at least within the region containing the closed loop in question. (A single-valued choice in the entire parameter space will generally be impossible. This is a subtle topic, reserved for Exercise (21.1.15).) So let us start with such a basis $|n(t)\rangle$ and make a switch to another one $|n'(t)\rangle = e^{i\chi(t)} |n(t)\rangle$. Since the new basis is by assumption single-valued, so must be the additional phase factor. In other words, $(\chi(T) - \chi(0)) = 2m\pi$, where m is an integer. This in turn means that the prefactor $e^{i\gamma} \equiv \exp[i \oint \langle n(t)| i(d/dt)|n(t)\rangle dt]$ arising in a closed circuit cannot be altered by a choice of basis. Note also that since dt cancels out in any of the integrals, we cannot shake this phase by slowing down the rate of change of the parameter. The phase factor depends only on the path in parameter space, which explains the name “geometric phase.” Note that we have not shown that $e^{i\gamma} \neq 1$, but only that its value is not affected by redefinition of phases for the state vectors.

So let us suppose we have a nonzero γ . What exactly does it do? To see this, let us consider a problem where the box is not really a box, but the potential of some heavy object. For example, let R be the coordinate of some nucleus and r that of an electron that is orbiting around it. In this discussion we will deviate from our usual notation: capital letters will stand for nuclear coordinates and momenta (classical or quantum) and lowercase letters will represent the electron. We will also temporarily ignore the vector nature of these variables. The box here is the Coulomb well created by the nucleus. As the nucleus moves, the box moves, rather than change size, but the issues are the same. As the nucleus crawls from place to place, the nimble electron stays in the instantaneous eigenstate. Even though we have paid no

attention to the dynamics of the nucleus, we shall see one is generated by the Berry phase. Let us rewrite the phase factor as follows:

$$\exp\left(-\int_0^t \langle n(t') | \frac{d}{dt'} |n(t')\rangle dt'\right)$$

$$= \exp\left(\frac{i}{\hbar} \int_0^t \langle n(t') | \frac{d}{dt'} |n(t')\rangle dt'\right) \quad (21.1.62)$$

$$= \exp\left(\frac{i}{\hbar} \int_0^t i\hbar \langle n(R(t')) | \frac{d}{dR} |n(R(t'))\rangle \frac{dR}{dt'} dt'\right) \quad (21.1.63)$$

$$= \exp\left(\frac{i}{\hbar} \int_0^t A^n(R) \frac{dR}{dt'} dt'\right) \text{ where} \quad (21.1.64)$$

$$A^n(R) = i\hbar \langle n(R) | \frac{d}{dR} |n(R)\rangle \quad (21.1.65)$$

Thus we see that the slow nuclear degree of freedom has a velocity coupling to a vector potential $A^n(R)$, called the *Berry potential*. The potential depends on which quantum state $|n\rangle$ the electronic degree of freedom is in. When the state vectors are redefined by phase transformations, this vector potential undergoes a gauge transformation:

$$|n(R)\rangle \rightarrow e^{i\chi(R)} |n(R)\rangle \quad (21.1.66)$$

$$A^n(R) \rightarrow A^n(R) - \hbar \frac{d\chi}{dR} \quad (21.1.67)$$

However, its line integral around a closed loop is gauge invariant and could be nonzero. To ignore this would be to get the wrong dynamics for the nucleus.

Now some of you may feel a little unhappy and say: “I know how the vector potential is supposed to enter the Lagrangian or action, but you pulled it out of a phase factor in the wave function of the fast coordinates.” This is a fair objection and in answering it in some detail we will learn that there is also a scalar potential besides the vector potential.

We begin by constructing a path integral for the nuclear degrees of freedom. What resolution of the identity should we use? The one appropriate to our problem is this:

$$I = \int dR \sum_n |R, n(R)\rangle \langle n(R), R| \quad (21.1.68)$$

where $|R, n(R)\rangle \equiv |R\rangle \otimes |n(R)\rangle$. In other words, at each R , we pick a basis for the electrons that diagonalizes the instantaneous electronic Hamiltonian $H_e(R, r, p)$

$$H_e(R, r, p)|R, n(R)\rangle = E_n(R)|R, n(R)\rangle \quad (21.1.69)$$

Of course, you can pick a basis for the electrons that has no correlation to the nuclear coordinates. While this is mathematically correct, it is not wise for the adiabatic approximation. For the latter, we now make the approximation that if the electron starts out at some value of n , it stays there and all other values can be ignored. Thus we write:

$$I \simeq \int dR |R, n(R)\rangle \langle n(R), R| \quad (21.1.70)$$

where the sum on n has been dropped. The derivation of the configuration space path integral in R proceeds as usual. A typical factor in the path-integrand will be

$$\langle n(R(t + \varepsilon)), R(t + \varepsilon) | \exp\left[-\frac{i\varepsilon}{\hbar} H(R, P)\right] \exp\left[-\frac{i\varepsilon}{\hbar} H_e(R, r, p)\right] |n(R(t)), R(t)\rangle \quad (21.1.71)$$

The nuclear part, sandwiched between nuclear coordinate eigenstates, will give the usual factor

$$\begin{aligned} & \langle R(t + \varepsilon) | \exp\left[-\frac{i\varepsilon}{\hbar} H(R, P)\right] |R(t)\rangle \\ &= \sqrt{\frac{m}{2\pi\hbar i\varepsilon}} \exp\left[\frac{i\varepsilon}{\hbar} \left[\frac{m}{2\varepsilon^2} (R(t + \varepsilon) - R(t))^2 - V(R) \right] \right] \end{aligned} \quad (21.1.72)$$

while the electronic exponential will act on its eigenket to the right and give a factor $\exp[-(i\varepsilon/\hbar)E_n(R)]$ which will change the nuclear potential by $E_n(R)$. This is how Born and Oppenheimer analyzed molecules, where there is a clear separation of fast (electronic) and slow (nuclear) degrees of freedom: fix the slow ones, solve for the fast ones at this value, and use the fast eigenenergies as an additional potential for the slow problem which is then solved.

But this is not the full story. After the electronic exponential has acted on its eigenket to the right, yielding the exponential phase factor $\exp[-(i\varepsilon/\hbar)E_n(R)]$, we are still left with the following dot product which multiplies everything:

$$\langle n(R(t + \varepsilon)) | n(R(t)) \rangle \equiv \langle n(R') | n(R) \rangle \quad (21.1.73)$$

All the results will follow from an analysis of this factor. First, it is true that when $R = R'$ this factor equals unity. We are going to perform a Taylor expansion of this product in the difference $R - R' = \eta$. How far should we go? The answer is clear if we recall Chapter 8 where we derived the Schrödinger equation from the path integral

by considering the propagator for infinitesimal times, i.e., one time slice of width ε . I reproduce the relevant formula Eq. (8.5.7) with two changes. I drop all interactions and keep just the free particle propagator but I append the dot product $\langle n(R')|n(R) \rangle$. This yields for the nucleus

$$\psi(R', \varepsilon) = \left(\frac{m}{2\pi\hbar i\varepsilon} \right)^{1/2} \int_{-\infty}^{\infty} e^{im\eta^2/2\hbar\varepsilon} \langle n(R')|n(R' + \eta) \rangle \psi(R' + \eta, 0) d\eta \quad (21.1.74)$$

The exponential allows η to fluctuate by (recall Eq. (8.5.6))

$$|\eta| \simeq \left(\frac{2\pi\hbar\varepsilon}{m} \right)^{1/2} \quad (21.1.75)$$

This means we must go to order η^2 since we want to go to order ε to derive the Schrödinger equation. So we expand ψ and $\langle n(R')|n(R' + \eta) \rangle$ to this order:

$$\psi(R' + \eta, 0) = \psi(R', 0) + \eta \frac{\partial\psi}{\partial\eta} + \frac{\eta^2}{2} \frac{\partial^2\psi}{\partial\eta^2} + \dots \quad (21.1.76)$$

$$\langle n(R')|n(R' + \eta) \rangle = 1 + \eta \langle n|\hat{\partial}n \rangle + \frac{\eta^2}{2} \langle n|\hat{\partial}^2n \rangle + \dots \quad (21.1.77)$$

where all derivatives are taken at the point R' and $|\hat{\partial}n\rangle$ is the derivative of $|n\rangle$ with respect to R' and so on. If we now inject these expansions into Eq. (21.1.74), and keep just the even powers of η as we did in Chapter 8, we find upon doing the Gaussian integrals and dropping the prime on R'

$$i\hbar(\psi(R, \varepsilon) - \psi(R, 0)) = \varepsilon \left[-\frac{\hbar^2}{2m} \frac{\partial^2\psi}{\partial R^2} - \frac{\hbar^2}{m} \langle n|\hat{\partial}n \rangle \frac{\partial\psi}{\partial R} - \frac{\hbar^2}{2m} \langle n|\hat{\partial}^2n \rangle \psi \right] \quad (21.1.78)$$

*Exercise 21.1.7.** Provide the missing steps leading to the above equation.

The Hamiltonian can be read off the above:

$$H = \frac{1}{2m} (P - A'')^2 + \Phi'' \quad (21.1.79)$$

$$A'' = i\hbar \langle n|\hat{\partial}n \rangle \quad (21.1.80)$$

$$\Phi'' = \frac{\hbar^2}{2m} [\langle \hat{\partial}n|\hat{\partial}n \rangle - \langle \hat{\partial}n|n \rangle \langle n|\hat{\partial}n \rangle] \quad (21.1.81)$$

*Exercise 21.1.8.** Providing the missing steps. Use $\langle n|\hat{\partial}n \rangle = -\langle \hat{\partial}n|n \rangle$ which follows from $\partial\langle n|n \rangle = 0$. The potential Φ'' arises from adding and subtracting the $(A'')^2$ term which isn't there to begin with.

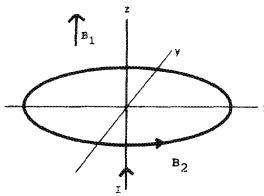


Figure 21.1. The field B_2 and electron motion are along the circle. The particle spin is up or down the local magnetic field which is the sum of B_1 and B_2 . The current I produces B_2 .

The (discretized) action function which will give exactly these results will have the $v \cdot A''$ term (with A'' evaluated at the midpoint) and the extra scalar potential Φ'' . We will not write that down since we have the Hamiltonian. The following exercise considers this point more carefully.

Exercise 21.1.9. Suppose we do not derive the Hamiltonian as above (by invoking the wave function) but want to determine the correct discretized action function starting with Eq. (21.1.71) and expanding $\langle n(R')|n(R)\rangle$ to quadratic order in $R' - R$ as per Eq. (21.1.77) and exponentiating the result. Do all of the above and show that the argument of the vector potential that arises is not at the midpoint to begin with, as it should to represent the effect correctly [Exercise (8.6.4)]. Fix this with a Taylor series, combine the term quadratic in $R' - R$ that arises, with the one you had to begin with, to obtain (for one time slice)

$$\begin{aligned} S = & \frac{im(R' - R)^2}{2\hbar\varepsilon} + \frac{i}{\hbar}(R' - R)A''\left(\frac{R + R'}{2}\right) \\ & - \frac{(R' - R)^2}{2}\langle \hat{c}_n | (I - |n\rangle\langle n|) | \hat{c}_n \rangle \end{aligned} \quad (21.1.82)$$

Let us now ask what continuum form this describes. Multiplying and dividing by ε converts the first term into the kinetic energy and the middle term to the vector potential coupling. The last term needs to be multiplied and divided by ε^2 to become the square of the velocity. But this would leave it with an extra ε in the continuum action. Despite this, the term is important since the square of the velocity is very singular. The effect of the term is best revealed by noting that the factor $(R' - R)^2$ is going to be replaced by $i\varepsilon\hbar/m$ when the functional integral is done, (because of the kinetic energy term in the action that controls the variance of $R - R'$), make this replacement now, and convert this term to the scalar potential Φ'' , which we know describes the right Hamiltonian. The role of such terms, naively vanishing in the continuum limit has been discussed by Klauder, Klauder and Skagerstam (1985).

It should be clear that the preceding results generalize with R and A replaced by vectors \mathbf{R} , \mathbf{A} , more fast and slow degrees of freedom, etc.

We turn to a simple problem where the Berry potential makes a difference.[‡] Consider the situation in Fig. 21.1.

A spinless, electrically neutral particle of mass M is restricted to move on a circle of radius a . This motion is going to be the slow degree of freedom in our problem. The orbit is penetrated by a flux due to a field $B_1 k$ along the z -axis. In addition, a wire carrying some current along the z -axis is introduced at the center.

[‡] I thank Ady Stern for suggesting a variant of this example. He is not responsible for any errors in my presentation.

It produces an azimuthal field of strength B_2 . The total field makes an angle

$$\theta = \arctan B_2/B_1$$

with respect to the z -axis and has a magnitude $B = \sqrt{B_1^2 + B_2^2}$. When the particle coordinate is ϕ , the field B_2 is tangent to the circle, i.e., has an azimuthal angle $\phi + \pi/2$ in \mathbf{B} -space. The Hamiltonian of the particle (not yet coupled to \mathbf{B}) is:

$$H = \frac{L^2}{2I} \quad (21.1.83)$$

where $I = Ma^2$ is the moment of inertia, set equal to $1/2$ from now on and $L = -i\hbar \partial/\partial\phi$ is the angular momentum operator. The energy eigenvalues are

$$E_m = \hbar^2 m^2 \quad m = 0, \pm 1, \pm 2 \dots \quad (21.1.84)$$

We now bring in the fast degree of freedom. Imagine that the particle has spin $1/2$. As the particle goes around the circle, the spin will see a varying magnetic field, \mathbf{B} , which is the vector sum of the fixed field B_1 along the z -axis and the azimuthal field B_2 . We modify H as follows:

$$H = L^2 - C\sigma \cdot \mathbf{B}(\phi) \quad (21.1.85)$$

where C and hence the splitting between the two spin states is assumed to be so large (as is the frequency associated with the splitting) that the spin is truly a fast degree of freedom which will not jump between its states as the particle crawls around the loop.

What will the allowed energies be? The naive answer is

$$E_m = \hbar^2 m^2 \mp CB \quad (21.1.86)$$

where $B = \sqrt{B_1^2 + B_2^2}$ and the two signs correspond to the spin pointing up/down the local magnetic field as the particle goes round and round. This is however wrong and one must take into account the Berry potentials $A(\phi)$ and Φ . Let us focus on the lower-energy solution in which the spin points up the local field. We choose the spinor to be

$$|\theta\phi\rangle = \begin{bmatrix} \cos \frac{\theta}{2} \\ i \sin \frac{\theta}{2} e^{i\phi} \end{bmatrix} \quad (21.1.87)$$

(The additional i in the lower component is due to the fact that orbital angle ϕ differs from the azimuthal angle of the field by $\pi/2$ as is clear from Fig. 21.1.) It is

readily found that

$$A^+(\phi) = i\hbar \langle \theta \phi | \frac{\partial}{\partial \phi} | \theta \phi \rangle = -\hbar \sin^2 \frac{\theta}{2} \quad (21.1.88)$$

which is independent of ϕ , and that the scalar Berry potential is

$$\Phi = \frac{\hbar^2 \sin^2 \theta}{4} \quad (21.1.89)$$

which is independent of whether the spin is pointing up or down the local field. Since θ is fixed in this problem, Φ can be eliminated by a choice of reference energy, and we no longer consider it.

Exercise 21.1.10. Prove the above equations for the vector and scalar potentials.

Since the effect of the vector potential is $L \rightarrow L - A^+$, it follows that if we solve

$$\left[-i\hbar \frac{\partial}{\partial \phi} - A^+ \right] \psi = \lambda \psi \quad (21.1.90)$$

the energy is given by

$$E^+ = \lambda^2 - BC \quad (21.1.91)$$

The orbital eigenfunctions are once again

$$\psi = e^{im\phi} \quad m = 0, \pm 1, \pm 2, \dots \quad (21.1.92)$$

so that

$$\lambda = m\hbar - A^+ = \left(m + \sin^2 \frac{\theta}{2} \right) \hbar \quad (21.1.93)$$

and the energy of the spin up state is

$$E^+ = \left(m + \sin^2 \frac{\theta}{2} \right)^2 \hbar^2 - BC \quad (21.1.94)$$

It is evident that without the vector potential we would get the wrong answer. For example, without it, there would be a twofold degeneracy under $m \rightarrow -m$.

Exercise 21.1.11. Find the potential for the other (spin down) state and the energy eigenvalues.

Let us rederive the scalar and vector potentials of Eqs. (21.1.79–21.1.81) without path integrals, by extracting the effective Hamiltonian that acts on the slow degrees of freedom R . Now the latter need not be in an eigenstate of position, it could be in a superposition $\psi(R)$:

$$|\psi\rangle = \int \psi(R)|R, n(R)\rangle dR \quad (21.1.95)$$

Note that $|\psi\rangle$ is a ket in the direct product space of the slow and fast degrees of freedom. Usually the coefficients in such a superposition would depend on both labels. But in our problem the fast degree of freedom is slaved to the slow one, so that the amplitude for the slow one to be in $|R\rangle$ is the same as the amplitude for the entire system to be in $|R, n(R)\rangle$. We are going to find the Hamiltonian in the coordinate representation by calculating

$$(H\psi)(R') \equiv \langle R', n(R') | H | \psi \rangle \quad (21.1.96)$$

$$= \int \langle R', n(R') | H | R, n(R) \rangle \langle R, n(R) | \psi \rangle dR \quad (21.1.97)$$

$$= \int \langle R', n(R') | H | R, n(R) \rangle \psi(R) dR \quad (21.1.98)$$

Let

$$H = P^2/2M + V(R) + H_f(r, p, R) \quad (21.1.99)$$

It is evident that the fast Hamiltonian H_f , acting to the right on its eigenket, will give $E_n(R)$ and that this will join with $V(R)$ to provide a potential energy term. We focus therefore on just the $P^2/2M$ since here is where the action is. Let us recall that

$$\langle R' | \frac{P^2}{2M} | R \rangle = -\frac{\hbar^2}{2M} \delta''(R' - R) \quad (21.1.100)$$

and insert it into Eq. (21.1.98) to obtain

$$\begin{aligned} (H\psi)(R') &= -\frac{\hbar^2}{2M} \int \langle n(R') | n(R) \rangle \delta''(R' - R) \psi(R) dR \\ &= -\frac{\hbar^2}{2M} \langle n(R') | [|\partial^2 n(R)\rangle \psi(R) + 2|\partial n(R)\rangle \partial \psi(R) + |n(R)\rangle \partial^2 \psi(R)]_{R=R'} \\ &= -\frac{\hbar^2}{2M} [\langle n | \partial^2 n \rangle \psi(R') + 2\langle n | \partial n \rangle \partial \psi(R') + \partial^2 \psi(R')] \end{aligned} \quad (21.1.101)$$

where ∂ denotes derivatives respect to R' . It is now straightforward to show that the operator on the right-hand side is indeed the one in Eq. (21.1.79). The details are left to the following exercise.

Exercise 21.1.12. Provide the missing details. Suggestion: Start with Eq. (21.1.79) and expand out the $(P - A)^2$. Note that when P comes to the left of A , it differentiates both A and the wave function ψ that is imagined to be sitting to the right of the Hamiltonian. Now go to Eq. (21.1.101), add and subtract the A^2 term and regroup the terms using relations like $\partial\langle n|\partial n\rangle = \langle\partial n|\partial n\rangle + \langle n|\partial^2 n\rangle$.

Now that we accept the reality of the Berry vector potential, let us understand it a little better. Normally when we have a vector potential, we take its curl and the corresponding magnetic field has as its origin some current. Had there been magnetic monopoles, the source could have been a monopole. What is producing the Berry potential? Let us first appreciate that the source of the potential does not lie in the configuration space of the fast degree of freedom, but in the space of parameters that are slowly varying in the fast Hamiltonian H_f . Of course, this slow parameter could itself be a real live degree of freedom (as in our ring example) but this is not our focus. We simply treat the slow variables as external parameters that define H_f . Thus if we consider a spin-1/2 object with

$$H = -\boldsymbol{\sigma} \cdot \mathbf{B} \quad (21.1.102)$$

then the Berry potential lives in \mathbf{B} space. (Since we focus on just the fast variables, we drop the subscript on H_f .) To ease our thinking we are going to rename \mathbf{B} space as \mathbf{R} space, but you should not forget this fact. So we write

$$H = -\boldsymbol{\sigma} \cdot \mathbf{R} \quad (21.1.103)$$

Every point in \mathbf{R} space defines a possible spin Hamiltonian. We have managed to define in this space a vector potential. It is derived from the n th quantum state of the above Hamiltonian and is given by

$$\mathbf{A}'' = i\hbar\langle n(\mathbf{R}) | \nabla | n(\mathbf{R}) \rangle \quad (21.1.104)$$

What is its curl? To figure this out, we need a little groundwork. Using

$$0 = \nabla \langle n | H | m \rangle \quad m \neq n \quad (21.1.105)$$

we find on differentiating all three factors and shifting a derivative from bra to ket at the cost of sign change (thanks to $\nabla \langle n | m \rangle = 0$),

$$\langle n | \nabla | m \rangle = \frac{\langle n | (\nabla H) | m \rangle}{E_m - E_n} \quad (21.1.106)$$

It is now easy to find a formula for the field tensor F_{ij} associated with the Berry potential:

$$\begin{aligned}
 F_{ij}^n &= \partial_i A_j^n - \partial_j A_i^n \\
 &= i\hbar[\partial_i \langle n | \partial_j n \rangle - \partial_j \langle n | \partial_i n \rangle] \\
 &= i\hbar \sum_{m \neq n} \frac{\langle n | (\partial_i H) | m \rangle \langle m | (\partial_j H) | n \rangle - \langle n | (\partial_j H) | m \rangle \langle m | (\partial_i H) | n \rangle}{(E_m - E_n)^2} \\
 &\quad \left(\partial_i = \frac{\partial}{\partial R_i} \right)
 \end{aligned} \tag{21.1.107}$$

where m labels a complete set of states we introduce along the way. (The $m=n$ terms drop out due to a cancellation.) This formula is valid in general (for any H) and we now apply it to our problem.

In our problem there are many simplifying features:

- $\partial H / \partial R_j = -\sigma_j$
- There are only two states and hence only one term in the sum over m . The energy denominator squared is $4R^2$ since $2R$ is the difference between up and down spin states. (Remember R is now the magnitude of the magnetic field!)
- So we pull out this denominator, which is independent of m , add a term with $m=n$ (which vanishes by antisymmetry in i and j), use completeness to eliminate the intermediate states, use the commutation relations for the Pauli matrices, and finally the fact that $\langle n | \sigma_i | n \rangle = \pm \hat{\mathbf{R}}$ (for the states up/down the field).

Rather than state the field in terms of the tensor F_{ij}^n , we write in terms of the more familiar magnetic field defined by $\mathcal{B}_k^n = F_{ij}^n$ (where the indices i, j, k run cyclically):

$$\mathcal{B}^n = \mp \hbar \frac{\hat{\mathbf{R}}}{2R^2} \tag{21.1.108}$$

This is the field of a monopole of strength $-\hbar/2$ sitting at the origin, which is the point of degeneracy of the Hamiltonian.

Exercise 21.1.13. Furnish the missing steps in the above derivation.

Note that there are two different magnetic fields in the problem. The first is a real one \mathbf{B} which couples to the electron spin and resides in real space. It is produced by currents in real space. (There are no known monopole sources for such fields.) The second field is the curl of the Berry vector potential that resides in parameter space. Its components are denoted by \mathcal{B}_k^n which happens, in our problem, to describe a monopole in parameter space. We will now see that the Berry monopole will arise in any problem where the Hamiltonian (not necessarily containing magnetic fields) becomes doubly degenerate.

Assuming the parameter space is three-dimensional, let us focus on just the two nearly degenerate levels. Now, any 2×2 Hermitian operator can be written as

$$H = \sum_{\mu=0}^3 \sigma_\mu f_\mu \quad (21.1.109)$$

where $\sigma_0 = I$ is the fourth partner to the Pauli matrices, and f_μ are four functions of the three independent coordinates of parameter space. The eigenvalues of H are clearly

$$E = f_0 \pm \sqrt{f_x^2 + f_y^2 + f_z^2}. \quad (21.1.110)$$

The degeneracy occurs at $f_x = f_y = f_z = 0$ which we choose to be the origin of coordinates. We also shift the overall zero of energy so that the degenerate eigenvalue $f_0(0)$ vanishes. Let us now use the three f 's themselves as the new coordinates in which case f_0 will be some function of these coordinate and vanish at the origin. Thus

$$H = f_0(\mathbf{f})I + \boldsymbol{\sigma} \cdot \mathbf{f} \quad (21.1.111)$$

in obvious notation. Note that f_0 vanishes at the origin but not necessarily elsewhere. Let us repeat the same analysis we used in the spin problem, starting with

$$\partial_i H = \partial_i f_0 I + \sigma_i \quad (21.1.112)$$

If we next evaluate the field tensor as per Eq. (21.1.107), we see that the part proportional to the identity does not matter (since $\langle m|n \rangle = 0$ for $m \neq n$), the problem becomes isomorphic to the one in Eq. (21.1.103) and we get just the monopole at the origin.

Exercise 21.1.14. Take another look at the problem we studied, of a particle moving around in a loop with fields in the azimuthal and z -directions. As the particle goes once around the circle, the line integral of the vector potential A^+ is

$$\oint A^+ d\phi = -2\pi\hbar \sin^2 \frac{\theta}{2}$$

Let us now look at the same closed orbit in \mathbf{B} -space where it is a loop of fixed radius B_2 at a fixed height B_1 above the $B_x - B_y$ plane. Thus it defines the co-latitude (at angle θ measured from the north pole) of a sphere of radius $\sqrt{B_1^2 + B_2^2}$. In this space we have a monopole of strength $-\hbar/2$ at the origin according to Eq. (21.1.108). The flux through this loop is then the monopole flux penetrating the area of the cap bounded by this latitude. Using Stoke's theorem show that this flux equals $-2\pi\hbar \sin^2 \theta/2$ as it should. (Note that the Berry vector potential is different in real space and parameter space. Its line integral over a closed loop, which measures the accumulated phase change per revolution, is of course the same. Consider in general a map from manifold X with points labeled x , to Y with points labeled y , such that each x goes into a unique y . If $A(y)$ is a vector potential in Y , we can import it to X by defining a vector potential $A(x)$ such that (suppressing indices)

$$A(x) dx = A(y) dy \quad (21.1.113)$$

By construction, closed loops in X go to closed loops in Y . The line integral of $A(x)$ around a closed loop in X will then equal the line integral of $A(y)$ around the image loop in Y .)

Excise 21.1.15. Let us discuss the question of assigning phases to state vectors in parameter space through an example. Let $\mathbf{R} \equiv (R, \theta, \phi)$ be the coordinate in parameter space. Consider the Hamiltonian $H = -\boldsymbol{\sigma} \cdot \mathbf{R}$. Let us write down the ground state for this problem for all points. It is the one where the spin points radially outward everywhere. A choice for the spinor is

$$|+, \theta, \phi\rangle = \begin{bmatrix} \cos \frac{\theta}{2} \\ \sin \frac{\theta}{2} e^{i\phi} \end{bmatrix}$$

This is just the ket we used in the problem of the electron going around in a loop (except for the factor i in the lower component which arose due the $\pi/2$ difference between the azimuthal angles in real and parameter space). Since the spinor has no R dependence let us look at it on a unit sphere $R=1$. Observe that the lower component does not approach a unique value as we approach the south pole from different directions. (This problem does not exist at the north pole since $\sin \theta/2 = 0$ there.) Thus we really have not defined the spinor globally. If we multiply the whole spinor by the single-valued phase factor $e^{-i\phi}$, we now have a spinor well defined near the south pole, but singular at the north pole. It follows that we can only define the spinor in patches of parameter space. In our problem two patches will do, one excluding the north pole and one excluding the south.

Since we found the Berry potential by taking derivatives of the ket, it follows that the former is also defined only in the patches and not globally. In other words, Eq. (21.1.88) for A^+ is to be used away from $\theta = \pi$. To describe the south pole, we can use, for example, the potential coming from the spinor with good behavior at the south pole, but bad behavior at the north pole.

I will now argue that attempts to find a global vector potential in the presence of a monopole are doomed. Say we had a global nonsingular vector potential. Consider its line integral along the direction of increasing ϕ on a latitude near the north pole on a unit sphere surrounding the monopole. By Stokes's theorem this equals the flux through the cap above this latitude. If we enlarge the loop and go past the equator, the line integral will monotonically increase. Finally, let us shrink the loop to an infinitesimal one around the south pole. As this loop shrinks, the line integral does not vanish; it equals the full monopole flux. It follows there must be a singularity at the south pole since the integral of a nonsingular potential around an infinitesimal loop must be infinitesimal and vanish with loop size. (It is also possible that the singularity is elsewhere on the sphere, but it has to exist by similar reasoning.)

Starting with the gradient in spherical coordinates, show that the vector potential associated with $|+, \theta, \phi\rangle$ is given by

$$\mathbf{A} = -\frac{\hbar}{2} \mathbf{e}_\phi \frac{(1 - \cos \theta)}{R \sin \theta}$$

Observe the singularity at the south pole. This is called the *Dirac string*. Show that its line integral around a tiny loop surrounding the south pole is the full monopole flux. What is happening is this. This vector potential describes not a monopole at the origin, but one where a tiny tube (the Dirac string) comes up the negative z -axis, smuggling in the entire flux to the

origin, from which point it emanates radially. The string flux is the reason the tiny loop around the south pole gives a nonzero answer equal to the total flux.

Now there is nothing special about the south pole when we look at the monopole, since it is spherically symmetric. This is reflected in the fact that the Dirac string can be moved around by a gauge transformation. Calculate the vector potential \mathbf{A}' with the spinor obtained by multiplying both components of $|+, \theta, \phi\rangle$ by $e^{-i\phi}$. Show that it has troubles at the north pole and that the two vector potentials are related by the gauge transformation associated with the redefinition $|+, \theta, \phi\rangle \rightarrow e^{-i\phi} |+, \theta, \phi\rangle$.

If we are allowed to use words instead of equations, we can describe the effect of the monopole without any strings: when the charged particle goes around in a loop, it picks up a phase proportional to the solid angle the loop subtends at the origin (where the monopole is). The vector potential is the analytical way to generate the solid angle via Stokes's theorem, but it cannot do it globally.

Now Dirac ran into this problem trying to ask how we would describe a real (not Berry) monopole of charge g in real space. It has a radial field that falls off as g/R^2 . No problem there. *But quantum mechanics forces us to work with vector potentials.* Now any vector potential we can come up with has a string. As usual, Dirac turned a potential disaster into a dazzling prediction by arguing that *if there is a monopole and we have no choice but to describe it with a vector potential, it must be that the string is unobservable*. The line integral of the vector potential around the string at the south pole is $4\pi g$, the total flux of the monopole. For a particle of charge q , this will enter the dynamics via the factor

$$e^{4\pi i q g / \hbar c}$$

as per Eq. (18.4.38). (Think of an Aharonov–Bohm experiment in which a particle goes on either side of the string.) If this factor is to be unobservable we require that

$$q = \frac{\hbar n c}{2g}$$

where n is any integer. This remarkable argument tells us that *even if there is a single monopole in the universe, it forces all electric charges to be multiples of $\hbar c/2g$* . This explains, for example, why the proton and electron have exactly the same charge. However no monopole has yet been seen. But, the argument is so attractive I for one am sure at least one monopole exists. If not, nature would have missed a wonderful opportunity, to paraphrase Einstein.

In modern treatments, one uses two patches, say one without the south pole and one without the north pole, with a different vector potential in each. By demanding that where the patches overlap, say the equator, the two potentials differ by a single-valued gauge transformation, one recovers Dirac's quantization condition. (You may provide the proof yourself if you remember that (1) the difference of the line integrals of the two patch potentials around the equator is the integral over the whole sphere of the outgoing flux; (2) when the wave function of a particle of charge q is changed by a phase factor $\psi \rightarrow e^{i\chi} \psi$, vector potential changes as per $\mathbf{A} \rightarrow \mathbf{A} + \hbar c/q \partial \chi$; (3) the change in χ around a closed loop must be an integral multiple of 2π .)

In the Berry phase problem we looked at, the vector potential had q/c , the factor multiplying \mathbf{A} in the Hamiltonian, equal to unity, $g = \hbar/2$, and hence $n = 1$.

As another application of the Berry phase, let us return to the Hall effect. Laughlin proposed that the excited state (above the ground state), called the *quasihole*

state, be given by

$$u_{qh} = \prod_{i=1}^{vN} (z_i - z_0) u_v \quad (21.1.114)$$

Clearly this describes a situation where the wave function is modified in the vicinity of z_0 . We say it describes a quasihole centered at z_0 . Note that electrons avoid the point z_0 due to the extra zeros of the form $z - z_0$. This means the charge density near this point is below normal. If one integrates the charge deficit due to this modification in the wave function (which is the charge of the quasihole) one finds it is vq , where q is the elementary charge e . Thus a theory with elementary charges that are integers (electrons) has excitations which have fractional charge! The fractional charge can also be demonstrated as follows. First note that the location z_0 of the quasihole is arbitrary. Assume there is some substrate potential underneath the electron gas whose minimum selects out some preferred location. Suppose we slowly vary the potential and drag the coordinate z_0 in u_{qh} around some closed loop and calculate the accumulated Berry phase for this closed orbit. (Since we know the wave function explicitly for any z_0 , this is easily done.) This must equal the flux (due to the external magnetic field B that produces the Landau levels) enclosed times $\bar{q}/\hbar c$ where \bar{q} is the quasihole charge. The calculation gives a charge v times the elementary charge. Similarly, one may show that the quasiholes are neither bosons nor fermions, but *anyons* (a term coined by Wilczek; see Bibliography): they acquire a phase factor $e^{iv\pi}$ under exchange, by taking a state with two quasiholes (located at z_0 and z'_0) and adiabatically exchanging them (i.e., their centers) and computing the Berry phase change in the wave function. The adiabatic analysis is valid since the quasihole states are separated by a gap from other states. For details, see Shapere and Wilczek (1990).

We conclude with some history.

Why did Born and Oppenheimer miss the Berry phase? The reason was quite subtle. They were working with a real Hamiltonian whose wave functions could be chosen real. They assumed such a choice had been made and that the choice was nonsingular. While this is correct for any open curve in parameter space, there exists the possibility that in closed curves, one could be forced to return to minus the starting wave function. Berry considered complex Hamiltonians (isomorphic to the spin example) which allowed a continuum of possible values for the phase (instead of just ± 1) and made the phenomenon more transparent.

Finally, although we have discussed the Berry phase in connection with quantum mechanics, it was discovered in optics many decades earlier by Pancharatnam (1958) who considered a polarized beam of light rather than a quantum state going on a closed path in parameter space (see Bibliography). For a fascinating review of even earlier precursors, see Berry's article in *Physics Today* (see Bibliography).

Coherent State Path Integral

Now we discuss yet another resolution of the identity and the associated path integral. These are based on *coherent states* defined to be eigenstates of the destruction operator in the harmonic oscillator problem.

Each coherent state carries a complex label z and is given by

$$|z\rangle = \exp[z\alpha^\dagger] |0\rangle \quad (21.1.115)$$

where $|0\rangle$ is the ground state of the oscillator. If we recall that

$$|n\rangle = \frac{(a^\dagger)^n}{\sqrt{n!}} |0\rangle \quad (21.1.116)$$

we see that

$$|z\rangle = \sum_0^{\infty} \frac{z^n}{\sqrt{n!}} |n\rangle \quad (21.1.117)$$

States labeled by different values of z are not orthonormal. We should have expected nonorthogonality since the basis $|n\rangle$ labeled by the positive integers n forms a complete basis and here we have one state for every complex number z ! So they couldn't all be orthogonal. It is also possible that despite their large number, they are not a complete set. We shall, however, see that they are an *overcomplete basis*, i.e., a basis with enough vectors to expand any vector but with more than the smallest number one could have gotten away with.

Now we will establish the key property

$$a|z\rangle = z|z\rangle \quad (21.1.118)$$

as follows:

$$a|z\rangle = a \sum_0^{\infty} \frac{z^n}{\sqrt{n!}} |n\rangle \quad (21.1.119)$$

$$= \sum_1^{\infty} \frac{z^{n-1}}{\sqrt{(n-1)!}} |n-1\rangle \quad (21.1.120)$$

$$= z|z\rangle \quad (21.1.121)$$

where, in going to the last line, we have redefined a dummy label $n' = n - 1$ which runs from 0 to ∞ .

Likewise, by taking the adjoint of Eq. (21.1.118), the coherent state bra

$$\langle z| = \langle 0| \exp[z^* a] \quad (21.1.122)$$

is seen to obey

$$\langle z| a^\dagger = \langle z| z^* \quad (21.1.123)$$

Let us now consider the inner product

$$\langle z_2|z_1\rangle = \langle 0| \exp[z_2^* a] \exp[z_1 a^\dagger] |0\rangle \quad (21.1.124)$$

If we use the identity

$$e^A e^B = e^B e^A e^{[A,B]} \quad (21.1.125)$$

which is valid if $[A, B]$ commutes with A and B , we see

$$\langle z_2 | z_1 \rangle = e^{z_2^* z_2} \quad (21.1.126)$$

upon noting that when the exponentials are exchanged and expanded out, only the first term with no a 's acting to the right or a^\dagger 's acting to the left survives.

Completeness is shown by proving the following resolution of the identity

$$I = \int \frac{dx dy}{\pi} |z\rangle \langle z| e^{-z^* z} \equiv \int \frac{dz dz^*}{2\pi i} |z\rangle \langle z| e^{-z_2^* z_1} \quad (21.1.127)$$

where $z = x + iy$ and $z^* = x - iy$. Note that the integral is over the entire $x - y$ plane, and after replacing every z and z^* in the integrand by $x \pm iy$, may be carried out using any other coordinates. For example, in Exercise (21.1.16) polar coordinates are recommended in verifying the above completeness relation. One can also formally go from (x, y) to (z, z^*) (after inserting a Jacobian $1/2i$), but integration over (z, z^*) is a subtle question we will not get into. We indicate that measure in terms of (z, z^*) anyway (now and later) so you will know what it means if you ever run into it again.

To show Eq. (21.1.127), one uses

$$|z\rangle = \sum_0^\infty \frac{z^n}{\sqrt{n!}} |n\rangle \quad (21.1.128)$$

and its adjoint, does the $dx dy$ integral in polar coordinates, and recovers the usual sum over $|n\rangle \langle n|$.

Exercise 21.1.16. Verify the above resolution of the identity. Consult Appendix A2 for the Gamma function integral.

Since the coherent states are right eigenstates of a and left eigenstates of a^\dagger ,

$$\langle z_2 | : H(a^\dagger, a) : | z_1 \rangle = \langle z_2 | H(z_2^*, z_1) | z_1 \rangle \quad (21.1.129)$$

where: H : is any *normal ordered expression* i.e., an expression with all the destruction operators to the right and creation operators to the left. Thus, $a^\dagger a^2$ is a normal ordered expression while $a^2 a^\dagger$ is not. Given any expression we can always normal order it by pushing the a 's to the right, keeping track of commutators.

Exercise 21.1.17. Show that $a^2 a^\dagger = :a^2 a^\dagger: + 2a$. (Push one power of a at a time to the right, or use $[AB, C] = A[B, C] + [A, C]B$.)

We now prove the following remarkable result: if H is the oscillator Hamiltonian,

$$H = \hbar\omega a^\dagger a \quad (21.1.130)$$

(we drop the constant zero-point energy for this discussion), then

$$U(t)|z\rangle = U(t) \exp[a^\dagger z] U^\dagger(t) U(t)|0\rangle = \exp[a^\dagger e^{-i\omega t} z]|0\rangle = |z e^{-i\omega t}\rangle \quad (21.1.131)$$

where we have used the Heisenberg equations of motion for a^\dagger . (In the Heisenberg picture $U^\dagger(t)\Omega U(t)=\Omega(t)$. Here $U^\dagger(t)=U(-t)$ appears in place of $U(t)$. We use the result $a^\dagger(t)=a^\dagger(0)e^{i\omega t}$ and reverse the sign of t .)

It is remarkable that *under time evolution the coherent state remains a coherent state, but with a new label*. This was one of the reasons one got interested in them in the first place. They have far too many interesting properties for us to discuss them all here. Instead you are directed to the reference on this subject.

Exercise 21.1.18. Show that the wave function of the coherent state is

$$\psi_z(x) = \langle x|z\rangle = \left(\frac{m\omega}{\pi\hbar}\right)^{1/4} e^{-z^2/2} e^{-(m\omega/2\hbar)x^2} e^{\sqrt{(2m\omega/\hbar)}zx} \quad (21.1.132)$$

Start by using $a|z\rangle=z|z\rangle$ in the coordinate representation. Fix the normalization by demanding that $\langle z'|z\rangle=e^{z^*z}$. Read off its mean momentum and position. Show that these evolve with time like classical coordinates given that $|z\rangle\rightarrow|z e^{-i\omega t}\rangle$. Suggestion: Look at Eq. (9.3.7) and parametrize z as $z=\sqrt{(m\omega/2\hbar)}x_0+i\sqrt{(1/2m\omega\hbar)}p_0$.

It is very easy to find the propagator for the oscillator in this basis:

$$U(z_N, z_0, t) = \langle z_N|U(t)|z_0\rangle = \langle z_N|z_0 e^{-i\omega t}\rangle = \exp[z_N^* z_0 e^{-i\omega t}] \quad (21.1.133)$$

where the subscripts on the end point anticipates the following discussion.

Consider the path integral representation for the propagator. Let us first imagine that there are just three intermediate time slices (so that $\varepsilon=t/4$) and three resolutions of the identity operator are used, giving us

$$\begin{aligned} & \langle z_4|U^4(t/4)|z_0\rangle \\ &= \int [\mathcal{D}z \mathcal{D}z^*] \langle z_4| \left(I - \frac{i\varepsilon}{\hbar} H(a^\dagger a)\right) |z_3\rangle e^{-z_3^* z_3} \langle z_3| \left(I - \frac{i\varepsilon}{\hbar} H(a^\dagger a)\right) |z_2\rangle e^{-z_2^* z_2} \langle z_2| \\ & \quad \times \left(I - \frac{i\varepsilon}{\hbar} H(a^\dagger a)\right) |z_1\rangle e^{-z_1^* z_1} \langle z_1| \left(I - \frac{i\varepsilon}{\hbar} H(a^\dagger a)\right) |z_0\rangle \end{aligned}$$

where

$$[\mathcal{D}z \mathcal{D}z^*] = \prod_{i=1}^{N-1} \frac{dz_i dz_i^*}{2\pi i} = \prod_{i=1}^{N-1} \frac{dx_i dy_i}{\pi} \quad (21.1.134)$$

A typical factor we run into is as follows:

$$\langle z_{n+1} | \left(I - \frac{i\varepsilon}{\hbar} H(a^\dagger a) \right) | z_n \rangle = \exp \left(- \frac{i\varepsilon}{\hbar} H(z_{n+1}^*, z_n) \right) \langle z_{n+1} | z_n \rangle \quad (21.1.135)$$

$$= \exp \left(- \frac{i\varepsilon}{\hbar} H(z_{n+1}^*, z_n) \right) \exp(z_{n+1}^* z_n) \quad (21.1.136)$$

where we have treated ε as infinitesimal since eventually it will be, as we let $N \rightarrow \infty$. If we assemble all the exponential factors together, there will be a piece related to the Hamiltonian which clearly gives a factor

$$\exp \left(- \frac{i}{\hbar} \int_0^t \hbar \omega z^*(t) z(t) dt \right) \quad (21.1.137)$$

in the continuum notation, where z_n has become $z(t=n\varepsilon)$. (We also made the approximation $H(z^*(t+\varepsilon), z(t)) \approx H(z^*(t), z(t))$.)

The other factor in the exponent is

$$z_4^* z_3 - z_3^* z_3 + z_3^* z_2 - z_2^* z_2 + z_2^* z_1 - z_1^* z_1 + z_1^* z_0 \quad (21.1.138)$$

$$= (z_4^* - z_3^*) z_3 + (z_3^* - z_2^*) z_2 + (z_2^* - z_1^*) z_1 + z_1^* z_0 \quad (21.1.139)$$

which we write in continuum notation as

$$\frac{i}{\hbar} \left[\int_0^t (-i\hbar) \frac{dz^*}{dt'} z dt' \right] + z^*(0) z(0) \quad (21.1.140)$$

where $z(0) = z_0$ and $z^*(0) = \lim_{\varepsilon \rightarrow 0} z^*(\varepsilon)$. In other words, in the discretized version z_0 was defined but not z_0^* . Only in the continuum picture, where we focus on smooth trajectories, is this object defined as the above limit.

The sum in Eq. (21.1.139) can also be rearranged to give

$$\left[\frac{i}{\hbar} \int_0^t (i\hbar) \left(z^* \frac{dz}{dt} \right) dt \right] + z^*(t) z(t) \quad (21.1.141)$$

where $z(t)$ is again extraneously introduced as a limit $z(t) = \lim_{\varepsilon \rightarrow 0} z(t-\varepsilon)$.

One usually sees the two schemes averaged to give the following final form of the continuum result:

$$\langle z_f | U(t) | z_i \rangle = \exp \left[\frac{z_f^* z_f + z_i^* z_i}{2} + \frac{i}{\hbar} \int_0^t \left[\frac{i\hbar}{2} \left(z^* \frac{dz}{dt} - \frac{dz^*}{dt} z \right) - H(z^*, z) \right] dt \right] \quad (21.1.142)$$

We will use the asymmetric form obtained by doing an integration by parts:

$$\langle z_f | U(t) | z_i \rangle = \exp \left[z_f^* z_f + \frac{i}{\hbar} \left[\int_0^t \left[i\hbar z^* \frac{dz}{dt} - H(z^*, z) \right] dt \right] \right] \quad (21.1.143)$$

The warning that this is just a schematic for the previous discretized expression is all the more true here since there is very little in the action to guarantee smooth paths. However, in the limit $\hbar \rightarrow 0$, the integral is asymptotically approximated by smooth paths. Let us evaluate this integral in such a limit by finding the stationary point of the action, i.e., the classical solution. It is clear from the action, which has the phase space form $(p\dot{x} - \mathcal{H})$ that z and $i\hbar z^*$ are canonically conjugate variables. Given this action, if one were asked to quantize, one would promote them to operators obeying commutation relations

$$[Z, i\hbar Z^\dagger] = i\hbar \quad (21.1.144)$$

which we see are just the commutation rules for a and a^\dagger . Of course, we are not trying to construct the quantum theory from the classical one, but the reverse. The Hamiltonian equation is

$$\dot{z} = \frac{\partial(\hbar\omega z^* z)}{\partial(i\hbar z^*)} = -i\omega z \quad (21.1.145)$$

which is solved to give

$$z(t) = z(0) e^{-i\omega t} \quad (21.1.146)$$

Similarly, we find

$$z^*(t) = z^*(0) e^{i\omega t} \quad (21.1.147)$$

To evaluate

$$\langle z_f | U(T) | z_i \rangle \quad (21.1.148)$$

in the semiclassical approximation, we need to find a solution that obeys

$$z(0) = z_i \quad (21.1.149)$$

$$z^*(T) = z_f^* \quad (21.1.150)$$

Now we see a problem that we did not have in the configuration space version: since the equations here are first order in time, z_i determines $z(t)$ for all times. How can we get $z^*(T)$ to equal an independently given z_f^* ? The answer is that we must regard

z and z^* as independent and restrict $z(t)$ at $t=0$ and $z^*(t)$ at $t=T$. The solutions then are

$$z(t) = z_i e^{-i\omega t} \quad (21.1.151)$$

$$z^*(t) = z_f^* e^{i\omega(t-T)} \quad (21.1.152)$$

Note that $z^*(T)$ is not the complex conjugate of $z(T)$. This means that x and y invoked in the definition $z=x+iy$ are not real on this trajectory. However, a Gaussian integral is given by its saddle point even if the point is off the original axis of integration. This point is explained in Faddeev's lectures (see Bibliography).

If we feed this solution into the action we find that the t -integral gives zero due to a cancellation between the two terms in the integrand and the only piece that survives is

$$z^*(T)z(T) = z_f^* z_i e^{-i\omega T}$$

giving us

$$\langle z_f | U(T) | z_i \rangle = \exp(z_f^* z_i e^{-i\omega T}) \quad (21.1.153)$$

which is the exact answer!

Exercise 21.1.19. Evaluate the action for the above path and check the answer given.

Exercise 21.1.20. Consider the Gaussian integrals in Eqs. (A.2.4–A.2.5.) Show that if we want just the exponential dependence of the answer, it is given by finding the exponential where the exponent is stationary. This is a general feature of Gaussian integrals.

Exercise 21.1.21. A good take-home problem. Rederive the oscillator propagator $\langle x_2 | U(T) | x_1 \rangle$ given $\langle z_f | U(T) | z_i \rangle = \exp[z_f^* z_i e^{-i\omega T}]$. Introduce two resolutions of the identity on either side of $U(T)$ in $\langle x_2 | U(T) | x_1 \rangle$. Use the suitably normalized wave functions $\langle x | z \rangle$ from Exercise (21.1.18). You will have to do a Gaussian integral over the two pairs of intermediate coherent state variables. Do the integral by saddle point, i.e., find the stationary point of the action and evaluate the integrand there. Focus on just the exponential factor and show that you get the answer to Exercise (8.6.2).

21.2. Imaginary Time Formalism

Consider the *imaginary time propagator*

$$U(\tau) = \exp\left(-\frac{1}{\hbar} H\tau\right) \quad (21.2.1)$$

This is obtained by setting

$$t = -i\tau \quad (21.2.2)$$

in the usual propagator. In other words, if the Schrödinger equation had been

$$-\hbar \frac{d}{d\tau} |\psi(t)\rangle = H |\psi(\tau)\rangle \quad (21.2.3)$$

this would have been the propagator.

The reasons for looking at this operator will be clear as we go along. But first let us note that we can write down the formula for it at once:

$$U(\tau) = \sum |n\rangle \langle n| \exp\left(-\frac{1}{\hbar} E_n \tau\right) \quad (21.2.4)$$

where

$$H|n\rangle = E_n |n\rangle \quad (21.2.5)$$

The main point to note is that even *though the time is now imaginary, the eigenvalues and eigenfunctions that enter into the formula for $U(\tau)$ are the usual ones*. Conversely, if we knew $U(\tau)$, we could extract the former.

Path Integral for the Imaginary Time Propagator

Consider the matrix element

$$\langle x | U(\tau) | x' \rangle \quad (21.2.6)$$

We can write down a path integral for it following exactly the same steps as before. The final answer in continuum notation is

$$\langle x | U(\tau) | x' \rangle = U(x, x', \tau) = \int [\mathcal{D}x] \exp\left[-\frac{1}{\hbar} \int_0^\tau \mathcal{L}_E(x, \dot{x}) d\tau\right] \quad (21.2.7)$$

$$\int [\mathcal{D}x] = \lim_{N \rightarrow \infty} \left(\frac{m}{2\pi\hbar\varepsilon}\right)^{1/2} \prod_0^{N-1} \left(\frac{m}{2\pi\hbar\varepsilon}\right)^{1/2} dx_i \quad (21.2.8)$$

$$\mathcal{L}_E = \frac{m}{2} \left(\frac{dx}{d\tau}\right)^2 + V(x) \quad (21.2.9)$$

where $\varepsilon = \tau/N$ and \mathcal{L}_E is called the *euclidean Lagrangian*. The adjective “euclidean” means that space and time now behave alike—the minus signs of Minkowski space in the formula for invariants are gone. For example, the invariant $x^2 - c^2 t^2$ now becomes $x^2 + c^2 \tau^2$. Notice that \mathcal{L}_E is the *sum* of the euclidean kinetic energy and real-time potential energy. *Thus the particle obeying the euclidean equations of motion will see the potential turned upside down.* This will be exploited later.

We have emphasized that the continuum form of the path integral is a shorthand for the discrete version. It is true here also, but of all the path integrals, this is the best behaved. Rapidly varying paths are suppressed by the falling (rather than rapidly oscillating) exponential factor.

Suppose we want to calculate the euclidean path integral for a free particle. We can proceed as we did in Chapter 8 and obtain

$$\langle x| U(\tau)|x'\rangle = \left(\frac{m}{2\pi\hbar\tau} \right)^{1/2} \exp \left[-\frac{m(x-x')^2}{2\hbar\tau} \right] \quad (21.2.10)$$

If someone gave us this propagator, we could get the Minkowski space answer by setting

$$\tau = it \quad (21.2.11)$$

This is called *analytic continuation*.

A very important feature of euclidean quantum mechanics is that the operator $U(\tau)$ is not unitary but Hermitian. Thus the norm of the state is not preserved in time. In fact what happens is that after a long time every state evolves into the ground state $|0\rangle$:

$$\lim_{\tau \rightarrow \infty} \langle x| U(\tau)|x'\rangle = \lim_{\tau \rightarrow \infty} \sum \langle x|n\rangle \langle n|x'\rangle \exp \left(-\frac{1}{\hbar} E_n \tau \right) \quad (21.2.12)$$

$$\simeq \langle x|0\rangle \langle 0|x'\rangle \exp \left(-\frac{1}{\hbar} E_0 \tau \right) \quad (21.2.13)$$

$$= \psi_0(x) \psi_0^*(x') \exp \left(-\frac{1}{\hbar} E_0 \tau \right) \quad (21.2.14)$$

Thus all states lead to the ground state as long as the starting point has some overlap with it. This is one way to find the ground state in any problem: take any initial state and let it evolve for a long time. You should hit the ground state unless you had chosen an initial state orthogonal to the ground state. (Sometimes you may do this on purpose to find the first excited state. For instance if the problem has parity invariance and you choose an initial state odd under parity, you will hit an excited state.)

For example, the propagator for the oscillator is

$$U(x, x', \tau) = A(\tau) \exp \left(-\frac{m\omega}{2\hbar \sinh \omega \tau} [(x^2 + x'^2) \cosh \omega \tau - 2xx'] \right) \quad (21.2.15)$$

obtained, say by analytic continuation from real times of the answer in Exercise (8.6.2). Note that as $\tau \rightarrow \infty$ this becomes proportional to the product of ground state wave functions. The prefactor is left to the following exercise.

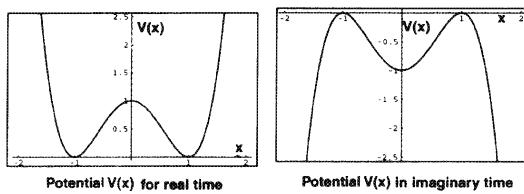


Figure 21.2. The double-well potential in real and imaginary time.

Exercise 21.2.1. Obtain $A(t)$ from Exercise (8.6.3) and continue to imaginary time, and verify that in the large τ limit, it yields the right prefactor.

Tunneling by Path Integrals: Well, well!

We now consider one application of the euclidean formalism. We have seen how one can derive the WKB wave function for nonbound states by using path integrals. This procedure does not work for tunneling amplitudes across barriers since we cannot find a classical path that goes over the barrier. On the other hand, in the euclidean dynamics the potential is turned upside down and what is forbidden in Minkowski space is suddenly allowed in the euclidean region!

Here is a problem that illustrates this point and many more. Consider a particle in a double-well potential

$$V(x) = A^2(x^2 - a^2)^2 \quad (21.2.16)$$

The classical minima are at

$$x_{L/R} = \pm a \quad (21.2.17)$$

Figure 21.2 shows a graph in Minkowski and euclidean space for the case $a = 1$.

Notice that in the euclidean problem the double-well has been inverted into the double-hill.

What is the ground state of the system? The classical ground state is doubly-degenerate: the particle can be sitting at either of the two minima. In the semiclassical approximation, we can broaden these out to Gaussians that are ground states $|\pm a\rangle$ in the harmonic oscillatorlike potential around each minimum at $x = \pm a$. This will shift each degenerate ground state by $\frac{1}{2}\hbar\omega$ where ω measures the curvature of the potential near the minimum. We can go to higher-order approximations that recognize that the bottom of the well is not exactly quadratic and shift the ground state energies by higher powers of \hbar . However, none of this will split the degeneracy of the ground states since whatever we find at the left minimum we will find at the right by symmetry under reflection. Lifting of the degeneracy will happen only if we take into account tunneling between the two wells. So we study this problem in the following stripped-down version. First we drop all but the degenerate ground states $|\pm a\rangle$. (The Gaussians centered around the two minima are not quite orthogonal.)

$$H = \begin{bmatrix} E_0 & 0 \\ 0 & E_0 \end{bmatrix} \quad (21.2.18)$$

Let us shift our reference energy so that $E_0=0$.

Note that there are no off-diagonal matrix elements. If this were an exact result, it should mean that if a particle starts out in one well it will never be found at the other. But we know from the wave function approach that if it starts at one side, it can tunnel to the other. This means that there is effectively a nonzero matrix off-diagonal matrix element $H_{+-}=H_{-+}=\langle a|H|-a\rangle$ in this basis. The challenge is to find that element in the semiclassical approximation. Once we find it, it is evident that the energy levels will be split into

$$E = \pm H_{+-} \quad (21.2.19)$$

and the eigenstates will be $|S/A\rangle$, the sum and difference of $|\pm a\rangle$.

Consider

$$\langle a|U(\tau)|-a\rangle = \langle a| \exp\left(-\frac{1}{\hbar} H\tau\right) | -a \rangle \quad (21.2.20)$$

In this discussion of tunneling, $U(\tau)$ is the propagator from $-\tau/2$ to $\tau/2$ and not from 0 to τ . Note that the term linear in τ gives us the off-diagonal matrix element:

$$\langle a| \exp\left(-\frac{1}{\hbar} H\tau\right) | -a \rangle \simeq 0 - \frac{1}{\hbar} \tau \langle a|H|-a\rangle + \mathcal{O}\tau^2 \quad (21.2.21)$$

We shall calculate $\langle a| e^{-(1/\hbar)H\tau} | -a \rangle$ by the semiclassical approximation to the euclidean path integral and extract the approximate matrix element H_{+-} . *Once again, as in the real-time semiclassical approximation, we focus on just the exponential factor and ignore all prefactors.* In the semiclassical approximation,

$$\langle a| \exp\left(-\frac{1}{\hbar} H\tau\right) | -a \rangle \simeq \exp\left(-\frac{1}{\hbar} S_{\text{cl}}\right) \quad (21.2.22)$$

where S_{cl} is the euclidean action for the classical path connecting the left hill to the right. *The key point, of course, is that in the double-hill potential of euclidean mechanics the classical ground states are not separated by a barrier, so that there will be no problem finding a classical path going from one hill to the other.*

The euclidean equations of motion are the same as the real times ones, *except for the reversal of the potential*. Thus there will be a conserved energy E_e given by

$$E_e = \frac{m}{2} \left(\frac{dx}{d\tau} \right)^2 - V(x) \quad (21.2.23)$$

Using this we can solve for the trajectory by quadrature:

$$\int_{x_1}^{x_2} \frac{\sqrt{m} dx}{\sqrt{2(E_e + V(x))}} = \int_{t_1}^{t_2} d\tau \quad (21.2.24)$$

Now we want the tunneling from the state $| -a \rangle$ to the state $| a \rangle$. These are not eigenstates of position, but Gaussians centered at $x = \mp a$. We shall however calculate the amplitude to tunnel from the *position eigenstate* $x = -a$ to the position eigenstate $x = a$. Except for the overlaps $\langle x = a | a \rangle$ and $\langle -a | x = -a \rangle$ this is the same as $\langle a | U | -a \rangle$. These overlaps know nothing about the tunneling barrier. They will constitute undetermined prefactors in front of the exponential dependence on the barrier which alone we are after. To extract the extreme low energy physics we must let $\tau \rightarrow \infty$. To this end, let us consider the trajectory that has $E_e = 0$. It is given by doing the above integral with $E_e = 0$:

$$x(\tau) = a \tanh \left[\sqrt{\frac{2}{m}} A \tau \right] \quad (21.2.25)$$

Notice that in this trajectory the particle starts out at the left *maximum* (Fig. 21.2) at $\tau \rightarrow -\infty$ and rolls down the hill and only reaching of the right maximum as $\tau \rightarrow \infty$. If the starting point and ending point are exactly $x = \mp a$, tunneling takes infinite time since only in this limit does the tanh take its limiting value of $\pm a$. Physically, it takes forever since the particle must start from rest at the left end to have zero euclidean energy. On the other hand, for points which are slightly below the maximum at each end, the time of travel will be finite since the particle can start with nonzero velocity. Since these points will also have roughly the same overlap with the states $| \pm a \rangle$ we can start with them instead of $x = \pm a$ in which case the tunneling will take place in finite time. This will be understood in what follows.

The action for the above solution is (using $T = V$ for the zero energy solution),

$$S_{\text{cl}} = \int (T + V) d\tau = \int 2T d\tau = \int_{-a}^a p(x) dx = \int_{-a}^a \sqrt{2mV(x)} dx \quad (21.2.26)$$

and the tunneling amplitude is (ignoring prefactors)

$$\langle a | U | -a \rangle \simeq \exp \left(-\frac{1}{\hbar} \int_{-a}^a \sqrt{2mV(x)} dx \right) \quad (21.2.27)$$

in agreement with tunneling result in the Schrödinger approach, Eq. (16.2.24) with $E = 0$.

Now, onward to extract the matrix element by looking for the term linear in τ in the answer. But we see no such explicit τ dependence in the answer! The resolution can be stated in two ways.

- The first is tied to the fact that in the limit of large τ , the problem becomes translationally invariant in time. In other words, if we stare at the classical solution above, we see that the tanh is close to $\pm a$ most of the time and jumps rapidly from $-a$ to a in a short time centered around $\tau=0$. Pictorially, the particle takes a long time to roll off the top, but once it gets going, it rolls down very quickly to a point close to the other end point. (For this reason this solution is called an *instanton*, a term coined by 't Hooft: except for the brief “instant” when tunneling takes place, the system is essentially in one of its classical ground states.) If we draw a new trajectory in which the same tunneling takes place in the same time interval, but is centered around a time $\tau=\tau_0\neq 0$, this too will be close to being a minimum of the action. (It will have exactly the same action as $\tau\rightarrow\infty$.) In other words, the solution we found has many companions, all of nearly the same action, but different tunneling instants τ_0 . We must sum over all these paths, i.e., integrate over the instant of tunneling τ_0 . Since they all have nearly the same action, the effect is to multiply the answer by τ since τ_0 is forced to lie within the period $-\tau/2 < \tau_0 < \tau/2$.
- The second way to argue is that once we find one classical path, we must integrate the functional over all fluctuations $\delta x(\tau) = x(\tau) - x_{\text{cl}}(\tau)$. (See Section 8.6.) If we expand the action near x_{cl} , there will be no linear term since the action is stationary here and we will start with a quadratic expression in $\delta x(\tau)$. By diagonalizing this quadratic form we can get the answer as a product of Gaussian integrals. Consider the one-dimensional example of some function approximated by a Gaussian centered at $x=0$:

$$I(a) = \int_{x_1}^{x_2} e^{-ax^2} dx \quad (21.2.28)$$

If $a > 0$ we can assume the limits can be pushed to infinity and the answer approximated by

$$I(a) \simeq \sqrt{\pi/a} \quad (21.2.29)$$

What happens when $a \rightarrow 0$? The approximate answer diverges but we know the real answer is

$$I = \lim_{a \rightarrow 0} \int_{x_1}^{x_2} e^{-ax^2} dx = x_2 - x_1 \quad (21.2.30)$$

This is essentially what happens in the functional integral. Say $x(\tau)$ is a classical solution. Then $x(\tau - \tau_0)$ is also a solution, and

$$\delta x(\tau) = x(\tau - \tau_0) - x(\tau) \quad (21.2.31)$$

is a fluctuation that costs no extra action, i.e., the Gaussian that is supposed to damp out this fluctuation has $\alpha \rightarrow 0$. The Gaussian integral is then replaced by the range of integration corresponding to this degree of freedom, which is just $\int d\tau_0 \simeq \tau$.

So we have argued for a prefactor of τ which came from considering a fluctuation about the classical solution. We were forced to consider it since it reflected an exact symmetry (under time-translation) as result of which it had no α in the Gaussian to cut it off. We do, however, ignore the Gaussian integrals over the rest of the fluctuations since they cut off by nonzero α s.

With the prefactor τ in front of

$$\langle -a | U(\tau) | a \rangle = \tau \exp\left(-\frac{1}{\hbar} S_{\text{cl}}\right) \quad (21.2.32)$$

we are ready to compare to

$$\langle a | \exp\left(-\frac{1}{\hbar} H\tau\right) | -a \rangle \simeq 0 - \frac{1}{\hbar} \tau \langle a | H | -a \rangle + \mathcal{O}r^2 \quad (21.2.33)$$

and read off

$$H_{-+} \simeq -\exp\left(-\frac{1}{\hbar} S_{\text{cl}}\right) \quad (21.2.34)$$

where once again we have dropped all prefactors except for the sign which is important. (All euclidean transition amplitudes are positive since the functional is positive. The minus sign comes from $e^{-(1/\hbar)H\tau}$.)

It is now clear that with H_{-+} negative, the new eigenstates and energies are as follows:

$$|S\rangle = \sqrt{\frac{1}{2}} [|+a\rangle + |-a\rangle] \quad E_S = -\exp\left(-\frac{1}{\hbar} S_{\text{cl}}\right) \quad (21.2.35)$$

$$|A\rangle = \sqrt{\frac{1}{2}} [|+a\rangle - |-a\rangle] \quad E_A = \exp\left(-\frac{1}{\hbar} S_{\text{cl}}\right) \quad (21.2.36)$$

Spontaneous Symmetry Breaking

Why are we interested in a term that vanishes exponentially fast as $\hbar \rightarrow 0$ when we ignored all the perturbative corrections to the states $|\pm a\rangle$ which vanished as finite powers of \hbar ? The reason is that the exponentially small term is the leading term in the *splitting* of the two classically degenerate ground states.

But there is another very significant implication of the tunneling calculation. This has to do with the phenomenon of *spontaneous symmetry breaking* which will now be described.

Consider a Hamiltonian which has a symmetry, say under parity. *If the lowest energy state of the problem is itself not invariant under the symmetry, we say symmetry is spontaneously broken.*

Spontaneous symmetry breaking occurs quite readily in classical mechanics. Consider the single-well oscillator. The Hamiltonian is invariant under parity. The ground state is a particle sitting at the bottom of the well. This state respects the symmetry: the effect of parity on this state gives back the state. Now consider the double-well with minima at $x = \pm a$. There are two lowest energy configurations available to the particle: sitting still at the bottom of either well. No matter which choice it makes, it breaks the symmetry. The breakdown is spontaneous in that there was nothing in the Hamiltonian that tilted the scales. Once the particle has made a choice (based on accidents of initial conditions) the other option does not enter its dynamics. Let us note the twin signatures of symmetry breaking: there is more than one ground state, and these states are not invariant under the symmetry (some observable, not invariant under the symmetry has a nonzero value), but instead get mapped into each other by the symmetry operation.

Now consider the quantum case of the double well, but with an infinite barrier between the wells. (I mean a barrier across which tunneling is impossible either in the path integrals or wave function approach. So a delta function spike is not such a barrier.) Once again the particle has two choices, these being Gaussian-like functions centered at the two troughs: $| \pm a \rangle$. They show the twin features of symmetry breaking: they are degenerate and noninvariant under parity ($\langle X \rangle \neq 0$). But here is a twist. In quantum theory a particle can be in two places at the same time. In particular, we can form the combinations of these degenerate eigenvectors

$$|S/A\rangle = \frac{|+a\rangle \pm |-a\rangle}{\sqrt{2}} \quad (21.2.37)$$

$$\Pi |S/A\rangle = \pm |S/A\rangle \quad (21.2.38)$$

which are eigenstates of parity. Indeed, in quantum theory the relation

$$[\Pi, H] = 0 \quad (21.2.39)$$

guarantees that such parity eigenstates *can* be formed. But *should* they be formed? The answer is negative in this problem due to the infinite barrier. The reason is this. Suppose the particle in question is sighted in one side during a measurement. Then there is no way for its wave function to develop any support in the other side. (One says the motion is not ergodic.) Even in quantum theory, where energy can be nonconserved over small times, barrier penetration is forbidden if the barrier is infinite. This means in particular that the symmetric and antisymmetric functions will never be realized by any particle that has ever been seen on either side. The correct thing to do then is to build a Hilbert space of functions with support on just one side. That every state so built has a degenerate partner in the inaccessible well across the barrier, is academic. The particle will not even know a parallel universe just like its

own exists. Real life will not be symmetric in such a problem and the symmetric and antisymmetric wave functions (with zero $\langle X \rangle$) represent unrealizable situations. Symmetry is spontaneously broken.

Now for the more typical problem with a finite barrier. In this case, a particle once seen in the left side can later be seen in the right side and vice versa. Symmetric and antisymmetric wave functions are physically sensible and we can choose energy eigenstates which are also parity eigenstates. These states will no longer be degenerate. In normal problems, the symmetric state, or more generally the state with eigenvalue unity for the symmetry operation, the one invariant under the symmetry operation, will be the unique ground state. Recall that in the oscillator problem the ground state not only had definite parity, it was invariant under parity. Likewise, in the hydrogen atom, the ground state not only had definite angular momentum, the angular momentum was zero and was invariant under rotations. However, in both these problems there was no multiplicity of classical ground states and no real chance of symmetry breakdown. (The oscillator had just one classical ground state at the bottom of the well, and the hydrogen atom had one infinitely deep within the Coulomb well.) What the instanton calculation tells us is that the double well, despite having two classical ground states that break symmetry, has, in the quantum theory, a unique, symmetric, ground state.

Thus, even though the tunneling calculation was very crude and approximate, it led to a very profound conclusion: the symmetry of the Hamiltonian is the symmetry of the ground state, symmetry breaking does not take place in the double-well problem.

This concept of symmetry restoration by tunneling (which in turn is tied to the existence of classical euclidean solutions with finite action going from one putative degenerate ground state to another) is very deep and plays a big role in many problems. There have been problems (quantum chromodynamics) where one did not even realize that the minimum one had assumed was unique for years was one of an infinite family of degenerate minima, till an instanton (of finite action) connecting the two classical minima was found and interpreted. We discuss a simpler example to illustrate the generality of the notion: a particle in a periodic potential $V(x) = 1 - \cos 2\pi x$. The minima are at $x = n$, where n is any integer. The symmetry of the problem is the discrete translation $x \rightarrow x + 1$. The approximate states, $|n\rangle$, which are Gaussians centered around the classical minima, break the symmetry and are converted to each other by T , the operator that translates $x \rightarrow x + 1$

$$T|n\rangle = |n+1\rangle \quad (21.2.40)$$

However, adjacent classical minima are connected by a nonzero tunneling amplitude of the type we just calculated and H has off-diagonal amplitudes between $|n\rangle$ and $|n \pm 1\rangle$. (There are also solutions describing tunneling to next-nearest-neighbor minima, but these have roughly double the action as the nearest-neighbor tunneling process and lead to an off-diagonal matrix element that is roughly the square of the one due to nearest-neighbor tunneling.) Suppose the one-dimensional world were finite and formed a closed ring of size N , so that there were N degenerate classical minima. These would evolve into N nondegenerate levels (the analogs of $|S/A\rangle$) due to the mixing due to tunneling. The ground state would be a symmetric

combination:

$$|S\rangle = \frac{1}{\sqrt{N}} \sum_1^N |n\rangle \quad (21.2.41)$$

The details are left to the following exercise.

Exercise 21.2.2. (Very important)

Assume that

$$H = \sum_1^N E_0 |n\rangle \langle n| - t(|n\rangle \langle n+1| + |n+1\rangle \langle n|) \quad (21.2.42)$$

describes the low-energy Hamiltonian of a particle in a periodic potential with minima at integers n . The integers n go from 1 to N since it is assumed the world is a ring of length N so that the $N+1$ th point is the first. Thus the problem has symmetry under translation by one site despite the finite length of the world. The first term in H represents the energy of the Gaussian state centered at $x=n$. The second represents the tunneling to adjacent minima with tunneling amplitude t . Consider the state

$$|\theta\rangle = \frac{1}{\sqrt{N}} \sum_1^N e^{in\theta} |n\rangle \quad (21.2.43)$$

Show that it is an eigenstate of T . Find the eigenvalue. Use the condition $T^N = I$ to restrict the allowed values of θ and make sure that we still have just N states. Show that $|\theta\rangle$ is an eigenstate of H with eigenvalue $E(\theta) = E_0 - 2t \cos \theta$. Consider $N=2$ and regain the double-well result. (You might have some trouble with a factor of 2 in front of the $\cos \theta$ term. Remember that in a ring with just two sites, each site is both ahead and behind the other and H couples them twice.)

Will the ground state always be invariant under the symmetry operation that commutes with H ? The answer is yes, as long as the barrier height is finite, or more precisely, as long as there is a finite action solution to the euclidean equations of motion linking classical minima. This is usually the case for quantum mechanics of finite number of degrees of freedom with finite parameters in the Hamiltonian. On the other hand, if $V_0 \rightarrow \infty$ in the periodic potential, there really will be N degenerate minima with particles living in any one minimum trapped there forever. In quantum field theory, where there are infinitely many degrees of freedom, even if the parameters are finite, the barrier is often infinitely high if all degrees of freedom try to jump over a barrier. In other words, symmetry breaking can take place.

For a more complete discussion of the tunneling question, you must consult the Bibliography, especially the works by Coleman and Rajaraman. These references will also answer other questions you might have such as: What about solutions where the particle rattles back and forth between the two hilltops in the inverted double-well potential? (These give contributions where the prefactors go as higher powers of τ .) Is there a way to read off the splitting between $|S/A\rangle$ directly from $\langle a|U(\tau)|-a\rangle$ without picking off the term linear in τ ? (Yes, by summing over an infinite amount of rattling back and forth.) You will find many interesting points to

ponder, but our result will prove to be correct to leading order in the exponentially small quantity $e^{-(1/\hbar)S_0}$.

Imaginary Time Path Integrals and Quantum Statistical Mechanics

We now discuss two other reasons for studying imaginary time path integrals. The first concerns quantum statistical mechanics and the second classical statistical mechanics.

Consider the partition function for a quantum system:

$$Z = \sum_n e^{-\beta E_n} \quad (21.2.44)$$

where the temperature T and Boltzmann's constant k appear in the combination $\beta = 1/kT$ and where E_n is the energy of the n th eigenstate of the Hamiltonian H . We can rewrite this as

$$Z = \text{Tr } e^{-\beta H} \quad (21.2.45)$$

where the trace is taken in the eigenbasis of H . Now we exploit the fact that the trace is invariant under a unitary change of basis and switch to the x -basis to obtain

$$Z = \int_{-\infty}^{\infty} \langle x | e^{-\beta H} | x \rangle dx \quad (21.2.46)$$

The integrand is of course familiar to us now:

$$\langle x | \exp(-\beta H) | x \rangle = \langle x | \exp\left(-\frac{1}{\hbar} \beta \hbar H\right) | x \rangle = U(x, x, \beta \hbar) \quad (21.2.47)$$

In other words, Z is the sum over amplitudes to go from the point x back to the point x in imaginary time $\tau = \beta \hbar$, in other words, over closed paths.

Exercise 21.2.3. Starting with $U(x, x, \tau)$ for the oscillator (see Eq. (21.2.15) and Exercise (21.2.1)) do the integral over x to obtain Z . Compare this to the sum

$$Z = \sum_0^\infty e^{-\beta \hbar \omega (n + 1/2)} \quad (21.2.48)$$

This connection between quantum statistical mechanics and imaginary time quantum mechanics is the starting point for a whole industry. Some applications are discussed in the book by Feynman and Hibbs. It would take us too far astray to get into any of these in depth. I will merely show how we take the classical limit of this formula. Consider a single particle of mass m in a potential $V(x)$. Then

$$Z(\beta) = \int dx \int_x^x [\mathcal{D}x] \exp\left[-\frac{1}{\hbar} \int_0^{\beta \hbar} \left[\frac{m}{2} \left(\frac{dx}{d\tau}\right)^2 + V(x(\tau))\right] d\tau\right] \quad (21.2.49)$$

where the limits on the functional integral remind us to consider paths starting and ending at the same point x , which is then integrated over, via the ordinary integral. Consider the limit $\beta\hbar \rightarrow 0$ either due to high temperatures or vanishing \hbar (the classical limit). Look at any one value of x . We need to sum over paths that start at x , go somewhere and come back to x in a very short time $\beta\hbar$. If the particle wanders off a distance Δx , the typical kinetic energy is $m(\Delta x/\beta\hbar)^2$ and the suppression factor is

$$\simeq \exp\left(-\frac{1}{\hbar} m(\Delta x/\beta\hbar)^2 \beta\hbar\right) \quad (21.2.50)$$

from which it follows that

$$\Delta x \simeq \sqrt{\frac{\beta}{m}} \hbar \quad (21.2.51)$$

If the potential does not vary over such a length scale [called the *thermal wavelength*, see Exercise (21.2.4)] we can approximate it by a constant equal to its value at the starting point x and write

$$\begin{aligned} Z(\beta) &\simeq \int dx e^{-\beta V(x)} \int_x^x [\mathcal{D}x] \exp\left[-\frac{1}{\hbar} \int_0^{\beta\hbar} \left[\frac{m}{2} \left(\frac{dx}{d\tau}\right)^2\right] d\tau\right] \\ &= \int dx e^{-\beta V(x)} \sqrt{\frac{m}{2\pi\hbar\beta\hbar}} \end{aligned} \quad (21.2.52)$$

where in the last step we have used the fact that with $V(x)$ pulled out, the functional integral is just the amplitude for a free particle to go from x to x in time $\beta\hbar$. How does this compare with classical statistical mechanics? There the sum over states is replaced by an integral over phase space:

$$Z = A \int dx \int dp \exp\left[-\beta \left(\frac{p^2}{2m} + V(x)\right)\right] \quad (21.2.53)$$

where the arbitrary prefactor A reflects one's freedom to multiply Z by a constant without changing anything physical since Z is a sum over relative probabilities and any prefactor will drop out in any averaging process. Equivalently it corresponds to the fact that the number of classical states in a region $dx dp$ of phase space is not uniquely defined. If we do the p integral and compare to the classical limit of the path integral we see that quantum theory fixes

$$A = \frac{1}{2\pi\hbar} \quad (21.2.54)$$

in accordance with the uncertainty principle which associates an area of order $\Delta X \Delta P \simeq \hbar$ in phase space with each quantum state.

Exercise 21.2.4. Consider a particle at temperature T , with mean energy of order kT . Assuming all the energy is kinetic, estimate its momentum and convert to the de Broglie wavelength. Show that this gives us a number of the order of the thermal wavelength. This is the minimum size over which the particle can be localized.

Relation to Classical Statistical Mechanics

So far we have discussed the relation of the imaginary time path integral to quantum statistical mechanics. Now we consider its relation to classical statistical mechanics. Consider a classical system with $N+1$ sites and a degree of freedom x_n at each site. The variables at the end of the chain, called x_0 and x_N , are fixed. Then

$$Z = \int_{-\infty}^{\infty} \prod_{i=1}^{N-1} dx_i \exp\left(-\frac{1}{kT} E(x_0, \dots, x_N)\right) \quad (21.2.55)$$

where E is the energy function and we have written β in terms of the more familiar temperature variable as $\beta = (1/kT)$. Let E have the form

$$E = \sum_{i=1}^{N-1} [K_1(x_n - x_{n-1})^2 + K_2 x_n^2] \quad (21.2.56)$$

where the first term represents the springlike coupling between nearest neighbors that forces them to maintain a fixed separation and the second one provides a quadratic potential that discourages each x from wandering off its neutral position $x=0$. If we compare this to the discretized imaginary time Feynman path integral for the quantum oscillator

$$U(x_0, x_N, \tau) = \int_{-\infty}^{\infty} \prod_{i=1}^{N-1} dx_i \exp\left[-\frac{1}{\hbar} \sum_{i=1}^{N-1} \varepsilon \left(\frac{m}{2} \frac{(x_n - x_{n-1})^2}{\varepsilon^2} + \frac{m\omega^2}{2} x_n^2\right)\right] \quad (21.2.57)$$

we see the following correspondence:

- The Feynman path integral from x_0 to x_N is identical in form to a classical partition function of a system of $N+1$ coordinates x_n with the boundary condition that the first and last be fixed at x_0 and x_N . The variables x_n are interpreted as intermediate state labels of the quantum problem (in the repeated resolution of the identity) and as the classical variables summed over in the partition function.
- The role of the action in the Feynman integral is played by the energy in the partition function.
- The role of \hbar is played by T . In particular, as either variable goes to zero, the sum over configurations is dominated by the minimum of action or energy and fluctuations are suppressed.
- The parameters in the classical and quantum problems can be mapped into each other. For example, $\beta K_1 = m/2\hbar\varepsilon$ and $\beta K_2 = m\omega^2\varepsilon/2\hbar$.
- Since $\varepsilon \rightarrow 0$ in the quantum problem, the parameters of the classical problem must take some limiting values ($K_1 \rightarrow \infty$ and $K_2 \rightarrow 0$ in a special way) to really be in correspondence with the quantum problem with $H = P^2/2m + m\omega^2x^2/2$.

- The single quantum degree of freedom is traded for a one dimensional array of classical degrees of freedom. This is a general feature: the dimensionality goes up by 1 as we go from the quantum to the classical problem. For example, a one-dimensional array of *quantum* oscillators would map on to the partition function of a two-dimensional array of classical variables. The latter array would be labeled by the time slice n as well as the quantum oscillator whose intermediate state label it stands for.

Our emphasis has been on the notion that the quantum oscillator problem can be written as a path integral which we now see is also a classical partition function. It is just as interesting to take a classical problem and translate it back to the operator version. In the classical problem we are interested in the free energy and thermal averages over the Boltzmann distribution, i.e., correlation functions like

$$\langle x_{12}x_{78} \rangle = \frac{\int_{-\infty}^{\infty} \prod_1^{N-1} dx_i x_{12}x_{78} e^{-\beta E(x_0, \dots, x_N)}}{\int_{-\infty}^{\infty} \prod_1^{N-1} dx_i e^{-\beta E(x_0, \dots, x_N)}} \quad (21.2.58)$$

where we use wedgy brackets to represent thermal averages as we did quantum averages, hoping you will be able to keep track of what is meant from the context. In the quantum theory we are interested in eigenstates of H , especially the ground state, Heisenberg operators, etc. We now develop the dictionary between the two approaches. Rather than use the oscillator, we turn to a problem with a simpler Hilbert space: that of a spin-1/2 problem.

For this purpose consider the *Ising model* in one dimension. The lattice now is an array of $N+1$ dots numbered 0 to N . At each point lies an *Ising spin* which can take only two values, $s = \pm 1$. The partition function is

$$Z = \sum_{s_i = \pm 1} \exp \left[\sum_{i=0}^{N-1} K(s_i s_{i+1} - 1) \right] \quad (21.2.59)$$

where K contains the factor $-\beta$. For the case we are interested in, $K > 0$, the Boltzmann weight is large when $s_i = s_{i+1}$ and small when $s_i = -s_{i+1}$. Thus the nearest-neighbor coupling represents the ferromagnetic tendency of the spins to be aligned with their neighbors. The additional, spin independent energy of minus $-K$ per site is a shift in energy made for convenience. Given this formula for Z , we can answer all thermodynamic questions. This is our classical problem. We will first solve for the free energy and correlation function viewing the problem classically. Then we will map this into a quantum problem and rederive the same results and our dictionary.

Let us first keep s_0 fixed at one value and define a relative variable:

$$t_i = s_i s_{i+1} \quad (21.2.60)$$

It is clear that given s_0 and t_i , we can reconstruct the state of the system. Thus, we can write

$$Z = \sum_{t_i} \exp \left[\sum_{i=0}^{N-1} K(t_i - 1) \right] = \prod_{t_i} e^{K(t_i - 1)} \quad (21.2.61)$$

Since the exponential factorizes into a product over i , we can do the sums over each t_i and obtain (after appending a factor of 2 for the two possible choices of s_0)

$$Z = 2(1 + e^{-2K})^N \quad (21.2.62)$$

One is generally interested in the free energy per site in the *thermodynamic limit* $N \rightarrow \infty$:

$$f(K) = \lim_{N \rightarrow \infty} \frac{1}{N} \ln Z \quad (21.2.63)$$

(This definition of f differs by a factor $-\beta$ from the more traditional one. I use the present one to reduce the clutter.) We see that

$$f(K) = \ln(1 + e^{-2K}) \quad (21.2.64)$$

where we have dropped $\ln 2/N$ in the thermodynamic limit. Had we chosen to fix s_0 at one of the two values, the factor 2 would have been missing in Eq. (21.2.62) but there would have been no difference in Eq. (21.2.64) for the free energy per site. Boundary conditions are unimportant in the thermodynamic limit in this sense.

Consider next the *correlation function* (which measure the likelihood that spins s_i and s_j are parallel):

$$\langle s_j s_i \rangle = \frac{\sum_{s_k} s_j s_i \exp \left[\sum_k K(s_k s_{k+1} - 1) \right]}{Z} \quad (21.2.65)$$

for $j > i$. Using the fact that $s_i^2 \equiv 1$, we can write

$$s_j s_i = s_i s_{i+1} s_{i+2} \dots s_{j-1} s_j = t_i t_{i+1} \dots t_{j-1} \quad (21.2.66)$$

Thus

$$\langle s_i s_j \rangle = \langle t_i \rangle \langle t_{i+1} \rangle \dots \langle t_{j-1} \rangle \quad (21.2.67)$$

where the answer factorizes over i since the Boltzmann weight factorizes over i when written in terms of t_i . The average for any one t is easy

$$\langle t \rangle = \frac{1}{2} \frac{e^{0 \cdot K} - 1}{e^{0 \cdot K} + e^{-2K}} = \tanh K \quad (21.2.68)$$

so that finally

$$\langle s_j s_i \rangle = (\tanh K)^{|j-i|} = \exp[(j-i) \ln \tanh K] \quad (21.2.69)$$

Note that the result depends on just the difference in coordinates. *This is not a generic result but a peculiarity of this model.* The reason is that the problem of $N+1$ points (for any finite N) is not translationally invariant. Correlations between two spins could, and generally do, depend on where the two points are in relation to the ends. On the other hand, in all models we expect that as $N \rightarrow \infty$, we will see translational invariance far from the ends and deep in the interior. To have translational invariance in a finite system, we must use periodic boundary conditions: now the world has the shape of a ring and every point is equivalent to every other. Correlation functions will now depend only on the difference between the two coordinates but they will not decay monotonically with separation! This is because as one point starts moving away from the other, it eventually starts approaching the first point from the other side! Thus the correlation function will be a sum of two terms, one of which grows as $|j-i|$ increases to values of order N . However, if we promise never to consider separations comparable to N , this complication can be ignored [see Exercise (21.2.9)]. (Our calculation of correlations in terms of t_i must be amended in the face of periodic boundary conditions to ensure that the sum over t_i is restricted to configurations for which the product of t_i 's over the ring equals unity.)

The *correlation length* ξ is defined by the formula

$$\lim_{|j-i| \rightarrow \infty} \langle s_i s_j \rangle \rightarrow e^{-(j-i)/\xi} \quad (21.2.70)$$

Thus in our problem

$$\xi^{-1} = -\ln \tanh K \quad (21.2.71)$$

(We have assumed $j > i$ in our analysis. In general $|j-i|$ is to be replaced by $|j-i|$ in these definitions. Also the model in question shows the exponential behavior for all separations and not just in the limit $|j-i| \rightarrow \infty$. This too is peculiar to our model and stems from the fact that the model is in one spatial dimension and the Ising spin can take only two values.)

We will now rederive these results in the quantum version. If Z stands for a path integral, the Ising variables must be the intermediate state labels that occur in the resolution of the identity for a quantum problem. Clearly the quantum problem is that of a spin-1/2.

To proceed, let us take another look at

$$Z = \sum_{s_i} \prod_i e^{K(s_i s_{i+1} - 1)} \quad (21.2.72)$$

Each exponential factor is labeled by two discrete indices which can take two values each. Furthermore, the second label for any factor is the first label for the next.

Finally, these labels are being summed over. It is clear that we are seeing here a matrix product. (We are simply undoing the resolution of the identity.) So we write

$$Z = \sum_{s_1} T_{s_N s_{N-1}} \cdots T_{s_2 s_1} T_{s_1 s_0} \quad (21.2.73)$$

where we have introduced a 2×2 matrix T whose rows and columns are labeled by a pair of spins and whose element $T_{ss'}$ equals the Boltzmann weight associated with a pair of neighboring spins in the state s, s' . Thus

$$T_{++} = T_{--} = 1, \quad T_{+-} = T_{-+} = \exp(-2K)$$

Thus this matrix, called the *Transfer Matrix*, is given by

$$T = I + e^{-2K} \sigma_1 \quad (21.2.74)$$

and

$$Z = \langle s_N | T^N | s_0 \rangle \quad (21.2.75)$$

for the case of fixed boundary conditions (which we will focus on) where the first spin is fixed at s_0 and the last at s_N . If we sum over the end spins (free boundary conditions)

$$Z = \sum_{s_0 s_N} \langle s_N | T^N | s_0 \rangle \quad (21.2.76)$$

If we consider periodic boundary conditions where $s_0 = s_N$ and one sums over these,

$$Z = \text{Tr } T^N \quad (21.2.77)$$

We will now show the insensitivity of the free energy per site to boundary conditions in the thermodynamic limit. Suppose we used fixed boundary conditions. Then if we write

$$T = \lambda_0 |0\rangle\langle 0| + \lambda_1 |1\rangle\langle 1| \quad (21.2.78)$$

where $|i\rangle, \lambda_i [i=0, 1]$ are the eigenvectors (assumed orthonormal) and eigenvalues of T , then

$$T^N = \lambda_0^N |0\rangle\langle 0| + \lambda_1^N |1\rangle\langle 1| \quad (21.2.79)$$

Assuming λ_0 is the bigger of the two eigenvalues,

$$T^N \lim_{N \rightarrow \infty} \simeq \lambda_0^N |0\rangle\langle 0| \left(1 + \mathcal{O} \left(\frac{\lambda_1}{\lambda_0} \right)^N \right) \quad (21.2.80)$$

and

$$Z \approx \langle s_N | 0 \rangle \langle 0 | s_0 \rangle \lambda_0^N \left(1 + \mathcal{O} \left(\frac{\lambda_1}{\lambda_0} \right)^N \right) \quad (21.2.81)$$

and the free energy per site in the infinite volume limit,

$$f = \lambda_0 + \frac{1}{N} \ln(\langle s_N | 0 \rangle \langle 0 | s_0 \rangle) + \dots \quad (21.2.82),$$

is clearly independent of the boundary spins as long as $\langle 0 | s_0 \rangle$ and $\langle s_N | 0 \rangle$ do not vanish.

Exercise 21.2.5. Check this claim for periodic boundary conditions starting with Eq. (21.2.75).

Let us rewrite T as follows. Consider the identity

$$e^{K^* \sigma_1} = \cosh K^* + \sinh K^* \sigma_1 \quad (21.2.83)$$

$$= \cosh K^* (I + \tanh K^* \sigma_1) \quad (21.2.84)$$

where K^* is presently unrelated to K ; in particular, it is not the conjugate! If we choose

$$\tanh K^* = e^{-2K} \quad (21.2.85)$$

we see from Eq. (21.2.74) that up to a prefactor $\cosh K^*$,

$$T = e^{K^* \sigma_1} \quad (21.2.86)$$

We will temporarily drop this prefactor but remember to subtract $\ln \cosh K^*$ from the free energy per site. It does not, however, affect the correlation function which will be seen to depend only on the ratios of eigenvalues of T . Note that K^* , called the *dual* of K , is large when K is small and vice versa.

For later reference, let us note that in the present case, the eigenvalues of T are $e^{\pm K^*}$ and the corresponding eigenvectors are

$$|0\rangle, |1\rangle = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ \pm 1 \end{bmatrix} \quad (21.2.87)$$

Suppose we write

$$T = e^{-H} \quad (21.2.88)$$

Then T can be interpreted as the time evolution operator for one time step in the imaginary time direction. The spatial site index i of the classical problem has become

the discrete imaginary time index for the quantum problem. The free energy is simply related to E_0 , the ground state energy of H :

$$H = -K^* \sigma_1 \quad (21.2.89)$$

$$f = -E_0 = K^* \quad (21.2.90)$$

Exercise 21.2.6. Show that f above agrees with Eq. (21.2.64) upon remembering to subtract $\ln \cosh K^*$ and using the definition of K^* .

Consider next the correlation function $\langle s_j s_i \rangle$ for $j > i$. I claim that if the boundary spins are fixed at s_0 and s_N ,

$$\langle s_j s_i \rangle = \frac{\langle s_N | T^{N-j} \sigma_3 T^{j-i} \sigma_3 T^i | s_0 \rangle}{\langle s_N | T^N | s_0 \rangle} \quad (21.2.91)$$

To see the correctness of this, look at the numerator. Retrace our derivation by introducing a complete set of σ_3 eigenstates between every factor of T . Reading from right to left, we get just the Boltzmann weights till we get to site i . There the σ_3 acting on its eigenstate, gives s_i , the value of the spin there. Then we proceed as usual to j , repeat this and go to the N th site. (The dependence of $\langle s_j s_i \rangle$ on the boundary conditions will be seen to disappear in the thermodynamic limit.) Let us rewrite Eq. (21.2.89) another way. Define *Heisenberg operators*

$$\sigma_3(n) = T^{-n} \sigma_3 T^n \quad (21.2.92)$$

In terms of these

$$\langle s_j s_i \rangle = \frac{\langle s_N | T^N \sigma_3(j) \sigma_3(i) | s_0 \rangle}{\langle s_N | T^N | s_0 \rangle} \quad (21.2.93)$$

Consider now the limit as $N \rightarrow \infty$, i and j fixed at values far from the end points labeled 0 and N so that $N-j$ and i are large, and we may approximate

$$T^\alpha \simeq |0\rangle \langle 0| \lambda_0^\alpha \quad \alpha = N, N-j, i \quad (21.2.94)$$

In this limit, we have from Eq. (21.2.91)

$$\langle s_j s_i \rangle = \frac{\langle s_N | 0 \rangle \langle 0 | \lambda_0^{N-j} \sigma_3 T^{j-i} \sigma_3 \lambda_0^i | 0 \rangle \langle 0 | s_0 \rangle}{\langle s_N | 0 \rangle \lambda_0^N \langle 0 | s_0 \rangle} = \langle 0 | \sigma_3(j) \sigma_3(i) | 0 \rangle \quad (21.2.95)$$

and the dependence on the boundary has dropped out. For the case $i > j$, we will get the operators in the other order. In general then,

$$\langle s_j s_i \rangle = \langle 0 | \mathcal{T}(\sigma_3(j) \sigma_3(i)) | 0 \rangle \quad (21.2.96)$$

where the *time-ordering symbol* \mathcal{T} will order the operators with time increasing from the right to left:

$$\mathcal{T}(\sigma_3(j)\sigma_3(i)) = \theta(j-i)(\sigma_3(j)\sigma_3(i)) + \theta(i-j)(\sigma_3(i)\sigma_3(j)) \quad (21.2.97)$$

We will pursue the evaluation of this correlation function using the eigenvectors of T . But first let us replace $\sigma_3(j)$ by the unit operator in the above derivation to obtain the mean magnetization as

$$\langle s_i \rangle = \langle 0 | \sigma_3(0) | 0 \rangle \quad (21.2.98)$$

In our example, $|0\rangle$ is the eigenket of σ_1 so that there is no mean magnetization. The only exception is at zero temperature or zero K^* : now the eigenvalues are equal and we can form linear combinations corresponding to either of the fully ordered (up or down) σ_3 eigenstates.

Let us compare symmetry breaking and its restoration in the Ising problem to what happened in the double well.

- In the limit $\hbar \rightarrow 0$, the particle in the double-well seeks the minimum of the euclidean action:

$$S_E = \int \left(\frac{m}{2} (dx/d\tau)^2 + V(x(\tau)) \right) d\tau \quad (21.2.99)$$

which is given by $(dx/d\tau) = 0$, $x = \pm a$, the minima of the double-well potential. There is degeneracy and symmetry breaking in the ground state. A particle that starts out in one well will not ever go to the other in the course of time. Even though Π commutes with H , we do not form parity eigenstates, instead we form eigenstates of position (or more accurately, well index, left or right). In the Ising problem, in the limit of zero temperature, the partition function is dominated by the state of minimum energy, with all spins up or all spins down on all sites (which can be viewed as discrete points in imaginary time of the spin-1/2 problem). In the operator language, T and H commute with σ_1 in general and eigenstates of H are chosen to be eigenstates of σ_1 as well. But at zero K^* , the two eigenstates become degenerate and we form combinations which are chosen to be eigenstates of σ_3 . This is because a state starting out up/down with respect to σ_3 will stay that way forever. (In classical statistical mechanics terms, if the spin at one of the chain is up/down, all will be up/down at zero temperature.)

- For nonzero \hbar , there is tunneling between the wells, degeneracy is lifted and symmetry is restored in the ground state. This is thanks to an instanton configuration that has finite action and connects the two classical ground states. In the Ising problem, for nonzero K^* , i.e., nonzero temperature, there exist instantonlike configurations in which the spin starts out up at one end of the chain (i.e., the distant past in the imaginary time interpretation) and at some point flips down and vice versa. This has finite energy (only one pair of nearest-neighbor spins is antiparallel and the additional energy cost is $2K$). The eigenstates of the transfer matrix (or the spin Hamiltonian) are now the symmetric and antisymmetric

combinations of the up and down states, i.e., eigenstates of σ_1 . The ground state is unique and symmetric.

Exercise 21.2.7. Consider the Hamiltonian of the spin-1/2 problem that arises in the transfer matrix treatment of the Ising chain

$$H = -K^* \sigma_1 \quad (21.2.100)$$

The off-diagonal matrix element (after pulling out the sign), i.e., K^* , must represent the tunneling amplitude (for going from up to down ground state) per unit time in the low-temperature limit (which you recall is like the $\hbar \rightarrow 0$ limit). The preceding discussion tells us it is just e^{-2K} where $2K$ is the energy cost of the interface of the up and down ground states. Verify that these two results agree for low temperatures by going back to the definition of K^* .

Let us return to Eq. (21.2.96). Even though it appears that everything depends on just the ground state, a knowledge of all states is required even in the infinite volume limit to evaluate the correlation. Going to Eq. (21.2.96) for the case $j > i$, let us insert the complete set of (two) eigenvectors of T between the Pauli matrices. When we insert $|0\rangle\langle 0|$ we get $\langle s \rangle^2$, the square of the magnetization which happens to vanish here. Moving it to the left-hand side, we get the *connected correlation function*

$$\langle s_j s_i \rangle_c \equiv \langle s_j s_i \rangle - \langle s \rangle^2 = \langle 0 | T^{-j} \sigma_3(0) T^{j-i} | 1 \rangle \langle 1 | \sigma_3(0) T^i | 0 \rangle \quad (21.2.101)$$

$$= \left(\frac{\lambda_1}{\lambda_0} \right)^{j-i} |\langle 0 | \sigma_3 | 1 \rangle|^2 \quad (21.2.102)$$

$$= e^{-2K^*(j-i)} |\langle 0 | \sigma_3 | 1 \rangle|^2 \quad (21.2.103)$$

Let us note that

- The correlation depends only on ratios of the eigenvalues of T and falls exponentially with distance with a coefficient $2K^*$. Now $2K^*$ is just the gap to the first excited state of the Hamiltonian H defined by $T = e^{-H}$ which in our example is $-K^* \sigma_1$. The result

$$\xi^{-1} = E_1 - E_0 \equiv m \quad (21.2.104)$$

is also very general. The reason one uses the symbol m for the gap (called the mass gap) is that in a field theory the lowest energy state above the vacuum is a single particle at rest and this has energy m (in units where $c=1$).

- The connected correlation function is determined by the matrix element of the operator in question (σ_3) between the ground state and the next excited state. This is also a general feature. If this matrix element vanishes, we must go up in the levels till we find a state that is connected to the ground state by the action of the operator. (In this problem we know $|\langle 0 | \sigma_3 | 1 \rangle|^2 = 1$ since σ_3 is the spin-flip operator for the eigenstates of σ_1 .)

This simple example has revealed most of the general features of the problem. The only difference is that for a bigger transfer matrix, the sum over states will have more than two terms. Thus the correlation function will be a sum of decaying exponentials and a unique correlation length will emerge only asymptotically when the smallest mass gap dominates. Also in the more complex problems (in higher dimensions) there may not be any finite action instantons connecting the multiple classical minima and there can be many ground states of H with broken symmetry. Assuming this happens, as it does in the two-dimensional Ising model (below some temperature T_c), you can ask: how does the ground state choose between spin up and spin down since there is no bias in the Boltzmann weight to make the choice? The answer is that indeed, if we do not set any bias, the system will always pick a mean magnetization of zero. How then do we know that the system is ready to magnetize? We use a principle called *clustering*. It states that as i and j separate, $\langle s_j s_i \rangle \rightarrow \langle s_j \rangle \langle s_i \rangle$. The idea is that if i lies in our galaxy and j lies in another they become statistically independent. Consider now the two-dimensional Ising model below T_c . In zero field we will find that $\langle s_j s_i \rangle$ does not approach $\langle s_i \rangle \langle s_j \rangle$ (which is zero since we gave the system no reason to choose one value of magnetization over its opposite) but that instead $\langle s_j s_i \rangle$ approaches the square of the magnetization the system will have if you would only give it the slightest reason for choosing one sign over the other. At this point, having seen the breakdown of clustering for the spin variable, you are to modify the partition function to restore clustering in two equivalent ways. One is to limit the sum over states to those with a net positive (or negative) magnetization. Then $\langle s \rangle \neq 0$ any more and you will find that $\langle s_i \rangle \langle s_j \rangle \rightarrow \langle s \rangle^2$. The other option is to apply a small field, calculate the magnetization, and let the field go to zero. (This too essentially kills half the states in the sum. Both recipes reflect the fact that a magnet below its T_c will not be able to dynamically evolve from pointing up to pointing down. Recall the particle trapped on one side of the infinite barrier between the two wells. Thus summing over things the system cannot do is a mistake.) Now, the magnetization is the derivative of the free energy with respect to the applied field h . It is easy to show that it is an even function of h . [See Exercise 21.2.8.] If the system does not want to magnetize, you will find that $f \sim h^2$, so that $df/dh \rightarrow 0$ as $h \rightarrow 0$. On the other hand if it wants to magnetize you will find $f \sim |h|$ and $df/dh \sim \text{sign } h$.

Exercise 21.2.8. Consider the Ising model in a magnetic field by adding a term $h \sum s_i$ to the exponent in Eq. (21.2.59). Show that $Z(h) = Z(-h)$. Show that the transfer matrix $T = e^{K^* \sigma_1} e^{h \sigma_3} \equiv T_K T_h$ reproduces the Boltzmann weight. Note that T is not Hermitian. By splitting the coupling to h into two factors, show that $T_h^{1/2} T_K T_h^{1/2}$ is just as good and also Hermitian. Find its eigenvalues and eigenvectors and show that there is degeneracy only for $h = K^* = 0$. Find the magnetization as a function of h by evaluating $\langle s \rangle = \langle 0 | \sigma_3 | 0 \rangle$. Starting with the partition function, show that

$$\langle s \rangle = \frac{1}{N} \frac{\partial \ln Z}{\partial h} = \frac{\partial f}{\partial h}$$

Evaluate f from the largest eigenvalue of T and regain the answer for $\langle s \rangle$ found from $\langle s \rangle = \langle 0 | \sigma_3 | 0 \rangle$.

Exercise 21.2.9. Consider the correlation function for the problem with periodic boundary conditions and write it as a ratio of two traces. Saturate the denominator with the largest

eigenket, but keep both eigenvectors in the numerator and show that the answer is invariant under $j-i \leftrightarrow N-(j-i)$. Using the fact that σ_3 exchanges $|0\rangle$ and $|1\rangle$ should speed things up. Provide the interpretation. Argue that as long as $j-i$ is much smaller than N , only one term is needed.

Exercise 21.2.10. Recall the remarkable fact that the correlation function $\langle s_i s_j \rangle$ in the Ising model was translationally invariant in the finite open chain with one end fixed at s_0 . Derive this result using the transfer matrix formalism as follows.

Explicitly evaluate $\sigma_3(j)$ by evaluating $T^{-j} \sigma_3 T^j$ in terms of σ_3 and σ_1 . Show that $\sigma_3(j) \sigma_3(i)$ is a function only of $j-i$ by using some identities for hyperbolic functions. Keep going till you explicitly have the correlation function. It might help to use $\sum_{s_N} |s_N\rangle = (I + \sigma_1)|s_0\rangle$.

21.3. Spin and Fermion Path Integrals

Now we turn to path integrals for two systems with no classical limit: a spin S system and a fermionic oscillator, to be described later. The fermion problem will be somewhat abstract at this stage, but it is in here because you are likely to see it in many different branches of physics.

Spin Coherent States and Path Integral

Consider a spin S degree of freedom. The Hilbert space is $2S+1$ dimensional. Choosing S_z eigenstates as our basis we can write the propagator $U(t)$ as a sum over configurations by using the resolution

$$I = \sum_{-S}^S |S_z\rangle \langle S_z| \quad (21.3.1)$$

The intermediate states will have discrete labels (as in the Ising model).

We consider here an alternate scheme in which an overcomplete basis is used. Consider the *spin coherent state*

$$|\Omega\rangle \equiv |\theta, \phi\rangle = U(R(\Omega))|SS\rangle \quad (21.3.2)$$

where $|\Omega\rangle$ denotes the state obtained by rotating the normalized, fully polarized state, $|SS\rangle$ by an angle θ around the x -axis and then by ϕ around the z -axis using the unitary rotation operator $U(R(\Omega))$.

Given that

$$\langle SS|S|SS\rangle = \mathbf{k}S \quad (21.3.3)$$

it is clear (say by considering $U^\dagger \mathbf{S} U$) that

$$\langle \Omega | \mathbf{S} | \Omega \rangle = S(\mathbf{i} \sin \theta \cos \phi + \mathbf{j} \sin \theta \sin \phi + \mathbf{k} \cos \theta) \quad (21.3.4)$$

Note that our spin operators are not defined with an \hbar . Thus for spin-1, the eigenvalues of S_z are 0, ± 1 .

Exercise 21.3.1. Show the above result by invoking Eq. (12.4.13).

The *coherent state* is one in which the spin operator has a nice expectation value: equal to a classical spin of length S pointing along the direction of Ω . It is not an eigenvector of the spin operator (not expected anyway since the three components of spin do not commute) and higher powers of the spin operators do not have expectation values equal to the corresponding powers of the classical spin. For example, $\langle \Omega | S_x^2 | \Omega \rangle \neq S^2 \sin^2 \theta \cos^2 \phi$. However, the difference between this wrong answer and the right one is of order S . Generally the n th power of the spin operator will have an expectation value equal to the n th power of the expectation value of that operator plus corrections that are of order S^{n-1} . If S is large, they may be ignored. This is so when one usually uses the present formalism.

Let us now examine the equation

$$\langle \Omega_2 | \Omega_1 \rangle = \left(\cos \frac{\theta_2}{2} \cos \frac{\theta_1}{2} + e^{i(\phi_1 - \phi_2)} \sin \frac{\theta_2}{2} \sin \frac{\theta_1}{2} \right)^{2S} \quad (21.3.5)$$

The result is obviously true for $S=1/2$, given that the up spinor along the direction $\theta\phi$ is

$$|\Omega\rangle \equiv |\theta\phi\rangle = \cos \frac{\theta}{2} |1/2, 1/2\rangle + e^{i\phi} \sin \frac{\theta}{2} |1/2, -1/2\rangle \quad (21.3.6)$$

As for higher spin, imagine $2S$ spin-1/2 particles joining to form a spin S state. There is only one direct product state with $S_z=S$: where all the spin-1/2's are pointing up. Thus the normalized fully polarized state is

$$|SS\rangle = |1/2, 1/2\rangle \otimes |1/2, 1/2\rangle \otimes \cdots \otimes |1/2, 1/2\rangle \quad (21.3.7)$$

If we now rotate this state, it becomes a tensor product of rotated states and when we form the inner product in the left-hand side of Eq. (21.3.5), we obtain the right-hand side.

The resolution of the identity in terms of these states is

$$I = \frac{2S+1}{4\pi} \int d\Omega |\Omega\rangle \langle \Omega| \quad (21.3.8)$$

where $d\Omega = d\cos \theta d\phi$. The proof can be found in the references. You are urged to do the following exercise that deals with $S=1/2$.

Exercise 21.3.2. Prove the completeness relation for $S=1/2$ by carrying out the integral over Ω using Eq. (21.3.6).

When we work out the path integral we will get a product of factors like the following:

$$\cdots \langle \Omega(t+\varepsilon) | I - \frac{i\varepsilon}{\hbar} H(\mathbf{S}) | \Omega(t) \rangle \cdots \quad (21.3.9)$$

We work to order ε . Since H already has a factor of ε in front of it, we set

$$\langle \Omega(t+\varepsilon) | -\frac{i\varepsilon}{\hbar} H(\mathbf{S}) | \Omega(t) \rangle \simeq -\frac{i\varepsilon}{\hbar} \langle \Omega(t) | H(\mathbf{S}) | \Omega(t) \rangle \equiv -i\varepsilon \mathcal{H}(\Omega) \quad (21.3.10)$$

If the Hamiltonian is linear in S , we simply replace the quantum spin operator by the classical vector pointing along θ, ϕ and if not, we can replace the operator by the suitable expectation value in the state $|\Omega(t)\rangle$. This is what we called $\hbar\mathcal{H}(\Omega)$ in the preceding equation.

Next we turn to the product

$$\langle \Omega(t+\varepsilon) | \Omega(t) \rangle \simeq 1 - i\varepsilon S(1 - \cos \theta) \dot{\phi} \simeq e^{iS(\cos \theta - 1)\dot{\phi}\varepsilon} \quad (21.3.11)$$

where we have expanded Eq. (21.3.5) to first order in $\Delta\theta$ and $\Delta\phi$. This gives us the following representation of the propagator in the continuum limit:

$$\langle \Omega_f | U(t) | \Omega_i \rangle = \int \mathcal{D}\Omega \exp \left[i \int_{t_1}^{t_2} [S \cos \theta \dot{\phi} - \mathcal{H}(\Omega)] dt \right] \quad (21.3.12)$$

where a total derivative in ϕ has been dropped and $\int \mathcal{D}\Omega$ is the measure with all factors of π in it.

Even by the standards of continuum functional integrals we have hit a new low, when we replaced differences by derivatives as if the paths are smooth. In the configuration path integral, we saw that between one time and the next the fluctuation in x was of the order $\varepsilon^{1/2}$ which is why we had to expand $\langle n(R')|n(R) \rangle$ to order $(R'-R)^2$ in the Berry calculation of the effective interaction. The factor that provided any kind of damping on the variation in the coordinate was the kinetic energy term $\exp[i m(x' - x)^2 / 2\hbar\varepsilon]$. *In the present problem there is no such term.* There is no reason why the difference in Ω from one time to another should be treated as a small quantity. Thus although the discretized functional integral is never wrong (since all we use is the resolution of the identity) any further assumptions about the smallness of the change in Ω from one time to the next are suspect. There is one exception. Suppose $S \rightarrow \infty$. Then we see from Eq. (21.3.5) that the overlap is unity if the two states are equal and falls rapidly if they are different. (It is easier to consider the case $\phi_2 = \phi_1$.) This is usually the limit ($S \rightarrow \infty$) in which one uses this formalism.

We now consider two simple applications. First let

$$H = \hbar S_z \quad (21.3.13)$$

We know the allowed eigenvalues are $\hbar(-S, -S+1, \dots, S)$. Let us derive this from the continuum path integral.

Given $\langle \Omega | H | \Omega \rangle = \hbar S \cos \theta$, it follows that $\mathcal{H} = S \cos \theta$, and that the functional integral is

$$\left[\int \mathcal{D} \cos \theta \mathcal{D} \phi \right] \exp \left[iS \int (\cos \theta \dot{\phi} - \cos \theta) dt \right] \quad (21.3.14)$$

We note that

- This is a phase space path integral with $\cos \theta$ as the momentum conjugate to ϕ !
- Phase space is compact here (the unit sphere), as compared to the problem of a particle moving on a sphere for which configuration space is compact but all momenta are allowed and phase space is infinite in extent.
- The spin S plays the role of $1/\hbar$.
- The Hamiltonian for the dynamics is $\cos \theta$ since we pulled out the S to the front. In particular, this means that $\cos \theta$ is a constant of motion, i.e., the orbits will be along fixed latitude.

Recall the WKB quantization rule

$$\oint p dq = 2\pi n \hbar \quad (21.3.15)$$

for a problem with no turning points. In our problem, $p = \cos \theta$ is just the conserved energy E . Of all the classical orbits along constant latitude lines, the ones chosen by WKB obey

$$\oint E d\phi = 2\pi n S^{-1} \quad (21.3.16)$$

since S^{-1} plays the role of \hbar . The allowed energies are

$$E_n = \frac{n}{S} \quad [-S \leq n \leq S] \quad (21.3.17)$$

Note that there is exactly enough room in this compact phase space for $2S+1$ orbits and that the allowed values of E translate into the allowed values of H when we reinstate the factor of $\hbar S$ that was pulled out along the way.

So we got lucky with this problem. In general, if H is more complicated we cannot hope for much luck unless S is large. Now you may ask why we bother with this formalism given that spins of real systems are very small. Here is at least one reason, based on a problem I am familiar with. In nuclear physics one introduces a *pseudospin* formalism in which a proton is called spin up and the neutron is called spin down. A big nucleus can have a large pseudospin, say 25. The Hamiltonian for the problem can be written in terms of the pseudospin operators and they can be

50×50 matrices. Finding the energy levels analytically is hopeless. But we can turn the large S in our favor by doing a WKB quantization using the appropriate H .

Coherent states are also very useful in the study of interacting quantum spins. For example, in the one-dimensional Heisenberg model, the Hamiltonian is a sum of dot products of nearest neighbor spin operators on a line of points. Since each spin operator appears linearly, the Hamiltonian in the action is just the quantum one with \mathbf{S} replaced by a classical vector of length S . Even though the spin is never very large in these problems, one studies the large S limit to get a feeling for the subject and to make controlled approximations in $1/S$.

Fermion Oscillator and Coherent States

Let us recall that in the case of the harmonic oscillator the fact that the energy levels were uniformly spaced

$$E = n\hbar\omega \quad (21.3.18)$$

(dropping zero point motion) allowed one to introduce the notion of quanta. Rather than saying the oscillator was in the n th state we could say there was one quantum level of energy $\hbar\omega$ and there were n quanta in it. This is how phonons, photons, etc., are viewed, and it is a very seminal idea.

That the level could be occupied by any number of quanta meant they were bosons. Indeed our perception of a classical electric or magnetic field is thanks to this feature.

Consider now a variant of the problem wherein the quanta are fermions. Thus the level can contain one or no quanta. There can be no macroscopic field associated with this state, which is why the fermion problem is unfamiliar to us at first. We now develop the theory of a fermionic oscillator.

We start by writing down the Hamiltonian:

$$H_0 = \Psi^\dagger \Psi \Omega_0 \quad (21.3.19)$$

What distinguishes this problem from the bosonic one are the *anticommutation relations*:

$$\{\Psi^\dagger, \Psi\} = \Psi^\dagger \Psi + \Psi \Psi^\dagger = 1 \quad (21.3.20)$$

$$\{\Psi, \Psi\} = \{\Psi^\dagger, \Psi^\dagger\} = 0 \quad (21.3.21)$$

Note that the last equation tells us

$$\Psi^{\dagger 2} = \Psi^2 = 0 \quad (21.3.22)$$

This equation will be used all the time without explicit warning. We shall see that it represents the Pauli principle forbidding double occupancy. The *number operator*

$$N = \Psi^\dagger \Psi \quad (21.3.23)$$

obeys

$$N^2 = \Psi^\dagger \Psi \Psi^\dagger \Psi = \Psi^\dagger (1 - \Psi^\dagger \Psi) \Psi = N \quad (21.3.24)$$

Thus the eigenvalues of N can only be 0 or 1. The corresponding normalized eigenstates obey

$$N|0\rangle = 0|0\rangle \quad (21.3.25)$$

$$N|1\rangle = 1|1\rangle \quad (21.3.26)$$

We will now prove that

$$\Psi^\dagger |0\rangle = |1\rangle \quad (21.3.27)$$

$$\Psi |1\rangle = |0\rangle \quad (21.3.28)$$

As for the first,

$$N\Psi^\dagger |0\rangle = \Psi^\dagger \Psi \Psi^\dagger |0\rangle = \Psi^\dagger (1 - \Psi^\dagger \Psi) |0\rangle = \Psi^\dagger |0\rangle \quad (21.3.29)$$

which shows that $\Psi^\dagger |0\rangle$ has $N=1$. Its norm is unity:

$$|\Psi^\dagger |0\rangle|^2 = \langle 0 | \Psi \Psi^\dagger |0\rangle = \langle 0 | (1 - \Psi^\dagger \Psi) |0\rangle = \langle 0 | 0 \rangle = 1 \quad (21.3.30)$$

It can be similarly shown that $\Psi |1\rangle = |0\rangle$ after first verifying that $\Psi |1\rangle$ is not a null vector, that it has unit norm.

There are no other vectors in the Hilbert space: any attempts to produce more states are thwarted by $\Psi^2 = \Psi^{\dagger 2} = 0$. In other words, the Pauli principle rules out more vectors: the state is either empty or singly occupied.

Thus the Fermi oscillator Hamiltonian

$$H_0 = \Omega_0 \Psi^\dagger \Psi \quad (21.3.31)$$

has eigenvalues 0 and Ω_0 .

We will work not with H_0 but with

$$H = H_0 - \mu N \quad (21.3.32)$$

where μ is called the *chemical potential*. For the oscillator, since

$$H = (\Omega_0 - \mu) \Psi^\dagger \Psi \quad (21.3.33)$$

this merely amounts to measuring all energies relative to the chemical potential. The role of the chemical potential will be apparent soon.

Let us now turn to thermodynamics. The central object here is the *grand partition function*, defined to be

$$Z = \text{Tr } e^{-\beta(H_0 - \mu N)} = e^{A(\mu, \beta)} \quad (21.3.34)$$

where the trace is over any complete set of eigenstates, β is the inverse temperature $1/kT$, and A is the free energy (different from the traditional one by a factor $-\beta$). The term *grand partition function* signifies that we are summing over states with a different number of particles or quanta. For this reason the free energy is denoted by A and not f . Just as β controls the amounts of energy the system takes from the reservoir, μ controls the number of particles. (This description is also possible for the bosonic oscillator. Instead of saying that we have just one oscillator which can be in any state labeled by n , and viewing the sum over states as the partition function of one oscillator, we can focus on the quanta and say that we are summing over states with variable number of quanta and interpret the usual sum over states as a grand partition function.)

If we use the N basis, this sum is trivial:

$$Z = 1 + e^{-\beta(\Omega_0 - \mu)} \quad (21.3.35)$$

All thermodynamic quantities can be deduced from this function. For example, it is clear from Eq. (21.3.34) that the mean occupation number is

$$\langle N \rangle = \frac{1}{\beta Z} \frac{\partial Z}{\partial \mu} = \frac{1}{\beta} \frac{\partial \ln Z}{\partial \mu} = \frac{1}{e^{\beta(\Omega_0 - \mu)} + 1} \quad (21.3.36)$$

Exercise 21.3.3. Prove the formula for $\langle N \rangle$ in general, starting with Eq. (21.3.34). (Write out the trace in a basis common to H and N , as a sum over energy levels at any one N , followed by a sum over N .)

At zero temperature we find from Eq. (21.3.36)

$$\langle N \rangle = \theta(\mu - \Omega_0) \quad (21.3.37)$$

i.e., the fermion is present if its energy is below chemical potential and absent if it is not. At finite temperatures the mean number varies more smoothly with μ .

We will now develop a path integral formula for the partition function.

We proceed in analogy with the bosonic oscillator by trying to find a *fermion coherent state* $|\psi\rangle$ which is an eigenstate of the destruction operator

$$\Psi|\psi\rangle = \psi|\psi\rangle \quad (21.3.38)$$

The eigenvalue ψ is a peculiar beast because if we act once more with Ψ we find

$$\psi^2 = 0 \quad (21.3.39)$$

since $\Psi^2 = 0$. Any ordinary variable whose square is zero is itself zero. But this ψ is no ordinary variable, it is a *Grassmann variable*. These variables *anticommute* with

each other and with all fermionic creation and destruction operators. (They will therefore commute with a string containing an even number of such operators.) That is how they are defined. The variable ψ is rather abstract and defined by its anticommuting nature. There are no big or small Grassmann variables. You will get used to them and even learn to love them just as you did the complex numbers. (Surely when you first heard it, you did not readily embrace the notion that $4i$ was an honest solution to the question “What number times itself gives -16 ?“ You probably felt that it may be the right answer, but it sure wasn’t a number.)

We now write down the coherent state. It is

$$|\psi\rangle = |0\rangle - \psi|1\rangle \quad (21.3.40)$$

where ψ is a Grassmann number. This state obeys:

$$\Psi|\psi\rangle = \Psi|0\rangle - \Psi\psi|1\rangle \quad (21.3.41)$$

$$= 0 + \psi\Psi|1\rangle \quad (21.3.42)$$

$$= \psi|0\rangle \quad (21.3.43)$$

$$= \psi(|0\rangle - \psi|1\rangle) \quad (21.3.44)$$

$$= \psi|\psi\rangle \quad (21.3.45)$$

where we have appealed to the fact that ψ anticommutes with Ψ and that $\psi^2 = 0$. If we act on both sides of Eq. (21.3.45) with Ψ , the left vanishes due to $\Psi^2 = 0$ and the right due to $\psi^2 = 0$.

It may be similarly verified that

$$\langle\bar{\psi}|\Psi^\dagger = \langle\bar{\psi}|\bar{\psi} \quad (21.3.46)$$

where

$$\langle\bar{\psi}| = \langle 0| - \langle 1| \bar{\psi} = \langle 0| + \bar{\psi}\langle 1| \quad (21.3.47)$$

Please note two points. First, the coherent state vectors are not the usual vectors from a complex vector space since they are linear combinations with Grassmann coefficients. Second, $\bar{\psi}$ is not in any sense the complex conjugate of ψ and $\langle\bar{\psi}|$ is not the adjoint of $|\psi\rangle$. You should therefore be prepared to see a change of Grassmann variables in which ψ and $\bar{\psi}$ undergo totally unrelated transformations.

The inner product of two coherent states is

$$\langle\bar{\psi}|\psi\rangle = (\langle 0| - \langle 1| \bar{\psi})(|0\rangle - \psi|1\rangle) \quad (21.3.48)$$

$$= \langle 0|0\rangle + \langle 1|\bar{\psi}\psi|1\rangle \quad (21.3.49)$$

$$= 1 + \bar{\psi}\psi \quad (21.3.50)$$

$$= e^{\bar{\psi}\psi} \quad (21.3.51)$$

Any function of a Grassmann variable can be expanded as follows:

$$F(\psi) = F_0 + F_1 \psi \quad (21.3.52)$$

there being no higher powers possible.

We will now define integrals over Grassmann numbers. (Don't throw up your hands: it will be over in no time.) These have no geometric significance (as areas or volumes) and are formally defined. We just have to know how to integrate 1 and ψ since that takes care of all possible functions. Here is the list of integrals:

$$\int \psi \, d\psi = 1 \quad (21.3.53)$$

$$\int 1 \, d\psi = 0 \quad (21.3.54)$$

That's it! As you can see, a table of Grassmann integrals is not going to be a best-seller. (For those of you have trouble remembering all these integrals, here is a useful mnemonic: the integral of any function is the same as the derivative! Verify this.) There are no limits on these integrals. Integration is assumed to be a linear operation. The differential $d\psi$ is also a Grassmann number. Thus $\int d\psi \psi = -1$. The integrals for $\bar{\psi}$ or any other Grassmann variable are identical. These integrals are simply assigned these values. They are very important since we see for the first-time ordinary numbers on the right-hand side. Anything numerical we calculate in this theory goes back to these integrals.

A result we will use often is this:

$$\int \bar{\psi} \psi \, d\psi \, d\bar{\psi} = 1 \quad (21.3.55)$$

Note that if the differentials or variables come in any other order there can be a change of sign. For example, we will also invoke the result

$$\int \bar{\psi} \psi \, d\bar{\psi} \, d\psi = -1 \quad (21.3.56)$$

Let us now consider some Gaussian integrals. You are urged to show the following:

$$\int e^{-a\psi\psi} \, d\bar{\psi} \, d\psi = a \quad (21.3.57)$$

$$\int e^{-\bar{\psi}M\psi} [d\bar{\psi} \, d\psi] = \det M \quad (21.3.58)$$

where in the second formula M is a 2-by-2 matrix, ψ is a column vector with entries ψ_1 and ψ_2 , $\bar{\psi}$ a column vector with entries $\bar{\psi}_1$ and $\bar{\psi}_2$ and $[d\bar{\psi} d\psi] = d\bar{\psi}_1 d\psi_1 d\bar{\psi}_2 d\psi_2$. This result is true for matrices of any size. To prove these simply expand the exponential and do the integrals.

Exercise 21.3.4. Prove the above two equations.

Consider next the “averages” over the Gaussian measure:

$$\langle \bar{\psi} \psi \rangle = \frac{\int \bar{\psi} \psi e^{a\bar{\psi}\psi} d\bar{\psi} d\psi}{\int e^{a\bar{\psi}\psi} d\bar{\psi} d\psi} = \frac{1}{a} = -\langle \psi \bar{\psi} \rangle \quad (21.3.59)$$

The proof is straightforward and left as an exercise.

Exercise 21.3.5. Provide the missing details in the evaluation of the above integral.

Exercise 21.3.6. Jacobians for Grassmann change of variables are the inverses of what you expect. Start with $\int a\phi d\phi = a$. Define $\chi = a\phi$, write $d\phi = J(\phi/\chi) d\chi$ and show that $J(\phi/\chi) = a$ and not $1/a$. (Treat the Jacobian as a constant that can be pulled out of the integral.) Evaluate Eq. (21.3.57) by introducing $\chi = a\psi$. Remember there is no need to change $\bar{\psi}$.

Consider now two sets of Grassmann variables (labeled 1 and 2). It is readily shown that

$$\langle \bar{\psi}_i \psi_j \rangle = \frac{\int \bar{\psi}_i \psi_j e^{a_1 \bar{\psi}_1 \psi_1 + a_2 \bar{\psi}_2 \psi_2} d\bar{\psi}_1 d\psi_1 d\bar{\psi}_2 d\psi_2}{\int e^{a_1 \bar{\psi}_1 \psi_1 + a_2 \bar{\psi}_2 \psi_2} d\bar{\psi}_1 d\psi_1 d\bar{\psi}_2 d\psi_2} \quad (21.3.60)$$

$$= \frac{\delta_{ij}}{a_i} \equiv \langle \bar{i}j \rangle \quad (21.3.61)$$

Exercise 21.3.7. Prove the above result.

Exercise 21.3.8. Show that

$$\langle \bar{\psi}_i \bar{\psi}_j \psi_k \psi_l \rangle = \frac{\int \bar{\psi}_i \bar{\psi}_j \psi_k \psi_l e^{a_1 \bar{\psi}_1 \psi_1 + a_2 \bar{\psi}_2 \psi_2} d\bar{\psi}_1 d\psi_1 d\bar{\psi}_2 d\psi_2}{\int e^{a_1 \bar{\psi}_1 \psi_1 + a_2 \bar{\psi}_2 \psi_2} d\bar{\psi}_1 d\psi_1 d\bar{\psi}_2 d\psi_2} \quad (21.3.62)$$

$$= \frac{\delta_{il}}{a_i} \frac{\delta_{jk}}{a_j} - \frac{\delta_{ik}}{a_i} \frac{\delta_{jl}}{a_j} \quad (21.3.63)$$

$$\equiv \langle \bar{i}l \rangle \langle \bar{j}k \rangle - \langle \bar{i}k \rangle \langle \bar{j}l \rangle \quad (21.3.64)$$

This is called *Wick's theorem* and is very useful in field theory and many-body theory.

We need two more results before we can write down the path integral. The first is the resolution of the identity:

$$I = \int |\psi\rangle\langle\bar{\psi}| e^{-\bar{\psi}\psi} d\bar{\psi} d\psi \quad (21.3.65)$$

In the following proof of this result we will use all the previously described properties and drop terms that are not going to survive integration. (Recall that only $\bar{\psi}\psi = -\psi\bar{\psi}$ has a nonzero integral.)

$$\begin{aligned} \int |\psi\rangle\langle\bar{\psi}| e^{-\bar{\psi}\psi} d\bar{\psi} d\psi &= \int |\psi\rangle\langle\bar{\psi}|(1 - \bar{\psi}\psi) d\bar{\psi} d\psi \\ &= \int (|0\rangle - \psi|1\rangle)(\langle 0| - \langle 1|\bar{\psi})(1 - \bar{\psi}\psi) d\bar{\psi} d\psi \\ &= \int (|0\rangle\langle 0| + \psi|1\rangle\langle 1|\bar{\psi})(1 - \bar{\psi}\psi) d\bar{\psi} d\psi \\ &= |0\rangle\langle 0| \int (-\bar{\psi}\psi) d\bar{\psi} d\psi + |1\rangle\langle 1| \int \psi\bar{\psi} d\bar{\psi} d\psi \\ &= I \end{aligned} \quad (21.3.66)$$

The final result we need is that for any bosonic operator (an operator made of an even number of Fermi operators)

$$\text{Tr } \Omega = \int \langle -\bar{\psi}|\Omega|\psi\rangle e^{-\bar{\psi}\psi} d\bar{\psi} d\psi \quad (21.3.67)$$

The proof is very much like the one just given and is left as an exercise.

Exercise 21.3.9. Prove the above formula for the trace.

The Fermionic Path Integral

Consider the partition function for a single oscillator:

$$Z = \text{Tr } e^{-\beta(\Omega_0 - \mu)\Psi^\dagger\Psi} \quad (21.3.68)$$

$$= \int \langle -\bar{\psi}|e^{-\beta(\Omega_0 - \mu)\Psi^\dagger\Psi}|\psi\rangle e^{-\bar{\psi}\psi} d\bar{\psi} d\psi \quad (21.3.69)$$

You cannot simply replace Ψ^\dagger and Ψ by $-\bar{\psi}$ and ψ , respectively, in the exponential. This is because when we expand out the exponential not all the Ψ 's will be acting to the right on their eigenstates and neither will all Ψ^\dagger 's be acting to the left on their eigenstates. (Remember that we are now dealing with operators, not Grassmann

numbers. The exponential will have an infinite number of terms in its expansion.) We need to convert the exponential to its *normal ordered form* in which all the creation operators stand to the left and all the destruction operators to the right. Luckily we can write down the answer by inspection:

$$e^{-\beta(\Omega_0 - \mu)\Psi^\dagger\Psi} = 1 + (e^{-\beta(\Omega_0 - \mu)} - 1)\Psi^\dagger\Psi \quad (21.3.70)$$

whose correctness we can verify by considering the two possible values of $\Psi^\dagger\Psi$. (Alternatively, you can expand the exponential and use the fact that $N^k = N$ for any nonzero k .) Now we may write

$$Z = \int \langle -\bar{\psi}|1 + (e^{-\beta(\Omega_0 - \mu)} - 1)\Psi^\dagger\Psi|\psi\rangle e^{-\bar{\psi}\psi} d\bar{\psi} d\psi \quad (21.3.71)$$

$$= \int \langle -\bar{\psi}|\psi\rangle (1 + (e^{-\beta(\Omega_0 - \mu)} - 1)(-\bar{\psi}\psi)) e^{-\bar{\psi}\psi} d\bar{\psi} d\psi \quad (21.3.72)$$

$$= \int (1 - (e^{-\beta(\Omega_0 - \mu)} - 1)\bar{\psi}\psi) e^{-2\bar{\psi}\psi} d\bar{\psi} d\psi \quad (21.3.73)$$

$$= 1 + e^{-\beta(\Omega_0 - \mu)} \quad (21.3.74)$$

as expected. While this is the right answer, this is not the path integral approach. It does, however, confirm the correctness of all the Grassmannian integration and minus signs. As for the path integral approach the procedure is the usual one. Consider

$$Z = \text{Tr } e^{-\beta H} \quad (21.3.75)$$

where H is a normal ordered operator $H(\Psi^\dagger, \Psi)$. We write the exponential as follows:

$$e^{-\beta H} = \lim_{N \rightarrow \infty} \left(\exp \left(-\frac{\beta}{N} H \right) \right)^N \quad (21.3.76)$$

$$= \underbrace{(1 - \varepsilon H) \dots (1 - \varepsilon H)}_{N \text{ times}} \quad \varepsilon = \beta/N \quad (21.3.77)$$

take the trace as per Eq. (21.3.67) by integrating over $\bar{\psi}_0\psi_0$, and introduce the resolution of the identity $N-1$ times:

$$\begin{aligned} Z &= \int \langle -\bar{\psi}_0|(1 - \varepsilon H)|\psi_{N-1}\rangle e^{-\bar{\psi}_{N-1}\psi_{N-1}} \langle \bar{\psi}_{N-1}|(1 - \varepsilon H)|\psi_{N-2}\rangle e^{-\bar{\psi}_{N-2}\psi_{N-2}} \\ &\quad \times \langle \bar{\psi}_{N-2} | \dots | \psi_1 \rangle e^{-\bar{\psi}_1\psi_1} \langle \bar{\psi}_1|(1 - \varepsilon H)|\psi_0\rangle e^{-\bar{\psi}_0\psi_0} \prod_{i=0}^{N-1} d\bar{\psi}_i d\psi_i \end{aligned} \quad (21.3.78)$$

Now we may legitimately make the replacement

$$\begin{aligned}\langle \bar{\psi}_{i+1} | 1 - \varepsilon H(\Psi^*, \Psi) | \psi_i \rangle &= \langle \bar{\psi}_{i+1} | 1 - \varepsilon H(\bar{\psi}_{i+1}, \psi_i) | \psi_i \rangle \\ &= e^{\bar{\psi}_{i+1} \psi_i} e^{-\varepsilon H(\bar{\psi}_{i+1}, \psi_i)}\end{aligned}\quad (21.3.79)$$

where in the last step we are anticipating the limit of infinitesimal ε . Let us now define an additional pair of variables (not to be integrated over)

$$\bar{\psi}_N = -\bar{\psi}_0 \quad (21.3.80)$$

$$\psi_N = -\psi_0 \quad (21.3.81)$$

The first of these equations allows us to replace the leftmost bra in Eq. (21.3.78), $\langle -\bar{\psi}_0 |$, by $\langle \bar{\psi}_N |$. The reason for introducing ψ_N will follow soon.

Putting together all the factors (including the overlap of coherent states) we end up with

$$Z = \int \prod_{i=0}^{N-1} e^{\bar{\psi}_{i+1} \psi_i} e^{-\varepsilon H(\bar{\psi}_{i+1}, \psi_i)} e^{-\bar{\psi}_i \psi_i} d\bar{\psi}_i d\psi_i \quad (21.3.82)$$

$$= \int \prod_{i=0}^{N-1} \exp \left[\left[\left(\frac{(\bar{\psi}_{i+1} - \bar{\psi}_i)}{\varepsilon} \psi_i - H(\bar{\psi}_{i+1}, \psi_i) \right) \right] \varepsilon \right] d\bar{\psi}_i d\psi_i \quad (21.3.83)$$

$$\simeq \int \exp \left(\int_0^\beta \bar{\psi}(\tau) \left(-\frac{\partial}{\partial \tau} - \Omega_0 + \mu \right) \psi(\tau) d\tau \right) [\mathcal{D}\bar{\psi} \mathcal{D}\psi] \quad (21.3.84)$$

where the last step needs some explanation. With all the factors of ε in place we do seem to get the continuum expression in the last formula. However, the notion of replacing differences by derivatives is purely symbolic for Grassmann variables. There is no sense in which $\bar{\psi}_{i+1} - \bar{\psi}_i$ is small, in fact the objects have no numerical values. What this really means here is the following. In a while we will trade $\psi(\tau)$ for $\psi(\omega)$ related by Fourier transformation. At that stage we will replace $-\partial/\partial\tau$ by $i\omega$ while the exact answer is $e^{i\omega} - 1$. If we do not make this replacement, the Grassmann integral, when evaluated in terms of ordinary numbers, will give exact results for anything one wants to calculate, say the free energy. With this approximation, only quantities insensitive to high frequencies will be given correctly. The free energy will come out wrong but the correlation functions will be correctly reproduced. (This is because the latter are given by derivatives of the free energy and these derivatives make the integrals sufficiently insensitive to high frequencies.) Notice also that we are replacing $H(\bar{\psi}_{i+1}, \psi_i) = H(\bar{\psi}(\tau + \varepsilon), \psi(\tau))$ by $H(\bar{\psi}(\tau), \psi(\tau))$ in the same spirit.

Now turn to the Fourier expansions alluded to above. Let us write

$$\bar{\psi}(\tau) = \sum_n \frac{e^{i\omega_n \tau}}{\beta} \bar{\psi}(\omega) \quad (21.3.85)$$

$$\psi(\tau) = \sum_n \frac{e^{-i\omega_n \tau}}{\beta} \psi(\omega) \quad (21.3.86)$$

where the allowed frequencies, called *Matsubara frequencies*, are chosen to satisfy the antisymmetric boundary conditions in Eqs. (21.3.80–21.3.81). Thus

$$\omega_n = \frac{(2n+1)\pi}{\beta} \quad (21.3.87)$$

where n is an integer. Note that we have chosen the Fourier expansions as if ψ and $\bar{\psi}$ were complex conjugates, which they are not. This choice, however, makes the calculations easy.

For future reference note that if $\beta \rightarrow \infty$, it follows from Eq. (21.3.87) that when n increases by unity, ω_n changes by $d\omega = 2\pi/\beta$. Thus

$$\frac{1}{\beta} \sum_n \rightarrow \int \frac{d\omega}{2\pi} \quad (21.3.88)$$

The inverse transformations are

$$\psi(\omega) = \int_0^\beta \psi(\tau) e^{i\omega_n \tau} d\tau \quad (21.3.89)$$

$$\bar{\psi}(\omega) = \int_0^\beta \bar{\psi}(\tau) e^{-i\omega_n \tau} d\tau \quad (21.3.90)$$

where we use the orthogonality property

$$\int_0^\beta e^{i\omega_n \tau} e^{-i\omega_m \tau} d\tau = \frac{e^{i(\omega_n - \omega_m)\beta} - 1}{i(\omega_n - \omega_m)} = \beta \delta_{mn} \quad (21.3.91)$$

Performing the Fourier transforms in the action and changing the functional integration variables to $\psi(\omega)$ and $\bar{\psi}(\omega)$ (the Jacobian is unity) and going to the limit $\beta \rightarrow \infty$, which converts sums over discrete frequencies to integrals over a continuous ω , as per Eq. (21.3.88), we end up with

$$Z = \int \exp \left[\int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \bar{\psi}(\omega) (i\omega - \Omega_0 + \mu) \psi(\omega) \right] [\mathcal{D}\bar{\psi}(\omega) \mathcal{D}\psi(\omega)] \quad (21.3.92)$$

Although β has disappeared from the picture it will appear as $2\pi\delta(0)$, which we know stands for the total imaginary time β . (Recall Fermi's golden rule calculations.) An example will follow shortly.

Let us first note that the frequency space correlation function is related to the integral over just a single pair of variables [Eq. (21.3.59)] and is given by:

$$\begin{aligned} & \langle \bar{\psi}(\omega_1) \psi(\omega_2) \rangle \\ &= \frac{\int \bar{\psi}(\omega_1) \psi(\omega_2) \exp \left[\int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \bar{\psi}(\omega) (i\omega - \Omega_0 + \mu) \psi(\omega) \right] [\mathcal{D}\bar{\psi}(\omega) \mathcal{D}\psi(\omega)]}{\int \exp \left[\int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \bar{\psi}(\omega) (i\omega - \Omega_0 + \mu) \psi(\omega) \right] [\mathcal{D}\bar{\psi}(\omega) \mathcal{D}\psi(\omega)]} \\ &= \frac{2\pi \delta(\omega_1 - \omega_2)}{i\omega_1 - \Omega_0 + \mu} \end{aligned} \quad (21.3.93)$$

In particular,

$$\langle \bar{\psi}(\omega) \psi(\omega) \rangle = \frac{2\pi \delta(0)}{i\omega - \Omega_0 + \mu} = \frac{\beta}{i\omega - \Omega_0 + \mu} \quad (21.3.94)$$

Exercise 21.3.10. Try to demonstrate the above two equations. Note first of all that unless $\omega_1 = \omega_2$, we get zero since only a $\bar{\psi}\psi$ pair has a chance of having a nonzero integral. This explains the δ -function. As for the 2π , go back to the stage where we had a sum over frequencies and not an integral, i.e., go against the arrow in Eq. (21.3.88) and use it in the exponent of Eq. (21.3.93).

Let us now calculate the mean occupation number $\langle N \rangle$:

$$\langle N \rangle = \frac{1}{\beta Z} \frac{\partial Z}{\partial \mu} \quad (21.3.95)$$

$$= \frac{1}{\beta} \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \langle \bar{\psi}(\omega) \psi(\omega) \rangle \quad (21.3.96)$$

$$= \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{e^{i\omega\mu}}{i\omega - \Omega_0 + \mu} \quad (21.3.97)$$

$$= \theta(\mu - \Omega_0) \quad (21.3.98)$$

as in the operator approach.

Notice that we had to introduce the factor $e^{i\omega\mu}$ into the ω integral. We understand this as follows. If we had done the calculation using time τ instead of frequency ω , we would have calculated the average of $\Psi^\dagger \Psi$. This would automatically have turned into $\bar{\psi}(\tau + \varepsilon)\psi(\tau)$ when introduced into the path integral since the coherent state bra to the left of the operator would have come from the next time slice compared to the ket at the right. [Remember how $H(\Psi^\dagger, \Psi)$ turned into $H(\bar{\psi}(i+1), \psi(i))$.] Notice that the integral over ω was not convergent, varying as $d\omega/\omega$. It was therefore sensitive to the high frequencies and we had to intervene

with the factor $e^{i\omega_0 \tau}$. This factor allows us to close the contour in the upper half-plane. If $\mu > \Omega_0$, the pole of the integrand lies in that half-plane and makes a contribution. If not we get zero. In correlation functions that involve integrals that have two or more powers of ω in the denominator and are hence convergent, we will not introduce this factor.

Exercise 21.3.11. Advanced

In field theory and many-body physics one is interested in the Green's function:

$$G(\tau) = \langle \mathcal{T}(\Psi(\tau)\Psi^\dagger(0)) \rangle \quad (21.3.99)$$

where $\langle \rangle$ denotes the average with respect to Z ,

$$\Omega(\tau) = e^{H\tau} \Omega e^{-H\tau}$$

is the Heisenberg operator, and \mathcal{T} the time-ordering symbol for fermionic operators:

$$\mathcal{T}(\Psi(\tau)\Psi^\dagger(0)) = \theta(\tau)\Psi(\tau)\Psi^\dagger(0) - \theta(-\tau)\Psi^\dagger(0)\Psi(\tau) \quad (21.3.100)$$

Note the minus sign when the order of operators is reversed. Show that $\Psi(\tau) = \Psi e^{-(\Omega_0 - \mu)\tau}$ for our problem of the single oscillator.

Show, using the operator formalism that in our problem

$$G(\tau) = \frac{\theta(\tau) e^{-(\Omega_0 - \mu)\tau} - \theta(-\tau) e^{-(\Omega_0 - \mu)(\tau + \beta)}}{1 + e^{-\beta(\Omega_0 - \mu)}} \quad (21.3.101)$$

and that in the zero-temperature limit this reduces to

$$G(\tau) = \theta(\tau) e^{-(\Omega_0 - \mu)\tau} \quad \mu < \Omega_0 \quad (21.3.102)$$

$$= -\theta(-\tau) e^{-(\Omega_0 - \mu)\tau} \quad \mu > \Omega_0 \quad (21.3.103)$$

Let us define the pair of transforms:

$$G(\omega) = \int_{-\infty}^{\infty} G(\tau) e^{i\omega\tau} d\tau \quad (21.3.104)$$

$$G(\tau) = \int_{-\infty}^{\infty} G(\omega) e^{-i\omega\tau} \frac{d\omega}{2\pi} \quad (21.3.105)$$

Show that

$$G(\omega) = \frac{1}{\Omega_0 - \mu - i\omega} \quad (21.3.106)$$

independent of which of Ω_0 or μ is greater.

We saw in the study of the Ising model that the two-point correlation function in the functional integral translates into ground state average of the time-ordered product (for infinitely long system in the imaginary time direction) and vice versa. (If the parenthetical condition

is not met, there will not be enough time for the system to relax into the ground state before we stick in the operators being averaged.)

It is likewise true here that

$$\langle \mathcal{T}(\Psi(\tau)\Psi^\dagger(0)) \rangle = \langle \bar{\psi}(\tau)\psi(0) \rangle \quad (21.3.107)$$

where the average on the right-hand side is done by the Grassmann functional integral. Working at zero temperature, verify this for the frequency transform of both sides. (In the right-hand side write $\psi(\tau)$ in terms of $\psi(\omega)$, etc., using the zero temperature version of Eqs. (21.3.85–21.3.86) and Eq. (21.3.93).)

This brings us to the end of the discussion of fermionic path integrals. Clearly this is just the beginning and our discussion has been just an introduction.

21.4. Summary

Let us survey what has been done in this chapter. We started by learning how to use different resolutions of the identity to derive different path integrals. We looked at the configuration space, phase space, and coherent state path integrals. We realized that, while the introduction of the resolution of the identity is not an approximation, any assumption that changes in the coordinates being integrated over were small between time slices was to be carefully examined. In configuration space integrals the kinetic energy term provided a damping of fluctuations to something of order $\epsilon^{1/2}$. In other integrals there was no such assurance. In particular, the continuum forms of the action were purely formal objects and only the discrete version defined the path integral, assuming the limit of infinite number of integrals existed. Despite this, the path integrals were very useful for seeing the theory as a whole before us, as a constructive solution to the quantum evolution problem. In particular, in the classical limit the smallness of \hbar allowed us to think in terms of smooth paths. The study of the LLL (in connection with the QHE) and the Berry phase analysis illustrated some correct uses of the path integral.

We then turned to imaginary time quantum mechanics. We showed that from it one could extract the real-time energies and wave functions. In addition, imaginary time path integrals directly defined quantum statistical mechanics and were formally similar to classical statistical mechanics. The transfer matrix played the role of the discrete imaginary time evolution operator. Symmetry breaking was analyzed from many angles.

Finally, we studied two systems with no classical limit: the quantum spins and fermion oscillators. Although we studied just one fermionic oscillator, the generalization to many is direct and you should have no trouble following that topic when you get to it. Grassmann integrals are undoubtedly the most abstract notion in this book. But there is no doubt that as you use them (comparing them to the operator solution as a check) you will soon learn to think directly in terms of them. But remember this: there is no real notion of a semiclassical analysis here since the action is not a number-valued object and cannot be said to be stationary at any point. Note also that every Grassmann integral you write is eventually equal to an ordinary number though the integrand and integration measure are not. These numbers

correspond to physical entities like the ground energy or correlation function of a fermion system.

The only functional integral we evaluated was the Gaussian integral. This is essentially all we know how to do. What if the action is not quadratic but has quartic terms? Then we do perturbation theory. We bring down the quartic term from the exponential (in the form of an infinite series) and evaluate term by term since we know how to integrate x^n times a Gaussian. Recall Appendix A.2 as well as the Wick's theorem for fermions in Exercise (21.3.64). But that's another story.

Bibliography

- M. Berry, *Proc. R. Soc. Lond., Ser. A* 392 **45**, 1984. A fascinating account of the history of the subject may be found in M. Berry, *Physics Today*, **43**, 12 1990.
- S. Coleman, *Aspects of Symmetry* (Cambridge University Press, New York, 1985). Included here for the article on instantons.
- L. D. Faddeev, in *Methods in Field Theory*, Les Houches Lectures, 1975 (R. Balian and J. Zinn-Justin, eds.) (North-Holland/World Scientific, Singapore, 1981). Look in particular at the discussion of holomorphic form of the functional integral, i.e., the coherent state integral.
- M. Gutzwiller, *Chaos in Classical and Quantum Mechanics* (Springer Verlag, New York, 1990). For a solid introduction to many aspects of path integrals, especially the semiclassical limit.
- J. R. Klauder and B. Skagerstam, eds., *Coherent States* (World Scientific, Singapore, 1985). Everything you ever wanted to know about coherent states.
- S. Pancharatnam, *Ind. Acad. Sci.*, **44**(5), Sec. A, 1958. Reprinted in Shapere and Wilczek (1989) (see below). Discusses the geometric phase in the context of optics.
- R. E. Prange and S. M. Girvin, eds., *The Quantum Hall Effect* (Springer, New York, 1987). Has many contributed papers by leaders in the field as well as helpful commentaries.
- R. Rajaraman, *Instantons and Solitons* (North Holland, New York, 1982). Extremely clear discussion of the subject, usually starting with a warmup toy example from elementary quantum mechanics. Very few details are "left to the reader."
- L. S. Schulman, *Techniques and Applications of Path Integrals* (Wiley Interscience, New York, 1981). A very readable and clear discussion of functional integrals, and pitfalls and fine points (such as the midpoint prescription for vector potential coupling).
- A. Shapere and F. Wilczek, eds., *Geometric Phases in Physics* (World Scientific, Singapore, 1989). Collection of all key papers and some very good introductions to each subtopic. Saves countless trips to the library.
- M. Stone, ed., *The Quantum Hall Effect* (World Scientific, Singapore, 1992). A nice collection of reprints with commentary.
- 't Hooft, *Phys. Rev. Lett.*, **37**, 8 (1976).
- F. Wilczek, ed., *Fractional Statistics and Anyon Superconductivity* (World Scientific, Singapore, 1990). Referenced here for its applications to the Quantum Hall Effect. However, other topics like fractional statistics discussed there should be within your reach after this book.

Appendix

A.1. Matrix Inversion

This brief section is included only to help you understand Eq. (1.8.5) in the main text and is by no means comprehensive.

Consider the inversion of a 3×3 matrix

$$M = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix} \quad (\text{A.1.1})$$

The elements of M have been named in this way rather than as M_{ij} , for in the following discussion we will treat the rows as components of the vectors \mathbf{A} , \mathbf{B} , and \mathbf{C} , i.e., in the notation of vector analysis (which we will follow in this section),

$$\mathbf{A} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k} \text{ and so on}$$

Consider next a triplet of vectors

$$\begin{aligned} \mathbf{A}_R &= \mathbf{B} \times \mathbf{C} \\ \mathbf{B}_R &= \mathbf{C} \times \mathbf{A} \\ \mathbf{C}_R &= \mathbf{A} \times \mathbf{B} \end{aligned} \quad (\text{A.1.2})$$

which are said to be *reciprocal* to \mathbf{A} , \mathbf{B} , and \mathbf{C} . In general,

$$\mathbf{A} \cdot \mathbf{A}_R \neq 0, \quad \mathbf{A} \cdot \mathbf{B}_R = \mathbf{A} \cdot \mathbf{C}_R = 0 \quad \text{and cyclic permutations} \quad (\text{A.1.3})$$

If we construct now a matrix $\bar{\mathbf{M}}$ (called the *cofactor transpose* of \mathbf{M}) whose *columns* are the reciprocal vectors,

$$\bar{\mathbf{M}} = \begin{bmatrix} (a_R)_1 & (b_R)_1 & (c_R)_1 \\ (a_R)_2 & (b_R)_2 & (c_R)_2 \\ (a_R)_3 & (b_R)_3 & (c_R)_3 \end{bmatrix}$$

then

$$\mathbf{M} \cdot \bar{\mathbf{M}} = \begin{bmatrix} \mathbf{A} \cdot \mathbf{A}_R & \mathbf{A} \cdot \mathbf{B}_R & \mathbf{A} \cdot \mathbf{C}_R \\ \mathbf{B} \cdot \mathbf{A}_R & \mathbf{B} \cdot \mathbf{B}_R & \mathbf{B} \cdot \mathbf{C}_R \\ \mathbf{C} \cdot \mathbf{A}_R & \mathbf{C} \cdot \mathbf{B}_R & \mathbf{C} \cdot \mathbf{C}_R \end{bmatrix} = \begin{bmatrix} \mathbf{A} \cdot \mathbf{A}_R & 0 & 0 \\ 0 & \mathbf{B} \cdot \mathbf{B}_R & 0 \\ 0 & 0 & \mathbf{C} \cdot \mathbf{C}_R \end{bmatrix} \quad (\text{A.1.4})$$

Now all three diagonal elements are equal:

$$\begin{aligned} \mathbf{A} \cdot \mathbf{A}_R &= \mathbf{A} \cdot (\mathbf{B} \times \mathbf{C}) = \mathbf{B} \cdot (\mathbf{C} \times \mathbf{A}) = \mathbf{B} \cdot \mathbf{B}_R = \mathbf{C} \cdot (\mathbf{A} \times \mathbf{B}) = \mathbf{C} \cdot \mathbf{C}_R \\ &= \det \mathbf{M} \end{aligned} \quad (\text{A.1.5})$$

where the last equality follows from the fact that the cross product may be written as a determinant:

$$\mathbf{B} \times \mathbf{C} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} \quad (\text{A.1.6})$$

(We shall follow the convention of using two vertical lines to denote a determinant.)
Hence the inverse of the matrix \mathbf{M} is given by

$$\mathbf{M}^{-1} = \frac{\bar{\mathbf{M}}}{\det \mathbf{M}} \quad (\text{A.1.7})$$

When does $\det \mathbf{M}$ vanish? If one of the vectors, say \mathbf{C} , is a linear combination of the other two; for if

$$\mathbf{C} = \alpha \mathbf{A} + \beta \mathbf{B}$$

then

$$\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C}) = \mathbf{A} \cdot (\mathbf{B} \times \alpha \mathbf{A}) + \mathbf{A} \cdot (\mathbf{B} \times \beta \mathbf{B}) = \mathbf{B} \cdot (\alpha \mathbf{A} \times \mathbf{A}) = 0$$

Thus the determinant vanishes if the rows of the matrix are not linearly independent (LI) and vice versa. If the matrix is used to represent three simultaneous equations, it means not all three equations are independent. The method can be generalized for inverting $n \times n$ matrices, with real or complex elements. One defines a cross product of $n - 1$ vectors as

$$\mathbf{A}_1 \times \mathbf{A}_2 \times \cdots \times \mathbf{A}_{n-1} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} & \dots \\ (a_1)_1 & (a_1)_2 & \dots & \\ & \vdots & & \\ (a_{n-1})_1 & (a_{n-1})_2 & \dots & (a_{n-1})_n \end{vmatrix} \quad (\text{A.1.8})$$

The resulting vector is orthogonal to the ones in the product, changes sign when we interchange any two of the adjacent ones, and so on, just like its three-dimensional counterpart. If we have a matrix M , whose n rows may be identified with n vectors, $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_n$, then the cofactor transpose has as its columns the reciprocal vectors $\mathbf{A}_{1R}, \dots, \mathbf{A}_{nR}$, where

$$\mathbf{A}_{jR} = \mathbf{A}_{j+1} \times \mathbf{A}_{j+2} \times \cdots \times \mathbf{A}_n \times \mathbf{A}_1 \times \cdots \times \mathbf{A}_{j-1} \quad (\text{A.1.9})$$

One tricky point: the cross product is defined to be orthogonal to the vectors in the product with respect to an inner product

$$\mathbf{A} \cdot \mathbf{B} = \sum A_i B_i$$

and *not*

$$\mathbf{A}^* \cdot \mathbf{B} = \sum A_i^* B_i$$

even when the components of \mathbf{A} are complex. There is no contradiction here, for the vectors $\mathbf{A}_1, \dots, \mathbf{A}_n$ are fictitious objects that enter a mnemonic and not the elements of the space $\mathbb{V}^n(C)$ on which the operator acts.

Exercise A.1.1. Using the method described above, show that

$$\begin{bmatrix} 2 & 1 & 3 \\ 0 & 1 & 2 \\ -1 & 1 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -2 & 1 \\ 2 & -5 & 4 \\ -1 & 3 & -2 \end{bmatrix}$$

and

$$\begin{bmatrix} 2 & 1 & 3 \\ 4 & 1 & 2 \\ 0 & -1 & 2 \end{bmatrix}^{-1} = \frac{1}{12} \begin{bmatrix} -4 & 5 & 1 \\ 8 & -4 & -8 \\ 4 & -2 & 2 \end{bmatrix}$$

Theorem A.1.1. If $\Omega|V\rangle=|0\rangle$ implies $|V\rangle=|0\rangle$ then Ω^{-1} exists.

Proof. Let $|V_1\rangle, \dots, |V_n\rangle$ be a LI basis in \mathbb{V}^n . Then another LI basis is generated by the action of Ω , i.e., $\Omega|V_1\rangle, \dots, \Omega|V_n\rangle$ is also a LI basis. To see this, let us assume the contrary, that there exists a relation of the form

$$\sum_i \alpha_i \Omega|V_i\rangle = 0$$

with not all $\alpha_i=0$. Upon pulling out Ω , because it is linear, we get

$$\Omega\left(\sum_i \alpha_i|V_i\rangle\right) = 0$$

which, when combined with the assumed property of Ω , implies that

$$\sum_i \alpha_i|V_i\rangle = |0\rangle$$

with not all $\alpha_i=0$, which is not true. So we can conclude that every vector $|V'\rangle$ in \mathbb{V}^n may be written as a *unique* linear combination in the new basis generated by Ω as

$$|V'\rangle = \sum_i \alpha_i \Omega|V_i\rangle$$

In terms of $|V\rangle = \sum_i \alpha_i|V_i\rangle$, we see that *every* $|V'\rangle$ in \mathbb{V}^n may be written as

$$|V'\rangle = \Omega|V\rangle$$

where $|V\rangle$ is *unique*. In other words, we can think of *every* $|V'\rangle$ in \mathbb{V}^n as arising from a *unique* source $|V\rangle$ in \mathbb{V}^n under the action of Ω . Define an operator Λ whose action on any vector $|V'\rangle$ in \mathbb{V}^n is to take it back to its unique source $|V\rangle$. (If the source of $|V'\rangle$ were not unique—say, because there are two vectors $|V_1\rangle$ and $|V_2\rangle$ that are mapped into $|V'\rangle$ by Ω —then we could not define Λ , for acting on $|V'\rangle$, it would not know whether to give $|V_1\rangle$ or $|V_2\rangle$.) The action of Λ is then

$$\Lambda|V'\rangle = |V\rangle, \quad \text{where } |V'\rangle = \Omega|V\rangle$$

We may identify Λ as the inverse of Ω ,

$$\Lambda = \Omega^{-1} \quad \text{or} \quad \Lambda\Omega = I$$

since for any $|V'\rangle$ in \mathbb{V}^n

$$\Lambda|V'\rangle = \Lambda\Omega|V\rangle = |V\rangle \quad \text{Q.E.D.}$$

A.2. Gaussian Integrals

We discuss here all the Gaussian integrals that we will need. Consider

$$I_0(\alpha) = \int_{-\infty}^{\infty} e^{-\alpha x^2} dx, \quad \alpha > 0 \quad (\text{A.2.1})$$

This integral cannot be evaluated by conventional methods. The trick is to consider

$$I_0^2(\alpha) = \int_{-\infty}^{\infty} e^{-\alpha x^2} dx \int_{-\infty}^{\infty} e^{-\alpha y^2} dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\alpha(x^2 + y^2)} dx dy$$

Switching to polar coordinates in the $x-y$ plane,

$$\begin{aligned} I_0^2(\alpha) &= \int_0^{\infty} \int_0^{2\pi} e^{-\alpha\rho^2} \rho d\rho d\phi \\ &= \pi/\alpha \end{aligned}$$

Therefore

$$I_0(\alpha) = (\pi/\alpha)^{1/2} \quad (\text{A.2.2})$$

By differentiating with respect to α we can get all the integrals of the form

$$I_{2n}(\alpha) = \int_{-\infty}^{\infty} x^{2n} e^{-\alpha x^2} dx$$

For example,

$$\begin{aligned} I_2(\alpha) &= \int_{-\infty}^{\infty} x^2 e^{-\alpha x^2} dx = -\frac{\partial}{\partial \alpha} \int_{-\infty}^{\infty} e^{-\alpha x^2} dx \\ &= -\frac{\partial}{\partial \alpha} I_0(\alpha) = \frac{1}{2\alpha} \left(\frac{\pi}{\alpha}\right)^{1/2} \end{aligned} \quad (\text{A.2.3})$$

The integrals $I_{2n+1}(\alpha)$ vanish because these are integrals of odd functions over an even interval $-\infty$ to $+\infty$. Equations (A.2.2) and (A.2.3) are valid even if α is purely imaginary.

Consider next

$$I_0(\alpha, \beta) = \int_{-\infty}^{\infty} e^{-\alpha x^2 + \beta x} dx \quad (\text{A.2.4})$$

By completing the square on the exponent, we get

$$I_0(\alpha, \beta) = e^{\beta^2/4\alpha} \int_{-\infty}^{\infty} e^{-\alpha(x - \beta/2\alpha)^2} dx = e^{\beta^2/4\alpha} \left(\frac{\pi}{\alpha} \right)^{1/2} \quad (\text{A.2.5})$$

These results are valid even if α and β are complex, provided $\operatorname{Re} \alpha > 0$. Finally, by applying to both sides of the equation

$$\int_0^{\infty} e^{-ar} dr = \frac{1}{a}$$

the operator $(-d/d\alpha)^n$, we obtain

$$\int_0^{\infty} r^n e^{-ar} dr = \frac{n!}{a^{n+1}}$$

Consider this integral with $\alpha = 1$ and n replaced by $z - 1$, where z is an arbitrary complex number. This defines the *gamma function* $\Gamma(z)$

$$\Gamma(z) = \int_0^{\infty} r^{z-1} e^{-r} dr$$

For real, positive and integral z ,

$$\Gamma(z) = (z - 1)!$$

A.3. Complex Numbers

A complex variable z can be written in terms of two real variables x and y , and $i = (-1)^{1/2}$, as

$$z = x + iy \quad (\text{A.3.1})$$

Its *complex conjugate* z^* is defined to be

$$z^* = x - iy \quad (\text{A.3.2})$$

One may invert these two equations to express the *real and imaginary parts*, x and y , as

$$x = \frac{1}{2}(z + z^*), \quad y = (z - z^*)/2i \quad (\text{A.3.3})$$

The *modulus squared* of z , defined to be zz^* , equals

$$zz^* \equiv |z|^2 = (x + iy)(x - iy) = x^2 + y^2 \quad (\text{A.3.4})$$

You may verify that $z = z'$ implies that $x = x'$ and $y = y'$ by considering the modulus of $z - z'$.

From the power-series expansions

$$\sin x = x - x^3/3! + x^5/5! - \dots$$

$$\cos x = 1 - x^2/2! + x^4/4! - \dots$$

one can deduce that

$$e^{ix} = \cos x + i \sin x \quad (\text{A.3.5})$$

It is clear that e^{ix} has unit modulus (x is real).

The expression $z = x + iy$ gives z in *Cartesian form*. The *polar form* is

$$\begin{aligned} z = x + iy &= (x^2 + y^2)^{1/2} \left[\frac{x}{(x^2 + y^2)^{1/2}} + i \frac{y}{(x^2 + y^2)^{1/2}} \right] \\ &= \rho(\cos \theta + i \sin \theta) \\ &= \rho e^{i\theta} \end{aligned}$$

where

$$\rho = (x^2 + y^2)^{1/2} \quad \text{and} \quad \theta = \tan^{-1}(y/x) \quad (\text{A.3.6})$$

Clearly

$$|z| = \rho \quad (\text{A.3.7})$$

Each complex number $z = x + iy$ may be visualized as a point (x, y) in the x - y plane. This plane is also called the *complex z plane*.

A.4. The *iε* Prescription

We will now derive and interpret the formula

$$\frac{1}{x \mp i\varepsilon} = \mathcal{P} \frac{1}{x} \pm i\pi\delta(x) \quad (\text{A.4.1})$$

where $\varepsilon \rightarrow 0$ is a positive infinitesimally small quantity. Consider an integral of the form

$$I = \lim_{\varepsilon \rightarrow 0} \int_{-\infty}^{\infty} \frac{f(x) dx}{x - i\varepsilon}. \quad (\text{A.4.2})$$

Viewing this as the integral on the real axis of the complex $z = x + iy$ plane, we see that the integrand has an explicit pole at $z = i\varepsilon$ in addition to any singularities f might have. We assume f has no singularities on or infinitesimally close to the real axis. As long as ε is fixed, there is no problem with the integral. For example, if f has some poles in the upper half-plane and vanishes fast enough to permit our closing the contour in the upper half-plane, the integral equals $2\pi i$ times the sum of the residues of the poles of f and the pole at $z = i\varepsilon$. Likewise, if we change the sign of the ε term, we simply drop the contribution from the explicit pole, which is now in the lower half-plane.

What if $\varepsilon \rightarrow 0$? Now the pole is going to ram (from above) into our contour which runs along the x -axis. So we prepare for this as follows. Since the only singularity near the real axis is the explicit pole at $z = i\varepsilon$, we make the following deformation of the contour without changing the value of I : the contour runs along the real axis from $-\infty$ to $-\varepsilon'$, (ε' is another positive infinitesimal) goes around counterclockwise, below the origin in a semicircle of radius ε' , and resumes along the real axis from $x = \varepsilon'$ to ∞ . The nice thing is that we can now set $\varepsilon = 0$, which brings the pole to the origin. The three parts of the integration contour contribute as follows:

$$\begin{aligned} I &= \lim_{\varepsilon' \rightarrow 0} \left[\int_{-\infty}^{-\varepsilon'} \frac{f(x) dx}{x} + \int_{\varepsilon'}^{\infty} \frac{f(x) dx}{x} + i\pi f(0) \right] \\ &\equiv \mathcal{P} \int_{-\infty}^{\infty} \frac{f(x) dx}{x} + i\pi f(0). \end{aligned} \quad (\text{A.4.3})$$

The sum of the two integrals in the limit $\varepsilon' \rightarrow 0$ is defined as the *principal value integral* denoted by the symbol \mathcal{P} . In the last term, which is restricted to the infinitesimal neighbourhood of the origin, we have set the argument of the smooth function f to zero and done the integral of dz/z counterclockwise around the *semicircle* to get $i\pi$.

Eq. (A.4.1) is a compact way to say all this. It is understood that Eq. (A.4.1) is to be used inside an integral only and that inside an integral the factor $1/(x - i\varepsilon)$ leads to two terms: the first, $\mathcal{P}(1/x)$, leads to the principal value integral, and the second, $i\pi\delta(x)$, leads to $i\pi f(0)$.

It is clear that if we reverse the sign of the ε term, we change the sign of the delta function since the semicircle now goes around the pole in the clockwise direction. The principal part is not sensitive to this change of direction and is unaffected.

It is clear that if we replace x by $x - a$ the pole moves from the origin to $x = a$ and $f(0)$ gets replaced by $f(a)$ so that we may write

$$\frac{1}{(x - a) \mp i\varepsilon} = \mathcal{P} \frac{1}{(x - a)} \pm i\pi\delta(x - a) \quad (\text{A.4.4})$$

It is clear that the limits on x need not be $\pm\infty$ for the formula to work.

Finally, note that according to Eq. (A.4.4) the difference between the integrals with two signs of ε is just $2\pi i f(a)$. This too agrees with the present analysis in terms of the integral I in Eq. (A.4.2) since in the difference of the two integrals the contribution along the real axis cancels due to opposite directions of travel except for the part near the pole where the difference of the two semicircles (one going above and going below the pole) is a circle around the pole.

Answers to Selected Exercises

Chapter 1

$$1.8.1. \quad (1) \quad |\omega=1\rangle \rightarrow \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad |\omega=2\rangle \rightarrow \frac{1}{(30)^{1/2}} \begin{bmatrix} -5 \\ -2 \\ 1 \end{bmatrix}, \quad |\omega=4\rangle \rightarrow \frac{1}{(10)^{1/2}} \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}$$

(2) No, no.

$$1.8.2. \quad (1) \quad \text{Yes}$$

$$(2) \quad |\omega=0\rangle \rightarrow \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad |\omega=1\rangle \rightarrow \frac{1}{2^{1/2}} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad |\omega=-1\rangle \rightarrow \frac{1}{2^{1/2}} \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

$$1.8.10. \quad \omega = 0, 0, 2; \lambda = 2, 3, -1.$$

Chapter 4

$$4.2.1. \quad (1) \quad 1, 0, -1$$

$$(2) \quad \langle L_x \rangle = 0, \langle L_x^2 \rangle = 1/2, \Delta L_x = 1/2^{1/2}$$

$$(3) \quad |L_x=1\rangle \rightarrow \begin{bmatrix} 1/2 \\ 1/2^{1/2} \\ 1/2 \end{bmatrix}, \quad |L_x=0\rangle \rightarrow \begin{bmatrix} -1/2^{1/2} \\ 0 \\ 1/2^{1/2} \end{bmatrix},$$
$$|L_x=-1\rangle \rightarrow \begin{bmatrix} 1/2 \\ -1/2^{1/2} \\ 1/2 \end{bmatrix}$$

$$(4) \quad P(L_x=1)=1/4, \quad P(L_x=0)=1/2, \quad P(L_x=-1)=1/4$$

$$(5) \quad |\psi\rangle \rightarrow \frac{1}{(1/4+1/2)^{1/2}} \begin{bmatrix} 1/2 \\ 0 \\ 1/2^{1/2} \end{bmatrix} = \text{projection of } |\psi\rangle \text{ on the } L_z^2=1 \text{ eigen-space. } P(L_z^2=1)=3/4. \text{ If } L_z \text{ is measured } P(L_z=1)=1/3, P(L_z=-1)=2/3. \text{ Yes, the state changes.}$$

(6) No. To see this right away note that if $\delta_1=\delta_2=\delta_3=0$, $|\psi\rangle=1|L_x=1\rangle$ and if $\delta_1=\delta_3=0$ and $\delta_2=\pi$, $|\psi\rangle=|L_x=-1\rangle$. [See answer to part (3).] The vectors $|\psi\rangle$ and $e^{i\theta}|\psi\rangle$ are physically equivalent only in the sense that they generate the same probability distribution for any observable. This does not mean that when the vector $|\psi\rangle$ appears as a part of a linear combination it can be multiplied by an arbitrary phase factor. In our example one can only say, for instance, that

$$\begin{aligned} |\psi'\rangle &= e^{-i\delta_1}|\psi\rangle \\ &= \frac{1}{2}|L_z=1\rangle + \frac{e^{i(\delta_2-\delta_1)}}{2^{1/2}}|L_z=0\rangle + \frac{e^{i(\delta_3-\delta_1)}}{2}|L_z=-i\rangle \end{aligned}$$

is physically equivalent to $|\psi\rangle$. Although $|\psi'\rangle$ has different coefficients from $|\psi\rangle$ in the linear expansion, it has the same “direction” as $|\psi\rangle$. In summary, then, the relative phases $\delta_2-\delta_1$ and $\delta_3-\delta_1$ are physically relevant but the overall phase is not, as you will have seen in the calculation of $P(L_x=0)$.

Chapter 5

$$5.4.2. \quad (a) \quad R=(maV_0)^2/(\hbar^4k^2+m^2a^2V_0^2); \quad T=1-R$$

$$(b) \quad T=(\cosh^2 2\kappa a + \alpha^2 \sinh^2 2\kappa a)^{-1} \text{ where } i\kappa \text{ is the complex wave number for } |x| \leq a \text{ and } \alpha=(V_0-2E)/[4E(V_0-E)]^{1/2}.$$

Chapter 7

$$7.4.2. \quad 0, \quad 0, \quad (n+1/2)\hbar/m\omega, \quad (n+1/2)m\omega\hbar, \quad (n+1/2)\hbar. \quad \text{Note that the recipe } m\omega \rightarrow (m\omega)^{-1} \text{ is at work here.}$$

$$7.4.5. \quad (1) \quad (1/2^{1/2})(|0\rangle e^{-i\omega t/2} + |1\rangle e^{-3i\omega t/2})$$

$$(2) \quad \langle X(t) \rangle = (\hbar/2m\omega)^{1/2} \cos \omega t, \quad \langle P(t) \rangle = -(m\omega\hbar/2)^{1/2} \sin \omega t$$

$$(3) \quad \langle \dot{X}(t) \rangle = (i\hbar)^{-1} \langle [X, H] \rangle = \langle P(t) \rangle / m, \quad \langle \dot{P}(t) \rangle = -m\omega^2 \langle X(t) \rangle. \quad \text{By eliminating } \langle \dot{P} \rangle \text{ we can get an equation for } \langle X(t) \rangle \text{ and vice versa and solve it using the initial values } \langle X(0) \rangle \text{ and } \langle P(0) \rangle, \text{ e.g., } \langle X(t) \rangle = \langle X(0) \rangle \cos \omega t + [\langle P(0) \rangle / m\omega] \sin \omega t.$$

Chapter 10

10.3.2. $3^{-1/2}[|334\rangle + |343\rangle + |433\rangle]$

Chapter 12

12.6.1. $E = -\hbar^2/2\mu a_0^2$, $V = -\hbar^2/\mu a_0 r$

Chapter 13

13.3.1. Roughly 200 MeV.

13.3.2. Roughly 1 Å.

Chapter 14

14.3.5. $M = \left(\frac{\alpha + \delta}{2}\right)I + \left(\frac{\beta + \gamma}{2}\right)\sigma_x + i\left(\frac{\beta - \gamma}{s}\right)\sigma_y + \left(\frac{\alpha - \delta}{2}\right)\sigma_z$

14.3.7. (1) $2^{1/4}(\cos \pi/8 + i(\sin \pi/8)\sigma_x)$.

(2) $2/3I - 1/3\sigma_x$.

(3) σ_x

14.4.4. Roughly 2×10^{-9} second.

14.4.6. $(e\hbar/2mc) \tanh(e\hbar B/2mckT)\mathbf{k}$

14.5.2. (1) Roughly one part in a million.

(2) 10^{10} G.

14.5.3. 1/2, 1/4, 0.

14.5.4. $\left(\frac{1 + \cos \theta}{2}\right)^2$

Chapter 15

15.2.2. (1) $\langle 1\ 1, 1/2(-1/2)|3/2\ 1/2\rangle = (1/3)^{1/2}$

$\langle 1\ 0, 1/2\ 1/2|3/2\ 1/2\rangle = (2/3)^{1/2}$

$\langle 1\ 1, 1/2(-1/2)|1/2\ 1/2\rangle = (2/3)^{1/2}$

$\langle 1\ 0, 1/2\ 1/2|1/2\ 1/2\rangle = -(1/3)^{1/2}$

$$(2) \quad |jm\rangle = |2, 1\rangle = 2^{-1/2}|m_1=1, m_2=0\rangle + 2^{-1/2}|m_1=0, m_2=1\rangle \\ |2, 0\rangle = 6^{-1/2}|1, -1\rangle + (\frac{2}{3})^{1/2}|0, 0\rangle + (\frac{1}{6})^{1/2}|-1, 1\rangle \\ |1, 1\rangle = 2^{-1/2}|1, 0\rangle - 2^{-1/2}|0, 1\rangle \\ |1, 0\rangle = 2^{-1/2}|1, -1\rangle - 2^{-1/2}|-1, 1\rangle \\ |0, 0\rangle = 3^{-1/2}|1, -1\rangle - 3^{-1/2}|0, 0\rangle + 3^{-1/2}|-1, 1\rangle$$

The others are either zero, obvious, or follow from Eq. (15.2.11).

$$15.2.6. \quad \mathbb{P}_+ = \frac{(2\mathbf{L} \cdot \mathbf{S})/\hbar^2 + l + 1}{2l + 1}, \quad \mathbb{P}_- = \frac{l - (2\mathbf{L} \cdot \mathbf{S})/\hbar^2}{2l + 1}$$

Chapter 16

$$16.1.2. \quad E(a_0) = 10E_0/\pi^2$$

$$16.1.3. \quad -ma_0^2V_0^2/\pi\hbar^2$$

$$16.1.4. \quad E(a_0) = \frac{1}{2}\hbar\omega(\frac{12}{11})^{1/2}$$

$$16.2.4. \quad \text{Roughly } 1.5 \times 10^{17} \text{ seconds or } 10^{10} \text{ years.}$$

Table of Constants

$$\hbar c = 1973.3 \text{ eV Å}$$

$$\alpha = e^2 / \hbar c = 1/137.04$$

$$mc^2 = 0.511 \text{ MeV} \quad (m \text{ is the electron mass})$$

$$Mc^2 = 938.28 \text{ MeV} \quad (M \text{ is the proton mass})$$

$$a_0 = \hbar^2 / mc^2 = 0.511 \text{ Å}$$

$$e\hbar / 2mc = 0.58 \times 10^{-8} \text{ eV/G} \quad (\text{Bohr magneton})$$

$$k = 8.62 \times 10^{-5} \text{ eV/K}$$

$$kT \approx 1/40 \text{ eV at } T = 300 \text{ K} \quad (\text{room temperature})$$

$$1 \text{ eV} = 1.6 \times 10^{-12} \text{ erg}$$

Mnemonics for Hydrogen

In the ground state,

$$v/c \equiv \beta = \alpha$$

$$E_1 = -T = -\frac{1}{2}mv^2 = -\frac{1}{2}mc^2\alpha^2$$

$$mva_0 = \hbar$$

In higher states, $E_n = E_1/n^2$.

Index

- Absorption spectrum, 368
Accidental degeneracy
 free-particle case, 426
 harmonic oscillator case, 352, 423
 hydrogen atom case, 359, 422
Actinides, 371
Active transformations, 29, 280
Adjoint, 13, 25, 26
Aharonov–Bohm effect, 497
Angular momentum
 addition of
 $\mathbf{J} + \mathbf{J}$, 408
 $\mathbf{L} + \mathbf{S}$, 414
 $\mathbf{S} + \mathbf{S}$, 403
 commutation rules, 319
 eigenfunctions, 324, 333
 eigenvalue problem of, 321
 spin, 373
 in three dimensions, 318
 in two dimensions, 308
Anticommutation relations, 640
Anti-Hermitian operators, 27
Antisymmetric states, 261
Anyons, 607
- Balmer series, 367
Basis, 6
Berry phase, 592
Berry potential, 603
Bohr magneton, 389
Bohr model, 364
Bohr radius, 244, 357
Bohr-Sommerfeld quantization rule, 448
Born approximation
 time-dependent, 529
 time-independent, 534
 validity of, 543
- Bose-Einstein statistics, 271
Bosons, 263
Bound states, 160, 445
 energy quantization in, 160
Bra, 11
Breit-Wigner form, 551
de Broglie waves, 112, 366
Double well, 616
 tunneling in, 616
- Canonical commutation rule, 131
Canonically conjugate operators, 69
Canonical momentum, 80
 electromagnetic case, 84
Canonical transformations
 active, 97
 introduction to, 92
 point transformations, 94
 regular, 97
Center of mass (CM), 85
Centrifugal barrier, 340
Characteristic equation, 33
Characteristic polynomial, 33
Chemical potential, 641
Classical limit, 179
Classical radius of electron, 364
Clebsch–Gordan coefficients, 412
Cofactor matrix, 656
Coherent states
 fermionic, 642
 oscillator, 607
 spin, 636
Collapse of state vector, 122, 139
Commutator, 20
Compatible variables, 129
Completeness relation, 23, 59
Complete set of observables, 133

- Complex numbers, 660
- Compton scattering, 123
- Compton wavelength, 246
 - electronic, 363
- Condon–Shortley convention, 410
- Configuration space, 76
- Consistency test
 - for three-dimensional rotations, 318
 - for translations, 306, 312
 - for translations and rotations, 310
- Coordinates
 - canonical, 94
 - center-of-mass, 85
 - cyclic, 81
 - relative, 85
- Correlation function, 628
 - connected, 634
- Correlation length, 629
- Correspondence principle, 197
- Coulomb scattering, 531
- Coupled mass problem, 46
- Creation operator, 205
- Cross section
 - in CM frame, 557
 - differential, 526, 529
 - for Gaussian potential, 533
 - for hard sphere, 549
 - in lab frame, 559
 - partial, 548
 - photoelectric, 506
 - Rutherford, 531
 - for Yukawa potential, 531
- Cyclotron frequency, 588
- Dirac equation
 - electromagnetic, 566
 - free particle, 565
- Dirac monopole, 605
- Dirac notation, 3
- Dirac string, 605
- Direct product
 - of operators, 250
 - spaces, 249
- Double-slit experiment, 108
 - quantum explanation of, 175
- Dual spaces, 11
- Ehrenfest's theorem, 180
- Eigenket, 30
- Eigenspace, 37
- Eigenvalue problem, 30
- Eigenvector, 30
- Einstein temperature, 220
- Electromagnetic field
 - interactions with matter, 83, 90, 499
 - quantization of, 506
 - review of, 492
- Ensemble
 - classical, 125
 - mixed, 133
 - quantum, 125
- Euclidean Lagrangian, 614
- Euler angles, 333
- Euler-Lagrange equations, 79
- Exchange operator, 278
- Exclusion principle, 264
- Expectation value, 127
- Dalgarno and Lewis method, 462
- Darwin term, 572
- Degeneracy, 38, 44, 120
- Density matrix, 133
- Derivative operator, 63
 - eigenvalue problem for, 66
 - matrix elements of, 64
- Destruction operator, 205
- Determinant, 29
- Diagonalization
 - of Hermitian operator, 40
 - simultaneous, 43
- Differential cross section, 526, 529
- Dipole approximation, 502
- Dipole moment, 463
- Dipole selection rule, 465
- Dirac delta function, 60
 - definition of, 60
 - derivatives of, 61
- Gaussian approximation for, 61
 - integral representation of, 63
- three-dimensional, 342
- Fermi-Dirac statistics, 270
- Fermionic oscillator, 640
 - thermodynamics of, 642
- Fermi's golden rule, 483
- Fermions, 263
- Field, 2
- Filling factor, 591
- Fine-structure constant, 362
- Fine-structure correction, 367, 466
- Fourier transform, 62
- Free-particle problem
 - cartesian coordinates, 151
 - spherical coordinates, 426
- Functional, 77
- Functions of operators, 54
- Gauge
 - Coulomb, 494
 - invariance, 493, 496
 - transformations, 493, 496
- Gaussian integrals, 659

- Gaussian potential, 533
- Generalized force, 80
- Generalized potential, 84
- Geometric phase, 593
- Gram–Schmidt theorem, 14
- Grassmann numbers, 642
- Green's function, 534
- Gyromagnetic ratio, 386

- Hamiltonian formulation, 86
- Hamilton's equations, 88
- Harmonic oscillator
 - classical, 83
 - fermionic, 640
 - isotropic, 260, 351
 - quantum, 185
 - in the coordinate basis, 189
 - in the energy basis, 202
 - energy levels of, 194
 - propagator for, 196
 - wave functions of, 195, 202
 - thermodynamics of, 219
 - three-dimensional, 260, 351
 - two-dimensional, 316
- Heisenberg picture, 147, 490
- Hermite polynomials, 490
- Hermitian operators, 27
 - diagonalization of, 40
 - simultaneous, 43
 - eigenbases of, 36
 - eigenvalues of, 35
 - eigenvectors of, 36
 - infinite-dimensional, 65
- Hilbert space, 67
 - bosonic, 265
 - fermionic, 265
 - normal mode problem in, 70
 - for two particles, 265
- 't Hooft, 619
- Hydrogen atom
 - degeneracy of, 359
 - energy levels of, 356
 - 21-cm line, 408
 - wave functions of, 356, 357
- Hyperfine interaction, 407

- Ideal measurement, 122
- Identical particles
 - bosons, 263
 - definition of, 260
 - fermions, 263
- Identity operator, 19
- Impact parameter, 523
- Improper vectors, 67
- Incompatible variables, 128
- Induced emission, 521

- Inelasticity, 554
- Infinite-dimensional spaces, 57
- Inner product, 8
- Inner product space, 7
- Inverse of operator, 20, 655
- Ionic bond, 370
- Irreducible space, 330
- Irreducible tensor operator, 418
- Ising model, 627

- Ket, 3
- Klein–Gordon equation, 564
- Kronecker's delta, 10

- Lagrangian, 76
 - for electromagnetic interactions, 83
- Laguerre polynomial, 356
- Lamb shift, 574
- Landau Level, 587, 588
- Laughlin wave function, 592
- Laughlin quasihole, 607
- Least action principle, 77
- Legendre transform, 87
- Linear independence, 4
- Linear operators, 18
- Lorentz spinor, 566
- Lowering operator
 - angular momentum, 322
 - for harmonic oscillator, 205
 - see also* Destruction operator
- Lowest Landau Level, 588
- Lyman series, 367

- Magnetic moment, 385
- Magnetic quantum number, 314
- Matrix elements, 20
- Matrix inversion, 655
- Mendeleev, 370
- Metastable states, 553
- Minimum uncertainty packet, 241
- Multielectron atoms, 369

- Negative absolute temperature, 394
- Norm, 9
- Normal modes, 52
- Number operator, 207
- Numerical estimates, 361

- Operators, 18
 - adjoint of, 25
 - anti-Hermitian, 27
 - conjugate, 69
 - derivatives of, 55
 - functions of, 54

- Hermitian, 27
- identity, 22
- infinite-dimensional, 63
- inverse of, 20
- linear, 18
- matrix elements of, 21
- product of, 20
- projection, 22
- unitary, 28
- Optical theorem, 548, 555
- Orthogonality, 9
- Orthogonal matrix, 28
- Orthonormality, 9
- Outer product, 23
- Paramagnetic resonance, 392
- Parity invariance, 297
- Partial wave
 - amplitude, 545
 - expansion, 545
- Particle in a box, 157, 259
- Paschen series, 367
- Passive transformation, 29, 280
- Path integral
 - coherent state, 607, 610
 - configuration space, 582
 - definition, 223
 - fermionic, 646
 - free particle, 225, 582
 - imaginary time, 614
 - phase space, 586
 - recipe, 223
 - and Schrödinger's equation, 229
 - statistical mechanics, 624
- Pauli equation, 568
- Pauli exclusion principle, 264
- Pauli matrices, 381
- Periodic table, 370
- Perturbations
 - adiabatic, 478
 - periodic, 482
 - sudden, 477
 - time-independent, 451
- Phase shift, 546
- Phase space, 88
- Phonons, 198
- Photoelectric effect, 111, 499
- Photons, 110, 198
 - quantum theory of, 516
- Physical Hilbert space, 67
- Pictures
 - Heisenberg, 147, 490
 - interaction (Dirac), 485
 - Schrödinger, 147, 484
- Planck's constant, 111
- Poisson brackets, 92
 - invariance of, 96
- Polarizability, 464
- P operator, 116
- Population inversion, 395
- Postulates, 115, 211
- Probability amplitude, 111, 121
- Probability current density, 166
- Probability density, 121
- Product basis, 403
- Projection operator, 23
- Propagator
 - for coupled masses, 51
 - Feynman's, 578
 - for free particle, 153
 - for Gaussian packet, 154
 - for harmonic oscillator, 615
 - for (classical) string, 72
- Proper vectors, 67
- Pseudospin, 639
- Quadrupole tensor, 425
- Quanta, 197
- Quantization of energy, 160
- Quantum Hall Effect (QHE), 589
- Radial equation
 - in three dimensions, 339
 - in two dimensions, 316
- Radial part of wave function
 - in three dimensions, 339
 - in two dimensions, 316
- Raising operator
 - for angular momentum, 222
 - for harmonic oscillator, 205
- Range of potential, 525
- Rare earth elements, 371
- Ray, 118
- Recursion relation, 193
- Reduced mass, 86
- Reduced matrix element, 420
- Reflection coefficient, 168
- Resonances, 550
- Rotations
 - generators of (classical), 100
 - generators of (quantum), 308
 - invariance under (classical), 100
 - invariance under (quantum), 310
- Runge–Lenz vector, 360, 422
- Rutherford cross section, 531
- Rydberg, 355
- Scattering
 - general theory, 523
 - of identical particles, 560
 - from step potential, 167
 - of two particles, 555
- Scattering amplitude, 527

- Schrödinger equation
 equivalence to path integral, 229
 time-dependent, 116, 143
 time-independent, 145
 Schrödinger picture, 147, 484
 Schwartz inequality, 16
 Selection rule
 angular momentum, 458, 459
 dipole, 459
 general, 458
 Shell, 370
 Singlet, 405
 S matrix
 definition of, 529
 partial wave, 547
 Spectroscopic notation, 350
 modified, 415
 Spherical Bessel functions, 348
 Spherical Hankel functions, 348
 Spherical harmonics, 335, 336
 Spherical Neumann functions, 348
 Spin, 325, 373
 Spinor, 375
 Spin-orbit interaction, 468
 Spin statistics theorem, 264
 Spontaneous decay, 517
 Spontaneous emission, 521
 Square-well potential, 164
 Stark effect, 459, 465
 Stationary states, 146
 Statistics, 264
 determination of, 269
 Stern-Gerlach experiment, 399
 Subspaces, 17
 Superposition principle, 117
 Symmetric states, 263
 Symmetries
 classical, 98
 quantum, 279
 spontaneous breakdown of, 620
- Tensor
 antisymmetric (ε_{ijk}), 319
 cartesian, 417
 irreducible, 418
 operator, 417
 quadrupole, 425
 second rank, 418
 spherical, 417
 Thermal wavelength, 625
 Thomas factor, 468, 571
 Thomas-Reiche-Kuhn rule, 457
 Time-ordered integral, 148
 Time-ordering symbol, 633, 651
 Time-reversal symmetry, 301
 Time translation invariance, 294
 Top state, 410
- Total S basis, 405
 Trace, 30
 Transformations, 29
 active, 29, 97, 280
 canonical, 92
 generator of, 99, 283
 identity, 98
 passive, 29, 280, 284
 point, 94
 regular, 97
 unitary, 27
 Translated state, 280
 Translation
 finite, 289
 generator of, 100, 283
 operator, 280
 Translational invariance
 implications of, 98, 292
 in quantum theory, 279
 Transmission coefficient, 168
 Transverse relaxation time, 395
 Triangle inequality, 116, 412
 Triplets, 405
 Tunneling, 175, 616
 Two-particle Hilbert space, 247
- Uncertainty, 128
 Uncertainty principle
 applications of, 198
 derivation of, 237
 energy-time, 245
 physical basis of, 140
 Uncertainty relation, 138
 Unitarity bound, 548
 Unitary operator, 27
 eigenvalues of, 39
 eigenvectors of, 39
- Variational method, 429
 Vector addition coefficients, 412
 Vectors
 components of, 6
 improper, 67
 inner product of, 8
 norm of, 9
 orthogonality of, 9
 outer product of, 25
 proper, 67
 Vector operator, 313
 Vector space
 axioms for, 2
 basis for, 6
 dimensionality of, 5
 field of, 2
 of Hilbert, 67
 infinite dimensional, 57
 subspace of, 17

- Virial theorem, 212
 - for hydrogen, 359, 471
- Wave functions, 121
- Wave-particle duality, 113
- Waves
 - interference of, 108
 - matter, 112
 - plane, 108
- Wick's theorem, 645
- Wigner-Eckart theorem, 420
- WKB approximation
 - and bound states, 445
- introduction to, 435
- and path integrals, 438
- three-dimensional, 449
- and tunneling, 444
- X* operator, 68
 - matrix elements of, 68
- Yukawa potential, 531
- Zeeman effect, 398
- Zero point energy, 198