21

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 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

$$H=\frac{P^2}{2m}+V(X) \tag{21.1.1}$$

The propagator is defined by

$$U(xt;x'0)\equiv U(x,x',t)=\langle x|\exp\left(-\frac{i}{\hbar}Ht\right)|x'\rangle \tag{21.1.2}$$

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

$$\exp\left(-\frac{i}{\hbar}Ht\right)=\left[\exp\left(-\frac{i}{\hbar}H\frac{t}{N}\right)\right]^N \tag{21.1.3}$$

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

$$\varepsilon=\frac{t}{N} \tag{21.1.4}$$

and consider the limit $N\to\infty$. Now we can write

$$\exp\left(-\frac{i\varepsilon}{\hbar}(P^2/2m+V(X))\right)\simeq\exp\left(-\frac{i\varepsilon}{2m\hbar}P^2\right)\cdot\exp\left(-\frac{i\varepsilon}{\hbar}V(X)\right) \tag{21.1.5}$$

because of the fact that

$$e^Ae^B=e^{A+B+1/2[A,B]+\cdots} \tag{21.1.6}$$

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 $\varepsilon\to 0$ 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

$$\underbrace{\langle x|\exp\left(-\frac{i\varepsilon}{2m\hbar}P^2\right)\cdot\exp\left(-\frac{i\varepsilon}{\hbar}V(X)\right)\cdot\exp\left(-\frac{i\varepsilon}{2m\hbar}P^2\right)\cdot\exp\left(-\frac{i\varepsilon}{\hbar}V(X)\right)\ldots|x'\rangle}_{N\text{ times}} \tag{21.1.7}$$

The next step is to introduce the resolution of the identity:

$$I=\int_{-\infty}^{\infty}dx|x\rangle\langle x| \tag{21.1.8}$$

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)

$$\begin{aligned}U(x_3,x_0,t)=\int\prod_{n=1}^{2}dx_n&\langle x_3|\exp\left(-\frac{i\varepsilon}{2m\hbar}P^2\right)\exp\left(-\frac{i\varepsilon}{\hbar}V(X)\right)|x_2\rangle\\&\times\langle x_2|\exp\left(-\frac{i\varepsilon}{2m\hbar}P^2\right)\exp\left(-\frac{i\varepsilon}{\hbar}V(X)\right)|x_1\rangle\\&\times\langle x_1|\exp\left(-\frac{i\varepsilon}{2m\hbar}P^2\right)\exp\left(-\frac{i\varepsilon}{\hbar}V(X)\right)|x_0\rangle\end{aligned} \tag{21.1.9}$$

Consider now the evaluation of the matrix element

$$\langle x_n|\exp\left(-\frac{i\varepsilon}{2m\hbar}P^2\right)\cdot\exp\left(-\frac{i\varepsilon}{\hbar}V(X)\right)|x_{n-1}\rangle \tag{21.1.10}$$

When the rightmost exponential operates on the ket to its right, the operator X gets replaced by the eigenvalue x_{n-1}. Thus,

$$\begin{aligned}&\langle x_n|\exp\left(-\frac{i\varepsilon}{2m\hbar}P^2\right)\cdot\exp\left(-\frac{i\varepsilon}{\hbar}V(X)\right)|x_{n-1}\rangle\\&=\langle x_n|\exp\left(-\frac{i\varepsilon}{2m\hbar}P^2\right)|x_{n-1}\rangle\exp\left(-\frac{i\varepsilon}{\hbar}V(x_{n-1})\right)\end{aligned} \tag{21.1.11}$$

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

$$\langle x_n|\exp\left(\frac{-i\varepsilon}{2m\hbar}P^2\right)|x_{n-1}\rangle=\left[\frac{m}{2\pi i\hbar\varepsilon}\right]^{1/2}\exp\left[\frac{im(x_n-x_{n-1})^2}{2\hbar\varepsilon}\right] \tag{21.1.12}$$

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

$$I=\int_{-\infty}^{\infty}\frac{dp}{2\pi\hbar}\,|p\rangle\langle p| \tag{21.1.13}$$

where the plane wave states have a wave function given by

$$\langle x|p\rangle=e^{ipx/\hbar} \tag{21.1.14}$$

which explains the measure for the p integration.

Resuming our derivation, we now have

$$\begin{aligned}\langle x_n|\exp\left(-\frac{i\varepsilon}{2m\hbar}P^2\right)\cdot\exp\left(-\frac{i\varepsilon}{\hbar}V(X)\right)|x_{n-1}\rangle\\ =\left[\frac{m}{2\pi i\hbar\varepsilon}\right]^{1/2}\exp\left[\frac{im(x_n-x_{n-1})^2}{2\hbar\varepsilon}\right]\exp\left(-\frac{i\varepsilon}{\hbar}V(x_{n-1})\right)\end{aligned} \tag{21.1.15}$$

Collecting all such factors (there are just two more in this case with $N=3$), we can readily see that for general N

$$\begin{aligned}U(x_N,x_0,t)=\left(\frac{m}{2\pi i\hbar\varepsilon}\right)^{1/2}\left[\int\prod_{n=i}^{N-1}\left(\frac{m}{2\pi i\hbar\varepsilon}\right)^{1/2}dx_n\right]\\ \times\exp\left[\sum_{n=1}^{N}\frac{im(x_n-x_{n-1})^2}{2\hbar\varepsilon}-\frac{i\varepsilon}{\hbar}V(x_{n-1})\right]\end{aligned} \tag{21.1.16}$$

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 $\exp(iS/\hbar)$:

$$\begin{aligned}\exp\left[\sum_{n=1}^{N}\frac{im(x_n-x_{n-1})^2}{2\hbar\varepsilon}-\frac{i\varepsilon}{\hbar}V(x_{n-1})\right]\\ =\exp\frac{i}{\hbar}\varepsilon\sum_{n=1}^{N}\left[\frac{m(x_n-x_{n-1})^2}{2\varepsilon^2}-V(x_{n-1})\right]\end{aligned} \tag{21.1.17}$$

We can go back to the continuum notation and write all this as follows:

$$U(x, x', t) = \int [\mathscr{D}x] \exp\left[\frac{i}{\hbar}\int_0^t \mathscr{L}(x, \dot{x})\, dt\right] \tag{21.1.18}$$

where

$$\int [\mathscr{D}x] = \lim_{N\to\infty} \left(\frac{m}{2\pi i\hbar\varepsilon}\right)^{1/2} \int\left[\prod_{n=i}^{N-1} \left(\frac{m}{2\pi i\hbar\varepsilon}\right)^{1/2} dx_n\right] \tag{21.1.19}$$

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

$$\underbrace{\langle x_N| \exp\left(-\frac{i\varepsilon}{2m\hbar} P^2\right)\cdot \exp\left(-\frac{i\varepsilon}{\hbar} V(X)\right)\cdot \exp\left(-\frac{i\varepsilon}{2m\hbar} P^2\right)\cdot \exp\left(-\frac{i\varepsilon}{\hbar} V(X)\right)\ldots |x_0\rangle}_{N \text{ times}} \tag{21.1.20}$$

Let us now introduce resolutions of the identity between *every exponential* and the next. We need two versions

$$I = \int_{-\infty}^{\infty} dx |x\rangle\langle x| \tag{21.1.21}$$

$$I = \int_{-\infty}^{\infty} \frac{dp}{2\pi\hbar} |p\rangle\langle p| \tag{21.1.22}$$

where the plane wave states have a wave function given by

$$\langle x| p\rangle = e^{ipx/\hbar} \tag{21.1.23}$$

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

$$\begin{aligned} U(x_3, x_0, t) = \int [\mathscr{D}p\mathscr{D}x] &\langle x_3| \exp\left(-\frac{i\varepsilon}{2m\hbar} P^2\right)|p_3\rangle \\ &\times \langle p_3| \exp\left(-\frac{i\varepsilon}{\hbar} V(X)\right)|x_2\rangle\langle x_2| \exp\left(-\frac{i\varepsilon}{2m\hbar} P^2\right)|p_2\rangle \\ &\times \langle p_2| \exp\left(-\frac{i\varepsilon}{\hbar} V(X)\right)|x_1\rangle\langle x_1| \exp\left(-\frac{i\varepsilon}{2m\hbar} P^2\right)|p_1\rangle \\ &\times \langle p_1| \exp\left(-\frac{i\varepsilon}{\hbar} V(X)\right)|x_0\rangle \end{aligned} \tag{21.1.24}$$

where

$$\int [\mathscr{D}p\mathscr{D}x] = \underbrace{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}}_{2N-1 \text{ times}} \prod_{n=1}^{N}\frac{dp_n}{2\pi\hbar}\prod_{n=1}^{N-1} dx_n \tag{21.1.25}$$

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

$$U(x, x', t) = \int [\mathscr{D}p\mathscr{D}X] \exp\left[\sum_{i=1}^{N}\left(\frac{-i\varepsilon}{2m\hbar}p_n^2 + \frac{i}{\hbar}p_n(x_n - x_{n-1}) - \frac{i\varepsilon}{\hbar}V(x_{n-1})\right)\right] \tag{21.1.26}$$

This formula derived for $N=3$ is obviously true for any N. In the limit $N\to\infty$, i.e., $\varepsilon\to 0$, we write schematically in continuous time (upon multiplying and dividing the middle term by ε), the following continuum version:

$$U(x, x', t) = \int [\mathscr{D}p\mathscr{D}x] \exp\left[\frac{i}{\hbar}\int_0^t [p\dot{x} - \mathscr{H}(x, p)]\, dt\right] \tag{21.1.27}$$

where $\mathscr{H} = p^2/2m + V(x)$ 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

the part that depends on just p's and do the integrals:

$$\prod_1^N \int_{-\infty}^{\infty} \frac{dp_n}{2\pi\hbar} \exp\left[\left(\frac{-i\varepsilon}{2m\hbar} p_n^2 + \frac{i}{\hbar} p_n (x_n - x_{n-1})\right)\right]$$
$$= \prod_1^N \left(\frac{m}{2\pi i\hbar\varepsilon}\right)^{1/2} \exp\left[\frac{im(x_n - x_{n-1})^2}{2\hbar\varepsilon}\right] \tag{21.1.28}$$

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

$$\mathbf{A} = \frac{B}{2}(-y\mathbf{i} + x\mathbf{j}) \tag{21.1.29}$$

we obtained a Hamiltonian

$$H = \frac{[P_x + qYB/2c]^2}{2\mu} + \frac{[P_y - qXB/2c]^2}{2\mu} \tag{21.1.30}$$

You were asked to verify that

$$Q = \frac{(cP_x + qYB/2)}{qB} \qquad P = (P_y - qBX/2c) \tag{21.1.31}$$

were canonical variables with $[Q, P] = i\hbar$. It followed that H was given by the formula

$$H = \frac{P^2}{2\mu} + \frac{1}{2}\mu\omega_0^2 Q^2 \tag{21.1.32}$$

and had a harmonic oscillator spectrum with spacing $\hbar\omega_0$, where

$$\omega_0 = qB/\mu c \tag{21.1.33}$$

is the *cyclotron frequency*. In terms of

$$a=\left(\frac{\mu\omega_0}{2\hbar}\right)^{1/2}Q+i\left(\frac{1}{2\mu\omega_0\hbar}\right)^{1/2}P \tag{21.1.34}$$

and its adjoint, we can write

$$H=[a^\dagger a+\tfrac{1}{2}]\hbar\omega_0 \tag{21.1.35}$$

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

$$P'=\frac{(cP_x-qYB/2)}{qB} \qquad Q'=(P_y+qBX/2c) \tag{21.1.36}$$

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

$$a|0\rangle=0 \tag{21.1.37}$$

which becomes in the coordinate representation

$$\left[\frac{\partial}{\partial z^*}+\frac{qB}{4\hbar c}z\right]\psi_0(z,z^*)=0 \tag{21.1.38}$$

wherein we have switched to complex coordinates

$$z=x+iy \qquad z^*=x-iy \tag{21.1.39}$$

If we make the ansatz

$$\psi_0(z,z^*)=\exp\left[-\frac{qB}{4\hbar c}zz^*\right]u(z,z^*) \tag{21.1.40}$$

we find the beautiful result

$$\frac{\partial}{\partial z^*}u(z,z^*)=0 \tag{21.1.41}$$

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 $[z^m | m=0, 1, 2, \ldots]$ 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

$$\psi_{0,m}=z^m \exp\left[-\frac{qB}{4\hbar c} zz^*\right] \tag{21.1.42}$$

I now make the following assertions:

- For large m the probability density for the particle is concentrated at some radius $r_m=\sqrt{2m}r_0$ 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

$$N=\frac{\pi R^2 B}{\Phi_0} \tag{21.1.43}$$

where the numerator is the flux through the sample and the denominator is the flux quantum of Eq. (18.4.39):

$$\Phi_0=\frac{2\pi\hbar c}{q} \tag{21.1.44}$$

*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

$$S=\int\left[\frac{\mu}{2}(\dot{x}^2+\dot{y}^2)+\frac{qB}{2c}(-y\dot{x}+x\dot{y})-V(x,y)\right]dt \tag{21.1.45}$$

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 $\mu\to 0$.* (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

$$S_{LLL}=\int\left[\frac{qB}{c}x\dot{y}-V(x,y)\right]dt \tag{21.1.46}$$

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 $\partial\mathscr{L}/\partial\dot{y}=(qB/c)x\equiv\bar{x}$ 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:

$$\dot{y}=\frac{\partial V}{\partial x}\qquad \dot{\bar{x}}=-\frac{\partial V}{\partial y} \tag{21.1.47}$$

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 $\mu\to 0$ 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 νN of electrons where $1/\nu$ is an odd integer. (This is one of the cases where the experiments show surprising results.) We say the system has a *filling factor* ν meaning it has ν times the maximum allowed number of particles in the LLL. The fully filled LLL is thus given by $\nu = 1$. 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:

$$u_\nu = \prod_{i=1}^{\nu N} \prod_{j=1}^{i-1} (z_i - z_j)^{1/\nu} \tag{21.1.50}$$

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)

$$z_1^{1/\nu \cdot \nu N} = z_1^N$$

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 $1/\nu$ 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 $1/\nu$).

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 nth instantaneous eigenket of $H(R(t))$ at a time t if it started out in the nth eigenstate of $H(R(0))$.

What is the solution to the Schrödinger equation in this approximation? Here is a reasonable guess:

$$|\psi(t)\rangle = \exp\left(-\frac{i}{\hbar}\int_0^t E_n(t')\,dt'\right)|n(t)\rangle \tag{21.1.51}$$

where

$$H(t)|n(t)\rangle = E_n(t)|n(t)\rangle \tag{21.1.52}$$

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:

$$|\psi(t)\rangle = c(t) \exp\left(-\frac{i}{\hbar}\int_0^t E_n(t')\,dt'\right)|n(t)\rangle \tag{21.1.53}$$

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:

$$\left(i\hbar\frac{\partial}{\partial t} - H(t)\right)|\psi(t)\rangle = 0 \tag{21.1.54}$$

What 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):

$$\dot{c}(t) = -c(t)\langle n(t)|\frac{d}{dt}|n(t)\rangle \tag{21.1.55}$$

with a solution

$$c(t) = c(0)\exp\left(-\int_0^t \langle n(t')|\frac{d}{dt'}|n(t')\rangle\,dt'\right) = c(0)\,e^{i\gamma} \tag{21.1.56}$$

$$\gamma = i\int_0^t \langle n(t')|\frac{d}{dt'}|n(t')\rangle\,dt' \tag{21.1.57}$$

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

$$|n'(t)\rangle = e^{i\chi(t)}\,|n(t)\rangle \tag{21.1.58}$$

then we find

$$i\langle n'(t)|\,\frac{d}{dt}\,|n'(t)\rangle = i\langle n(t)|\,\frac{d}{dt}\,|n(t)\rangle - \frac{d\chi(t)}{dt} \tag{21.1.59}$$

which suggests that we can 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:

$$H(T) = H(0) \tag{21.1.60}$$

Now it is no longer obvious that we can get rid of the extra phase. We find

$$i\oint \langle n'(t)|\,\frac{d}{dt}\,|n'(t)\rangle\,dt = i\oint \langle n(t)|\,\frac{d}{dt}\,|n(t)\rangle\,dt - (\chi(T)-\chi(0)) \tag{21.1.61}$$

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:

$$\exp\left(-\int_0^t \langle n(t')|\frac{d}{dt'}|n(t')\rangle\, dt'\right)$$

$$=\exp\left(\frac{i}{\hbar}\, i\hbar \int_0^t \langle n(t')|\frac{d}{dt'}|n(t')\rangle\, dt'\right) \tag{21.1.62}$$

$$=\exp\left(\frac{i}{\hbar}\int_0^t i\hbar\langle n(R(t'))|\frac{d}{dR}|n(R(t'))\rangle\frac{dR}{dt'}\,dt'\right) \tag{21.1.63}$$

$$=\exp\left(\frac{i}{\hbar}\int_0^t A^n(R)\frac{dR}{dt'}\,dt'\right) \quad \text{where} \tag{21.1.64}$$

$$A^n(R)=i\hbar\langle n(R)|\frac{d}{dR}|n(R)\rangle \tag{21.1.65}$$

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:

$$|n(R)\rangle \to e^{i\chi(R)}|n(R)\rangle \tag{21.1.66}$$

$$A^n(R)\to A^n(R)-\hbar\frac{d\chi}{dR} \tag{21.1.67}$$

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:

$$I=\int dR\sum_n |R, n(R)\rangle\langle n(R), R| \tag{21.1.68}$$

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 $H_e(R, r, p)$

$$H_e(R, r, p)|R, n(R)\rangle = E_n(R)|R, n(R)\rangle \tag{21.1.69}$$

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:

$$I \simeq \int dR|R, n(R)\rangle\langle n(R), R| \tag{21.1.70}$$

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

$$\langle n(R(t+\varepsilon)), R(t+\varepsilon)| \exp\left[-\frac{i\varepsilon}{\hbar} H(R, P)\right] \exp\left[-\frac{i\varepsilon}{\hbar} H_e(R, r, p)\right] |n(R(t)), R(t)\rangle \tag{21.1.71}$$

The nuclear part, sandwiched between nuclear coordinate eigenstates, will give the usual factor

$$\langle R(t+\varepsilon)| \exp\left[-\frac{i\varepsilon}{\hbar} H(R, P)\right] |R(t)\rangle$$
$$= \sqrt{\frac{m}{2\pi\hbar i\varepsilon}} \exp\left[\frac{i\varepsilon}{\hbar}\left[\frac{M}{2\varepsilon^2}(R(t+\varepsilon) - R(t))^2 - V(R)\right]\right] \tag{21.1.72}$$

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:

$$\langle n(R(t+\varepsilon))|n(R(t))\rangle \equiv \langle n(R')|n(R)\rangle \tag{21.1.73}$$

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

$$\psi(R',\varepsilon)=\left(\frac{m}{2\pi\hbar i\varepsilon}\right)^{1/2}\int_{-\infty}^{\infty}e^{im\eta^2/2\hbar\varepsilon}\langle n(R')|n(R'+\eta)\rangle\psi(R'+\eta,0)\,d\eta \tag{21.1.74}$$

The exponential allows η to fluctuate by (recall Eq. (8.5.6))

$$|\eta|\simeq\left(\frac{2\pi\hbar\varepsilon}{m}\right)^{1/2} \tag{21.1.75}$$

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:

$$\psi(R'+\eta,0)=\psi(R',0)+\eta\frac{\partial\psi}{\partial\eta}+\frac{\eta^2}{2}\frac{\partial^2\psi}{\partial\eta^2}+\cdots \tag{21.1.76}$$

$$\langle n(R')|n(R'+\eta)\rangle=1+\eta\langle n|\partial n\rangle+\frac{\eta^2}{2}\langle n|\partial^2 n\rangle+\cdots \tag{21.1.77}$$

where all derivatives are taken at the point R' and $|\partial n\rangle$ 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'

$$i\hbar(\psi(R,\varepsilon)-\psi(R,0))=\varepsilon\left[-\frac{\hbar^2}{2m}\frac{\partial^2\psi}{\partial R^2}-\frac{\hbar^2}{m}\langle n|\partial n\rangle\frac{\partial\psi}{\partial R}-\frac{\hbar^2}{2m}\langle n|\partial^2 n\rangle\psi\right] \tag{21.1.78}$$

*Exercise 21.1.7.** Provide the missing steps leading to the above equation.

The Hamiltonian can be read off the above:

$$H=\frac{1}{2m}(P-A^n)^2+\Phi^n \tag{21.1.79}$$

$$A^n=i\hbar\langle n|\partial n\rangle \tag{21.1.80}$$

$$\Phi^n=\frac{\hbar^2}{2m}[\langle\partial n|\partial n\rangle-\langle\partial n|n\rangle\langle n|\partial n\rangle] \tag{21.1.81}$$

*Exercise 21.1.8.** Providing the missing steps. Use $\langle n|\partial n\rangle=-\langle\partial n|n\rangle$ which follows from $\partial\langle n|n\rangle=0$. The potential Φ^n arises from adding and subtracting the $(A^n)^2$ term which isn't there to begin with.

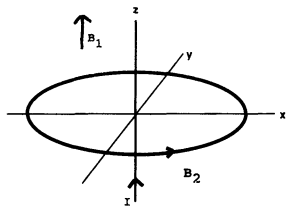

Figure 21.1. The field B_2 and electron motion are along the circle. The particle spin is up or down the local magnetic field which is the sum of B_1 and B_2. The current I produces B_2.

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)

$$S=\frac{im(R'-R)^2}{2\hbar\varepsilon}+\frac{i}{\hbar}(R'-R)A''\left(\frac{R+R'}{2}\right)$$
$$-\frac{(R'-R)^2}{2}\langle\partial n|(I-|n\rangle\langle n|)|\partial n\rangle \qquad (21.1.82)$$

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 **R**, **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 $B_1\mathbf{k}$ 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 **B**-space. Thus the particle's Hamiltonian is

$$H=\frac{L^2}{2I} \tag{21.1.83}$$

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

$$E_m=\hbar^2m^2 \qquad m=0, \pm1, \pm2 \ldots \tag{21.1.84}$$

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, **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:

$$H=L^2-C\boldsymbol{\sigma}\cdot\mathbf{B}(\phi) \tag{21.1.85}$$

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

$$E_m=\hbar^2m^2\mp CB \tag{21.1.86}$$

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 of course 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

$$|\theta\phi\rangle=\begin{bmatrix}\cos\dfrac{\theta}{2}\\ i\sin\dfrac{\theta}{2}\,e^{i\phi}\end{bmatrix} \tag{21.1.87}$$

(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

$$A^+(\phi) = i\hbar\langle\theta\phi|\frac{\partial}{\partial\phi}|\theta\phi\rangle = -\hbar\sin^2\frac{\theta}{2} \tag{21.1.88}$$

which is independent of ϕ, and that the scalar Berry potential is

$$\Phi = \frac{\hbar^2\sin^2\theta}{4} \tag{21.1.89}$$

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 $L \to L - A^+$, it follows that if we solve

$$\left[-i\hbar\frac{\partial}{\partial\phi} - A^+\right]\psi = \lambda\psi \tag{21.1.90}$$

the energy is given by

$$E^+ = \lambda^2 - BC \tag{21.1.91}$$

The orbital eigenfunctions are once again

$$\psi = e^{im\phi} \qquad m = 0, \pm 1, \pm 2, \ldots \tag{21.1.92}$$

so that

$$\lambda = m\hbar - A^+ = \left(m + \sin^2\frac{\theta}{2}\right)\hbar \tag{21.1.93}$$

and the energy of the spin up state is

$$E^+ = \left(m + \sin^2\frac{\theta}{2}\right)^2\hbar^2 - BC \tag{21.1.94}$$

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 $m \to -m$.

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 $\psi(R)$:

$$|\psi\rangle = \int \psi(R)|R, n(R)\rangle\, dR \tag{21.1.95}$$

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

$$(H\psi)(R') \equiv \langle R', n(R')|H|\psi\rangle \tag{21.1.96}$$

$$= \int \langle R', n(R')|H|R, n(R)\rangle\langle R, n(R)|\psi\rangle\, dR \tag{21.1.97}$$

$$= \int \langle R', n(R')|H|R, n(R)\rangle\psi(R)\, dR \tag{21.1.98}$$

Let

$$H = P^2/2M + V(R) + H_f(r, p, R) \tag{21.1.99}$$

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

$$\langle R'|\frac{P^2}{2M}|R\rangle = -\frac{\hbar^2}{2M}\delta''(R'-R) \tag{21.1.100}$$

and insert it into Eq. (21.1.98) to obtain

$$\begin{aligned}(H\psi)(R') &= -\frac{\hbar^2}{2M}\int \langle n(R')|n(R)\rangle\delta''(R'-R)\psi(R)\, dR \\ &= -\frac{\hbar^2}{2M}\langle n(R')|[|\partial^2 n(R)\rangle\psi(R) + 2|\partial n(R)\rangle\partial\psi(R) + |n(R)\rangle\partial^2\psi(R)]_{R=R'} \\ &= -\frac{\hbar^2}{2M}[\langle n|\partial^2 n\rangle\psi(R') + 2\langle n|\partial n\rangle\,\partial\psi(R') + \partial^2\psi(R')]\end{aligned} \tag{21.1.101}$$

where ∂ denotes derivatives respect to R'. It is now straightforward to show that the operator on the right-hand side is indeed the one on 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

$$H = -\boldsymbol{\sigma}\cdot\mathbf{B} \tag{21.1.102}$$

then the Berry potential lives in **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 **B** space as **R** space, but you should not forget this fact. So we write

$$H = -\boldsymbol{\sigma}\cdot\mathbf{R} \tag{21.1.103}$$

Every point in **R** space defines a possible spin Hamiltonian. We have managed to define in this space a vector potential. It is derived from the nth quantum state of the above Hamiltonian and is given by

$$\mathbf{A}^n = i\hbar\langle n(\mathbf{R})|\nabla|n(\mathbf{R})\rangle \tag{21.1.104}$$

What is its curl? To figure this out, we need a little groundwork. Using

$$0 = \nabla\langle n|H|m\rangle \qquad m\neq n \tag{21.1.105}$$

we find on differentiating all three factors and shifting a derivative from bra to ket at the cost of sign change (thanks to $\nabla\langle n|n\rangle = 0$),

$$\langle n|\nabla|m\rangle = \frac{\langle n|(\nabla H)|m\rangle}{E_m - E_n} \tag{21.1.106}$$

It is now easy to find a formula for the field tensor F_{ij} associated with the Berry potential:

$$\begin{aligned} F_{ij}^n &= \partial_i A_j^n - \partial_j A_i^n \\ &= i\hbar[\partial_i\langle n|\partial_j n\rangle - \partial_j\langle n|\partial_i n\rangle] \\ &= i\hbar \sum_{m\neq n} \frac{\langle n|(\partial_i H)|m\rangle\langle m|(\partial_j H)|n\rangle - \langle n|(\partial_j H)|m\rangle\langle m|(\partial_i H)|n\rangle}{(E_m - E_n)^2} \end{aligned}$$

$$\left(\partial_j = \frac{\partial}{\partial R_j}\right) \tag{21.1.107}$$

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 $\langle n|\boldsymbol{\sigma}|n\rangle = \pm\hat{\mathbf{R}}$ (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 $\mathscr{B}_k^n = F_{ij}^n$ (where the indices i, j, k run cyclically):

$$\mathscr{B}^n = \mp\hbar \frac{\hat{\mathbf{R}}_k}{2R^2} \tag{21.1.108}$$

which 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 **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 $\mathscr{B}_k^n$ 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

$$H=\sum_{\mu=0}^{3}\sigma_\mu f_\mu \tag{21.1.109}$$

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

$$E=f_0\pm\sqrt{f_x^2+f_y^2+f_z^2}. \tag{21.1.110}$$

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

$$H=f_0(\mathbf{f})U+\boldsymbol{\sigma}\cdot\mathbf{f} \tag{21.1.111}$$

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

$$\partial_i H=\partial_i f_0 I+\sigma_i \tag{21.1.112}$$

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 **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)

$$A(x)\,dx=A(y)\,dy \tag{21.1.113}$$

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.)

Execise 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

$$|+, \theta, \phi\rangle = \begin{bmatrix} \cos\dfrac{\theta}{2} \\ \sin\dfrac{\theta}{2}\, e^{i\phi} \end{bmatrix}$$

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

$$e^{4\pi iqg/\hbar c}$$

as per Eq. (18.4.38). (Think of an Aharaonov–Bohm experiment in which a particle goes on either side of the string.) If this factor is to be unobservable we require that

$$q=\frac{\hbar nc}{2g}$$

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 $A \rightarrow A + \hbar c/q\, \partial\chi$; (3) the change in χ around a closed loop must be an integral multiple of 2π.)

In the Berry phase problem the vector potential has qg/c, the factor multiplying $\mathbf{A}$ in the Hamiltonian, equal to unity; and the monopole we found corresponds to $n=-1/2$.

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

$$u_{qh} = \prod_{i=1,}^{\nu N} (z_i - z_0) u_\nu \tag{21.1.114}$$

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 νq, 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 ν 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 $e^{i\nu\pi}$ 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 Shapir 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

$$|z\rangle = \exp[z a^\dagger]|0\rangle \tag{21.1.115}$$

where $|0\rangle$ is the ground state of the oscillator. If we recall that

$$|n\rangle = \frac{(a^\dagger)^n}{\sqrt{n!}}|0\rangle \tag{21.1.116}$$

we see that

$$|z\rangle = \sum_0^\infty \frac{z^n}{\sqrt{n!}}|n\rangle \tag{21.1.117}$$

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

$$a|z\rangle = z|z\rangle \tag{21.1.118}$$

as follows:

$$a|z\rangle = a\sum_0^\infty \frac{z^n}{\sqrt{n!}}|n\rangle \tag{21.1.119}$$

$$= \sum_1^\infty \frac{z^n\sqrt{n}}{\sqrt{n!}}|n-1\rangle \tag{21.1.120}$$

$$= z|z\rangle \tag{21.1.121}$$

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

$$\langle z| = \langle 0|\exp[z^* a] \tag{21.1.122}$$

is seen to obey

$$\langle z|a^\dagger = \langle z|z^* \tag{21.1.123}$$

Let us now consider the inner product

$$\langle z_2|z_1\rangle = \langle 0|\exp[z_2^* a]\exp[z_1 a^\dagger]|0\rangle \tag{21.1.124}$$

If we use the identity

$$e^A e^B = e^B e^A e^{[A,B]} \tag{21.1.125}$$

which is valid if $[A, B]$ commutes with A and B, we see

$$\langle z_2 | z_1 \rangle = e^{z^*_2 z_1} \tag{21.1.126}$$

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

$$I = \int \frac{dx\, dy}{\pi} |z\rangle\langle z|\, e^{-z^* z} \equiv \int \frac{dz\, dz^*}{2\pi i} |z\rangle\langle z|\, e^{-z^* z} \tag{21.1.127}$$

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

$$|z\rangle = \sum_0^\infty \frac{z_n}{\sqrt{n!}} |n\rangle \tag{21.1.128}$$

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 A3 for the Gamma function integral.

Since the coherent states are right eigenstates of a and left eigenstates of $a^\dagger$,

$$\langle z_2 | : H(a^\dagger, a) : | z_1 \rangle = \langle z_2 | H(z_2^*, z_1) | z_1 \rangle \tag{21.1.129}$$

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,

$$H=\hbar\omega a^\dagger a \tag{21.1.130}$$

(we drop the constant zero-point energy for this discussion), then

$$U(t)|z\rangle = U(t)\exp[a^\dagger z]U^\dagger(t)U(t)|0\rangle = \exp[a^\dagger e^{-i\omega t}z]|0\rangle = |z\,e^{-i\omega t}\rangle \tag{21.1.131}$$

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

$$\psi_z(x)=\langle x|z\rangle=\left(\frac{m\omega}{\pi\hbar}\right)^{1/4} e^{-z^2/2}\,e^{-(m\omega/2\hbar)x^2}\,e^{\sqrt{(2m\omega/\hbar)}zx} \tag{21.1.132}$$

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 $|z\rangle\to|z\,e^{-i\omega t}\rangle$. 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:

$$U(z_N,z_0,t)=\langle z_N|U(t)|z_0\rangle=\langle z_N|z_0\,e^{-i\omega t}\rangle=\exp[z_N^*z_0\,e^{-i\omega t}] \tag{21.1.133}$$

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

$$\begin{aligned}&\langle z_4|U^4(t/4)|z_0\rangle\\ &=\int[\mathscr{D}z\mathscr{D}z^*]<z_4|\left(I-\frac{i\varepsilon}{\hbar}H(a^\dagger a)\right)|z_3\rangle\,e^{-z_3^*z_3}\,\langle z_3|\left(I-\frac{i\varepsilon}{\hbar}H(a^\dagger a)\right)|z_2\rangle\,e^{-z_2^*z_2}\,\langle z_2|\\ &\quad\times\left(I-\frac{i\varepsilon}{\hbar}H(a^\dagger a)\right)|z_1\rangle\,e^{-z_1^*z_1}\,\langle z_1|\left(I-\frac{i\varepsilon}{\hbar}H(a^\dagger a)\right)|z_0\rangle\end{aligned}$$

where

$$[\mathscr{D}z\mathscr{D}z^*]=\prod_1^{N-1}\frac{dz_i\,dz_i^*}{2\pi i}=\prod_1^{N-1}\frac{dx_i\,dy_i}{\pi} \tag{21.1.134}$$

A typical factor we run into is as follows:

$$\langle z_{n+1}|\left(I-\frac{i\varepsilon}{\hbar}H(a^\dagger a)\right)|z_n\rangle=\exp\left(-\frac{i\varepsilon}{\hbar}H(z^*_{n+1},z_n)\right)<z_{n+1}|z_n\rangle \tag{21.1.135}$$

$$=\exp\left(-\frac{i\varepsilon}{\hbar}H(z^*_{n+1},z_n)\right)\exp(z^*_{n+1}z_n) \tag{21.1.136}$$

where we have treated ε as infinitesimal since eventually it will be, as we let $N\to\infty$. If we assemble all the exponential factors together, there will be a piece related to the Hamiltonian which clearly gives a factor

$$\exp\left(-\frac{i}{\hbar}\int_0^t \hbar\omega z^*(t)z(t)\,dt\right) \tag{21.1.137}$$

in the continuum notation, where z_n has become $z(t=n\varepsilon)$. (We also made the approximation $H(z^*(t+\varepsilon),z(t))\simeq H(z^*(t),z(t))$.)

The other factor in the exponent is

$$z^*_4z_3-z^*_3z_3+z^*_3z_2-z^*_2z_2+z^*_2z_1-z^*_1z_1+z^*_1z_0 \tag{21.1.138}$$

$$=(z^*_4-z^*_3)z_3+(z^*_3-z^*_2)z_2+(z^*_2-z^*_1)z_1+z^*_1z_0 \tag{21.1.139}$$

which we write in continuum notation as

$$\frac{i}{\hbar}\left[\int_0^t(-i\hbar)\frac{dz^*}{dt'}z\,dt'\right]+z^*(0)z(0) \tag{21.1.140}$$

where $z(0)=z_0$ and $z^*(0)=\lim_{\varepsilon\to0}z^*(\varepsilon)$. 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

$$\left[\frac{i}{\hbar}\int_0^t(i\hbar)\left(z^*\frac{dz}{dt}\right)dt\right]+z^*(t)z(t) \tag{21.1.141}$$

where $z(t)$ is again extraneously introduced as a limit $z(t)=\lim_{\varepsilon\to0}z(t-\varepsilon)$.

One usually sees the two schemes averaged to give the following final form of the continuum result:

$$\langle z_f|U(t)|z_i\rangle=\exp\left[\frac{z^*_fz_f+z^*_iz_i}{2}+\frac{i}{\hbar}\int_0^t\left[\frac{i\hbar}{2}\left(z^*\frac{dz}{dt}-\frac{dz^*}{dt}z\right)-H(z^*,z)\right]dt\right] \tag{21.1.142}$$

We will use the asymmetric form obtained by doing an integration by parts:

$$\langle z_f|\, U(t)|z_i\rangle = \exp\left[z_f^* z_f + \frac{i}{\hbar}\left[\int_0^t \left[i\hbar z^* \frac{dz}{dt} - H(z^*, z)\right] dt\right]\right] \tag{21.1.143}$$

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 $\hbar \to 0$, 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 $(p\dot{x} - \mathscr{H})$ 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

$$[Z, i\hbar Z^\dagger] = i\hbar \tag{21.1.144}$$

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

$$\dot{z} = \frac{\partial(\hbar\omega z^* z)}{\partial(i\hbar z^*)} = -i\omega z \tag{21.1.145}$$

which is solved to give

$$z(t) = z(0)\, e^{-i\omega t} \tag{21.1.146}$$

Similarly, we find

$$z^*(t) = z^*(0)\, e^{i\omega t} \tag{21.1.147}$$

To evaluate

$$\langle z_f|\, U(T)|z_i\rangle \tag{21.1.148}$$

in the semiclassical approximation, we need to find a solution that obeys

$$z(0) = z_i \tag{21.1.149}$$

$$z^*(T) = z_f^* \tag{21.1.150}$$

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

$$z(t)=z_i\,e^{-i\omega t} \tag{21.1.151}$$

$$z^*(t)=z_f^*\,e^{i\omega(t-T)} \tag{21.1.152}$$

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

$$\langle z_f|\,U(T)|z_i\rangle=\exp(z_f^*z_i\,e^{-i\omega T}) \tag{21.1.153}$$

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 $\langle z_f|\,U(T)|z_i\rangle=\exp[z_fz_i\,e^{-i\omega T}]$. 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*

$$U(\tau)=\exp\left(-\frac{1}{\hbar}H\tau\right) \tag{21.2.1}$$

This is obtained by setting

$$t=-i\tau \tag{21.2.2}$$

in the usual propagator. In other words, if the Schrödinger equation had been

$$-\hbar \frac{d}{d\tau} |\psi(t)\rangle = H|\psi(\tau)\rangle \tag{21.2.3}$$

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:

$$U(\tau) = \sum |n\rangle\langle n| \exp\left(-\frac{1}{\hbar} E_n \tau\right) \tag{21.2.4}$$

where

$$H|n\rangle = E_n|n\rangle \tag{21.2.5}$$

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

$$U(x, x', \tau) = \langle x| U(\tau)|x'\rangle \tag{21.2.6}$$

We can write down a path integral for it following exactly the same steps as before. The final answer in continuum notation is

$$\langle x| U(\tau)|x'\rangle = U(x, x', \tau) = \int [\mathscr{D}x] \exp\left[-\frac{1}{\hbar}\int_0^t \mathscr{L}_E(x, \dot{x})\, d\tau\right] \tag{21.2.7}$$

$$\int [\mathscr{D}x] = \lim_{N\to\infty} \left(\frac{m}{2\pi\hbar\varepsilon}\right)^{1/2} \prod_0^{N-1} \left(\frac{m}{2\pi\hbar\varepsilon}\right)^{1/2} dx_i \tag{21.2.8}$$

$$\mathscr{L}_E = \frac{m}{2}\left(\frac{dx}{d\tau}\right)^2 + V(x) \tag{21.2.9}$$

where $\varepsilon = \tau/N$ and $\mathscr{L}_E$ 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 $x^2 - c^2t^2$ now becomes $x^2 + c^2\tau^2$. Notice that $\mathscr{L}_E$ is the *sum* of the euclidean kinetic energy and real-time potential energy. *Thus the particle obeying the euclidean equations of motion will see 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

$$\langle x|\, U(\tau)|x'\rangle = \left(\frac{m}{2\pi\hbar\tau}\right)^{1/2} \exp\left[-\frac{m(x-x')^2}{2\hbar\tau}\right] \tag{21.2.10}$$

If someone gave us this propagator, we could get the Minkowski space answer by setting

$$\tau = it \tag{21.2.11}$$

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 $|0\rangle$:

$$\lim_{\tau\to\infty} \langle x|\, U(\tau)|x'\rangle = \lim_{\tau\to\infty} \sum \langle x|n\rangle\langle n|x'\rangle \exp\left(-\frac{1}{\hbar} E_n\tau\right) \tag{21.2.12}$$

$$\simeq \langle x|0\rangle\langle 0|x'\rangle \exp\left(-\frac{1}{\hbar} E_0\tau\right) \tag{21.2.13}$$

$$= \psi_0(x)\psi_0^*(x') \exp\left(-\frac{1}{\hbar} E_0\tau\right) \tag{21.2.14}$$

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

$$U(x, x', \tau) = A(\tau) \exp\left(-\frac{m\omega}{2\hbar \sinh \omega\tau}\left[(x^2+x'^2)\cosh \omega\tau - 2xx'\right]\right) \tag{21.2.15}$$

obtained, say by analytic continuation from real times of the answer in Exercise (8.6.2). Note that as $\tau\to\infty$ this beomes proportional to the product of ground state wave functions. The prefactor is left to the following exercise.

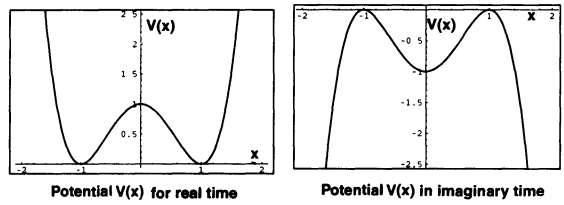

Figure 21.2. The double-well potential in real and imaginary time.

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

$$V(x) = A^2(x^2 - a^2)^2 \tag{21.2.16}$$

The classical minima are at

$$x_{L/R} = \pm a \tag{21.2.17}$$

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 $\hbar\omega$ 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.

Assume they have been orthogonalized by a Gram–Schmidt procedure.) The approximate Hamiltonian looks like this in this subspace:

$$H=\begin{bmatrix} E_0 & 0 \\ 0 & E_0 \end{bmatrix} \tag{21.2.18}$$

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

$$E=\pm H_{+-} \tag{21.2.19}$$

and the eigenstates will be $|S/A\rangle$, the sum and difference of $|\pm\rangle$.

Consider

$$\langle a|U(\tau)|-a\rangle=\langle a|\exp\left(-\frac{1}{\hbar}H\tau\right)|-a\rangle \tag{21.2.20}$$

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:

$$\langle a|\exp\left(-\frac{1}{\hbar}H\tau\right)|-a\rangle\simeq 0-\frac{1}{\hbar}\tau\langle a|H|-a\rangle+\mathcal{O}\tau^2 \tag{21.2.21}$$

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,

$$\langle a|\exp\left(-\frac{1}{\hbar}H\tau\right)|-a\rangle\simeq\exp\left(-\frac{1}{\hbar}S_{\text{cl}}\right) \tag{21.2.22}$$

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

$$E_e=\frac{m}{2}\left(\frac{dx}{d\tau}\right)^2-V(x) \tag{21.2.23}$$

Using this we can solve for the trajectory by quadrature:

$$\int_{x_1}^{x_2}\frac{\sqrt{m}\,dx}{\sqrt{2(E_e+V(x))}}=\int_{t_1}^{t_2}d\tau \tag{21.2.24}$$

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 that we are after.

Let us consider the trajectory that has $E_e=0$. It is given by doing the above integral with $E_e=0$:

$$x(\tau)=a\tanh\left[\sqrt{\frac{2}{m}}Aa\tau\right] \tag{21.2.25}$$

Notice that in this trajectory the particle starts out at the left *maximum* (Fig. 21.2) at $\tau\to-\infty$ and rolls down the hill and only reaching of the right maximum as $\tau\to\infty$. 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, if we consider points which are 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),

$$S_{\rm cl}=\int(T+V)\,d\tau=\int 2T\,d\tau=\int_{-a}^{a}p(x)\,dx=\int_{-a}^{a}\sqrt{2mV(x)}\,dx \tag{21.2.26}$$

and the tunneling amplitude is (ignoring prefactors)

$$\langle a|U|-a\rangle\simeq\exp\left(-\frac{1}{\hbar}\int_{-a}^{a}\sqrt{2mV(x)}\,dx\right) \tag{21.2.27}$$

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 $\tau\to\infty$.) 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 $x=0$:

$$I(a)=\int_{x_1}^{x_2} e^{-\alpha x^2}\,dx \tag{21.2.28}$$

If $\alpha>0$ we can assume the limtis can be pushed to infinity and the answer approximated by

$$I(a)\simeq\sqrt{\pi/\alpha} \tag{21.2.29}$$

What happens when $\alpha\to 0$? The approximate answer diverges but we know the real answer is

$$I=\lim_{\alpha\to 0}\int_{x_1}^{x_2} e^{-\alpha x^2}\,dx=x_2-x_1 \tag{21.2.30}$$

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

$$\delta x(\tau)=x(\tau-\tau_0)-x(\tau) \tag{21.2.31}$$

is a fluctuation that costs no extra action, i.e., the Gaussian that is supposed to damp out this fluctuation has $\alpha \to 0$. 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 αs.

With the prefactor τ in front of

$$\langle -a|\, U(\tau)|a\rangle = \tau \exp\left(-\frac{1}{\hbar} S_{\text{cl}}\right) \tag{21.2.32}$$

we are ready to compare to

$$\langle a|\exp\left(-\frac{1}{\hbar} H\tau\right)|-a\rangle \simeq 0 - \frac{1}{\hbar}\tau\langle a|H|-a\rangle + \mathcal{O}r^2 \tag{21.2.33}$$

and read off

$$H_{-+} \simeq -\exp\left(-\frac{1}{\hbar} S_{\text{cl}}\right) \tag{21.2.34}$$

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 $e^{-(1/\hbar)Ht}$.)

It is now clear that with H_{-+} negative, the new eigenstates and energies are as follows:

$$|S\rangle = \sqrt{\tfrac{1}{2}}[|+a\rangle + |-a\rangle] \qquad E_S = -\exp\left(-\frac{1}{\hbar} S_{\text{cl}}\right) \tag{21.2.35}$$

$$|A\rangle = \sqrt{\tfrac{1}{2}}\,[|+a\rangle - |-a\rangle] \qquad E_S = \exp\left(-\frac{1}{\hbar} S_{\text{cl}}\right) \tag{21.2.36}$$

Spontaneous Symmetry Breaking

Why are we interested in a term that vanishes exponentially fast as $\hbar \to 0$ 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

$$|S/A\rangle = \frac{[|+a\rangle \pm |-a\rangle]}{\sqrt{2}} \tag{21.2.37}$$

$$\Pi|S/A\rangle = \pm|S/A\rangle \tag{21.2.38}$$

which are eigenstates of parity. Indeed, in quantum theory the relation

$$[\Pi, H]=0 \tag{21.2.39}$$

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 violated over small times, barrier penetration is forbidden if the barrier is infinite. This means in particular that the symmetric and antisymmetric functions will never by 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 $x \to x + 1$. 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 $x \to x + 1$

$$T|n\rangle = |n+1\rangle \tag{21.2.40}$$

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 forms 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:

$$|S\rangle = \frac{1}{\sqrt{N}} \sum_{0}^{N} |n\rangle \tag{21.2.41}$$

The details are left to the following exercise.

Exercise 21.2.2. (Very important)
Assume that

$$H = \sum_{1}^{N} E_0 |n\rangle\langle n| - t(|n\rangle\langle n+1| + |n+1\rangle\langle n|) \tag{21.2.42}$$

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

$$|\theta\rangle = \frac{1}{\sqrt{N}} \sum_{0}^{N} e^{in\theta} |n\rangle \tag{21.2.43}$$

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 $V_0 \to \infty$ in the periodic potential, there really will be N degenerate minima with particles living in any one minimum trapped there for ever. 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 $e^{-(1/\hbar)S_{\text{cl}}}$.

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:

$$Z=\sum_n e^{-\beta E_n} \tag{21.2.44}$$

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 nth eigenstate of the Hamiltonian H. We can rewrite this as

$$Z=\text{Tr}\, e^{-\beta H} \tag{21.2.45}$$

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

$$Z=\int_{-\infty}^{\infty} \langle x|\, e^{-\beta H}\, |x\rangle\, dx \tag{21.2.46}$$

The integrand is of course familiar to us now:

$$\langle x|\, \exp(-\beta H)|x\rangle = \langle x|\, \exp\left(-\frac{1}{\hbar}\beta\hbar H\right)|x\rangle = U(x, x, \beta\hbar) \tag{21.2.47}$$

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

$$Z=\sum_0^{\infty} e^{-\beta\hbar\omega(n+1/2)} \tag{21.2.48}$$

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

$$Z(\beta)=\int dx \int_x^x [\mathscr{D}x] \exp\left[-\frac{1}{\hbar}\int_0^{\beta\hbar}\left[\frac{m}{2}\left(\frac{dx}{d\tau}\right)^2+V(x(\tau))\right]d\tau\right] \tag{21.2.49}$$

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 $\beta\hbar \to 0$ 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

$$\simeq \exp\left(-\frac{1}{\hbar} m(\Delta x/\beta\hbar)^2 \beta\hbar\right) \tag{21.2.50}$$

from which it follows that

$$\Delta x \simeq \sqrt{\frac{\beta}{m}}\,\hbar \tag{21.2.51}$$

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

$$Z(\beta) \simeq \int dx\, e^{-\beta V(x)} \int_x^x [\mathscr{D}x] \exp\left[-\frac{1}{\hbar}\int_0^{\beta\hbar}\left[\frac{m}{2}\left(\frac{dx}{d\tau}\right)^2\right] d\tau\right] \tag{21.2.52}$$

$$= \int dx\, e^{-\beta V(x)} \sqrt{\frac{m}{2\pi\hbar\beta\hbar}}$$

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:

$$Z = A \int dx \int dp \exp\left[-\beta\left(\frac{p^2}{2m} + V(x)\right)\right] \tag{21.2.53}$$

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

$$A = \frac{1}{2\pi\hbar} \tag{21.2.54}$$

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 is 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

$$Z=\int_{-\infty}^{\infty}\prod_{1}^{N-1} dx_i \exp\left(-\frac{1}{kT}E(x_0,\ldots,x_N)\right) \tag{21.2.55}$$

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

$$E=\sum_{1}^{N-1}[K_1(x_n-x_{n-1})^2+K_2x_n^2] \tag{21.2.56}$$

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

$$U(x_0,x_N,\tau)=\int_{-\infty}^{\infty}\prod_{1}^{N-1} dx_i \exp\left[-\frac{1}{\hbar}\sum_{1}^{N-1}\varepsilon\left(\frac{m}{2}\frac{(x_n-x_{n-1})^2}{\varepsilon^2}+\frac{m\omega^2}{2}x_n^2\right)\right] \tag{21.2.57}$$

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 $\varepsilon\to 0$ in the quantum problem, the parameters of the classical problem must take some limiting values ($K_1\to\infty$ and $K_2\to 0$ 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

$$\langle x_{12}x_{78}\rangle = \frac{\int_{-\infty}^{\infty}\prod_1^{N-1} dx_i\, x_{12}x_{78}\, e^{-\beta E(x_0,\ldots,x_N)}}{\int_{-\infty}^{\infty}\prod_1^{N-1} dx_i\, e^{-\beta E(x_0,\ldots,x_N)}} \tag{21.2.58}$$

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

$$Z=\sum_{s_i=\pm 1}\exp\left[\sum_{i=0}^{N-1} K(s_i s_{i+1}-1)\right] \tag{21.2.59}$$

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:

$$t_i = s_i s_{i+1} \tag{21.2.60}$$

It is clear that given s_0 and t_i, we can reconstruct the state of the system. Thus, we can write

$$Z=\sum_{t_i}\exp\left[\sum_{i=0}^{N-1}K(t_i-1)\right]=\sum_{t_i}\prod_i e^{K(t_i-1)} \tag{21.2.61}$$

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)

$$Z=2(1+e^{-2K})^N \tag{21.2.62}$$

One is generally interested in the free energy per site in the *thermodynamic limit* $N\to\infty$:

$$f(K)=\lim_{N\to\infty}\frac{1}{N}\ln Z \tag{21.2.63}$$

(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

$$f(K)=\ln(1+e^{-2K}) \tag{21.2.64}$$

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):

$$\langle s_j s_i\rangle=\frac{\sum_{s_k} s_j s_i \exp\left[\sum_k K(s_k s_{k+1}-1)\right]}{Z} \tag{21.2.65}$$

for $j>i$. Using the fact that $s_i^2\equiv 1$, we can write

$$s_j s_i=s_i s_{i+1}s_{i+1}s_{i+2}\cdots s_{j-1}s_j=t_i t_{i+1}\cdots t_{j-1} \tag{21.2.66}$$

Thus

$$\langle s_i s_j\rangle=\langle t_i\rangle\langle t_{i+1}\rangle\cdots\langle t_{j-1}\rangle \tag{21.2.67}$$

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

$$\langle t\rangle=\frac{1\,e^{0\cdot K}-1\,e^{-2K}}{e^{0\cdot K}+e^{-2K}}=\tanh K \tag{21.2.68}$$

so that finally

$$\langle s_j s_i \rangle = (\tanh K)^{j-i} = \exp[(j-i) \ln \tanh K] \tag{21.2.69}$$

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 $N \to \infty$, 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

$$\lim_{j-i \to \infty} \langle s_i s_j \rangle \to e^{-(j-i)/\xi} \tag{21.2.70}$$

Thus in our problem

$$\xi^{-1} = -\ln \tanh K \tag{21.2.71}$$

(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 $|j-i| \to \infty$. 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

$$Z = \sum_{s_i} \prod_i e^{K(s_i s_{i+1} - 1)} \tag{21.2.72}$$

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

$$Z=\sum_{s_i} T_{s_N s_{N-1}} \cdots T_{s_2 s_1} T_{s_1 s_0} \tag{21.2.73}$$

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

$$T_{++}=T_{--}=1,\ T_{+-}=T_{-+}=\exp(-2K)$$

Thus this matrix, called the *Transfer Matrix*, is given by

$$T=I+e^{-2K}\sigma_1 \tag{21.2.74}$$

and

$$Z=\langle s_N|T^N|s_0\rangle \tag{21.2.75}$$

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)

$$Z=\sum_{s_0 s_N} \langle s_N|T^N|s_0\rangle \tag{21.2.76}$$

If we consider periodic boundary conditions where $s_0=s_N$ and one sums over these,

$$Z=\mathrm{Tr}\ T^N \tag{21.2.77}$$

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

$$T=\lambda_0|0\rangle\langle 0|+\lambda_1|1\rangle\langle 1| \tag{21.2.78}$$

where $|i\rangle$, λ_i $[i=0, 1]$ are the eigenvectors (assumed orthonormal) and eigenvalues of T, then

$$T^N=\lambda_0^N|0\rangle\langle 0|+\lambda_1^N|1\rangle\langle 1| \tag{21.2.79}$$

Assuming λ_0 is the bigger of the two eigenvalues,

$$T^N \lim_{N\to\infty} \simeq \lambda_0^N|0\rangle\langle 0|\left(1+\mathcal{O}\left(\frac{\lambda_1}{\lambda_0}\right)^N\right) \tag{21.2.80}$$

and

$$Z \simeq \langle s_N|0\rangle\langle 0|s_0\rangle \lambda_0^N \left(1+\mathcal{O}\left(\frac{\lambda_1}{\lambda_0}\right)^N\right) \tag{21.2.81}$$

and the free energy per site in the infinite volume limit,

$$f=\lambda_0+\frac{1}{N}\ln(\langle s_N|0\rangle\langle 0|s_0\rangle)+\cdots \tag{21.2.82}$$

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

$$e^{K^*\sigma_1}=\cosh K^*+\sinh K^*\sigma_1 \tag{21.2.83}$$

$$=\cosh K^*(I+\tanh K^*\sigma_1) \tag{21.2.84}$$

where K^* is presently unrelated to K; in particular, it is not the conjugate! If we choose

$$\tanh K^*=e^{-2K} \tag{21.2.85}$$

we see from Eq. (21.2.72) that up to a prefactor $\cosh K^*$,

$$T=e^{K^*\sigma_1} \tag{21.2.86}$$

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

$$|0\rangle, |1\rangle=\frac{1}{\sqrt{2}}\begin{bmatrix}1\\ \pm 1\end{bmatrix} \tag{21.2.87}$$

Suppose we write

$$T=e^{-H} \tag{21.2.88}$$

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:

$$H=-K^*\sigma_1 \tag{21.2.89}$$

$$f=-E_0=K^* \tag{21.2.90}$$

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,

$$\langle s_j s_i\rangle=\frac{\langle s_N|T^{N-j}\sigma_3 T^{j-i}\sigma_3 T^i|s_0\rangle}{\langle s_N|T^N|s_0\rangle} \tag{21.2.91}$$

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 Nth 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*

$$\sigma_3(n)=T^{-n}\sigma_3 T^n \tag{21.2.92}$$

In terms of these

$$\langle s_j s_i\rangle=\frac{s_N|T^N\sigma_3(j)\sigma_3(i)|s_0\rangle}{\langle s_N|T^N|s_0\rangle} \tag{21.2.93}$$

Consider now the limit as $N\to\infty$, *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

$$T^\alpha\simeq|0\rangle\langle 0|\lambda_0^\alpha \qquad \alpha=N,\ N-j,\ i \tag{21.2.94}$$

In this limit, we have from Eq. (21.2.91)

$$\langle s_j s_i\rangle=\frac{\langle s_N|0\rangle\langle 0|\lambda_0^{N-j}\sigma_3 T^{j-i}\sigma_3\lambda_0^i|0\rangle\langle 0|s_0\rangle}{\langle s_N|0\rangle\lambda_0^N\langle 0|s_0\rangle}=\langle 0|\sigma_3(j)\sigma_3(i)|s_0\rangle \tag{21.2.95}$$

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,

$$\langle s_j s_i\rangle=\langle 0|\mathcal{T}(\sigma_3(j)\sigma_3(i))|0\rangle \tag{21.2.96}$$

where the *time-ordering symbol* $\mathcal{T}$ will order the operators with time increasing from the right to left:

$$\mathcal{T}(\sigma_3(j)\sigma_3(i)) = \theta(j-i)(\sigma_3(j)\sigma_3(i)) + \theta(i-j)(\sigma_3(i)\sigma_3(j)) \qquad (21.2.97)$$

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

$$\langle s_i \rangle = \langle 0|\sigma_3(0)|0\rangle \qquad (21.2.98)$$

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 $\hbar \to 0$, the particle in the double-well seeks the minimum of the euclidean action:

$$S_E = \int \left(\frac{m}{2} (dx/d\tau)^2 + V(x(\tau)) \right) d\tau \qquad (21.2.99)$$

  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 eignstates 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

$$H=-K^*\sigma_1 \tag{21.2.100}$$

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 $\hbar\to 0$ 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*

$$\langle s_j s_i\rangle_c \equiv \langle s_j s_i\rangle - \langle s\rangle^2 = \langle 0|T^{-j}\sigma_3(0)T^{j-i}|1\rangle\langle 1|\sigma_3(0)T^i|0\rangle \tag{21.2.101}$$

$$=\left(\frac{\lambda_1}{\lambda_0}\right)^{j-i}|\langle 0|\sigma_3|1\rangle|^2 \tag{21.2.102}$$

$$=e^{-2K^*(j-i)}|\langle 0|\sigma_3|1\rangle|^2 \tag{21.2.103}$$

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

$$\xi^{-1}=E_1-E_0\equiv m \tag{21.2.104}$$

  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, $\langle s_j s_i\rangle \to \langle s_j\rangle\langle s_i\rangle$. 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 $\langle s_i s_j\rangle$ 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 $\langle s_i\rangle\langle s_j\rangle \to \langle s\rangle^2$. 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 magnetic 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 $df/dh \to 0$ as $h \to 0$. 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 $\langle s_j s_i \rangle$ 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

$$I = \sum_{-S}^{S} |S_z\rangle\langle S_z| \tag{21.3.1}$$

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*

$$|\Omega\rangle \equiv |\theta, \phi\rangle = U(R(\Omega))|SS\rangle \tag{21.3.2}$$

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

$$\langle SS|\mathbf{S}|SS\rangle = \mathbf{k}S \tag{21.3.3}$$

it is clear (say by considering $U^\dagger \mathbf{S} U$) that

$$\langle \Omega|\mathbf{S}|\Omega\rangle = S(\mathbf{i}\sin\theta\cos\phi + \mathbf{j}\sin\theta + \mathbf{k}\cos\theta) \tag{21.3.4}$$

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 nth power of the spin operator will have an expectation value equal to the nth 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

$$\langle \Omega_2|\Omega_1\rangle = \left(\cos\frac{\theta_2}{2}\cos\frac{\theta_1}{2} + e^{i(\phi_1-\phi_2)}\sin\frac{\theta_2}{2}\sin\frac{\theta_1}{2}\right)^{2S} \tag{21.3.5}$$

The result is obviously true for $S=1/2$, given that the up spinor along the direction $\theta\phi$ is

$$|\Omega\rangle \equiv |\theta\phi\rangle = \cos\frac{\theta}{2}|1/2, 1/2\rangle + e^{i\phi}\sin\frac{\theta}{2}|1/2, -1/2\rangle \tag{21.3.6}$$

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

$$|SS\rangle = |1/2, 1/2\rangle \otimes |1/2, 1/2\rangle \otimes \cdots |1/2, 1/2\rangle \tag{21.3.7}$$

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

$$I = \frac{2S+1}{4\pi}\int d\Omega|\Omega\rangle\langle\Omega| \tag{21.3.8}$$

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:

$$\cdots\langle\Omega(t+\varepsilon)|I-\frac{i\varepsilon}{\hbar}H(\mathbf{S})|\Omega(t)\rangle\cdots \tag{21.3.9}$$

We work to order ε. Since H already has a factor of ε in front of it, we set

$$\langle\Omega(t+\varepsilon)|-\frac{i\varepsilon}{\hbar}H(\mathbf{S})|\Omega(t)\rangle\simeq-\frac{i\varepsilon}{\hbar}\langle\Omega(t)|H(\mathbf{S})|\Omega(t)\rangle\equiv-i\varepsilon\mathscr{H}(\Omega) \tag{21.3.10}$$

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 $\hbar\mathscr{H}(\Omega)$ in the preceding equation.

Next we turn to the product

$$\langle\Omega(t+\varepsilon)|\Omega(t)\rangle\simeq 1-i\varepsilon S(1-\cos\theta)\dot{\phi}\simeq e^{iS(\cos\theta-1)\dot{\phi}\varepsilon} \tag{21.3.11}$$

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:

$$\langle\Omega_f|U(t)|\Omega_i\rangle=\int\mathscr{D}\Omega\exp\left[i\int_{t_1}^{t_2}[S\cos\theta\dot{\phi}-\mathscr{H}(\Omega)]\,dt\right] \tag{21.3.12}$$

where a total derivative in ϕ has been dropped and $\int\mathscr{D}\Omega$ 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 $\exp[im(x'-x)^2/2\hbar\varepsilon]$. *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 $S\to\infty$. 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 ($S\to\infty$) in which one uses this formalism.

We now consider two simple applications. First let

$$H=\hbar S_z \tag{21.3.13}$$

We know the allowed eigenvalues are $\hbar(-S, -S+1, \ldots, S)$. Let us derive this from the continuum path integral.

Given $\langle\Omega|H|\Omega\rangle = \hbar S \cos\theta$, it follows that $\mathscr{H} = S\cos\theta$, and that the functional integral is

$$\left[\int \mathscr{D}\cos\theta \mathscr{D}\phi\right] \exp\left[iS\int(\cos\theta\dot{\phi} - \cos\theta)\,dt\right] \tag{21.3.14}$$

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 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

$$\oint p\,dq = 2\pi n\hbar \tag{21.3.15}$$

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

$$\oint E\,d\phi = 2\pi n S^{-1} \tag{21.3.16}$$

since S^{-1} *plays the role of* $\hbar$. The allowed energies are

$$E_n = \frac{n}{S} \qquad [-S \le n \le S] \tag{21.3.17}$$

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. Eventhough 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

$$E = n\hbar\omega \tag{21.3.18}$$

(dropping zero point motion) allowed one to introduce the notion of quanta. Rather than saying the oscillator was in the nth 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 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:

$$H_0 = \Psi^\dagger \Psi \Omega_0 \tag{21.3.19}$$

What distinguishes this problem from the bosonic one are the *anticommutation relations*:

$$\{\Psi^\dagger, \Psi\} = \Psi^\dagger \Psi + \Psi \Psi^\dagger = 1 \tag{21.3.20}$$

$$\{\Psi, \Psi\} = \{\Psi^\dagger, \Psi^\dagger\} = 0 \tag{21.3.21}$$

Note that the last equation tells us

$$\Psi^{\dagger 2} = \Psi^2 = 0 \tag{21.3.22}$$

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*

$$N = \Psi^\dagger \Psi \tag{21.3.23}$$

obeys

$$N^2 = \Psi^\dagger \Psi \Psi^\dagger \Psi = \Psi^\dagger (1 - \Psi^\dagger \Psi)^\dagger \Psi = N \tag{21.3.24}$$

Thus the eigenvalues of N can only be 0 or 1. The corresponding normalized eigenstates obey

$$N|0\rangle = 0|0\rangle \tag{21.3.25}$$

$$N|1\rangle = 1|1\rangle \tag{21.3.26}$$

We will now prove that

$$\Psi^\dagger|0\rangle = |1\rangle \tag{21.3.27}$$

$$\Psi|1\rangle = |0\rangle \tag{21.3.28}$$

As for the first,

$$N\Psi^\dagger|0\rangle = \Psi^\dagger \Psi \Psi^\dagger|0\rangle = \Psi^\dagger(1 - \Psi^\dagger \Psi)|0\rangle = \Psi^\dagger|0\rangle \tag{21.3.29}$$

which shows that $\Psi^\dagger|0\rangle$ has $N = 1$. Its norm is unity:

$$\|\Psi^\dagger|0\rangle\|^2 = \langle 0|\Psi\Psi^\dagger|0\rangle = \langle 0|(1 - \Psi^\dagger\Psi)|0\rangle = \langle 0|0\rangle = 1 \tag{21.3.30}$$

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

$$H_0 = \Omega_0 \Psi^\dagger \Psi \tag{21.3.31}$$

has eigenvalues 0 and Ω_0.

We will work not with H_0 but with

$$H = H_0 - \mu N \tag{21.3.32}$$

where μ is called the *chemical potential.* For the oscillator, since

$$H = (\Omega_0 - \mu)\Psi^\dagger \Psi \tag{21.3.33}$$

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

$$Z = \mathrm{Tr}\, e^{-\beta(H_0 - \mu N)} = e^{A(\mu,\beta)} \tag{21.3.34}$$

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:

$$Z = 1 + e^{-\beta(\Omega_0 - \mu)} \tag{21.3.35}$$

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

$$\langle N \rangle = \frac{1}{\beta Z}\frac{\partial Z}{\partial \mu} = \frac{1}{\beta}\frac{\partial \ln Z}{\partial \mu} = \frac{1}{e^{\beta(\Omega_0 - \mu)} + 1} \tag{21.3.36}$$

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)

$$\langle N \rangle = \theta(\mu - \Omega_0) \tag{21.3.37}$$

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

$$\Psi|\psi\rangle = \psi|\psi\rangle \tag{21.3.38}$$

The eigenvalue ψ is a peculiar beast because if we act once more with Ψ we find

$$\psi^2 = 0 \tag{21.3.39}$$

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

$$|\psi\rangle = |0\rangle - \psi|1\rangle \tag{21.3.40}$$

where ψ is a Grassmann number. This state obeys:

$$\Psi|\psi\rangle = \Psi|0\rangle - \Psi\psi|1\rangle \tag{21.3.41}$$

$$= 0 + \psi\Psi|1\rangle \tag{21.3.42}$$

$$= \psi|0\rangle \tag{21.3.43}$$

$$= \psi(|0\rangle - \psi|1\rangle) \tag{21.3.44}$$

$$= \psi|\psi\rangle \tag{21.3.45}$$

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

$$\langle\bar{\psi}|\Psi^\dagger = \langle\bar{\psi}|\bar{\psi} \tag{21.3.46}$$

where

$$\langle\bar{\psi}| = \langle 0| - \langle 1|\bar{\psi} = \langle 0| + \bar{\psi}\langle 1| \tag{21.3.47}$$

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

$$\langle\bar{\psi}|\psi\rangle = (\langle 0| - \langle 1|\bar{\psi})(|0\rangle - \psi|1\rangle) \tag{21.3.48}$$

$$= \langle 0|0\rangle + \langle 1|\bar{\psi}\psi|1\rangle \tag{21.3.49}$$

$$= 1 + \bar{\psi}\psi \tag{21.3.50}$$

$$= e^{\bar{\psi}\psi} \tag{21.3.51}$$

Any function of a Grassmann variable can be expanded as follows:

$$F(\psi)=F_0+F_1\psi \tag{21.3.52}$$

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:

$$\int \psi\, d\psi = 1 \tag{21.3.53}$$

$$\int 1\, d\psi = 0 \tag{21.3.54}$$

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:

$$\int \bar{\psi}\psi\, d\psi\, d\bar{\psi} = 1 \tag{21.3.55}$$

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

$$\int \bar{\psi}\psi\, d\bar{\psi}\, d\psi = -1 \tag{21.3.56}$$

Let us now consider some Gaussian integrals. You are urged to show the following:

$$\int e^{-a\bar{\psi}\psi}\, d\bar{\psi}\, d\psi = a \tag{21.3.57}$$

$$\int e^{-\bar{\psi}M\psi}\, [d\bar{\psi}\, d\psi] = \det M \tag{21.3.58}$$

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 $[d\bar{\psi}\, d\psi] = d\psi_1\, d\psi_1\, d\bar{\psi}_2\, d\psi_2$. 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:

$$\langle \bar{\psi}\psi \rangle = \frac{\int \bar{\psi}\psi\, e^{a\bar{\psi}\psi}\, d\bar{\psi}\, d\psi}{\int e^{a\bar{\psi}\psi}\, d\bar{\psi}\, d\psi} = \frac{1}{a} = -\langle \psi\bar{\psi} \rangle \tag{21.3.59}$$

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

$$\langle \bar{\psi}_i \psi_j \rangle = \frac{\int \bar{\psi}_i \psi_j\, e^{a_1\bar{\psi}_1\psi_1 + a_2\bar{\psi}_2\psi_2}\, d\bar{\psi}_1\, d\psi_1\, d\bar{\psi}_2\, d\psi_2}{\int e^{a_1\bar{\psi}_1\psi_1 + a_2\bar{\psi}_2\psi_2}\, d\bar{\psi}_1\, d\psi_1\, d\bar{\psi}_2\, d\psi_2} \tag{21.3.60}$$

$$= \frac{\delta_{ij}}{a_i} \equiv \langle \bar{i}j \rangle \tag{21.3.61}$$

Exercise 21.3.7. Prove the above result.

Exercise 21.3.8. Show that

$$\langle \bar{\psi}_i \bar{\psi}_j \psi_k \psi_l \rangle = \frac{\int \bar{\psi}_i \bar{\psi}_j \psi_k \psi_l\, e^{a_1\bar{\psi}_1\psi_1 + a_2\bar{\psi}_2\psi_2}\, d\bar{\psi}_1\, d\psi_1\, d\bar{\psi}_2\, d\psi_2}{\int e^{a_1\bar{\psi}_1\psi_1 + a_2\bar{\psi}_2\psi_2}\, d\bar{\psi}_1\, d\psi_1\, d\bar{\psi}_2\, d\psi_2} \tag{21.3.62}$$

$$= \frac{\delta_{il}}{a_i}\frac{\delta_{jk}}{a_j} - \frac{\delta_{ik}}{a_i}\frac{\delta_{jl}}{a_j} \tag{21.3.63}$$

$$\equiv \langle \bar{i}l \rangle \langle \bar{j}k \rangle - \langle \bar{i}k \rangle \langle \bar{j}l \rangle \tag{21.3.64}$$

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:

$$I=\int |\psi\rangle\langle\bar{\psi}|\, e^{-\bar{\psi}\psi}\, d\bar{\psi}\, d\psi \tag{21.3.65}$$

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.)

$$\begin{aligned}\int |\psi\rangle\langle\bar{\psi}|\, e^{-\bar{\psi}\psi}\, d\bar{\psi}\, d\psi &= \int |\psi\rangle\langle\bar{\psi}|(1-\bar{\psi}\psi)\, d\bar{\psi}\, d\psi \\ &= \int (|0\rangle - \psi|1\rangle)(\langle 0| - \langle 1|\bar{\psi})(1-\bar{\psi}\psi)\, d\bar{\psi}\, d\psi \\ &= \int (|0\rangle\langle 0| + \psi|1\rangle\langle 1|\bar{\psi})(1-\bar{\psi}\psi)\, d\bar{\psi}\, d\psi \\ &= |0\rangle\langle 0| \int (-\bar{\psi}\psi)\, d\bar{\psi}\, d\psi + |1\rangle\langle 1| \int \psi\bar{\psi}\, d\bar{\psi}\, d\psi \\ &= I \end{aligned} \tag{21.3.66}$$

The final result we need is that for any bosonic operator (an operator made of an even number of Fermi operators)

$$\mathrm{Tr}\,\Omega = \int \langle -\bar{\psi}|\Omega|\psi\rangle\, e^{-\bar{\psi}\psi}\, d\bar{\psi}\, d\psi \tag{21.3.67}$$

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:

$$Z = \mathrm{Tr}\, e^{-\beta(\Omega_0-\mu)\Psi^\dagger\Psi} \tag{21.3.68}$$

$$= \int \langle -\bar{\psi}|e^{-\beta(\Omega_0-\mu)\Psi^\dagger\Psi}|\psi\rangle\, e^{-\bar{\psi}\psi}\, d\bar{\psi}\, d\psi \tag{21.3.69}$$

You cannot simply replace $\Psi^\dagger$ and Ψ by $-\bar{\psi}$ and ψ, respectively, in the exponential. This is because when we expand out the exponential not all the Ψs will be acting to the right on their eigenstates and neither will all $\Psi^\dagger$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:

$$e^{-\beta(\Omega_0-\mu)\Psi^\dagger\Psi} = 1 + (e^{-\beta(\Omega_0-\mu)} - 1)\Psi^\dagger\Psi \tag{21.3.70}$$

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 $N^k = \mathrm{N}$ for any nonzero k.) Now we may write

$$Z = \int \langle -\bar{\psi}|1 + (e^{-\beta(\Omega_0-\mu)} - 1)\Psi^\dagger\Psi|\psi\rangle\, e^{-\bar{\psi}\psi}\, d\bar{\psi}\, d\psi \tag{21.3.71}$$

$$= \int \langle -\bar{\psi}|\psi\rangle(1 + (e^{-\beta(\Omega_0-\mu)} - 1)(-\bar{\psi}\psi))\, e^{-\bar{\psi}\psi}\, d\bar{\psi}\, d\psi \tag{21.3.72}$$

$$= \int (1 - (e^{-\beta(\Omega_0-\mu)} - 1)\bar{\psi}\psi)\, e^{-2\bar{\psi}\psi}\, d\bar{\psi}\, d\psi \tag{21.3.73}$$

$$= 1 + e^{-\beta(\Omega_0-\mu)} \tag{21.3.74}$$

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

$$Z = \mathrm{Tr}\, e^{-\beta H} \tag{21.3.75}$$

where H is a normal ordered operator $H(\Psi^\dagger, \Psi)$. We write the exponential as follows:

$$e^{-\beta H} = \lim_{N\to\infty} \left(\exp\left(-\frac{\beta}{N}H\right)\right)^N \tag{21.3.76}$$

$$= \underbrace{(1-\varepsilon H)\ldots(1-\varepsilon H)}_{N \text{ times}} \qquad \varepsilon = \beta/N \tag{21.3.77}$$

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:

$$Z = \int \langle -\bar{\psi}_0|(1-\varepsilon H)|\psi_{N-1}\rangle\, e^{-\bar{\psi}_{N-1}\psi_{N-1}} \langle \bar{\psi}_{N-1}|(1-\varepsilon H)|\psi_{N-2}\rangle\, e^{-\bar{\psi}_{N-2}\psi_{N-2}}$$

$$\times \langle \bar{\psi}_{N-2}|\cdots|\psi_1\rangle\, e^{-\bar{\psi}_1\psi_1} \langle \bar{\psi}_1|(1-\varepsilon H)|\psi_0\rangle\, e^{-\bar{\psi}_0\psi_0} \prod_{i=0}^{N-1} d\bar{\psi}_i\, d\psi_i \tag{21.3.78}$$

Now we may legitimately make the replacement

$$\langle \bar{\psi}_{i+1} | 1 - \varepsilon H(\Psi^\dagger, \Psi) | \psi_i \rangle = \langle \bar{\psi}_{i+1} | 1 - \varepsilon H(\bar{\psi}_{i+1}, \psi_i) | \psi_i \rangle$$
$$= e^{\bar{\psi}_{i+1}\psi_i} e^{-\varepsilon H(\bar{\psi}_{i+1}, \psi_i)} \tag{21.3.79}$$

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)

$$\bar{\psi}_N = -\bar{\psi}_0 \tag{21.3.80}$$

$$\psi_N = -\psi_0 \tag{21.3.81}$$

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

$$Z = \int \prod_{i=0}^{N-1} e^{\bar{\psi}_{i+1}\psi_i} e^{-\varepsilon H(\bar{\psi}_{i+1}, \psi_i)} e^{-\bar{\psi}_i \psi_i} d\bar{\psi}_i \, d\psi_i \tag{21.3.82}$$

$$= \int \prod_{i=0}^{N-1} \exp\left[\left[\left(\frac{(\bar{\psi}_{i+1} - \bar{\psi}_i)}{\varepsilon} \psi_i - H(\bar{\psi}_{i+1}, \psi_i)\right)\right]\varepsilon\right] d\bar{\psi}_i \, d\psi_i \tag{21.3.83}$$

$$\simeq \int \exp\left(\int_0^\beta \bar{\psi}(\tau)\left(-\frac{\partial}{\partial\tau} - \Omega_0 + \mu\right)\psi(\tau)\, d\tau\right)[\mathscr{D}\bar{\psi}\, \mathscr{D}\psi] \tag{21.3.84}$$

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

$$\bar{\psi}(\tau) = \sum_n \frac{e^{i\omega_n \tau}}{\beta} \bar{\psi}(\omega) \tag{21.3.85}$$

$$\psi(\tau) = \sum_n \frac{e^{-i\omega_n \tau}}{\beta} \psi(\omega) \tag{21.3.86}$$

where the allowed frequencies, called *Matsubara frequencies*, are chosen to satisfy the antisymmetric boundary conditions in Eqs. (21.3.80–21.3.81). Thus

$$\omega_n = \frac{(2n+1)\pi}{\beta} \tag{21.3.87}$$

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 $\beta \to \infty$, it follows from Eq. (21.3.87) that when n increases by unity, ω_n changes by $d\omega = 2\pi/\beta$. Thus

$$\frac{1}{\beta}\sum_n \to \int \frac{d\omega}{2\pi} \tag{21.3.88}$$

The inverse transformations are

$$\psi(\omega) = \int_0^\beta \psi(\tau)\, e^{i\omega_n \tau}\, d\tau \tag{21.3.89}$$

$$\bar{\psi}(\omega) = \int_0^\beta \bar{\psi}(\tau)\, e^{-i\omega_n \tau}\, d\tau \tag{21.3.90}$$

where we use the orthogonality property

$$\int_0^\beta e^{i\omega_n \tau}\, e^{-i\omega_m \tau}\, d\tau = \frac{e^{i(\omega_n - \omega_m)\beta} - 1}{i(\omega_n - \omega_m)} = \beta \delta_{mn} \tag{21.3.91}$$

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 $\beta \to \infty$, which converts sums over discrete frequencies to integrals over a continuous ω, as per Eq. (21.3.88), we end up with

$$Z = \int \exp\left[\int_{-\infty}^{\infty} \frac{d\omega}{2\pi}\, \bar{\psi}(\omega)(i\omega - \Omega_0 + \mu)\psi(\omega)\right][\mathcal{D}\bar{\psi}(\omega)\, \mathcal{D}\psi(\omega)] \tag{21.3.92}$$

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:

$$\langle\bar{\psi}(\omega_1)\psi(\omega_2)\rangle$$

$$=\frac{\int\bar{\psi}(\omega_1)\psi(\omega_2)\exp\left[\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\bar{\psi}(\omega)(i\omega-\Omega_0+\mu)\psi(\omega)\right][\mathscr{D}\bar{\psi}(\omega)\mathscr{D}\psi(\omega)]}{\int\exp\left[\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\bar{\psi}(\omega)(i\omega-\Omega_0+\mu)\psi(\omega)\right][\mathscr{D}\bar{\psi}(\omega)\mathscr{D}\psi(\omega)]}$$

$$=\frac{2\pi\delta(\omega_1-\omega_2)}{i\omega_1-\Omega_0+\mu} \tag{21.3.93}$$

In particular,

$$\langle\bar{\psi}(\omega)\psi(\omega)\rangle=\frac{2\pi\delta(0)}{i\omega-\Omega_0+\mu}=\frac{\beta}{i\omega-\Omega_0+\mu} \tag{21.3.94}$$

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 $\langle N\rangle$:

$$\langle N\rangle=\frac{1}{\beta Z}\frac{\partial Z}{\partial\mu} \tag{21.3.95}$$

$$=\frac{1}{\beta}\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\langle\bar{\psi}(\omega)\psi(\omega)\rangle \tag{21.3.96}$$

$$=\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\frac{e^{i\omega o^+}}{i\omega-\Omega_0+\mu} \tag{21.3.97}$$

$$=\theta(\mu-\Omega_0) \tag{21.3.98}$$

as in the operator approach.

Notice that we had to introduce the factor $e^{i\omega o^+}$ 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 $e^{i\omega o^+}$. 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:

$$G(\tau) = \langle \mathcal{T}(\Psi(\tau)\Psi^\dagger(0)) \rangle \tag{21.3.99}$$

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:

$$\mathcal{T}(\Psi(\tau)\Psi^\dagger(0)) = \theta(\tau)\Psi(\tau)\Psi^\dagger(0) - \theta(-\tau)\Psi^\dagger(0)\Psi(\tau) \tag{21.3.100}$$

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

$$G(\tau) = \frac{\theta(\tau)\, e^{-(\Omega_0-\mu)\tau} - \theta(-\tau)\, e^{-(\Omega_0-\mu)(\tau+\beta)}}{1+e^{-\beta(\Omega_0-\mu)}} \tag{21.3.101}$$

and that in the zero-temperature limit this reduces to

$$G(\tau) = \theta(\tau)\, e^{-(\Omega_0-\mu)\tau} \qquad \mu < \Omega_0 \tag{21.3.102}$$

$$= -\theta(-\tau)\, e^{-(\Omega_0-\mu)\tau} \qquad \mu > \Omega_0 \tag{21.3.103}$$

Let us define the pair of transforms:

$$G(\omega) = \int_{-\infty}^{\infty} G(\tau)\, e^{i\omega\tau}\, d\tau \tag{21.3.104}$$

$$G(\tau) = \int_{-\infty}^{\infty} G(\omega)\, e^{-i\omega\tau}\, \frac{d\omega}{2\pi} \tag{21.3.105}$$

Show that

$$G(\omega) = \frac{1}{\Omega_0 - \mu - i\omega} \tag{21.3.106}$$

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

$$\langle \mathcal{T}(\Psi(\tau)\Psi^{\dagger}(0))\rangle = \langle \bar{\psi}(\tau)\psi(0)\rangle \tag{21.3.107}$$

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 $\varepsilon^{1/2}$. 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.* A392 **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.