Beyond The Basics I

Chapter 5

From classical to quantum and back

In the first four chapters of this book we reviewed Newtonian mechanics and Einstein's expansion of it into the relativistic regime; introduced the variational calculus and used it to present the elegant formulation of mechanics known as Hamilton's principle of stationary action; and derived Lagrange's equations of motion.

However, in spite of its power, classical mechanics, even extending it into the domain of special relativity, has its limitations; it arises as a special case of the vastly more comprehensive theory of quantum mechanics. *Where* does classical mechanics fall short, and *why* is it limited?

The key turns out to be *Hamilton's principle*. We will show how this principle comes about as a special case of the larger theory. And so, since we can use Hamilton's principle to derive classical mechanics, we will reach a good understanding of when classical mechanics is valid and when we have to use the full apparatus of quantum theory. So this chapter is not part of classical mechanics *per se*; readers short on time can skip it without compromising their preparation for future chapters. But it would be too bad to skip this chapter forever, because here we will understand how classical mechanics comes about, and how it fits into the grand scheme of physics. We do this by showing how Hamilton's principle, that most compact, elegant statement of classical mechanics, which emerges rather mysteriously in Chapter 4, is in fact a natural consequence of quantum mechanics in a certain limit. This chapter sets classical mechanics in context.

We begin the chapter with the behavior of waves in classical physics, and then show results of some critical experiments with light and atoms, upsetting traditional notions of light as waves and atoms as particles. We proceed to give a brief review of Feynman's sum-over-paths formulation of quantum mechanics, which describes the actual behavior of light and atoms, and then show that Hamilton's Principle naturally emerges in a certain limiting case. Once equipped with Hamilton's Principle, we already know from Chapter 4 how to derive classical motion.

5.1 Classical waves

A string with uniform mass per unit length μ is held in a horizontal position under uniform tension T. What happens if we disturb the string? In particular, suppose that at time $t = 0$ we give the string some particular shape $y(x)$, and some velocity distribution $\partial y(x,t)/\partial t|_{\mathrm{t=0}}$, where x is the horizontal coordinate along the string and y is the transverse displacement (see Figure 5.1). Our goal is to find $y(x,t)$, the shape of the string at any later time.

Consider a very small slice of string of length Δx and mass $\Delta m = \mu \Delta x$, as shown in Figure 5.2(a). We ignore gravity, so the only forces acting on this piece are the string tensions to the right and to the left of it. If the string is displaced from equilibrium, the slice will generally be slightly curved. The two tension forces on the right and left therefore pull in slightly different directions, as shown in Figure 5.2(b), so the resulting unbalanced force causes Δm to accelerate. For small vertical displacements, the horizontal component of T remains essentially constant along the string, so Δm accelerates vertically, not horizontally. The vertical component of tension is $F_y = T \tan\theta$, where θ is the angle of the string at some point relative to the horizontal. The tangent of this angle, the slope of the string, can also be written as

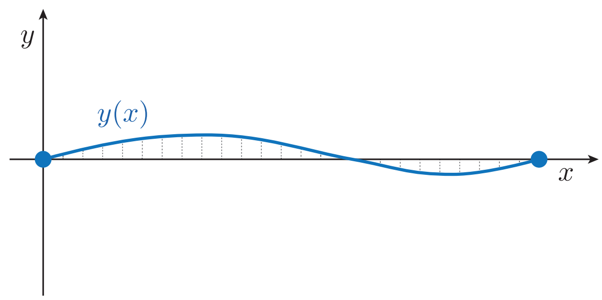


FIGURE 5.1 : A transverse small displacement of a string.

the partial derivative $\partial y(x,t)/\partial x$; we have to use the *partial* derivative here, because whereas y depends upon both the position x along the string and the time t, the slope at a fixed time t is the derivative of y with respect to x alone.

Let the left-hand end of Δm be located at x_0 and the right-hand end at $x_0 + \Delta x$. Using Taylor series, the slopes of the string at the left and right are related by

$$\left.\frac{\partial y}{\partial x}\right|_{\mathrm{x_0+\Delta x}} = \left.\frac{\partial y}{\partial x}\right|_{\mathrm{x_0}} + \left.\frac{\partial^2 y}{\partial x^2}\right|_{\mathrm{x_0}} \Delta x + \frac{1}{2!}\left.\frac{\partial^3 y}{\partial x^3}\right|_{\mathrm{x_0}} (\Delta x)^2 + \ldots \tag{5.1}$$

If $y(x)$ is smooth and Δx is sufficiently small, we can neglect all but the first two terms on the right. The vertical forces on the right-hand and left-hand sides of the slice Δm are

$$F_y(\text{right}) = T \left.\frac{\partial y}{\partial x}\right|_{\mathrm{x_0+\Delta x}} \quad \text{and} \quad F_y(\text{left}) = -T \left.\frac{\partial y}{\partial x}\right|_{\mathrm{x_0}}, \tag{5.2}$$

with the force at the left upward and the force at the right downward. The *net* vertical force on Δm is therefore

$$F_{\text{net}} = T\left[\left.\frac{\partial y}{\partial x}\right|_{\mathrm{x_0+\Delta x}} - \left.\frac{\partial y}{\partial x}\right|_{\mathrm{x_0}}\right] = T \left.\frac{\partial^2 y}{\partial x^2}\right|_{x_0} \Delta x. \tag{5.3}$$

(a) (b)

FIGURE 5.2 : (a) A small slice of string; (b) Tension forces on the slice.

The acceleration of Δm is $\partial^2 y/\partial t^2$, the second derivative of y with respect to *time*, keeping now the position x fixed. The left end of the slice of string moves up and down vertically, with position $y(x_0, t)$, velocity $\partial y/\partial t|_{\mathrm{x}_0}$, and acceleration $\partial^2 y/\partial t^2|_{\mathrm{x}_0}$. Newton's second law $F_{\mathrm{net}} = ma$ for Δm therefore becomes

$$T \left.\frac{\partial^2 y}{\partial x^2}\right|_{\mathrm{x}_0} \Delta x = \Delta m \left.\frac{\partial^2 y}{\partial t^2}\right|_{\mathrm{x}_0} = \mu \Delta x \left.\frac{\partial^2 y}{\partial t^2}\right|_{\mathrm{x}_0}, \tag{5.4}$$

so

$$\frac{\partial^2 y}{\partial x^2} - \left(\frac{\mu}{T}\right)\frac{\partial^2 y}{\partial t^2} = 0, \tag{5.5}$$

which is the *wave equation* of the string for small transverse displacements. It represents an infinite number of $F = ma$ equations for the infinite number of infinitesimal slices of the string.

For a single particle, an initial position and initial velocity determine the future position for given mass and forces by solving the ordinary differential equation $F = ma$. For a string, an initial shape $y(x, 0)$ and velocity distribution $\partial y/\partial t|_{\mathrm{t}=0}$ determine the future shape $y(x, t)$ for given mass density μ and tension T by solving a *partial* differential equation, the wave equation. It is most convenient to write the wave equation in more general form

$$\frac{\partial^2 y}{\partial x^2} - \frac{1}{v^2}\frac{\partial^2 y}{\partial t^2} = 0 \tag{5.6}$$

where v is a constant whose role will become apparent soon.

The wave equation is linear, which is a consequence of assuming that the displacements are small. As a result, it satisfies the **superposition principle**: two separate solutions of the equation can be simply added, and the result still satisfies the wave equation. A solution of the wave equation is *any* two functions $f(u)$ and $g(u)$ of the combination $u = x \pm v\,t$

$$y_{\rm sol} = f(x + v\,t) + g(x - v\,t) \tag{5.7}$$

as can be easily verified by substituting this expression into (5.6). This implies that, once $y(x, 0)$ and $y'(x, 0)$ are given, f and g are fixed. The wave equation simply tells that the profiles of f and g evolve in time, without distortion, at speed v towards negative and positive x respectively. An easy way to see this is to sketch the two functions at two instants in time.

Note also that the wave equation is *real* in addition to being linear. This means that we can solve it with complex functions as well, with the real and imaginary parts as separate solutions: the equation then splits into a real part and an imaginary part, each looking identical in form, applied to the real and imaginary parts of the complex solution.

Classical waves propagate also in fluids like air or water characterized by macroscopic properties such as mass density and pressure. If a fluid is locally perturbed, sound waves can be set up, in which both the local density and pressure oscillate, and the oscillations are propagated from the initial site throughout the material. In the case of small-amplitude waves, the density of the material has the form $\rho = \rho_0 + \Delta\rho$ and the pressure is $p = p_0 + \Delta p$, where ρ_0 and p_0 are the ambient density and pressure, and $\Delta\rho$ and Δp are the small perturbations that can propagate from place to place. We won't prove it here, but the disturbance $\Delta\rho$ obeys a wave equation which in three dimensions has the form

$$\nabla^2(\Delta\rho) - \left(\frac{\rho}{B}\right)\frac{\partial^2(\Delta\rho)}{\partial t^2} = 0, \tag{5.8}$$

where B is the *bulk modulus* of the material. If a change of pressure Δp at a point in the material causes a fractional change in density $\Delta\rho/\rho$, then

$$B \equiv \frac{\Delta p}{\Delta\rho/\rho}, \tag{5.9}$$

so B is a measure of stiffness: the less the fractional change in density for a given pressure change, the greater the stiffness, and the greater the bulk modulus.

The differential operator ∇^2 is the *Laplacian*, which In Cartesian coordinates takes the form

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}. \tag{5.10}$$

In the case of an infinite plane wave propagating in the x direction – that is, $\Delta\rho(x, y, z, t) = \Delta\rho(x, t)$, the wave equation becomes

$$\frac{\partial^2(\Delta\rho)}{\partial x^2} - \left(\frac{\rho}{B}\right)\frac{\partial^2(\Delta\rho)}{\partial t^2} = 0, \tag{5.11}$$

which has the same form as the wave equation for a string.

Another well-known classical wave is the *electromagnetic* wave, one of the famous consequences of Maxwell's equations of electrodynamics. In vacuum the equations can be combined to produce a wave equation

$$\nabla^2\boldsymbol{E} - \frac{1}{c^2}\frac{\partial^2\boldsymbol{E}}{\partial t^2} = 0 \tag{5.12}$$

for the electric field $\boldsymbol{E}$, with a similar equation for the magnetic field $\boldsymbol{B}$. The electric and magnetic fields propagate together at the speed of light c, so Maxwell was able to achieve a grand synthesis of electricity, magnetism, and optics by showing that light waves are in fact electromagnetic waves. Again, the one-dimensional form of these equations, corresponding to an electromagnetic plane wave propagating in the x direction, has the same form as waves on a string. A difference here is that the electromagnetic wave equation is valid for *any* classical electromagnetic wave, whatever its amplitude. No linear approximation has to be made in this case.

Among the infinite variety of solutions of the one-dimensional wave equation, whether for waves on a string, sound waves, or light waves, are **sinusoidal traveling waves** that propagate to the right or to the left. These have the form

$$y(x, t) = A_0 \sin\left(k\left(x \mp \frac{\omega}{k}t\right) - \varphi\right) \tag{5.13}$$

where A_0 is the amplitude, k is the wave number, ω is the angular frequency, and φ is the phase angle of the wave. Notice that this solution is indeed a function of $x \pm v\,t$ as argued earlier. The upper (minus) sign corresponds to a wave traveling to the right, and the lower (plus) sign corresponds to a wave traveling to the left. The wave number is related to the wavelength λ

by $k = 2\pi/\lambda$, and the angular frequency is related to the frequency ν (*i.e.*, , cycles/s) by $\omega = 2\pi\nu$. If the phase angle $\varphi = 0$, then $y(x,t) = y_0 \sin(kx - \omega t)$, so that initially the wave is a sine wave with $y = 0$ at $x = 0$, etc. The phase angle simply displaces the sinusoidal shape to the right (if φ is positive) or to the left (if φ is negative).

We can find the velocity of the wave by discovering how x changes as t increases, so as to keep the overall phase $\theta \equiv kx \mp \omega t - \varphi$ constant. That is, how far is a particular wave shape displaced during some interval of time? Setting $d\theta/dt = 0$ gives $v = |dx|/dt = \omega/k$, or in terms of wavelength and frequency, $v = \lambda\nu$. Substituting the wave form into the wave equation then shows that the traveling wave solves the wave equation for waves on a string if and only if

$$v = \omega/k = \sqrt{T/\mu}, \tag{5.14}$$

and for sound waves

$$v = \omega/k = \sqrt{B/\rho}. \tag{5.15}$$

The **intensity** I of a plane wave of sound or light, *i.e.*, , a wave solution that is independent of two of the Cartesian coordinates and moves in the third direction (x), is proportional to the square of its amplitude A_0; that is,

$$I = CA_0^2, \tag{5.16}$$

where C is a constant that depends upon the type of wave. The intensity is the energy/second passing through a square meter perpendicular to the wave velocity.

The wave equation also has complex exponential traveling-wave solutions of the form

$$y(x,t) = A_0 e^{i(kx \mp \omega t - \varphi)}, \tag{5.17}$$

which is simple to verify by substitution, but is quite obvious from the fact that the real and imaginary parts of the complex exponential are given by Euler's formula

$$e^{i\theta} = \cos\theta + i\sin\theta, \tag{5.18}$$

so that if we choose $\theta = kx \mp \omega t - \varphi$, it is in fact the sum of two sinusoidal traveling waves with the same amplitude, frequency, and wave number, but differing in phase by $\pi/2$. Complex exponential solutions are often used in part because they are easier to work with mathematically (the derivative or an exponential is an exponential, for example). Then we can always take the real (or imaginary) part of the final result to get the physical result, which in classical physics corresponds to an observable quantity and must therefore be real. In quantum mechanics, as we will introduce in this chapter, the complex exponential form turns out to be the natural form to use.

The intensity of a complex wave is proportional to the product of the wave amplitude and the complex conjugate of the wave amplitude; this gives the real quantity

$$I = Cy(x,t)y^*(x,t) = C\left[A_0 e^{i(kx \mp \omega t - \varphi)}\right]\left[A_0^* e^{-i(kx \mp \omega t - \varphi)}\right] = C|A_0|^2 \quad (5.19)$$

as expected.

EXAMPLE 5-1: Two-slit interference of waves

When two or more traveling waves combine, we observe interference effects. Direct a plane sinusoidal wave from left to right at a double-slit system, for example, as shown in Figure 5.3. Only waves that pass through one of the slits make it through to the right-hand side. The resulting wave is then detected on a detecting plane, which is a screen or bank of detectors much farther along to the right. What will be observed by the detectors? (We assume for simplicity that the detecting plane is very far from the slit system compared with the distance between the two slits, so the wave disturbances from each slit propagate essentially parallel to one another.) Using the complex exponential form at the position of the detector,

$$y(x,t) = y_1 + y_2 = A_0\left(e^{i(ks_1 - \omega t - \varphi)} + e^{i(ks_2 - \omega t - \varphi)}\right), \quad (5.20)$$

where s_1 is the distance of the detector from slit 1 and s_2 is the distance of the detector from slit 2. Here we have used the fact that the wave number, frequency, and phase angle of each part of the wave are the same (the phase angles are the same because the phase of both waves at the plane of the slits is the same.). We have also assumed that the amplitude of each wave as it reaches the detector is the same, which is an excellent approximation as long as the detecting plane is far away compared with the distance between the two slits.

The total wave amplitude at the right is then the sum of the two wave amplitudes,

$$y_T = A_0 e^{-i(\omega t + \phi)}\left(e^{iks_1} + e^{iks_2}\right). \quad (5.21)$$

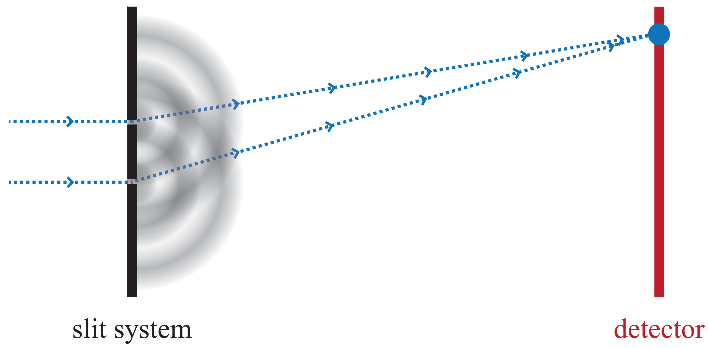


FIGURE 5.3 : Two paths for waves from slit system to detectors.

The intensity of the wave at the detecting plane is

$$\begin{aligned} I = Cy_T^* y_T &= C|A_0)|^2 \left(e^{-iks_1} + e^{-iks_2}\right)\left(e^{iks_1} + e^{iks_2}\right) \\ &= 2C|A_0)|^2 (1 + \cos(k(s_2 - s_1)) \\ &= 4C|A_0)|^2 \cos^2(k(s_2 - s_1)/2) \end{aligned} \tag{5.22}$$

using the identities $\cos q = (e^{iq} + e^{-iq})/2$ and $\cos^2(q/2) = (1/2)(1 + \cos q)$. The difference $s_2 - s_1$ of the path lengths from the two slits to a point on the detecting plane is $s_2 - s_1 = d\sin\theta$, as shown in Figure 5.4(a), where θ is the angle between the two rays and the forward direction. The phase difference between the two waves is $\Phi \equiv k(s_2 - s_1) = (2\pi d/\lambda)\sin\theta$, so the intensity at an arbitrary angle θ, in terms of the intensity I_0 in the forward direction $\theta = 0$, is

$$I(\theta) = I_0 \cos^2(\Phi/2) \quad \text{where} \quad \Phi = \frac{2\pi d}{\lambda}\sin\theta, \tag{5.23}$$

as illustrated in Figure 5.4(b). There are alternating maxima and minima, with the maxima occurring at angles θ for which $n\lambda = d\sin\theta$, with $n = 0, \pm 1, \pm 2, \ldots$

If we direct a plane wave of sound at a double-slit system, where the wavelength λ is smaller than the slit separation d, then we do observe the alternating maxima and minima predicted by equation (5.23). [1]

[1] We can detect the sound intensities by microphones placed along the detecting plane. The microphones must of course be large compared with the distance between molecules in the sound-transmitting medium; the wave equation for sound models the medium as a continuum.

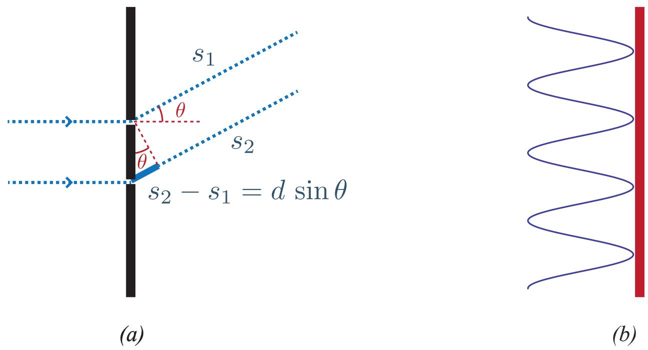


FIGURE 5.4 : (a) The relationship between $s_2 - s_1$, d, and θ; (b) The two-slit interference pattern.

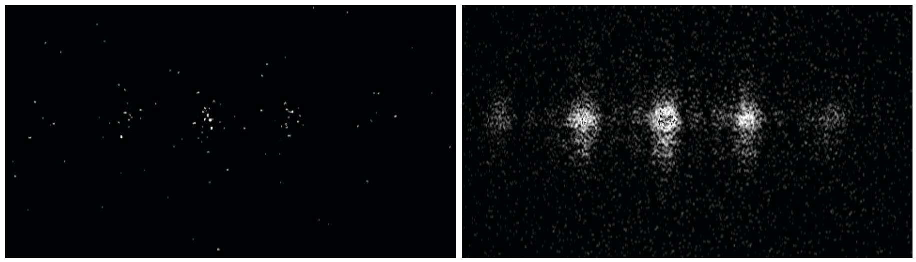
5.2 Two-slit experiments with light and atoms

According to Maxwell's equations, light is an electromagnetic wave, so if we direct a beam of light at a double slit we should observe wave interference. But if we direct a beam of *atoms* at a double slit, classical mechanics teaches us that we should observe a bunch of atoms downstream of each slit, much like what would happen if we tossed ball bearings at a pair of slits. Atoms are particles, after all, so should exhibit no interference at all. Now what about actual experiments?

LIGHT

Various light detectors can be used on the detecting plane, including photographic film, photomultipliers, CCDs, and others, depending upon the wavelength. If the wavelength of the light beam is smaller than the slit separation, a fairly bright light source is used, and fairly long exposures are made (the meaning of "fairly" here will soon become clear), the experimental intensities again show alternating maxima and minima, with maxima occurring where $n\lambda = d\sin\theta$.

But now crank the brightness of the light source way down, and observe what happens over short time intervals. Instead of seeing very low intensity light spread immediately over the detecting plane, as predicted by the interference/diffraction formula, one finds that at first the light arrives at



(a) (b)

FIGURE 5.5 : (a) At very low intensity light, individual photons appear to land on the screen randomly; (b) as the intensity is cranked up, the interference pattern emerges.

apparently random discrete locations. If the detector is a bank of photocells, for example, only certain cells will register the reception of light, while others (even at locations where the intensity should be a maximum) receive nothing at first. That is, light is seen to arrive in discrete lumps, or **photons** (see Figure 5.5(a)).

The remarkable fact is that even though the photons arrive one at a time at the detectors, if we wait long enough the large number of photons distribute themselves among the detectors exactly as the interference formula predicts (see Figure 5.5(b))! That is, in some sense light has both a particle nature (we observe single particles only in the detectors) and a wave nature (when huge numbers of photons have arrived at the detecting plane, the overall distribution shows the interference pattern predicted by wave theory.) If we close off one of the two slits, the pattern of photons shows no such interference.

By observing the number of photons arriving at the screen, and knowing the intensity, the wavelength λ, and frequency $\nu = c/\lambda$ of the beam, one finds that each photon must have an energy $E = h\nu$ and momentum $p = E/c = h/\lambda$, where h is *Planck's constant*, $h = 6.627 \times 10^{-34}$ J $\cdot$ s. It was Albert Einstein in 1905 who first realized that light is not a continuous, Maxwellian wave after all, but consists of discrete photons, and that each photon is massless and has energy $E = h\nu$ and momentum $p = h\nu/c = h/\lambda$.

The central puzzle is: If light consists of a stream of individual photons, so that in the case of two slits each photon presumably goes through one slit or the other slit and not both, how can they develop an interference pattern? How do photons "know" whether two slits are open or only one?

ATOMS

Now project a beam of *helium* atoms at a pair of slits and observe their distribution on the detecting plane. A double-slit system has slits of width $a = 1\ \mu$m and slit separation $d = 8\ \mu$m. Each helium atom has a mass $m = 6.68 \times 10^{-27}$ kg, and each can be detected by various counters as a discrete particle, where the detecting plane is a distance $D = 1.95$ m behind the slits. Our beam of helium atoms travels at speeds between 2.1 and 2.2 km/s. A distinct *interference* pattern is observed![2]

The atoms *do* arrive at the screen in discrete lumps, as expected, but the distribution shows interference effects similar to what we observe with light! Figure 5.6 shows the actual results of this experiment. The obvious question is: for helium atoms, as with photons, *what exactly is interfering?* The beam intensity can be turned so low that there is at most a single atom in flight at any given time, so atoms are not interfering with other atoms; each atom must be interfering with *itself* in some way. The interference distribution emerges only after many atoms have been detected. We can carry out similar experiments with atoms with different masses moving with different velocities. The results show that the wavelength λ deduced from a particular interference pattern on the screen is inversely proportional to both the atomic mass m and the velocity v of the atoms. That is,

$$\lambda = h/p \tag{5.24}$$

where $p = mv$ is the momentum of the nonrelativistic atom and h again is Planck's constant. This is exactly the same relation between λ and p as for photons.[3] If one of the two slits is blocked off, so atoms can only penetrate one of the slits, the two-slit interference pattern goes away.

[2]reference

[3]The wavelength $\lambda = h/p$ is called the *de Broglie wavelength*, because it was in his doctoral dissertation that the French physicist Louis de Broglie proposed that all particles have a wavelength $\lambda = h/p$. In the case of photons, we can increase the momentum by increasing the frequency of the light, since $p = h/\lambda = h\nu/c$. In the case of atoms, we can increase the momentum by increasing their velocity.

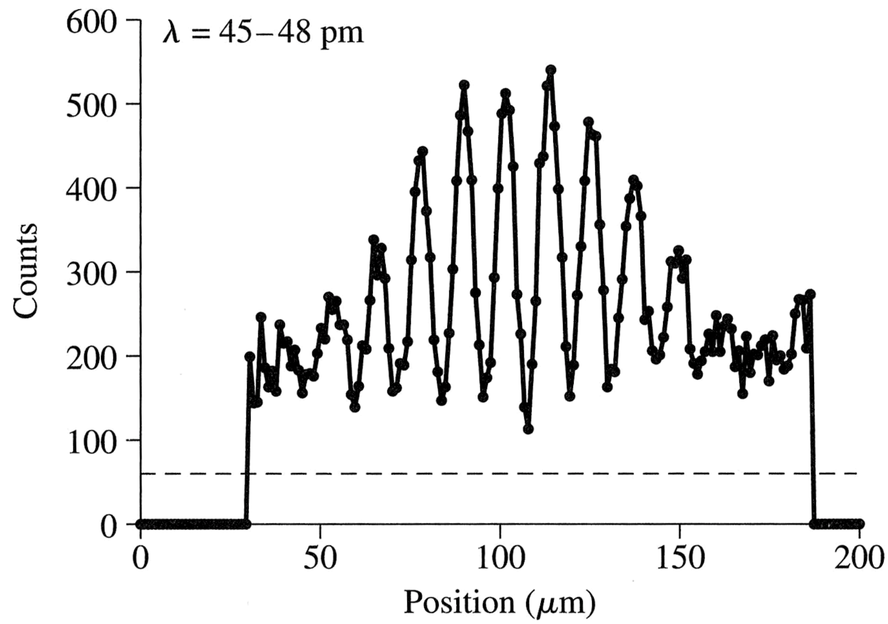


FIGURE 5.6 : Helium atoms with speeds between 2.1 and 2.2 km/s reaching the rear detectors, with both slits open. The detectors observe the arrival of individual atoms, *but the distribution shows a clear interference pattern as we would expect for waves!*. We see how the interference pattern builds up one atom at a time. The first data set is taken after 5 minutes of counting, while the last is taken after 42 hours of counting. The experiments were carried out by Ch. Kurtsiefer, T. Pfau, and J. Mlynek; see their article in *Nature* **386**, 150 (1997). (The "hotspot" in the data arises from an enhanced dark count due to an impurity in the microchannel plate detector.)

The central puzzle once again: Even though an atom is detected at a specific spot much like a classical particle, the interference patterns show that an individual atom somehow "knows" whether there are two slits open or only one. How does it know that? If both slits are open, does it somehow probe both paths? *Does it in some sense take both paths?*

5.3 Feynman sum-over-paths

Thirty-one years ago, Dick Feynman told me about his "sum over histories" version of quantum mechanics. "The electron does anything it likes," he said. "It just goes in any direction at any speed, . . . however it likes, and then you add up the amplitudes and it gives you the wave function." I said to him "You're crazy." But he wasn't.- Freeman Dyson, 1980.

According to the American physicist Richard Feynman (1918-1988) the answer to the question posed at the end of the preceding section is *yes*! In his sum-over-paths formulation of quantum mechanics, atoms (or electrons or molecules or photons or ball bearings or . . .) *do* take all available paths between two points. If both slits are open in the experiments we have described, the particle in some sense goes through *both slits*. If one of the slits is closed, that path is not available; or if we perform an experiment (perhaps at the slits themselves) showing which slit each atom passes through, then each atom takes only *one* of the paths, and no interference pattern is observed at the screen.

How do we predict what will be observed in each case? According to quantum mechanics, there is no way we can tell where a particular photon or helium atom will go. This is not because our measuring devices do not yet have sufficient precision; it is because a particle does not *have* a definite position or momentum at any given time, and it does not travel by any single classical path. The best we can do is find the *probability* P that a particle will be observed at any particular location.

How do we find the probability distribution? Here are the rules:

(1) The probability P that a particle will be observed at a particular location is given by the absolute square of a total complex *probability amplitude* z_T

to arrive there,

$$P = |z_T|^2 \equiv z_T^* z_T \tag{5.25}$$

where z_T^* is the complex conjugate of z_T.

(2) The total probability amplitude for a particle to go from a to b is simply the sum of the probability amplitudes to go by every path available to it,

$$z_T = z_1 + z_2 + \ldots \tag{5.26}$$

(3) The probability amplitude z to go from a source at a to a detector at b by some particular path is given by

$$z = z(t_0)e^{i\phi}, \tag{5.27}$$

where $z(t_0)$ is the amplitude of the source when the particle leaves, and ϕ is a phase that depends upon the path. The phase is the Lorentz-invariant quantity

$$\phi = \frac{1}{\hbar}\int_a^b \eta_{\mu\nu} p_\mu dx_\nu = \frac{1}{\hbar}\int_a^b (\boldsymbol{p}\cdot d\boldsymbol{s} - Edt) = \int_a^b (\boldsymbol{k}\cdot d\boldsymbol{s} - \omega dt) \tag{5.28}$$

where the coordinates at a are the initial position and time (x_0, y_0, z_0, t_0), and the coordinates at b are the final position and time (x, y, z, t), for a particular path. Here the magnitude of the wave number three-vector $\boldsymbol{k}$ is $k = 2\pi/\lambda = p/\hbar$ and $\omega = 2\pi\nu = E/\hbar$, where $\hbar \equiv h/2\pi$, p is the magnitude of the particle's momentum, and E is its energy. Note that both relationships $p = h/\lambda$ and $E = h\nu$ are valid for massless photons as well as massive particles like helium atoms.

These are the three simple rules for calculating the probability that a particle will be detected at time t.

Note from Rule 3 that if a particular path from a source at a to a detector at d is thought of as a sequence of path segments, for example, (1) $(a \to b)$, (2) $(b \to c)$, (3) $(c \to d)$, then the phase ϕ, being an integral over the entire path from a to d, is the sum of integrals for each segment of the path. That is, $\phi = \phi_1 + \phi_2 + \phi_3 + ...$, so the phase factor can be written

$$e^{i\phi} = e^{i(\phi_1+\phi_2+\phi_3+...)} = e^{i\phi_1}e^{i\phi_2}e^{i\phi_3}... \tag{5.29}$$

The amplitude to go by a particular path all the way from a source to a detector is therefore

$$z = z(t_0)e^{i\phi_1}e^{i\phi_2}e^{i\phi_3}..., \tag{5.30}$$

the amplitude at the source multiplied by the *product* of phase factors for each segment of the path.

Suppose for now that both the energy of the particle and the magnitude of its momentum are conserved and that they have the same value along each path. (Later on we will generalize to allow changes in each.) In that case $\int \boldsymbol{k} \cdot d\boldsymbol{s} = ks$, where s is the path length, and $\int \omega dt = \omega(t - t_0)$. The amplitude for a particular path can therefore be written

$$z = z(t_0)e^{i\phi} = z(t_0)\ e^{iks}\ e^{-i\omega(t-t_0)}. \tag{5.31}$$

The amplitude $z(t_0)$ when the particle leaves the source is itself generally complex, so

$$z(t_0) = |z(t_0)|e^{i\phi_0}, \tag{5.32}$$

where ϕ_0 is real. The amplitude for a particular path to reach the detector is therefore

$$z = z(t_0)e^{i\phi} = |z(t_0)|e^{i(\phi+\phi_0)}, \tag{5.33}$$

where $\phi = ks - \omega(t - t_0)$.

This probability amplitude z can be displayed as a two-dimensional vector called a **phasor** in the complex plane. The horizontal axis represents real numbers and the vertical axis imaginary numbers. Points not on either axis represent complex numbers with both a real and imaginary part. Placing the tail of the phasor at the origin, the length of the phasor is $|z(t_0)|$ and its angle with the real axis is the total phase $(\phi + \phi_0)$, where ϕ_0 is the angle of $z(t_0)$ relative to the real axis, as illustrated in Figure 5.7.

Now suppose the source emits a steady beam of particles, all with the same energy. The magnitude of the amplitude at the source therefore remains constant, but the phase at that point changes with time. That is because at the source itself we can set $s = 0$ in the amplitude, so at the source

$$z(t) = z(t_0)e^{-i\omega(t-t_0)}, \tag{5.34}$$

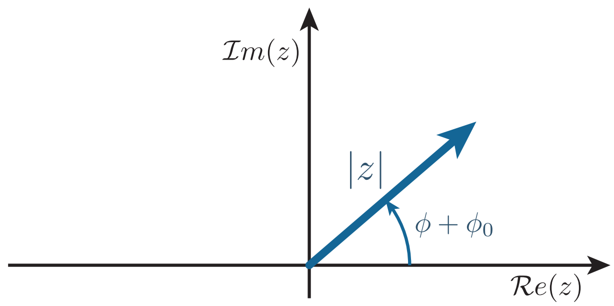


FIGURE 5.7 : A phasor $z(t_0)e^{i\phi} \equiv |z(t_0)|e^{i(\phi+\phi_0)}$ drawn in the complex plane. The real axis is horizontal and the imaginary axis is vertical. The absolute length of the phasor is $|z(t_0)|$ and the angle between the phasor and the real axis is the phase $(\phi + \phi_0)$, where ϕ_0 is the phase of $z(t_0)$ alone.

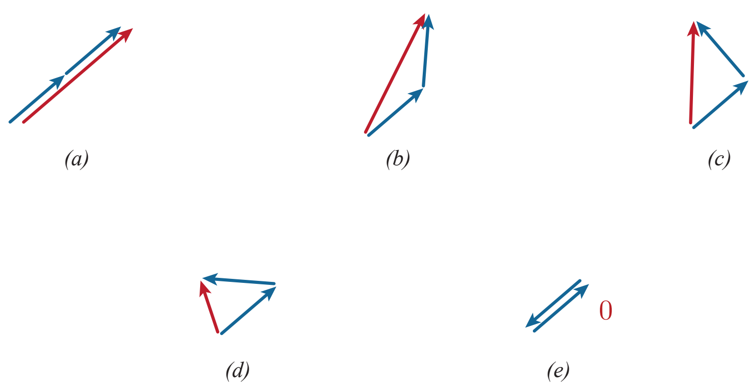


FIGURE 5.8 : The sum of two individual phasors with the same magnitudes $|z(t_0)|$ but different phases. The result is a phasor that extends from the tail of the first to the tip of the second, as in vector addition. The difference in their angles in the complex plane is the difference in their phase angles. Shown are examples with phase differences equal to (a) zero (b) $45°$ (c) $90°$ (d) $135°$ (e) $180°$.

which is a phasor that spins clockwise in the complex plane as time progresses. It is convenient to refer the amplitude to some standard time $t = 0$, so

$$z(0) = z(t_0)e^{i\omega t_0}. \tag{5.35}$$

In terms of $z(0)$, the amplitude for a path of length s is

$$z = z(t_0)e^{i(ks-\omega(t-t_0))} = z(0)e^{i(ks-\omega t)}, \tag{5.36}$$

so for a monoenergetic beam of particles the total amplitude for a particle to reach the detector at time t is

$$\begin{aligned} z_T &= z(0)\left(e^{i(ks_1-\omega t)} + e^{i(ks_2-\omega t)} + \ldots\right) \\ &= z(0)e^{-i\omega t}\left(e^{iks_1} + e^{iks_2} + \ldots\right). \end{aligned} \tag{5.37}$$

Therefore even though, for a monoenergetic beam of massive particles, those taking longer paths must have left the source earlier to arrive at the same time t, the starting time t_0 has been eliminated, so the total amplitude is exactly the same as it would be if the particles were all emitted from the source at $t = 0$ with energy $E = \hbar\omega$, no matter which path they take! Note that in summing over paths, we mean *every* path allowed by the physical circumstances. Paths can zig-zag, go back and forth, in circles, any way they like as long as there are no physical barriers to prevent them.

5.4 Two slits and two paths

We can now derive the probability distribution for particles passing through a double slit using the quantum rules. We will make a huge simplification for now, allowing particles to move along just two paths from the source to a detector, each path consisting of two straight-line segments joined at a slit, as illustrated in Figure 5.9. We assume also that both the source and the detector are far from the slit system, so the two paths from the source to the slits are essentially parallel to one another, and the two paths from the slits to the detector are also essentially parallel to one another.

The total probability amplitude is

$$z_T = z(0)e^{-i\omega t}\left(e^{iks_1} + e^{iks_2}\right), \tag{5.38}$$

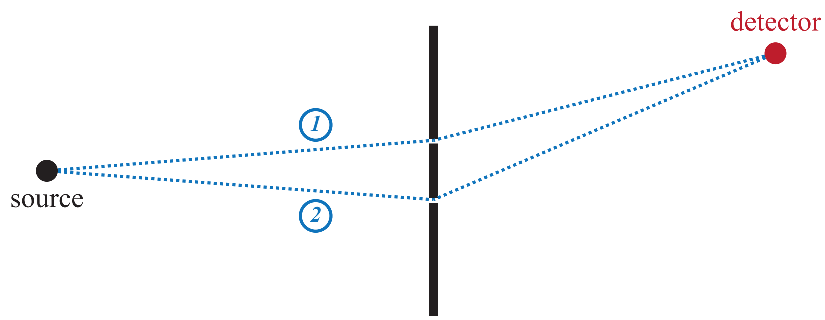


FIGURE 5.9 : Two paths from a source to a detector.

so the probability of observing a photon or atom at a particular detector is

$$\begin{aligned} P = z_T^* z_T &= |z(0)|^2 \left(e^{-iks_1} + e^{-iks_2}\right)\left(e^{iks_1} + e^{iks_2}\right) \\ &= 2|z(0)|^2 (1 + \cos(k(s_2 - s_1)) \\ &= 4|z(0)|^2 \cos^2(k(s_2 - s_1)/2). \end{aligned} \tag{5.39}$$

The probability that a particle is detected at arbitrary angle θ, in terms of the probability $P(0)$ of detecting it in the forward direction $\theta = 0$, is therefore

$$P(\theta) = P(0)\cos^2(\Phi/2) \quad \text{where} \quad \Phi = k(s_2 - s_1) = \frac{2\pi}{\lambda}(d \sin\theta) \tag{5.40}$$

using the same trig identities we used earlier for classical waves, where we found an *intensity* distribution $I(\theta) = I(0)\cos^2(\Phi/2)$. Naturally enough, if the probabilities of single-particle events obey the two-slit pattern, then if we collect a great many particles the intensity will have the same distribution. The formula agrees with the experimental results for photons or helium atoms whose wavelengths are not extremely small compared with the slit separation d, as shown in Figure 5.4.

Now what happens if the wavelength *is* extremely small, *i.e.*, , $\lambda \lll d$? From $\lambda = h/p$ it follows that if the particles happen to be nonrelativistic ball bearings, the momentum $p = mv$ is enormous because their masses are so large. If we toss ball bearings at a double-slit system, we expect bunches of balls to accumulate downstream of each slit, with no interference pattern at all. It is true that some balls might nick the slit edges and be deflected to one side or the other, yet we would certainly see no interference pattern. So something else must be going on to explain why in that case we do not see the two-slit interference pattern of equation (5.40). If the particles happen to be photons with $\lambda \ll d$, they have very large momenta and therefore very high energies; or if the particles happen to be nonrelativistic helium atoms, their velocities must be quite large to have very small wavelengths. If quantum mechanics applies to *everything*, including ball bearings, high-energy photons, and fast helium atoms, the sum-over-paths rules must still apply, even though two-slit interference is not evident.

Consider an actual experiment with fast helium atoms. As before, each slit has a width $a = 1\ \mu$m and the two slits are separated by a distance $d = 8\ \mu$m. Each atom has a mass $m = 6.68 \times 10^{-27}$ kg, and each can be detected by various counters as a discrete particle. For atoms with velocities above 30 km/s the results of actual experiments with both slits open are shown in Figure 5.10. They strike the screen with a bunch downstream of each slit, much like what we would find if we tossed ball bearings at a much larger slit system.

That is, any ball bearing that penetrated a two-slit system would go through either one slit or the other, and for those going through the top slit it would make no difference whether the bottom slit is open or not, and for those going through the bottom slit it would make no difference whether the top slit is open or not. The distribution with both slits open is simply the sum of the distributions with only one slit open at a time. And that is what we observe for these fast helium atoms.

In fact, the two-slit interference pattern is incomplete, because even with just two slits, many more than two paths are available. For example, because each individual slit has a finite width, there is an infinite number of nearby paths passing through each slit. These paths have slightly different phases, especially if the wavelength is small, so they interfere with one another. Consider a narrow slice of a single slit of width dy. If we measure y up from the bottom of the slit, then $s = s_0 + y \sin\theta$, where s_0 is the distance of the bottom of the slit from the detector, as shown in Figure 5.11(a). The amplitude dz of all paths passing through the narrow slice will be proportional to dy, the width of the slice, so

$$dz = (b\ dy)\ e^{iks_0}\ e^{i(ky\sin\theta - \omega t)} \tag{5.41}$$

where b is a constant. The total amplitude to reach the detector, passing

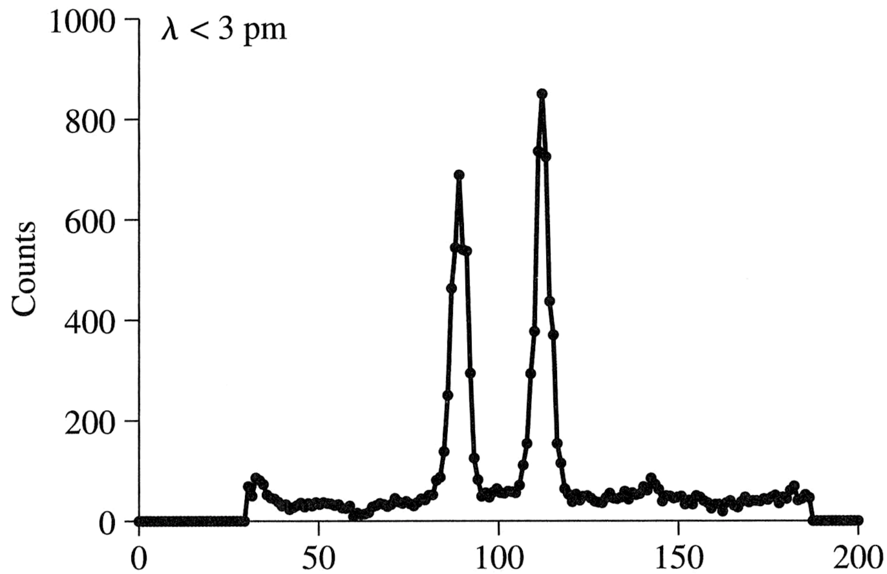


FIGURE 5.10 : High-velocity helium atoms, with speeds above 30 km/s, reaching the rear detectors, with both slits open. The detectors observe the arrival of individual atoms, and the distribution is what we would expect for classical particles. Experiments carried out by Ch. Kurtsiever, T. Pfau, and J. Mlynek, *Nature* **386**, 150 (1997)

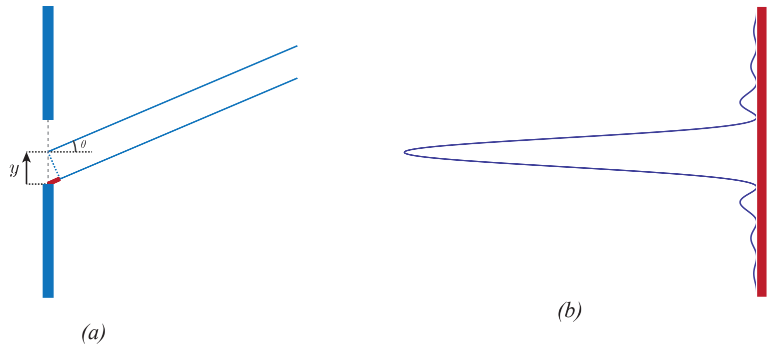


FIGURE 5.11 : (a) Path length as a function of position y within the slit. (b) The single-slit diffraction pattern.

through a single slit of width a, is therefore

$$z_T = b\ e^{i(ks_0-\omega t)} \int_0^a dy\ e^{iky\sin\theta} = b\ e^{i(ks_0-\omega t)} \left(\frac{e^{ika\sin\theta}-1}{ik\sin\theta}\right). \tag{5.42}$$

The probability is equal to the absolute square of z,

$$\begin{aligned} P = z_T^* z_T &= b^2 \left(\frac{e^{-iky\sin\theta}-1}{-ik\sin\theta}\right)\left(\frac{e^{iky\sin\theta}-1}{ik\sin\theta}\right) \\ &= \frac{2b^2}{k^2\sin^2\theta}(1-\cos(ka\sin\theta)) = 2b^2a^2\left(\frac{\sin^2\alpha}{\alpha^2}\right), \end{aligned} \tag{5.43}$$

where $\alpha \equiv (ka\sin\theta)/2$ and we have used the identity $\sin^2\alpha = (1/2)(1 - \cos 2\alpha)$. Now $(\sin^2\alpha)/(\alpha^2) \to 1$ as $\alpha \to 0$, which is the maximum value this ratio can achieve. This distribution is called *single slit diffraction.* The probability pattern in this case is

$$P = P_{(max)}\left(\frac{\sin^2\alpha}{\alpha^2}\right) \quad \text{where} \quad \alpha \equiv \frac{ka\sin\theta}{2} = \frac{\pi a}{\lambda}\sin\theta. \tag{5.44}$$

The diffraction pattern $P(\alpha)$ is shown in Figure 5.11. The distribution has a maximum in the middle where $\alpha = 0$, and the first minimum at each side corresponds to $\alpha = \pm\pi$. The half-width of the central peak is therefore $\Delta\alpha = \pi$, which occurs at an angle θ_0 for which $\sin\theta_0 = \lambda/a$.

So far, using Feynman's sum-over-paths, we have found both the two-slit interference pattern neglecting single-slit diffraction and the single-slit diffraction pattern in the absence of two-slit interference. Of course, the actual two-slit probability distribution at the screen must include *both* interference and diffraction.

Assume for now that the wavelength λ is small enough that $\lambda \ll a$. Then for the central diffraction peak, $\sin\theta \ll 1$, so $\sin\theta \cong \theta$. Let D be the distance from the slit system to the detecting screen, and x be the distance on the screen from the midpoint of the wave pattern on the screen, as shown in Figure 5.12. Now $x/D = \tan\theta \cong \theta$ if $\theta \ll 1$, so in this case the central peak of the single-slit diffraction pattern has a half-width on the screen of magnitude

$$\Delta x_{1/2} = D\theta = \frac{D\lambda}{a}. \tag{5.45}$$

Figure 5.13 illustrates interference/diffraction curves for a double slit system with $d = 4a$ and $d/D = 0.001$. In Figure 5.13(a) the wavelength of the particle beam obeys $\lambda/a = 0.01$, so the angle of the first minimum of the diffraction pattern is $\theta_{1/2} = \lambda/a = 0.01$, and the distance on the detecting plane from the center of the pattern to the first diffraction minimum is $x_{1/2} = D\theta_{1/2} = 0.01D$. This is ten times the distance d between the two slits, so the diffraction curves for the two slits essentially overlap. In this case the overall probability distribution is simply the product of the interference oscillation with the central diffraction envelope,

$$P = P_{\max}\cos^2\beta\left(\frac{\sin\alpha}{\alpha}\right)^2. \tag{5.46}$$

where $\beta \equiv \pi d\sin\theta/\lambda$ and $\alpha \equiv \pi a\sin\theta/\lambda$.

Now keep the same pair of slits and the same distance to the detecting plane, but decrease the beam wavelength by a factor of 20. The angle to the first minimum of each diffraction pattern, measured from the central peak, then becomes $\theta_{1/2} = \lambda/a = 0.0005$, so the distance on the detecting plane between the peak and first minimum is $x_{1/2} = d/2$. The entire pattern now looks as shown in Figure 5.13(b). The two peaks are quite well separated, as one would expect for classical particles. Clearly as the wavelength is further reduced, the pattern becomes closer and closer to that corresponding to two bunches of particles formed downstream of each of the two slits. In these

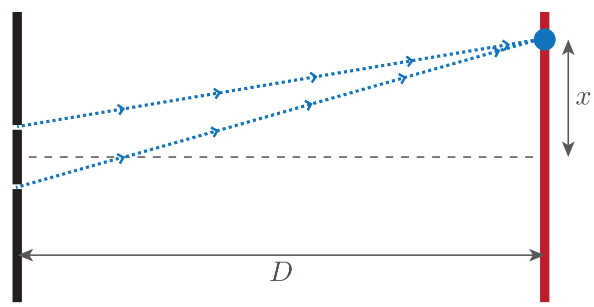


FIGURE 5.12 : The double slit, with a screen at distance D. We can view the intensity on the screen as a function of the transverse distance x.

experiments classical mechanics is shown to be a special case of quantum mechanics, corresponding to the limit of small deBroglie wavelength.

Figure 5.13(a) agrees very well with the observations of slower helium atoms, as we showed in Figure 5.4. We already found in Section 6.? that helium atoms with $v = 2.15$ km/s have the wavelength $\lambda = 4.63 \times 10^{-11}$ m. The ratio λ/a is therefore 4.63×10^{-11} m/ (10^{-6} m) $= 4.63 \times 10^{-5} \ll 1$, so the condition used to derive equation (5.46) for the half-width of the single-slit diffraction pattern is valid in this case. The results illustrated in Figure 5.13 used a distance $D = 1.95$ m from the slits to the screen, so the transverse half-width at the detectors is

$$\Delta x = D\theta = 1.95 \text{ m}(4.63 \times 10^{-5}) = 9.03 \times 10^{-5}\text{m}. \tag{5.47}$$

This is more than ten times the distance d between the two slits, so there is a strong overlap between the two single-slit diffraction patterns in the experiment shown in Figure 5.6?.

Now what about the experimental result shown in Figure 5.6 for velocities above 30 km/s? The wavelength of 30 km/s helium atoms is shorter than 2.15 km/s atoms by a factor $(2.15)/(30) = 0.0717$, so the half-width of the single-slit diffraction pattern on the screen is now

$$\Delta x = D\theta = (9.03 \times 10^{-5}\text{m})(0.0717) = 6.47 \times 10^{-6}\text{m}. \tag{5.48}$$

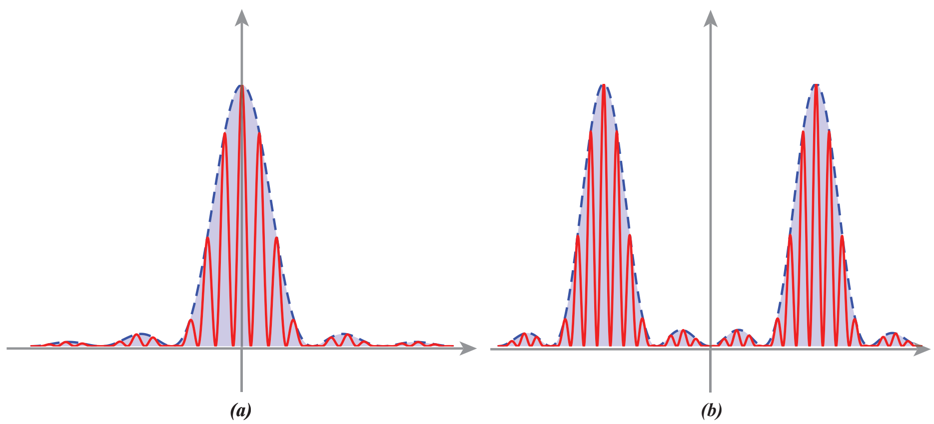


FIGURE 5.13 : Interference/diffraction patterns for a double slit with $a = d/4$ and $D = 1000d$. The diffraction curves, shown in dashed lines, serve as envelopes for the more rapidly oscillating interference pattern. (a) The pattern in the case $d = 0.1x_{1/2}$, where $x_{1/2}$ is the distance on the detecting plane between the center and the first minimum of the diffraction envelope. The diffraction curves of the two slits strongly overlap in this case, giving in effect a single diffraction envelope. (b) The pattern in the case $d = 2x_{1/2}$, showing that the two diffraction patterns have become separated, with the first minimum due to each slit at the same location in the center. This case corresponds to a wavelength smaller by a factor of 20 than the pattern shown in (a).

In this case the distance $2\Delta x$ is only a bit larger than d, so there is some separation in the peaks of the two diffraction patterns. Most of the atoms in this sample have even higher velocities and even shorter wavelengths, so their diffraction patterns are more distinctly separated. This is consistent with the experimental results shown in Figure 5.6. That is, the apparent classical behavior of fast helium atoms is "really" *well-separated diffraction patterns*. The two-slit interference within each diffraction curve is masked in the experiment because of the wide range of speeds and therefore wavelengths represented in the sample, with correspondingly different positions of the maxima and minima within the diffraction envelopes, and also by fuzziness in the detectors.

Our conclusion from atomic-beam experiments is that helium atoms generally behave like *neither* ball bearings nor sound waves: They are neither classical particles nor classical waves, but retain some properties of each. They are detected as localized units like particles, but they show interference patterns like waves. In fact, we find that their particle and wave properties are related by $p = h/\lambda \equiv \hbar k$ where h is Planck's constant and $\hbar \equiv h/2\pi$.

5.5 No barriers at all

Now remove the system of slits so there are no barriers at all between the source and detector, and consider all paths between them. According to the quantum rules, the total probability amplitude for the particle to be detected at time t_f is

$$z = z(0)e^{-i\omega t_f}(e^{iks_1} + e^{iks_2} + e^{iks_3} + \ldots), \tag{5.49}$$

the sum of an infinite number of terms, where we have assumed that all of the magnitudes $z(0)$ (which may be complex numbers) are equal.[4] The phases ks_1, ks_2, ks_3 ... are obviously proportional to the path lengths. Suppose that the path lengths (and therefore the phases) are all about the same for some particular set of paths. In that case the phasors associated with each term in the set add up to give a large total amplitude, as shown in Figure 5.14(a). If the path lengths are all quite different for another set of paths, then those phasors tend to cancel one another out, as in Figure 5.14(b).

[4] In fact, the amplitude for longer paths is less than the amplitude for shorter paths, but this turns out to make no difference in making the transition to classical mechanics,

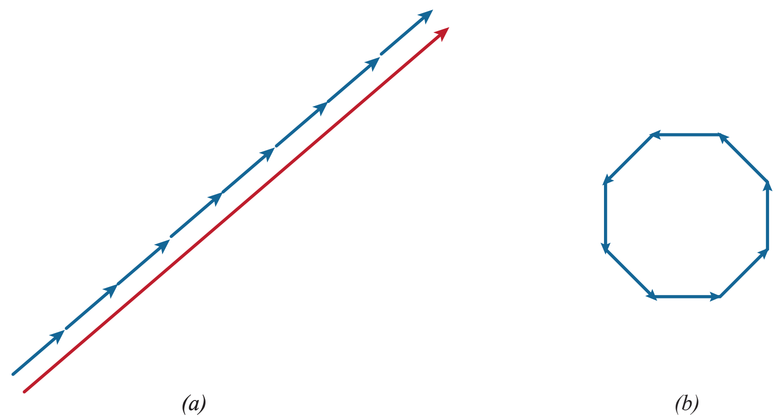


FIGURE 5.14 : The sum of a large number of phasors (a) that are about the same (b) that differ by constant amounts.

Under what circumstances will the phases be about the same for a set of paths? In the analogous case of a continuous *function* $y(x)$, the Taylor series about a point x_0 is

$$y(x) = y(x_0) + \left.\frac{dy(x)}{dx}\right|_{x_0}(x - x_0) + \frac{1}{2!}\left.\frac{d^2y(x)}{dx^2}\right|_{x_0}(x - x_0)^2 + \ldots \quad (5.50)$$

so the value of $y(x)$ is the same to first order in $\Delta x \equiv x - x_0$ if x_0 happens to be a maximum or minimum of the function. If x_0 does *not* correspond to a maximum or minimum, then $y(x)$ changes more rapidly as x varies. The same is true of the phases in the sum-over-paths. If the set of paths is nearby the path of minimum length, for example, then the phases will all be about the same and the corresponding phasors will add up to give a large total. If the set of paths is nearby some other arbitrary path, their phases will differ sufficiently from one another that the total phasor will be small. In the case of a free particle, the shortest path is a straight line, and the phase of nearby paths will be nearly the same, so the total probability amplitude will be large. That is, if a free particle travels from a to b by a straight-line path, the neighboring paths all have about the same length, so their phases add up constructively. But if the particle travels by some arbitrary path, the neighboring paths differ more markedly in length from it, so the phases for these surrounding paths tend to cancel one another out.

as we shall soon see.

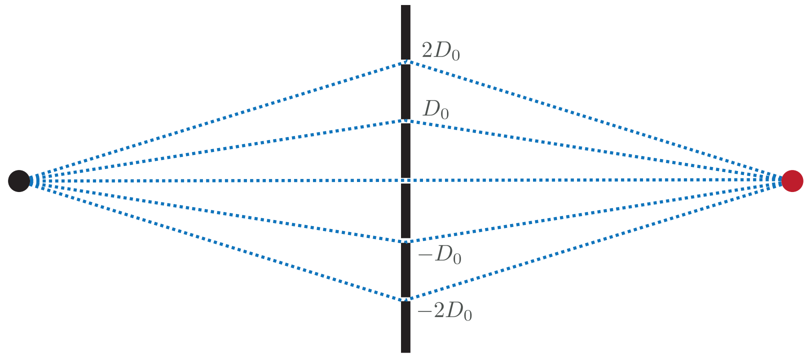


FIGURE 5.15 : A class of kinked paths between a source and detector. The straight line is the shortest path, and the midpoint of the others is a distance $D = |n|D_0$ from the straight line, where $(n = \pm1, \pm2, ...)$.

EXAMPLE 5-2: A class of paths near a straight-line path

We can show what happens for a special set of free-particle paths. Let s_0 be the shortest distance between the source and detector, corresponding to a straight-line path. We will sum the probability amplitudes for a certain class of paths near this path. These particular alternate paths consist of a straight line with a kink in the middle, where the kink is a distance $D = |n|D_0$ $(n = \pm1, \pm2, \ldots)$ from the straight line, as shown in Figure 5.15. If the paths have length s_n, then by the Pythagorean theorem

$$(s/2)^2 = (s_0/2)^2 + (nD_0)^2. \tag{5.51}$$

We will assume that $|n|D_0 \ll s_0$, so using the binomial approximation,

$$s = s_0\sqrt{1 + (2nD_0/s_0)^2} \cong s_0(1 + 2n^2D_0^2/s_0^2). \tag{5.52}$$

Therefore the probability amplitude to go by a particular path of length s is proportional to

$$e^{iks} = e^{iks_0}e^{i\theta_n} \tag{5.53}$$

where

$$\theta_n = \left(\frac{2kD_0^2}{s_0}\right)n^2 \qquad n = 0, 1, 2, \ldots \tag{5.54}$$

Note that θ_n is the angle of the associated phasor with respect to that of the straight-line path. As a particular case, suppose the phasor corresponding to the straight-line path is horizontal, and that $2kD_0^2/s_0 = \pi/200$. Then the angles θ_n are given in Table 6.1 for $n = 0$, $n = \pm1$, $n = \pm2$, . . . $.n = \pm25$, The sum of these phasors, all with the same length but in directions θ_n relative to the horizontal, will give the total phasor for these paths.

n	θ_n	n	θ_n	n	θ_n	n	θ_n	n	θ_n
0	$0°$	±6	$32.4°$	±12	$129.6°$	±18	$291.6°$	±24	$518.4°$
±1	$0.9°$	±7	$44.1°$	±13	$152.1°$	±19	$324.9°$	±25	$562.5°$
±2	$3.6°$	±8	$57.6°$	±14	$176.4°$	±20	$360.0°$	±26	$608.4°$
±3	$8.1°$	±9	$72.9°$	±15	$202.5°$	±21	$396.9°$	±27	$656.1°$
±4	$14.4°$	±10	$90.0°$	±16	$230.4°$	±22	$435.6°$	±28	$705.6°$
±5	$22.5°$	±11	$108.9°$	±17	$260.1°$	±23	$476.1°$	±29	$756.9°$

Table 5.1. The angles of bent-line segments in terms of the integer n that characterizes them.

The phasors are drawn in Figure 5.16. Those corresponding to $n = 0$ through $n = \pm5$ are more or less aligned, so the paths neighboring the straight-line path are enhancing it. As $|n|$ increases the angles between successive phasors gradually increase, so the phasors begin to loop around, winding up in tighter and tighter spirals so they no longer make any important contribution to the total amplitude. The shape of these phasors is called a **Cornu spiral**. The sum of all the phasors up to $n = \pm25$ is the long arrow shown. If we were to include additional phasors we would not change this sum very much. It is the straight-line path and its neighbors that contribute the most to the overall phasor, and therefore to the overall probability amplitude for the atom to go from the source to the detector. The classical path is the straight-line path in this case, but it is not the *only* path.

How do we *know* that the classical path is not the only path that a particle takes? The best way to show the importance of the other paths is to block them off. If a highly collimated beam of particles travels from a source to a detecting screen, detectors will find that the particles are confined to a narrow region on the screen, in accord with the idea that the particles follow a straight-line path from source to screen. Therefore if we were to introduce a narrow slit at a location directly between the source and screen, it should make no difference, because according to classical ideas the particles are taking only that path anyway. But it *does* make a difference. With the

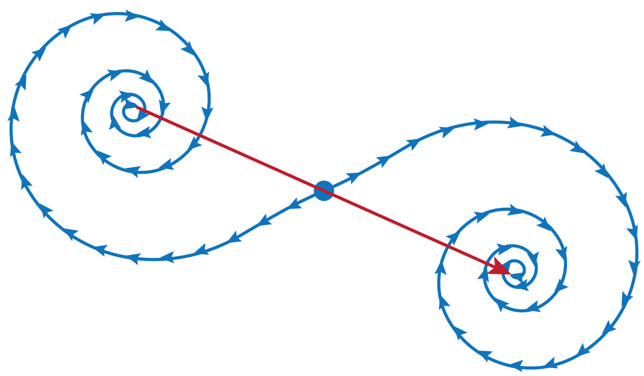


FIGURE 5.16 : Phasors up to $n = \pm 25$. The more distant paths wind up in spirals, contributing very little to the overall phasor sum.

narrow slit in place, the particles show a diffraction pattern on the screen. Therefore without the screen, particles must be taking more than the straight line path after all; in fact, according to quantum mechanics they take all paths from source to screen that are not blocked off by some barrier.

EXAMPLE 5-3: How classical is the path?

The phase of a free particle involves the product $ks = 2\pi s/\lambda$. If the wavelength is very small, so that $s/\lambda \gg 1$, then slight changes in s mean large changes in phase. The straight-line path between a and b is the minimum-distance path, so we already know that is the classical path. Neighboring paths have almost the same length, so they tend to add up in phase. But for a given path near the straight line, as the wavelength becomes smaller and smaller, the phase $2\pi s/\lambda$ changes more and more, so the corresponding phasors tend to spiral around and cancel out. That is, for very small wavelengths the set of mutually reinforcing paths becomes more and more constrained, closer and closer to the straight line; in the limit $s/\lambda \gg 1$, non-classical paths become less and less important in the overall sum, so classical mechanics becomes a better and better approximation to the true situation. Take for example an electron moving at speed v. Classical motion is valid in the limit

$$\frac{s}{\lambda} = \frac{p\,s}{h} = \frac{mvs}{h} \gg 1. \tag{5.55}$$

An electron is a cathode-ray tube (such as an old-fashioned TV picture tube), with $s = 0.5$ m and $v = 10^8$ m/s, has the ratio

$$\frac{mvs}{h} = \frac{9.1 \times 10^{-31}\ \mathrm{kg} \cdot 10^8\ \mathrm{m/s} \cdot 0.5\ \mathrm{m}}{6.6 \times 10^{-34}\ \mathrm{J\ s}} \sim 10^{11}, \tag{5.56}$$

so such an electron moves on a classical path (in a TV tube we obviously want the electrons to move along a classical, deterministic path.)

But consider the electron in a hydrogen atom, where it has a typical speed of 10^6 m/s and path length from one side of the atom to the other of order 10^{-10} m. The ratio in this case is

$$\frac{mvs}{h} = \frac{9.1 \times 10^{-31}\ \mathrm{kg} \cdot 10^{6}\ \mathrm{m/s} \cdot 10^{-10}\ \mathrm{m}}{6.6 \times 10^{-34}\ \mathrm{J\ s}} \sim 0.1, \tag{5.57}$$

so the path in this case is not at all classical, but quantum mechanically fuzzy. The electron in hydrogen does not move along a classical orbit anything like the motion of planets around the Sun. On the other hand, the ratio for Earth orbiting the Sun is

$$\frac{mvs}{h} = \frac{6 \times 10^{24}\ \mathrm{kg} \cdot 3 \times 10^{4}\ \mathrm{m/s} \cdot 1.5 \times 10^{11}\ \mathrm{m}}{6.6 \times 10^{-34}\ \mathrm{J\ s}} \sim 10^{73}, \tag{5.58}$$

which corresponds to supremely classical motion.

There are no sharp boundaries between classical and quantum behavior. All motion is really quantum mechanical, but classical behavior can take place as a special case. So quantum mechanics shows the role of classical mechanics and its range of validity. ————————

5.6 Path shapes for light rays and particles

We can now use the formalism of sum-over-paths to find the *shape* of paths taken by light rays and by non-relativistic massive particles. In Chapter 8 we will be able to use the methods derived here to find the path shapes of relativistic particles moving in electromagnetic fields.

LIGHT RAYS

When light passes from one medium to another, say air to a refractive medium like glass, its frequency (*i.e.*, , the energy/photon) remains constant, but the wavelength changes, because the velocity $v = \lambda\nu$ is less in glass than in air. If the medium has index of refraction n, where n may depend upon position, the speed of light in the medium is $v = c/n$, so with constant frequency the wavelength decreases by the factor $1/n$, and the wavenumber $k = 2\pi/\lambda$ increases by the factor n. So whereas in vacuum $\omega/k = c$ (or $k = \omega/c$), in a refractive medium $k = \omega/v = n\omega/c$. The quantum mechanical amplitude for photons to travel by a particular path therefore becomes

$$z = z(t_0)e^{i\phi} \tag{5.59}$$

where

$$\phi = \int_a^b kds - \omega(t - t_0) = (\omega/c) \int_a^b nds - \omega(t - t_0). \tag{5.60}$$

The corresponding phasors add up for paths near that path which extremizes the phase $\int_a^b n\ ds$. That is, they add up along the path that extremizes the optical path $\int_a^b n\ ds$. This is just Fermat's principle of stationary time, since the time for light to follow a particular path is

$$t = \int_a^b ds/v = \frac{1}{c} \int_a^b n\ ds. \tag{5.61}$$

According to Fermat, light "rays" take minimum-time paths between a and b.. According to Feynman, photons take *all* paths between a and b, but it is mainly paths near the stationary path that contribute to the total amplitude. Who is right? That is easy to test: although we may not think that photons take paths that differ from the stationary-time path, if we try to block off these alternative paths the signal at the detector changes, so they are really there. For example, if the source is in air and the detector is in glass, according to Fermat's principle and Snell's law, the light travels only by the path for which the angles of incidence and reflection obey $n_1 \sin\theta_1 = n_2 \sin\theta_2$, and the path that satisfies this law crosses the air-glass interface at only a single point. If so, it would make no difference if we blocked off all other points on the interface. But it *does* make a difference, because if we block off all other points, leaving only a tiny hole for the light to pass through, it will be diffracted, changing the amount of light observed at the detector, and allowing other detectors in the glass to receive light, even in directions for which Snell's law is not obeyed.

NONRELATIVISTIC MASSIVE PARTICLES

Up to now we have described the behavior of massive particles only when they are free, in which case their classical paths are straight lines. The probability amplitude for a path of length s in this case is

$$z = z(0)e^{(i/\hbar)p_\mu x_\mu} = z(0)e^{(i/\hbar)(ps-Et)}, \tag{5.62}$$

where $z(0)$ is the amplitude at the source at time $t = 0$, and p and E are the momentum magnitude and energy, both constant along the path.

For nonrelativistic particles encountering forces as they move from a to b, this expression cannot be correct, because even with a conserved energy

$E = T + U$, the momentum magnitude $p = \sqrt{2mT} = \sqrt{2m(E - U(\boldsymbol{r}))}$ is not conserved because the potential energy generally depends upon position. The expression for the amplitude $z = z(0)e^{(i/\hbar)(ps-Et)}$ therefore no longer makes sense, because p keeps changing; it has to be replaced by

$$z = z(0)e^{(i/\hbar)[(p_1\Delta s_1+p_2\Delta s_2+p_3\Delta s_3+\ldots)-Et]} = z(0)e^{(i/\hbar)(\int_a^b p\ ds-Et)}, \tag{5.63}$$

where the varying momentum is integrated over the path between the source and detector. The total probability amplitude of the particle beginning at point a and ending at b at time t_f is therefore

$$z_{\mathrm{T}} = z(0)e^{-iEt/\hbar}\sum e^{(i/\hbar)\int_a^b p\ ds}, \tag{5.64}$$

summing over all paths between a and b.

For any particular path the integral is

$$\int_a^b p\ ds = \int_a^b \sqrt{2m(E - U(\boldsymbol{r}))}\ ds. \tag{5.65}$$

In the case of a free particle, the classical path is the path for which the product ps is minimized, which means (because p is fixed in this case) that s is minimized, which of course is a straight line. Now that potential energies have been included, it is the *integral* $\int_a^b \sqrt{E - U(\boldsymbol{r})}\ ds$ that has to be minimized.

In classical mechanics the integral

$$J = \int_a^b \sqrt{E - U(\boldsymbol{r})}\ ds \tag{5.66}$$

is called the **Jacobi Action**, named for the German mathematician Carl Gustav Jacob Jacobi (1804-1851). Finding the path that makes J stationary is called the **Jacobi Principle of Stationary Action**: this is the classical path between a and b. Note that *time* does not appear in the Jacobi Action, so making J stationary provides the path *shape*, but not the dynamical equations of motion.

Without quantum mechanics, Jacobi's principle seems quite mysterious: Why should a particle choose that path that minimizes (or maximizes, or otherwise makes stationary) the Jacobi Action? Now we know why it works: According to Feynman's sum over paths, if the path s_0 happens to be the one that minimizes J, then nearby paths have nearly the same phase, so the corresponding phasors add up to give a large result. But paths surrounding some other, arbitrary path have more rapidly varying phases, so their phasors tend to cancel one another out. The integral $\int_a^b \sqrt{E - U(\boldsymbol{r})}\, ds$ is simply the integral $\int_a^b p\, ds$ in disguise.

EXAMPLE 5-4: Path shape for particles in uniform gravity

A nonrelativistic particle of mass m moves in a uniform gravitation field g. What is the shape of its classical path?

The potential energy is $U = mgy$, so the Jacobi Action is

$$J = \int_a^b \sqrt{E - mgy}\; ds = \int_a^b \sqrt{E - mgy}\; \sqrt{dx^2 + dy^2}. \tag{5.67}$$

We can choose either x or y as the independent variable: It is easier to choose y, because then the integral has the form

$$J = \int_a^b \sqrt{E - mgy}\sqrt{1 + (dx/dy)^2} dy = \int_a^b f(x, dx/dy, y) dy \tag{5.68}$$

where x is missing from f, and so x is a cyclic coordinate. Therefore from Euler's equation

$$\frac{\partial f}{\partial x} - \frac{d}{dy}\frac{\partial f}{\partial x'} = 0 \tag{5.69}$$

it follows that we have a first integral $\partial f/\partial x' = \text{constant}$ along the stationary path. That is,

$$\sqrt{E - mgy}\left(\frac{x'}{\sqrt{1 + x'^2}}\right) = \alpha, \quad \text{a constant.} \tag{5.70}$$

Solving for x',

$$x' \equiv \frac{dx}{dy} = \pm \frac{\alpha}{\sqrt{E - \alpha^2 - mgy}}. \tag{5.71}$$

Integrating both sides over y, choosing $x = 0$ and $y = y_0$ at the point where the path is horizontal (*i.e.*, ,where $x' = dx/dy = \infty$), and solving for $y(x)$ gives

$$y(x) = y_0 - \frac{mgx^2}{4(E - mgy_0)} \tag{5.72}$$

which is a parabola, as we knew it would be all along.[5]

[5] It is easy to show that this agrees with what is found from Newton's laws of motion; see problem 5-??

The classical path for a particle moving in a uniform gravitational field is a parabola, but according to quantum mechanics it is not the *only* path the particle takes. It actually takes all possible paths, but in the classical limit, where the deBroglie wavelength is very small compared with any physical dimension in the problem, the classical path is the only path for which the amplitudes of nearby paths all add up in phase. For other paths, nearby paths have substantially different phases, and they cancel one another out. ————————

5.7 Why Hamilton's principle?

We have already shown that quantum-mechanical amplitudes in Feynman's sum-over-paths add up *in phase* near the classical path, so that in the limit of small de Broglie wavelengths one observes only the classical path. We have assumed that the energy of a particle is the same whichever path it takes, so if it travels by a longer path it must have left the source earlier to arrive at the detector by the time the observation is made.

What we have *not* done so far is to show how *Hamilton's principle* comes about. This is the central question, because Lagrange's equations and all of the dynamical results that flow from them can be derived from making the action

$$S = \int_{\mathrm{t}_0}^{t_f} L dt \tag{5.73}$$

stationary. Where does this principle come from? And how does a classical particle "know" that it should follow a stationary-action path?

The key is to notice that the action is an integral from a fixed *initial* time to a fixed *final time*. Therefore we need to evaluate Feynman's sum-over-paths for a particle that not only ends at the final position at definite time t_f, but also starts at the source at a definite time t_0. For any path the particle takes, its starting and ending times must be the same, so in this case it is not the ending time and the energy that is the same for all paths, but the ending time and the starting time.

Suppose that a free particle leaves a source at point a at time $t = 0$; it has a particular energy E and so the magnitude of its momentum is also known. According to quantum Rule 3, the probability amplitude for it to reach its destination at b at time t is

$$z = z_0\, e^{(i/\hbar)(ps-Et)} = z_0\, e^{(i/\hbar)(\boldsymbol{p}\cdot\boldsymbol{r}-Et)} \tag{5.74}$$

where z_0 is the amplitude at the source at the starting time $t = 0$. We have the quantity $\boldsymbol{p} \cdot \boldsymbol{r} - Et = p_\mu \eta_{\mu\nu} x_\nu$ (using Einstein's summation convention), so

$$z = z_0 \; e^{(i/\hbar)(p_\mu \eta_{\mu\nu} x_\nu)} \equiv z_0 \; e^{i\phi} \tag{5.75}$$

in terms of a phase ϕ. In these expressions we have assumed that the energy and momentum components remain constant over the entire path, which is only true in a very special case. More generally, the components p_μ of the four-vector momentum vary along the path, so we need to write the phase ϕ as the integral

$$\phi = \frac{1}{\hbar} \int_{\mathrm{t}_0}^{t_f} p_\mu \eta_{\mu\nu} dx_\nu. \tag{5.76}$$

For a massive particle $p_\mu = m u_\mu$, where the four-vector velocity is $u_\mu = dx_\mu / d\tau$. We showed in Chapter 2 that $u_\mu u_\mu = -c^2$, so

$$\int p_\mu \eta_{\mu\nu} dx_\nu = \int (m u_\mu) \eta_{\mu\nu} u_\nu d\tau = -mc^2 \int d\tau = -mc^2 \tau \tag{5.77}$$

where τ is the proper time for the particle to travel by a particular path. (Recall that the proper time is the time interval read by a hypothetical clock carried along with the particle.) Therefore the total amplitude to reach the detector at a fixed time t_f is

$$z_T = z_0 \sum e^{-(imc^2/\hbar)\tau}. \tag{5.78}$$

Now since the particle begins at the source S at definite time t_0 and ends at the detector D at a definite time t_f, if it takes a shorter path its speed, and therefore also is momentum and energy are *less*; but if it takes a longer path its speed must be *larger*, so its momentum and energy are larger as well. The speed affects the proper time as well; the proper time for a particular path is

$$\tau = \int_{\mathrm{t}_0}^{t_f} dt \sqrt{1 - v^2/c^2}, \tag{5.79}$$

so τ is *maximized* for a path with the least possible speed v, which is the straight-line path between the source and detector. The speed along other paths must be larger, so other paths have smaller proper times.[6] The path about which all the amplitudes tend to add up (which is especially sharp in the classical regime where $mc^2\tau/\hbar \gg 1$) is the path corresponding to motion the extremum of τ – a straight line at constant speed.

5.8 Conclusions

Experiments with light and particle beams show the following features:

(1) Light, as photons, and massive particles as well, have both a particle and wave nature. Their energy and momenta are related to their frequency and wavelength by the de Broglie relations

$$E = h\nu \quad \text{and} \quad p = h/\lambda \tag{5.80}$$

where h is Planck's constant.

(2) Photons, electrons, atoms, and all other particles are detected as discrete lumps as though they were classical particles. If huge numbers of them are detected, they have an apparently continuous distribution.

(3) High-energy particles (photons, electrons, atoms.,..) sent through a double-slit system appear to have the same bimodal distribution at the detecting plane as that predicted for classical particles.

(4) Lower-velocity particles (lower frequency photons, slower electrons and atoms) show wavelike interference effects at the screen. Giving more speed to atoms makes the interference oscillations tighter, until they become so tight they cannot be separated, and they look just like the bimodal distribution predicted for classical particles.

(5) A particle takes all available paths from a source to a detector. The probability of reaching the detector is equal to the absolute square of the sum of amplitudes to travel by all possible paths. Near the classical path the amplitudes add up to give a strong total amplitude, especially if the de Broglie wavelength of the particle is small compared with the physical dimensions in an experiment.

(6) Fermat's principle for light and the Jacobi principle of stationary action

[6]Note that this is a way to view the twin paradox; the accelerating twin ends up younger when they get back together.

for massive particles can be used to find particle path shapes. They follow from the quantum rules.

(7) Hamilton's principle, which we have used to find the dynamical equations of motion of massive particles, also follows from the quantum mechanical sum-over-paths.

In Part I of this book we have encountered three strikingly different pictures of the nature of classical motion.

(i) In the Newtonian picture, including the extension of this picture into the relativistic regime, a massive particle moves along a path determined by its initial position and velocity and the forces encountered along the way. The particle knows nothing of the future, where it is going or how long it will take to get there.

(ii) In the picture accompanying Hamilton's principle, a massive particle moves along a single path that makes stationary the action S, a functional that depends upon the beginning points, end points, and potential energies encountered along the path. Rather than specifying the initial position and velocity, as in the Newtonian picture, with Hamilton's principle we specify the initial and final positions. This is not determinism based entirely upon the past and present, but determinism based upon both the past and future. A natural question for those with a Newtonian intuition, is: How does the particle "know" to choose that path which makes stationary the action between given endpoints?

(iii) In the quantum mechanical view, using the sum-over-paths formulation of Feynman, a particle, massive or massless, takes *all* paths between given initial and final points: That is how it "knows" the classical path. The probability that it will reach the final point at a certain time is given by the absolute square of a total complex probability amplitude, where the total amplitude is the sum of the amplitudes for all possible paths. The amplitudes reinforce one another along the path that makes stationary the action, especially if the de Broglie wavelength of the particle is small compared with any physical dimensions in the environment. Thus Hamilton's principle for massive particles emerges naturally from quantum mechanics in the case of small-wavelength particles. The picture of motion here is more "Darwinian" than deterministic: Every path is tried, but in the classical limit only the "fittest" (the stationary-action path) survives. The determinism of Newton is discovered to be only an illusion, emerging from the nondeterministic theory of quantum mechanics.

Problems

PROBLEM 5-1: Maxwell's equations for the electric field $\boldsymbol{E}$ and magnetic field $\boldsymbol{B}$ are

$$\nabla \cdot \boldsymbol{E} = \rho/\epsilon_0, \quad \nabla \cdot \boldsymbol{B} = 0, \quad \nabla \times \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t}, \quad \nabla \times \boldsymbol{B} = \mu_0 \boldsymbol{J} + \mu_0 \epsilon_0 \frac{\partial \boldsymbol{E}}{\partial t},$$

where ρ is the charge density, $\boldsymbol{J}$ is the current density, and ϵ_0 and μ_0 are (respectively) the permittivity and permeability of the vacuum, both constants. Derive the vacuum wave equations for $\boldsymbol{E}$ and for $\boldsymbol{B}$, and relate ϵ_0 and μ_0 to the speed of electromagnetic waves c. *Hint*: You can use the vector identity $\nabla \times (\nabla \times \boldsymbol{A}) = \nabla(\nabla \cdot \boldsymbol{A}) - \nabla^2 \boldsymbol{A}$ where $\boldsymbol{A}$ is any vector.

PROBLEM 5-2: Photons of wavelength 580 nm pass through a double-slit system, where the distance between the slits is $d = 0.16$ nm and the slit width is $a = 0.02$ nm. If the detecting screen is a distance $D = 60$ cm from the slits, what is the linear distance from the central maximum to the first minimum in the diffraction envelope?

PROBLEM 5-3: Photons are projected through a double-slit system. (a) What must be the ratio d/a of the slit separation to slit width, so that there will be exactly nine interference maxima within the central diffraction envelope? (b) Is any change observed on the detecting screen if the photon wavelength is changed from λ_0 to $2\lambda_0$? If so, what? (c) If 10^4 photons are counted within the central interference maximum, about how many do you expect will be counted within the last interference maximum that fits within the central diffraction envelope?

PROBLEM 5-4: A beam of monoenergetic photons is directed at a triple-slit system, where the distance between adjacent slits is d, and the photon wavelength is $\lambda = d/2$. Find the angles θ from the forward direction for which there are (a) interference maxima (b) interference minima. (c) Then show that some maxima have the same maximum probability as the central peak, but that others have a smaller maximum. Find the ratio of the larger to the smaller maximum probabilities.

PROBLEM 5-5: A beam of 10 keV photons is directed at a double-slit system and the interference pattern is measured on the detecting plane. The wavelength of these photons is less than the slit separation. Then electrons are accelerated so their (nonrelativistic) kinetic energies are also 10 keV; these electrons are then directed at the same double-slit system, and their interference pattern is measured on the same detecting plane. If the distance between two adjacent photon interference maxima on the detecting screen is y_0, what is the distance between two adjacent *electron* interference maxima? (Note that the mass energy of an electron is 0.5 MeV.)

PROBLEM 5-6: Consider a grating composed of four very narrow slits each separated by a distance d. (a) What is the probability that a photon strikes a detector centered at the central maximum if the probability that a photon is counted by this detector with a single slit open is r? (b) What is the probability that a photon is counted at the first minimum of this four-slit grating if the bottom two slits are closed? (From *Quantum Physics* by John S. Townsend.)

PROBLEM 5-7: Example 2 considered a set of kinked paths about a straight-line path. (a) Using the same set of alternative paths, suppose one considered the sum of phasors about the path with $n = 50$ instead of the sum about the $n = 0$ straight-line path. In particular, if one summed from $n = 25$ to $n = 75$, ± 25 about $n = 50$, how would the sum of phasors differ from the sum for paths about $n = 0$? What physical conclusion can you draw from this? (b) Now returning to the set of kinked paths about the straight-line $n = 0$ path, draw the phasor diagram if the wave number k of the particle were doubled (*i.e.*, , if the de Broglie wavelength λ were halved.) What can be concluded about the physical difference between this case and that used in Example 2?

PROBLEM 5-8: Example 2 considers a particular class of paths near a straight-line path. A different class of paths consists of a set of *parabolas* of the form $y = n\alpha(1 - x/x_0)$ fit to the endpoints of the straight line at $(x, y) = (0, 0)$ and $(x, y) = (x_0, 0)$. Here α is a (small) constant, and $n = 0, \pm 1, \pm 2, ...$ Let $\alpha = 0.1x_0$, and draw a careful phasor diagram including enough integers n to see the Cornu spiral behavior and obtain a good estimate of the sum of all these phasors.

PROBLEM 5-9: Judge whether or not the following situations are consistent with classical paths. (a) A nitrogen molecule moving with average kinetic energy $< 3/2kT >$ at room temperature $T = 300$ K (k is Boltzmann's constant.) (b) A typical hydrogen atom caught in a trap at temperature $T = 0.1$ K. (c) A typical electron in the center of the Sun, at temperature $T = 15 \times 10^6$ K.

PROBLEM 5-10: (a) What condition would have to be met so that the motion of a 135 g baseball would be inconsistent with a classical path? Is this a potentially feasible condition? (b) If we could adjust the value of Planck's constant, how large would it have to be so that the ball in a baseball game would fail to follow classical paths?

PROBLEM 5-11: According to the Heisenberg indeterminacy principle $\Delta x \; \Delta p \geq \hbar$, the uncertainty in position of a particle multiplied by the uncertainty in its momentum must be greater than Planck's constant divided by 2π. The neutrons in a particular atomic nucleus are confined to be within a nucleus of diameter 2.0 fm (1 fm $= 10^{-15}$ m). Can these neutrons be properly thought of as traveling along classical paths? Explain.

PROBLEM 5-12: Show from the Newtonian equations $x = v_{0\text{x}}t$ and $y = y_0 - (1/2)gt^2$ for a particle moving in a uniform gravitational field g, that the shape of its path is a parabola, given by

$$y = y_0 - \frac{mgx^2}{4(E - mgy_0)}, \tag{5.81}$$

the same result we found using the Jacobi principle of least action.

PROBLEM 5-13: A particle of mass m can move in two dimensions under the influence of a repulsive spring-like force in the x direction, $F = +kx$. Find the shape of its classical path in the x, y plane using the Jacobi action.

PROBLEM 5-14: An object of mass m can move in two dimensions in response to the simple harmonic oscillator potential $U = (1/2)kr^2$, where k is the force constant and r is the distance from the origin. Using the Jacobi action, find the shape of the orbits using polar coordinates r and θ; that is, find $r(\theta)$ for the orbit. Show that the shapes are ellipses and circles centered at the origin $r = 0$.

PROBLEM 5-15: A comet of mass m moves in two dimensions in response to the central gravitational potential $U = -k/r$, where k is a constant and r is the distance from the Sun. Using the Jacobi action and polar coordinates (r, θ), find the possible shapes of the comet's orbit. Show that these are (a) a parabola, if the energy of the comet is $E = 0$; (b) a hyperbola if $E > 0$; (c) an ellipse or a circle if $E < 0$, where in each case $r = 0$ at one of the foci.

PROBLEM 5-16: The action for a massive free particle is $S = -mc^2 \int_a^b d\tau$, where m is its mass and τ is its proper time. Now suppose a particle is subject is a four-scalar field Φ_0 which is constant in both space and time, and that its action then takes the simple form

$$S = -mc^2 \int_a^b d\tau + k \int_a^b d\tau\ \Phi_0 \tag{5.82}$$

where k is a coupling constant. (a) In that case demonstrate that the particle behaves exactly as before, but with an effective mass that is changed by the new term. (b) If Φ_0 is truly independent of position and time, then the particle can never be free. So perhaps the first term in S is not really there at all. What would k have to be in this case to give the usual free-particle action? (b) Now suppose we add a third term to the action involving a scalar field $\Phi_1(t, x, y, z)$ that *does* depend on position and time, so the total action becomes

$$S = -mc^2 \int_a^b d\tau + k_1 \int_a^b d\tau\ \Phi_0 + k_2 \int_a^b d\tau\ \Phi_1. \tag{5.83}$$

Find the differential equations of motion of a *nonrelativistic* particle in this case. Is the result consistent with supposing that Φ_1 is a gravitational potential?

Emmy Noether (1882 - 1935)

Amelie Emmy Noether was born in Erlangen, Germany. She completed her dissertation in mathematics at the university there, and then worked at the Mathematical Institute of Erlangen without pay: women were largely excluded from academic positions. In 1915 she was invited by David Hilbert and Felix Klein to join the mathematics department at the renowned University of Gottingen. Other faculty objected appointing a woman, but she finally became a *Privatdozent* in 1919, and remained at Gottingen until 1933, when her position was revoked when the Nazis came to power, because she was Jewish. She left for Bryn Mawr College in Pennsylvania, where she worked and taught until her death in 1935.

Noether was one of the most creative mathematicians of the twentieth century, and has been acclaimed by many (including Einstein) as the greatest woman mathematician of all time. Her early work was on the theories of algebraic invariants and number fields, and included her invention of "Noether's theorem", discussed in this chapter. Later on she concentrated on abstract algebra, leading to her 1921 paper "Theory of Ideals in Ring Domains" that revolutionized the field, and then worked on noncommutative algebras and united the representation theory of groups with the theory of modules and ideals.

The Association for Women in Mathematics holds a Noether Lecture in mathematics every year; the Association characterizes Noether as "one of the great mathematicians of her time, someone who worked and struggled for what she loved and believed in. Her life and work remain a tremendous inspiration."