

More Two-State Systems

11-1 The Pauli spin matrices

We continue our discussion of two-state systems. At the end of the last chapter we were talking about a spin one-half particle in a magnetic field. We described the spin state by giving the amplitude C_1 that the z-component of spin angular momentum is $+\hbar/2$ and the amplitude C_2 that it is $-\hbar/2$. In earlier chapters we have called these base states $|+\rangle$ and $|-\rangle$. We will now go back to that notation, although we may occasionally find it convenient to use $|+\rangle$ or $|1\rangle$, and $|-\rangle$ or $|2\rangle$, interchangeably.

We saw in the last chapter that when a spin one-half particle with a magnetic moment μ is in a magnetic field $\mathbf{B} = (B_x, B_y, B_z)$, the amplitudes $C_+(=C_1)$ and $C_-(=C_2)$ are connected by the following differential equations:

$$\begin{aligned} i\hbar \frac{dC_+}{dt} &= -\mu[B_z C_+ + (B_x - iB_y)C_-], \\ i\hbar \frac{dC_-}{dt} &= -\mu[(B_x + iB_y)C_+ - B_z C_-]. \end{aligned} \tag{11.1}$$

In other words, the Hamiltonian matrix H_{ij} is

$$\begin{aligned} H_{11} &= -\mu B_z, & H_{12} &= -\mu(B_x - iB_y), \\ H_{21} &= -\mu(B_x + iB_y), & H_{22} &= +\mu B_z. \end{aligned} \tag{11.2}$$

And Eqs. (11.1) are, of course, the same as

$$i\hbar \frac{dC_i}{dt} = \sum_j H_{ij} C_j, \tag{11.3}$$

where i and j take on the values $+$ and $-$ (or 1 and 2).

The two-state system of the electron spin is so important that it is very useful to have a neater way of writing things. We will now make a little mathematical digression to show you how people usually write the equations of a two-state system. It is done this way: First, note that each term in the Hamiltonian is proportional to μ and to some component of \mathbf{B} ; we can then—*purely formally*—write that

$$H_{ij} = -\mu[\sigma_{ij}^x B_x + \sigma_{ij}^y B_y + \sigma_{ij}^z B_z]. \tag{11.4}$$

There is no new physics here; this equation just means that the coefficients σ_{ij}^x , σ_{ij}^y , and σ_{ij}^z —there are $4 \times 3 = 12$ of them—can be figured out so that (11.4) is identical with (11.2).

Let's see what they have to be. We start with B_z . Since B_z appears only in H_{11} and H_{22} , everything will be O.K. if

$$\begin{aligned} \sigma_{11}^z &= 1, & \sigma_{12}^z &= 0, \\ \sigma_{21}^z &= 0, & \sigma_{22}^z &= -1. \end{aligned}$$

We often write the matrix H_{ij} as a little table like this:

$$H_{ij} = \begin{matrix} & \begin{matrix} \xrightarrow{j} \\ \left(\begin{matrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{matrix} \right) \end{matrix} \\ \begin{matrix} \downarrow i \\ \end{matrix} \end{matrix}$$

11-1 The Pauli spin matrices

11-2 The spin matrices as operators

11-3 The solution of the two-state equations

11-4 The polarization states of the photon

11-5 The neutral K-meson†

11-6 Generalization to N -state systems

Review: Chapter 35, Vol. I, Polarization

† This section should be omitted on the first reading of this book. It is more advanced than is appropriate in a first course.

For the Hamiltonian of a spin one-half particle in the magnetic field B_z , this is the same as

$$H_{ij} = \overset{i \downarrow}{\underset{j \rightarrow}{\begin{pmatrix} -\mu B_z & -\mu(B_x - iB_y) \\ -\mu(B_x + iB_y) & +\mu B_z \end{pmatrix}}}.$$

In the same way, we can write the coefficients σ_{ij}^z as the matrix

$$\sigma_{ij}^z = \overset{i \downarrow}{\underset{j \rightarrow}{\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}}}. \quad (11.5)$$

Working with the coefficients of B_x , we get that the terms of σ_x have to be

$$\begin{aligned} \sigma_{11}^x &= 0, & \sigma_{12}^x &= 1, \\ \sigma_{21}^x &= 1, & \sigma_{22}^x &= 0. \end{aligned}$$

Or, in shorthand,

$$\sigma_{ij}^x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \quad (11.6)$$

Finally, looking at B_y , we get

$$\begin{aligned} \sigma_{11}^y &= 0, & \sigma_{12}^y &= -i, \\ \sigma_{21}^y &= i, & \sigma_{22}^y &= 0; \end{aligned}$$

or

$$\sigma_{ij}^y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}. \quad (11.7)$$

Table 11-1
The Pauli spin matrices

$\sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$
$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$
$\sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$
$1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

With these three sigma matrices, Eqs. (11.2) and (11.4) are identical. To leave room for the subscripts i and j , we have shown which σ goes with which component of \mathbf{B} by putting x , y , and z as superscripts. Usually, however, the i and j are omitted—it's easy to imagine they are there—and the x , y , z are written as subscripts. Then Eq. (11.4) is written

$$H = -\mu[\sigma_x B_x + \sigma_y B_y + \sigma_z B_z]. \quad (11.8)$$

Because the sigma matrices are so important—they are used all the time by the professionals—we have gathered them together in Table 11-1. (Anyone who is going to work in quantum physics really has to memorize them.) They are also called the *Pauli spin matrices* after the physicist who invented them.

In the table we have included one more two-by-two matrix which is needed if we want to be able to take care of a system which has two spin states of the same energy, or if we want to choose a different zero energy. For such situations we must add $E_0 C_+$ to the first equation in (11.1) and $E_0 C_-$ to the second equation. We can include this in the new notation if we define the *unit matrix* "1" as δ_{ij} ,

$$1 = \delta_{ij} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad (11.9)$$

and rewrite Eq. (11.8) as

$$H = E_0 \delta_{ij} - \mu(\sigma_x B_x + \sigma_y B_y + \sigma_z B_z). \quad (11.10)$$

Usually, it is *understood* that any constant like E_0 is automatically to be multiplied by the unit matrix; then one writes simply

$$H = E_0 - \mu(\sigma_x B_x + \sigma_y B_y + \sigma_z B_z). \quad (11.11)$$

One reason the spin matrices are useful is that *any* two-by-two matrix at all can be written in terms of them. Any matrix you can write has four numbers in it, say,

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

It can always be written as a linear combination of four matrices. For example,

$$M = a \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + b \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + c \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} + d \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

There are many ways of doing it, but one special way is to say that M is a certain amount of σ_x , plus a certain amount of σ_y , and so on, like this:

$$M = \alpha I + \beta \sigma_x + \gamma \sigma_y + \delta \sigma_z,$$

where the “amounts” α , β , γ , and δ may, in general, be complex numbers.

Since any two-by-two matrix can be represented in terms of the unit matrix and the sigma matrices, we have all that we ever need for *any* two-state system. No matter what the two-state system—the ammonia molecule, the magenta dye, anything—the Hamiltonian equation can be written in terms of the sigmas. Although the sigmas seem to have a geometrical significance in the physical situation of an electron in a magnetic field, they can also be thought of as just useful matrices, which can be used for any two-state problem.

For instance, in one way of looking at things a proton and a neutron can be thought of as the same particle in either of two states. We say the *nucleon* (proton or neutron) is a two-state system—in this case, two states with respect to its charge. When looked at that way, the $|1\rangle$ state can represent the proton and the $|2\rangle$ state can represent the neutron. People say that the nucleon has two “isotopic-spin” states.

Since we will be using the sigma matrices as the “arithmetic” of the quantum mechanics of two-state systems, let’s review quickly the conventions of matrix algebra. By the “sum” of any two or more matrices we mean just what was obvious in Eq. (11.4). In general, if we “add” two matrices A and B , the “sum” C means that each term C_{ij} is given by

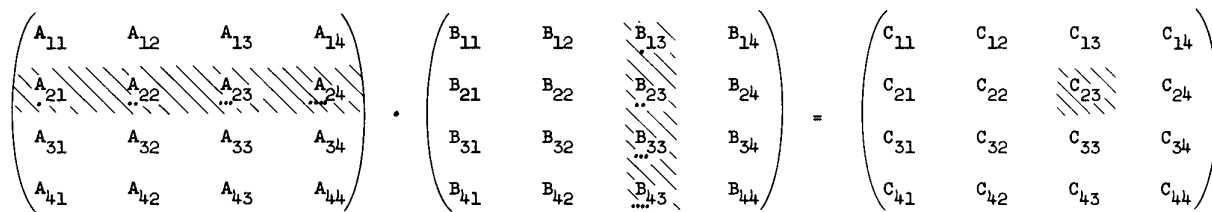
$$C_{ij} = A_{ij} + B_{ij}.$$

Each term of C is the sum of the terms in the same slots of A and B .

In Section 5–6 we have already encountered the idea of a matrix “product.” The same idea will be useful in dealing with the sigma matrices. In general, the “product” of two matrices A and B (in that order) is defined to be a matrix C whose elements are

$$C_{ij} = \sum_k A_{ik} B_{kj}. \quad (11.12)$$

It is the sum of products of terms taken in pairs from the i th row of A and the k th column of B . If the matrices are written out in tabular form as in Fig. 11–1, there is a good “system” for getting the terms of the product matrix. Suppose you are calculating C_{23} . You run your left index finger *along* the *second row* of A and your right index finger *down* the *third column* of B , multiplying each pair and adding as you go. We have tried to indicate how to do it in the figure.



$$C_{ik} = \sum_j A_{ij} B_{jk}$$

Example: $C_{23} = A_{21} B_{13} + A_{22} B_{23} + A_{23} B_{33} + A_{24} B_{43}$

Fig. 11–1. Multiplying two matrices.

It is, of course, particularly simple for two-by-two matrices. For instance, if we multiply σ_x times σ_x , we get

$$\sigma_x^2 = \sigma_x \cdot \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

which is just the unit matrix 1. Or, for another example, let's work out $\sigma_x\sigma_y$:

$$\sigma_x\sigma_y = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}.$$

Referring to Table 11-1, you see that the product is just i times the matrix σ_z . (Remember that a number times a matrix just multiplies each term of the matrix.) Since the products of the sigmas taken two at a time are important—as well as rather amusing—we have listed them all in Table 11-2. You can work them out as we have done for σ_x^2 and $\sigma_x\sigma_y$.

There's another very important and interesting point about these σ matrices. We can imagine, if we wish, that the three matrices σ_x , σ_y , and σ_z are analogous to the three components of a vector—it is sometimes called the “sigma vector” and is written σ . It is really a “matrix vector” or a “vector matrix.” It is three different matrices—one matrix associated with each axis, x , y , and z . With it, we can write the Hamiltonian of the system in a nice form which works in any coordinate system:

$$H = -\mu\sigma \cdot \mathbf{B}. \quad (11.13)$$

Although we have written our three matrices in the representation in which “up” and “down” are in the z -direction—so that σ_z has a particular simplicity—we could figure out what the matrices would look like in some other representation. Although it takes a lot of algebra, you can show that they change among themselves like the components of a vector. (We won't, however, worry about proving it right now. You can check it if you want.) You can use σ in different coordinate systems as though it is a vector.

You remember that the H is related to energy in quantum mechanics. It is, in fact, just equal to the energy in the simple situation where there is only one state. Even for two-state systems of the electron spin, when we write the Hamiltonian as in Eq. (11.13), it looks very much like the *classical* formula for the energy of a little magnet with magnetic moment μ in a magnetic field. \mathbf{B} Classically, we would say

$$U = -\mu \cdot \mathbf{B}, \quad (11.14)$$

where μ is the property of the object and \mathbf{B} is an external field. We can imagine that Eq. (11.14) can be converted to (11.13) if we replace the classical energy by the Hamiltonian and the classical μ by the matrix $\mu\sigma$. Then, after this purely formal substitution, we interpret the result as a matrix equation. It is sometimes said that to each quantity in classical physics there corresponds a matrix in quantum mechanics. It is really more correct to say that the Hamiltonian matrix corresponds to the energy, and any quantity that can be defined via energy has a corresponding matrix.

For example, the magnetic moment can be defined via energy by saying that the energy in an external field \mathbf{B} is $-\mu \cdot \mathbf{B}$. This *defines* the magnetic moment vector μ . Then we look at the formula for the Hamiltonian of a real (quantum) object in a magnetic field and try to identify whatever the matrices are that correspond to the various quantities in the classical formula. That's the trick by which *sometimes* classical quantities have their quantum counterparts.

You may try, if you want, to understand how a classical vector is equal to a matrix $\mu\sigma$, and maybe you will discover something—but don't break your head on it. That's not the idea—they are *not equal*. Quantum mechanics is a different kind of a theory to represent the world. It just happens that there are certain correspondences which are hardly more than mnemonic devices—things to remember with. That is, you remember Eq. (11.14) when you learn classical physics;

Table 11-2

Products of the spin matrices

$\sigma_x^2 = 1$
$\sigma_y^2 = 1$
$\sigma_z^2 = 1$
$\sigma_x\sigma_y = -\sigma_y\sigma_x = i\sigma_z$
$\sigma_y\sigma_z = -\sigma_z\sigma_y = i\sigma_x$
$\sigma_z\sigma_x = -\sigma_x\sigma_z = i\sigma_y$

then if you remember the correspondence $\mu \rightarrow \mu\sigma$, you have a handle for remembering Eq. (11.13). Of course, nature knows the quantum mechanics, and the classical mechanics is only an approximation; so there is no mystery in the fact that in classical mechanics there is some shadow of quantum mechanical laws—which are truly the ones underneath. To reconstruct the original object from the shadow is not possible in any direct way, but the shadow does help you to remember what the object looks like. Equation (11.13) is the truth, and Eq. (11.14) is the shadow. Because we learn classical mechanics first, we would like to be able to get the quantum formula from it, but there is no sure-fire scheme for doing that. We must always go back to the real world and discover the correct quantum mechanical equations. When they come out looking like something in classical physics, we are in luck.

If the warnings above seem repetitious and appear to you to be belaboring self-evident truths about the relation of classical physics to quantum physics, please excuse the conditioned reflexes of a professor who has usually taught quantum mechanics to students who hadn't heard about Pauli spin matrices until they were in graduate school. Then they always seemed to be hoping that, somehow, quantum mechanics could be seen to follow as a logical consequence of classical mechanics which they had learned thoroughly years before. (Perhaps they wanted to avoid having to learn something new.) You have learned the classical formula, Eq. (11.14), only a few months ago—and then with warnings that it was inadequate—so maybe you will not be so unwilling to take the quantum formula, Eq. (11.13), as the basic truth.

11-2 The spin matrices as operators

While we are on the subject of mathematical notation, we would like to describe still *another* way of writing things—a way which is used very often because it is so compact. It follows directly from the notation introduced in Chapter 8. If we have a system in a state $|\psi(t)\rangle$, which varies with time, we can—as we did in Eq. (8.31)—write the amplitude that the system would be in the state $|i\rangle$ at $t + \Delta t$ as

$$\langle i | \psi(t + \Delta t) \rangle = \sum_j \langle i | U(t, t + \Delta t) | j \rangle \langle j | \psi(t) \rangle.$$

The matrix element $\langle i | U(t, t + \Delta t) | j \rangle$ is the amplitude that the base state $|j\rangle$ will be converted into the base state $|i\rangle$ in the time interval Δt . We then *defined* H_{ij} by writing

$$\langle i | U(t, t + \Delta t) | j \rangle = \delta_{ij} - \frac{i}{\hbar} H_{ij}(t) \Delta t,$$

and we showed that the amplitudes $C_i(t) = \langle i | \psi(t) \rangle$ were related by the differential equations

$$i\hbar \frac{dC_i}{dt} = \sum_j H_{ij} C_j. \quad (11.15)$$

If we write out the amplitudes C_i explicitly, the same equation appears as

$$i\hbar \frac{d}{dt} \langle i | \psi \rangle = \sum_j H_{ij} \langle j | \psi \rangle. \quad (11.16)$$

Now the matrix elements H_{ij} are also amplitudes which we can write as $\langle i | H | j \rangle$; our differential equation looks like this:

$$i\hbar \frac{d}{dt} \langle i | \psi \rangle = \sum_j \langle i | H | j \rangle \langle j | \psi \rangle. \quad (11.17)$$

We see that $-i/\hbar \langle i | H | j \rangle$ is the amplitude that—under the physical conditions described by H —a state $|j\rangle$ will, during the time dt , “generate” the state $|i\rangle$. (All of this is implicit in the discussion of Section 8-4.)

Now following the ideas of Section 8–2, we can drop out the common term $\langle i |$ in Eq. (11.17)—since it is true for any state $|i\rangle$ —and write that equation simply as

$$i\hbar \frac{d}{dt} |\psi\rangle = \sum_j H |j\rangle \langle j | \psi\rangle. \quad (11.18)$$

Or, going one step further, we can also remove the j and write

$$i\hbar \frac{d}{dt} |\psi\rangle = H |\psi\rangle. \quad (11.19)$$

In Chapter 8 we pointed out that when things are written this way, the H in $H |j\rangle$ or $H |\psi\rangle$ is called an *operator*. From now on we will put the little hat ($\hat{\ }$) over an operator to remind you that it *is* an operator and not just a number. We will write $\hat{H} |\psi\rangle$. Although the two equations (11.18) and (11.19) *mean exactly the same thing* as Eq. (11.17) or Eq. (11.15), we can *think* about them in a different way. For instance, we would describe Eq. (11.18) in this way: “The time derivative of the *state vector* $|\psi\rangle$ is equal to what you get by operating with the Hamiltonian *operator* \hat{H} on each base state, multiplying by the amplitude $\langle j | \psi\rangle$ that ψ is in the state j , and summing over all j .” Or Eq. (11.19) is described this way. “The time derivative (times $i\hbar$) of a state $|\psi\rangle$ is equal to what you get if you operate with the Hamiltonian \hat{H} on the state vector $|\psi\rangle$.” It’s just a shorthand way of saying what is in Eq. (11.17), but, as you will see, it can be a great convenience.

If we wish, we can carry the “abstraction” idea one more step. Equation (11.19) is true for *any state* $|\psi\rangle$. Also the left-hand side, $i\hbar d/dt$, is also an operator—it’s the operation “differentiate by t and multiply by $i\hbar$.” Therefore, Eq. (11.19) can also be thought of as an equation between operators—the operator equation

$$i\hbar \frac{d}{dt} = \hat{H}.$$

The Hamiltonian operator (within a constant) produces the same result as does d/dt when acting on any state. Remember that this equation—as well as Eq. (11.19)—is *not* a statement that the \hat{H} operator is just the identical *operation* as d/dt . The equations are the dynamical law of nature—the law of motion—for a quantum system.

Just to get some practice with these ideas, we will show you another way we could get to Eq. (11.18). You know that we can write any state $|\psi\rangle$ in terms of its projections into some base set [see Eq. (8.8)],

$$|\psi\rangle = \sum_i |i\rangle \langle i | \psi\rangle. \quad (11.20)$$

How does $|\psi\rangle$ change with time? Well, just take its derivative:

$$\frac{d}{dt} |\psi\rangle = \frac{d}{dt} \sum_i |i\rangle \langle i | \psi\rangle. \quad (11.21)$$

Now the base states $|i\rangle$ do not change with time (at least *we* are always taking them as definite fixed states), but the amplitudes $\langle i | \psi\rangle$ are numbers which may vary. So Eq. (11.21) becomes

$$\frac{d}{dt} |\psi\rangle = \sum_i |i\rangle \frac{d}{dt} \langle i | \psi\rangle. \quad (11.22)$$

Since we know $d\langle i | \psi\rangle/dt$ from Eq. (11.16), we get

$$\begin{aligned} \frac{d}{dt} |\psi\rangle &= -\frac{i}{\hbar} \sum_i |i\rangle \sum_j H_{ij} \langle j | \psi\rangle \\ &= -\frac{i}{\hbar} \sum_{ij} |i\rangle \langle i | H | j\rangle \langle j | \psi\rangle = -\frac{i}{\hbar} \sum_j H |j\rangle \langle j | \psi\rangle. \end{aligned}$$

This is Eq. (11.18) all over again.

So we have many ways of looking at the Hamiltonian. We can think of the set of coefficients H_{ij} as just a bunch of numbers, or we can think of the “amplitudes” $\langle i | H | j \rangle$, or we can think of the “matrix” H_{ij} , or we can think of the operator” \hat{H} . It all means the same thing.

Now let’s go back to our two-state systems. If we write the Hamiltonian in terms of the sigma matrices (with suitable numerical coefficients like B_x , etc.), we can clearly also think of σ_{ij}^x as an amplitude $\langle i | \sigma_x | j \rangle$ or, for short, as the operator $\hat{\sigma}_x$. If we use the operator idea, we can write the equation of motion of a state $|\psi\rangle$ in a magnetic field as

$$i\hbar \frac{d}{dt} |\psi\rangle = -\mu(B_x \hat{\sigma}_x + B_y \hat{\sigma}_y + B_z \hat{\sigma}_z) |\psi\rangle. \quad (11.23)$$

When we want to “use” such an equation we will normally have to express $|\psi\rangle$ in terms of base vectors (just as we have to find the components of space vectors when we want specific numbers). So we will usually want to put Eq. (11.23) in the somewhat expanded form:

$$i\hbar \frac{d}{dt} |\psi\rangle = -\mu \sum_i (B_x \hat{\sigma}_x + B_y \hat{\sigma}_y + B_z \hat{\sigma}_z) |i\rangle \langle i | \psi\rangle. \quad (11.24)$$

Now you will see why the operator idea is so neat. To use Eq. (11.24) we need to know what happens when the $\hat{\sigma}$ operators work on each of the base states. Let’s find out. Suppose we have $\hat{\sigma}_z |+\rangle$; it is some vector $|?\rangle$, but what? Well, let’s multiply it on the left by $\langle + |$; we have

$$\langle + | \hat{\sigma}_z |+\rangle = \sigma_{11} = 1$$

(using Table 11-1). So we know that

$$\langle + | ? \rangle = 1. \quad (11.25)$$

Now let’s multiply $\hat{\sigma}_z |+\rangle$ on the left by $\langle - |$. We get

$$\langle - | \hat{\sigma}_z |+\rangle = \sigma_{21} = 0;$$

so

$$\langle - | ? \rangle = 0. \quad (11.26)$$

There is only one state vector that satisfies both (11.25) and (11.26); it is $|+\rangle$. We discover then that

$$\hat{\sigma}_z |+\rangle = |+\rangle. \quad (11.27)$$

By this kind of argument you can easily show that all of the properties of the sigma matrices can be described in the operator notation by the set of rules given in Table 11-3.

If we have products of sigma matrices, they go over into products of operators. When two operators appear together as a product, you carry out first the operation with the operator which is farthest to the right. For instance, by $\hat{\sigma}_x \hat{\sigma}_y |+\rangle$ we are to understand $\hat{\sigma}_x(\hat{\sigma}_y |+\rangle)$. From Table 11-3, we get $\hat{\sigma}_y |+\rangle = i |-\rangle$, so

$$\hat{\sigma}_x \hat{\sigma}_y |+\rangle = \hat{\sigma}_x(i |-\rangle). \quad (11.28)$$

Now any number—like i —just moves through an operator (operators work only on state vectors); so Eq. (11.28) is the same as

$$\hat{\sigma}_x \hat{\sigma}_y |+\rangle = i \hat{\sigma}_x |-\rangle = i |+\rangle.$$

If you do the same thing for $\hat{\sigma}_x \hat{\sigma}_y |-\rangle$, you will find that

$$\hat{\sigma}_x \hat{\sigma}_y |-\rangle = -i |-\rangle.$$

Looking at Table 11-3, you see that $\hat{\sigma}_x \hat{\sigma}_y$ operating on $|+\rangle$ or $|-\rangle$ gives just what you get if you operate with $\hat{\sigma}_z$ and multiply by $-i$. We can, therefore, say

Table 11-3
Properties of the $\hat{\sigma}$ -operator

$\sigma_z +\rangle = +\rangle$
$\sigma_z -\rangle = - -\rangle$
$\sigma_x +\rangle = -\rangle$
$\sigma_x -\rangle = +\rangle$
$\sigma_y +\rangle = i -\rangle$
$\sigma_y -\rangle = -i +\rangle$

that the operation $\hat{\sigma}_x \hat{\sigma}_y$ is identical with the operation $i\hat{\sigma}_z$, and write this statement as an operator equation:

$$\hat{\sigma}_x \hat{\sigma}_y = i\hat{\sigma}_z. \quad (11.29)$$

Notice that this equation is identical with one of our matrix equations of Table 11-2. So again we see the correspondence between the matrix and operator points of view. Each of the equations in Table 11-2 can, therefore, also be considered as equations about the sigma operators. You can check that they do indeed follow from Table 11-3. It is best, when working with these things, *not* to keep track of whether a quantity like σ or H is an operator or a matrix. All the equations are the same either way, so Table 11-2 is for sigma operators, or for sigma matrices, as you wish.

11-3 The solution of the two-state equations

We can now write our two-state equation in various forms, for example, either as

$$i\hbar \frac{dC_i}{dt} = \sum_j H_{ij} C_j \quad (11.30)$$

or

$$i\hbar \frac{d|\psi\rangle}{dt} = \hat{H}|\psi\rangle.$$

They both mean the same thing. For a spin one-half particle in a magnetic field, the Hamiltonian H is given by Eq. (11.8) or by Eq. (11.13).

If the field is in the z -direction, then—as we have seen several times by now—the solution is that the state $|\psi\rangle$, whatever it is, precesses around the z -axis (just as if you were to take the physical object and rotate it bodily around the z -axis) at an angular velocity equal to twice the magnetic field times μ/\hbar . The same is true, of course, for a magnetic field along any other direction, because the physics is independent of the coordinate system. If we have a situation where the magnetic field varies from time to time in a complicated way, then we can analyze the situation in the following way. Suppose you start with the spin in the $+z$ -direction and you have an x -magnetic field. The spin starts to turn. Then if the x -field is turned off, the spin stops turning. Now if a z -field is turned on, the spin precesses about z , and so on. So depending on how the fields vary in time, you can figure out what the final state is—along what axis it will point. Then you can refer that state back to the original $|+\rangle$ and $|-\rangle$ with respect to z by using the projection formulas we had in Chapter 10 (or Chapter 6). If the state ends up with its spin in the direction (θ, ϕ) , it will have an up-amplitude $\cos(\theta/2)e^{-i\phi/2}$ and a down-amplitude $\sin(\theta/2)e^{+i\phi/2}$. That solves any problem. It is a word description of the solution of the differential equations.

The solution just described is sufficiently general to take care of *any two-state system*. Let's take our example of the ammonia molecule—including the effects of an electric field. If we describe the system in terms of the states $|I\rangle$ and $|II\rangle$, the equations look like this:

$$\begin{aligned} i\hbar \frac{dC_I}{dt} &= +AC_I + \mu\mathcal{E}C_{II}, \\ i\hbar \frac{dC_{II}}{dt} &= -AC_{II} + \mu\mathcal{E}C_I. \end{aligned} \quad (11.31)$$

You say, "No, I remember there was an E_0 in there." Well, we have shifted the origin of energy to make the E_0 zero. (You can always do that by changing both amplitudes by the same factor— $e^{iE_0 T/\hbar}$ —and get rid of any constant energy.) Now if corresponding equations always have the same solutions, then we really don't have to do it twice. If we look at these equations and look at Eq. (11.1), then we can make the following identification. Let's call $|I\rangle$ the state $|+\rangle$ and $|II\rangle$ the state $|-\rangle$. That *does not mean* that we are lining-up the ammonia in space, or that $|+\rangle$ and $|-\rangle$ has anything to do with the z -axis. It is purely artificial.

We have an artificial space that we might “call the ammonia molecule representative space,” or something—a three-dimensional “diagram” in which being “up” corresponds to having the molecule in the state $|I\rangle$ and being “down” along this false z -axis represents having a molecule in the state $|II\rangle$. Then, the equations will be identified as follows. First of all, you see that the Hamiltonian can be written in terms of the sigma matrices as

$$H = +A\sigma_z + \mu\epsilon\sigma_x. \quad (11.32)$$

Or, putting it another way, μB_z in Eq. (11.1) corresponds to $-A$ in Eq. (11.32), and μB_x corresponds to $-\mu\epsilon$. In our “model” space, then, we have a constant B field along the z -direction. If we have an electric field ϵ which is changing with time, then we have a B field along the x -direction which varies in proportion. *So the behavior of an electron in a magnetic field with a constant component in the z -direction and an oscillating component in the x -direction is mathematically analogous and corresponds exactly to the behavior of an ammonia molecule in an oscillating electric field.* Unfortunately, we do not have the time to go any further into the details of this correspondence, or to work out any of the technical details. We only wished to make the point that *all* systems of two states can be made analogous to a spin one-half object precessing in a magnetic field.

11-4 The polarization states of the photon

There are a number of other two-state systems which are interesting to study, and the first new one we would like to talk about is the photon. To describe a photon we must first give its vector momentum. For a free photon, the frequency is determined by the momentum, so we don’t have to say also what the frequency is. After that, though, we still have a property called the polarization. Imagine that there is a photon coming at you with a definite monochromatic frequency (which will be kept the same throughout all this discussion so that we don’t have a variety of momentum states). Then there are two directions of polarization. In the classical theory, light can be described as having an electric field which oscillates horizontally or an electric field which oscillates vertically (for instance); these two kinds of light are called x -polarized and y -polarized light. The light can also be polarized in some other direction, which can be made up from the superposition of a field in the x -direction and one in the y -direction. Or if you take the x - and the y -components out of phase by 90° , you get an electric field that rotates—the light is elliptically polarized. (This is just a quick reminder of the classical theory of polarized light that we studied in Chapter 35, Vol. I.)

Now, however, suppose we have a *single* photon—just one. There is no electric field that we can discuss in the same way. All we have is *one photon*. But a photon has to have the analog of the classical phenomena of polarization. There must be at least two different kinds of photons. At first, you might think there should be an infinite variety—after all, the electric vector can point in all sorts of directions. We can, however, describe the polarization of a photon as a two-state system. A photon can be in the state $|x\rangle$ or in the state $|y\rangle$. By $|x\rangle$ we mean the polarization state of each one of the photons in a beam of light which *classically* is x -polarized light. On the other hand, by $|y\rangle$ we mean the polarization state of each of the photons in a y -polarized beam. And we can take $|x\rangle$ and $|y\rangle$ as our base states of a photon of given momentum pointing at you—in what we will call the z -direction. So there are two base states $|x\rangle$ and $|y\rangle$, and they are all that are needed to describe any photon at all.

For example, if we have a piece of polaroid set with its axis to pass light polarized in what we call the x -direction, and we send in a photon which we know is in the state $|y\rangle$, it will be absorbed by the polaroid. If we send in a photon which we know is in the state $|x\rangle$, it will come right through as $|x\rangle$. If we take a piece of calcite which takes a beam of polarized light and splits it into an $|x\rangle$ beam and a $|y\rangle$ beam, that piece of calcite is the complete analog of a Stern-Gerlach apparatus which splits a beam of silver atoms into the two states $|+\rangle$ and $|-\rangle$. So every-

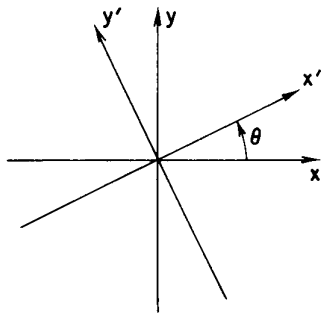


Fig. 11-2. Coordinates at right angles to the momentum vector of the photon.

thing we did before with particles and Stern-Gerlach apparatuses, we can do again with light and pieces of calcite. And what about light filtered through a piece of polaroid set at an angle θ ? Is that another state? Yes, indeed, it *is* another state. Let's call the axis of the polaroid x' to distinguish it from the axes of our base states. See Fig. 11-2. A photon that comes out will be in the state $|x'\rangle$. However, any state can be represented as a linear combination of base states, and the formula for the combination is, here,

$$|x'\rangle = \cos \theta |x\rangle + \sin \theta |y\rangle. \quad (11.33)$$

That is, if a photon comes through a piece of polaroid set at the angle θ (with respect to x), it can still be resolved into $|x\rangle$ and $|y\rangle$ beams—by a piece of calcite, for example. Or you can, if you wish, just analyze it into x - and y -components in your imagination. Either way, you will find the amplitude $\cos \theta$ to be in the $|x\rangle$ state and the amplitude $\sin \theta$ to be in the $|y\rangle$ state.

Now we ask this question: Suppose a photon is polarized in the x' -direction by a piece of polaroid set at the angle θ and arrives at a polaroid at the angle zero—as in Fig. 11-3; what will happen? With what probability will it get through? The answer is the following. After it gets through the first polaroid, it is definitely in the state $|x'\rangle$. The second polaroid will let the photon through if it is in the state $|x\rangle$ (but absorb it if it is the state $|y\rangle$). So we are asking with what probability does the photon appear to be in the state $|x\rangle$? We get that probability from the absolute square of amplitude $\langle x | x'\rangle$ that a photon in the state $|x'\rangle$ is also in the state $|x\rangle$. What is $\langle x | x'\rangle$? Just multiply Eq. (11.33) by $\langle x |$ to get

$$\langle x | x'\rangle = \cos \theta \langle x | x\rangle + \sin \theta \langle x | y\rangle.$$

Now $\langle x | y\rangle = 0$, from the physics—as they *must* be if $|x\rangle$ and $|y\rangle$ are base states—and $\langle x | x\rangle = 1$. So we get

$$\langle x | x'\rangle = \cos \theta,$$

and the probability is $\cos^2 \theta$. For example, if the first polaroid is set at 30° , a photon will get through $3/4$ of the time, and $1/4$ of the time it will heat the polaroid by being absorbed therein.

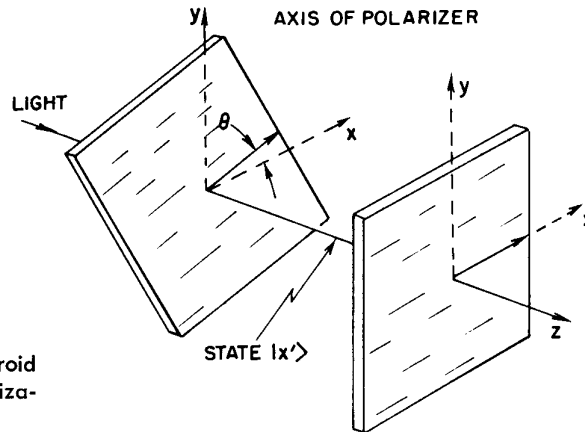


Fig. 11-3. Two sheets of polaroid with angle θ between planes of polarization.

Now let us see what happens classically in the same situation. We would have a beam of light with an electric field which is varying in some way or another—say “unpolarized.” After it gets through the first polaroid, the electric field is oscillating in the x' -direction with a size ε ; we would draw the field as an oscillating vector with a peak value ε_0 in a diagram like Fig. 11-4. Now when the light arrives at the second polaroid, only the x -component, $\varepsilon_0 \cos \theta$, of the electric field gets through. The *intensity* is proportional to the square of the field and, therefore, to $\varepsilon_0^2 \cos^2 \theta$. So the energy coming through is $\cos^2 \theta$ weaker than the energy which was entering the last polaroid.

The classical picture and the quantum picture give similar results. If you were to throw 10 billion photons at the second polaroid, and the average probability of each one going through is, say, 3/4, you would expect 3/4 of 10 billion would get through. Likewise, the energy that they would carry would be 3/4 of the energy that you attempted to put through. The classical theory says nothing about the statistics of the thing—it simply says that the energy that comes through will be precisely 3/4 of the energy which you were sending in. That is, of course, impossible if there is only one photon. There is no such thing as 3/4 of a photon. It is either *all* there, or it isn't there at all. Quantum mechanics tells us it is *all* there 3/4 of the time. The relation of the two theories is clear.

What about the other kinds of polarization? For example, right-hand circular polarization? In the classical theory, right-hand circular polarization has equal components in x and y which are 90° out of phase. In the quantum theory, a right-hand circularly polarized (RHC) photon has equal amplitudes to be polarized $|x\rangle$ or $|y\rangle$, and the *amplitudes* are 90° out of phase. Calling a RHC photon a state $|R\rangle$ and a LHC photon a state $|L\rangle$, we can write (see Vol. I, Section 33-1)

$$\begin{aligned} |R\rangle &= \frac{1}{\sqrt{2}} (|x\rangle + i|y\rangle), \\ |L\rangle &= -\frac{1}{\sqrt{2}} (|x\rangle - i|y\rangle). \end{aligned} \tag{11.34}$$

—the $1/\sqrt{2}$ is put in to get normalized states. With these states you can calculate any filtering or interference effects you want, using the laws of quantum theory. If you want, you can also choose $|R\rangle$ and $|L\rangle$ as base states and represent everything in terms of them. You only need to show first that $\langle R|L\rangle = 0$ —which you can do by taking the conjugate form of the first equation above [see Eq. (8.13)] and multiplying it by the other. You can resolve light into x - and y -polarizations, or into x' - and y' -polarizations, or into right and left polarizations as a basis.

Just as an example, let's try to turn our formulas around. Can we represent the state $|x\rangle$ as a linear combination of right and left? Yes, here it is:

$$\begin{aligned} |x\rangle &= \frac{1}{\sqrt{2}} (|R\rangle + |L\rangle), \\ |y\rangle &= -\frac{i}{\sqrt{2}} (|R\rangle - |L\rangle). \end{aligned} \tag{11.35}$$

Proof: Add and subtract the two equations in (11.34). It is easy to go from one base to the other.

One curious point has to be made, though. If a photon is right circularly polarized, it shouldn't have anything to do with the x - and y -axes. If we were to look at the same thing from a coordinate system turned at some angle about the direction of flight, the light would still be right circularly polarized—and similarly for left. The right and left circularly polarized light are the same for any such rotation; the definition is independent of any choice of the x -direction (except that the photon direction is given). Isn't that nice—it doesn't take any axes to define it. Much better than x and y . On the other hand, isn't it rather a miracle that when you *add* the right and left together you can find out which direction x was? If "right" and "left" do not depend on x in any way, how is it that we can put them back together again and get x ? We can answer that question in part by writing out the state $|R'\rangle$, which represents a photon RHC polarized in the frame x', y' . In that frame, you would write

$$|R'\rangle = \frac{1}{\sqrt{2}} (|x'\rangle + i|y'\rangle).$$

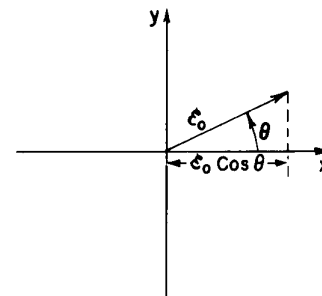


Fig. 11-4. The classical picture of the electric vector \mathcal{E} .

How does such a state look in the frame x, y ? Just substitute x' from Eq.(11. 33) and the corresponding $|y'\rangle$ —we didn't write it down, but it is $(-\sin \theta) |x\rangle + (\cos \theta) |y\rangle$. Then

$$\begin{aligned} |R'\rangle &= \frac{1}{\sqrt{2}} [\cos \theta |x\rangle + \sin \theta |y\rangle - i \sin \theta |x\rangle + i \cos \theta |y\rangle] \\ &= \frac{1}{\sqrt{2}} [(\cos \theta - i \sin \theta) |x\rangle + i(\cos \theta - i \sin \theta) |y\rangle] \\ &= \frac{1}{\sqrt{2}} (|x\rangle + i |y\rangle)(\cos \theta - i \sin \theta). \end{aligned}$$

The first term is just $|R\rangle$, and the second is $e^{+i\theta}$; our result is that

$$|R'\rangle = e^{-i\theta} |R\rangle. \quad (11.36)$$

The states $|R'\rangle$ and $|R\rangle$ are the same except for the phase factor $e^{-i\theta}$. If you work out the same thing for $|L'\rangle$, you get that†

$$|L'\rangle = e^{+i\theta} |L\rangle. \quad (11.37)$$

Now you see what happens. If we add $|R\rangle$ and $|L\rangle$, we get something different from what we get when we add $|R'\rangle$ and $|L'\rangle$. For instance, an x -polarized photon is [Eq. (11.35)] the sum of $|R\rangle$ and $|L\rangle$, but a y -polarized photon is the sum with the phase of one shifted 90° backward and the other 90° forward. That is just what we would get from the sum of $|R'\rangle$ and $|L'\rangle$ for the special angle $\theta = 90^\circ$, and that's right. An x -polarization in the *prime* frame is the same as a y -polarization in the original frame. So it is not exactly true that a circularly polarized photon looks the same for any set of axes. Its *phase* (the phase relation of the right and left circularly polarized states) keeps track of the x -direction.

11-5 The neutral K-meson‡

We will now describe a two-state system in the world of the strange particles—a system for which quantum mechanics gives a most remarkable prediction. To describe it completely would involve us in a lot of stuff about strange particles, so we will, unfortunately, have to cut some corners. We can only give an outline of how a certain discovery was made—to show you the kind of reasoning that was involved. It begins with the discovery by Gell-Mann and Nishijima of the concept of *strangeness* and of a new law of *conservation of strangeness*. It was when Gell-Mann and Pais were analyzing the consequences of these new ideas that they came across the prediction of a most remarkable phenomenon we are going to describe. First, though, we have to tell you a little about “strangeness.”

We must begin with what are called the *strong interactions* of nuclear particles. These are the interactions which are responsible for the strong nuclear forces—as distinct, for instance, from the relatively weaker electromagnetic interactions. The interactions are “strong” in the sense that if two particles get close enough to interact at all, they interact in a big way and produce other particles very easily.

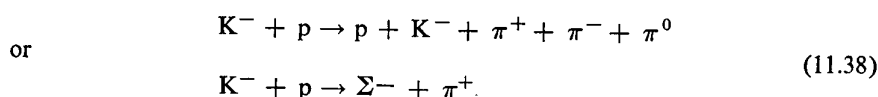
† It's similar to what we found (in Chapter 6) for a spin one-half particle when we rotated the coordinates about the z -axis—then we got the phase factors $e^{\pm i\phi/2}$. It is, in fact, exactly what we wrote down in Section 5-7 for the $|+\rangle$ and $|-\rangle$ states of a spin-one particle—which is no coincidence. The photon is a spin-one particle which has, however, no “zero” state.

‡ We now feel that the material of this section is longer and harder than is appropriate at this point in our development. We suggest that you skip it and continue with Section 11-6. If you are ambitious and have time you may wish to come back to it later. We leave it here, because it is a beautiful example—taken from recent work in high-energy physics—of what can be done with our formulation of the quantum mechanics of two-state systems.

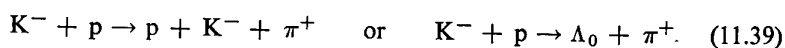
The nuclear particles have also what is called a “weak interaction” by which certain things can happen, such as beta decay, but always very slowly on a nuclear time scale—the weak interactions are many, many orders of magnitude weaker than the strong interactions and even much weaker than electromagnetic interactions.

When the strong interactions were being studied with the big accelerators, people were surprised to find that certain things that “should” happen—that were expected to happen—did not occur. For instance, in some interactions a particle of a certain type did not appear when it was expected. Gell-Mann and Nishijima noticed that many of these peculiar happenings could be explained at once by inventing a new conservation law: the *conservation of strangeness*. They proposed that there was a new kind of attribute associated with each particle—which they called its “strangeness” number—and that in any strong interaction the “quantity of strangeness” is conserved.

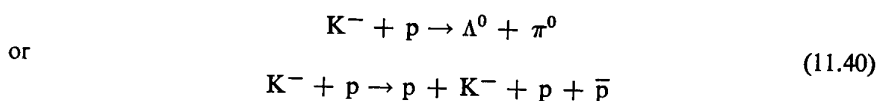
Suppose, for instance, that a high-energy negative K-meson—with, say, an energy of many Bev—collides with a proton. Out of the interaction may come many other particles: π -mesons, K-mesons, lambda particles, sigma particles—any of the mesons or baryons listed in Table 2-2 of Vol. I. It is observed, however, that only *certain combinations* appear, and never others. Now certain conservation laws were already known to apply. First, energy and momentum are always conserved. The total energy and momentum after an event must be the same as before the event. Second, there is the conservation of electric charge which says that the total charge of the outgoing particles must be equal to the total charge carried by the original particles. In our example of a K-meson and a proton coming together, the following reactions *do* occur:



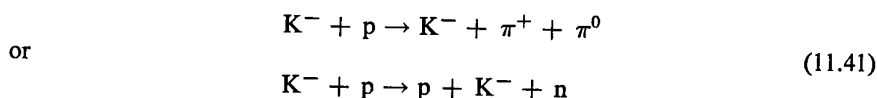
We would never get:



because of the conservation of charge. It was also known that the *number of baryons* is conserved. The number of baryons *out* must be equal to the number of baryons *in*. For this law, an *antiparticle* of a baryon is counted as *minus* one baryon. This means that we can—and *do*—see

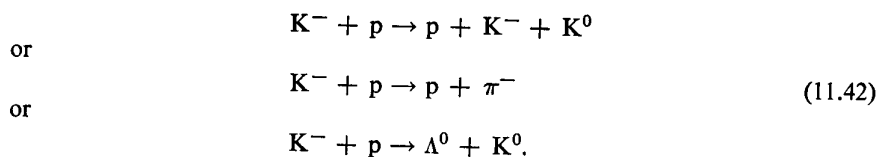


(where $\bar{\text{p}}$ is the antiproton, which carries a negative charge). But we *never* see



(even when there is plenty of energy), because baryons would not be conserved.

These laws, however, do *not* explain the strange fact that the following reactions—which do not immediately appear to be especially different from some of those in (11.38) or (11.40)—are also never observed:



The explanation is the conservation of strangeness. With each particle goes a number—its *strangeness S*—and there is a law that in any *strong* interaction, the

Table 11-4

The strangeness numbers of the strongly interacting particles

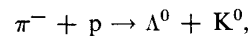
	S			
	-2	-1	0	+1
Baryons		Σ^+	p	
	Ξ^0	Λ^0, Σ^0	n	
	Ξ^-	Σ^-		
Mesons			π^+	K^+
		\bar{K}^0	π^0	K^0
		K^-	π^-	

Note: The π^- is the antiparticle of the π^+ (or *vice versa*).

total strangeness *out* must equal the total strangeness that went *in*. The proton and antiproton (p, \bar{p}), the neutron and antineutron (n, \bar{n}), and the π -mesons (π^+, π^0, π^-) all have the strangeness number *zero*; the K^+ and K^0 mesons have strangeness $+1$; the K^- and \bar{K}^0 (the anti- K^0),† the Λ^0 and the Σ -particles ($+, 0, -$) have strangeness -1 . There is also a particle with strangeness -2 —the Ξ -particle (capital “ksi”)—and perhaps others as yet unknown. We have made a list of these strangenesses in Table 11-4.

Let’s see how the strangeness conservation works in some of the reactions we have written down. If we start with a K^- and a proton, we have a total strangeness of $(-1 + 0) = -1$. The conservation of strangeness says that the strangeness of products *after* the reaction must also add up to -1 . You see that that is so for the reactions of (11.38) and (11.40). But in the reactions of (11.42) the strangeness of the right-hand side is *zero* in each case. Such reactions do not conserve strangeness, and do not occur. Why? Nobody knows. Nobody knows any more than what we have just told you about this. Nature just works that way.

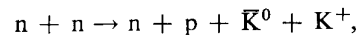
Now let’s look at the following reaction: a π^- hits a proton. You might, for instance, get a Λ^0 particle plus a neutral K-particle—two neutral particles. Now which neutral K do you get? Since the Λ -particle has a strangeness -1 and the π and p^+ have a strangeness zero, and since this is a fast production reaction, the strangeness must not change. The K-particle must have strangeness $+1$ —it must therefore be the K^0 . The reaction is



with

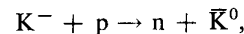
$$S = 0 + 0 = -1 + +1 \quad (\text{conserved}).$$

If the \bar{K}^0 were there instead of the K^0 , the strangeness on the right would be -2 —which nature does not permit, since the strangeness on the left side is zero. On the other hand, a \bar{K}^0 can be produced in other reactions, such as



$$S = 0 + 0 = 0 + 0 + +1 + -1$$

or



$$S = -1 + 0 = 0 + -1.$$

You may be thinking, “That’s all a lot of stuff, because how do you *know* whether it is a \bar{K}^0 or a K^0 ? They look exactly the same. They are antiparticles of each other, so they have exactly the same mass, and both have zero electric charge.

† Read as: “K-naught-bar,” or “K-zero-bar.”

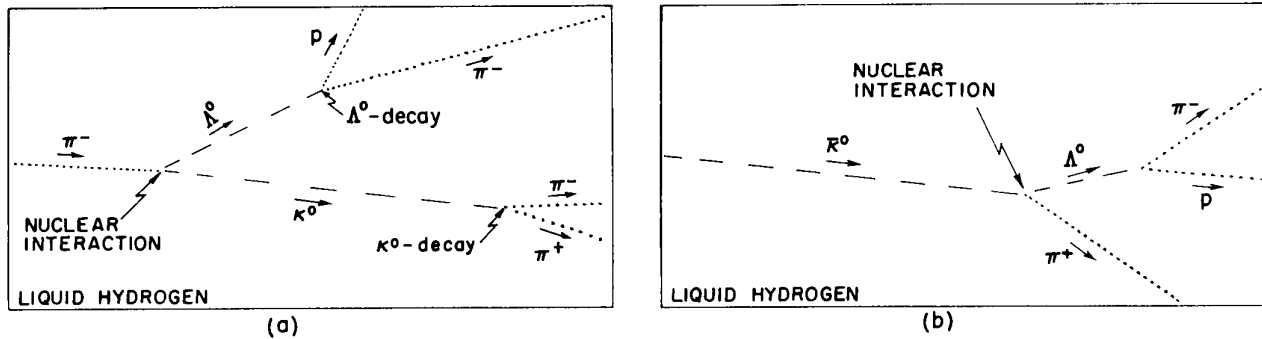
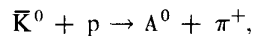


Fig. 11-5. High-energy events as seen in a hydrogen bubble chamber. (a) A π^- meson interacts with a hydrogen nucleus (proton) producing a Λ^0 particle and a K^0 meson. Both particles decay in the chamber. (b) A \bar{K}^0 meson interacts with a proton producing a π^+ meson and a Λ^0 particle which then decays. (The neutral particles leave no tracks. Their inferred trajectories are indicated here by light dashed lines.)

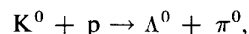
How do you distinguish them?" By the reactions *they* produce. For example, a \bar{K}^0 can interact with matter to produce a Λ -particle, like this:



but a K^0 cannot. There is *no* way a K^0 can produce a Λ -particle when it interacts with ordinary matter (protons and neutrons).† So the experimental distinction between the K^0 and the \bar{K}^0 would be that one of them will and one of them will not produce Λ 's.

One of the predictions of the strangeness theory is then this—if, in an experiment with high-energy pions, a Λ -particle is produced with a neutral K-meson, then *that* neutral K-meson going into other pieces of matter will never produce a Λ . The experiment might run something like this. You send a beam of π^- -mesons into a large hydrogen bubble chamber. A π^- track disappears, but somewhere else a pair of tracks appear (a proton and a π^-) indicating that a Λ -particle has disintegrated‡—see Fig. 11-5. Then you know that there is a K^0 somewhere which you cannot see.

You can, however, figure out where it is going by using the conservation of momentum and energy. [It could reveal itself later by disintegrating into two charged particles, as shown in Fig. 11-5(a).] As the K^0 goes flying along, it may interact with one of the hydrogen nuclei (protons), producing perhaps some other particles. The prediction of the strangeness theory is that it will *never* produce a Λ -particle in a simple reaction like, say,



although a \bar{K}^0 can do just that. That is, in a bubble chamber a \bar{K}^0 might produce the event sketched in Fig. 11-5(b)—in which the Λ^0 is seen because it decays—but a K^0 will not. That's the first part of our story. That's the conservation of strangeness.

The conservation of strangeness is, however, *not perfect*. There are very slow disintegrations of the strange particles—decays taking a long¶ time like 10^{-10} second in which the strangeness is *not* conserved. These are called the "weak" decays. For example, the K^0 disintegrates into a pair of π -mesons (+ and -)

† Except, of course, if it *also* produces *two* K^+ 's or other particles with a total strangeness of +2. We can think here of reactions in which there is insufficient energy to produce these additional strange particles.

‡ The free Λ -particle decays slowly via a *weak* interaction (so strangeness need not be conserved). The decay products are either a p and a π^- , or an n and a π^0 . The lifetime is 2.2×10^{-10} sec.

¶ A typical time for strong interactions is more like 10^{-23} sec.

with a lifetime of 10^{-10} second. That was, in fact, the way K-particles were first seen. Notice that the decay reaction

$$K^0 \rightarrow \pi^+ + \pi^-$$

does not conserve strangeness, so it cannot go “fast” by the strong interaction; it can only go through the weak decay process.

Now the \bar{K}^0 also disintegrates *in the same way*—into a π^+ and a π^- — and also with the same lifetime

$$\bar{K}^0 \rightarrow \pi^- + \pi^+.$$

Again we have a weak decay because it does not conserve strangeness. There is a principle that for any reaction there is the corresponding reaction with “matter” replaced by “antimatter” and *vice versa*. Since the \bar{K}^0 is the antiparticle of the K^0 , it should decay into the antiparticles of the π^+ and π^- , but the antiparticle of a π^+ is the π^- . (Or, if you prefer, *vice versa*. It turns out that for the π -mesons it doesn’t matter which one you call “matter.”) So as a consequence of the weak decays, the K^0 and \bar{K}^0 can go into the same final products. When “seen” through their decays—as in a bubble chamber—they look like the same particle. Only their strong interactions are different.

At last we are ready to describe the work of Gell-Mann and Pais. They first noticed that since the K^0 and the \bar{K}^0 can both turn into states of two π -mesons there must be some amplitude that a K^0 can turn into a \bar{K}^0 , and also that a \bar{K}^0 can turn into a K^0 . Writing the reactions as one does in chemistry, we would have

$$K^0 \rightleftharpoons \pi^- + \pi^+ \rightleftharpoons \bar{K}^0. \quad (11.43)$$

These reactions imply that there is some amplitude per unit time, say $-i/\hbar$ times $\langle \bar{K}^0 | W | K^0 \rangle$, that a K^0 will turn into a \bar{K}^0 through the weak interaction responsible for the decay into two π -mesons. And there is the corresponding amplitude $\langle K^0 | W | \bar{K}^0 \rangle$ for the reverse process. Because matter and antimatter behave in exactly the same way, these two amplitudes are numerically equal; we’ll call them both A :

$$\langle \bar{K}^0 | W | K^0 \rangle = \langle K^0 | W | \bar{K}^0 \rangle = A. \quad (11.44)$$

Now—said Gell-Mann and Pais—here is an interesting situation. What people have been calling two distinct states of the world—the K^0 and the \bar{K}^0 —should really be considered as *one* two-state system, because there is an amplitude to go from one state to the other. For a complete treatment, one would, of course, have to deal with more than two states, because there are also the states of 2π ’s, and so on; but since they were mainly interested in the relation of K^0 and \bar{K}^0 , they did not have to complicate things and could make the approximation of a two-state system. The other states *were* taken into account to the extent that their effects appeared implicitly in the amplitudes of Eq. (11.44).

Accordingly, Gell-Mann and Pais analyzed the neutral particle as a two-state system. They began by choosing as their two base states the states $|K^0\rangle$ and $|\bar{K}^0\rangle$. (From here on, the story goes very much as it did for the ammonia molecule.) Any state $|\psi\rangle$ of the neutral K-particle could then be described by giving the amplitudes that it was in either base state. We’ll call these amplitudes

$$C_+ = \langle K^0 | \psi \rangle, \quad C_- = \langle \bar{K}^0 | \psi \rangle. \quad (11.45)$$

The next step was to write the Hamiltonian equations for this two-state system. If there were no coupling between the K^0 and the \bar{K}^0 , the equations would be simply

$$\begin{aligned} i\hbar \frac{dC_+}{dt} &= E_0 C_+, \\ i\hbar \frac{dC_-}{dt} &= E_0 C_-. \end{aligned} \quad (11.46)$$

But since there is the amplitude $\langle \bar{K}^0 | W | K^0 \rangle$ for the K^0 to turn into a \bar{K}^0 there should be the additional term

$$\langle \bar{K}^0 | W | K^0 \rangle C_- = AC_-$$

added to the right-hand side of the first equation. And similarly, the term AC_+ should be inserted in the equation for the rate of change of C_- .

But that's not all. When the two-pion effect is taken into account there is an *additional* amplitude for the K^0 to turn into *itself* through the process

$$K^0 \rightarrow \pi^- + \pi^+ \rightarrow K^0.$$

The additional amplitude, which we would write $\langle K^0 | W | K^0 \rangle$, is just equal to the amplitude $\langle \bar{K}^0 | W | K^0 \rangle$, since the amplitudes to go to and from a pair of π -mesons are identical for the K^0 and the \bar{K}^0 . If you wish, the argument can be written out in detail like this. First write†

$$\langle \bar{K}^0 | W | K^0 \rangle = \langle \bar{K}^0 | W | 2\pi \rangle \langle 2\pi | W | K^0 \rangle$$

and

$$\langle K^0 | W | K^0 \rangle = \langle K^0 | W | 2\pi \rangle \langle 2\pi | W | K^0 \rangle.$$

Because of the symmetry of matter and antimatter

$$\langle 2\pi | W | K^0 \rangle = \langle 2\pi | W | \bar{K}^0 \rangle,$$

and also

$$\langle K^0 | W | 2\pi \rangle = \langle \bar{K}^0 | W | 2\pi \rangle.$$

It then follows that $\langle K^0 | W | K^0 \rangle = \langle \bar{K}^0 | W | K^0 \rangle$, and also that $\langle \bar{K}^0 | W | K^0 \rangle = \langle K^0 | W | \bar{K}^0 \rangle$, as we said earlier. Anyway, there are the two additional amplitudes $\langle K^0 | W | K^0 \rangle$ and $\langle \bar{K}^0 | W | \bar{K}^0 \rangle$, both equal to A , which should be included in the Hamiltonian equations. The first gives a term AC_+ on the right-hand side of the equation for dC_+/dt , and the second gives a new term AC_- in the equation for dC_-/dt . Reasoning this way, Gell-Mann and Pais concluded that the Hamiltonian equations for the $K^0 \bar{K}^0$ system should be

$$\begin{aligned} i\hbar \frac{dC_+}{dt} &= E_0 C_+ + AC_- + AC_+, \\ i\hbar \frac{dC_-}{dt} &= E_0 C_- + AC_+ + AC_-. \end{aligned} \tag{11.47}$$

We must now correct something we have said in earlier chapters: that two amplitudes like $\langle K^0 | W | \bar{K}^0 \rangle$ and $\langle \bar{K}^0 | W | K^0 \rangle$ which are the reverse of each other, are always complex conjugates. That was true when we were talking about particles that did not decay. But if particles can decay—and can, therefore, become “lost”—the two amplitudes are not necessarily complex conjugates. So the equality of (11.44) does not mean that the amplitudes are real numbers; they are in fact complex numbers. The coefficient A is, therefore, complex; and we can't just incorporate it into the energy E_0 .

Having played often with electron spins and such, our heroes knew that the Hamiltonian equations of (11.47) meant that there was *another* pair of base states which could also be used to represent the K -particle system and which would have especially simple behaviors. They said, “Let's take the sum and difference of these two equations. Also, let's measure all our energies from E_0 , and use units for

† We are making a simplification here. The 2π -system can have many states corresponding to various momenta of the π -mesons, and we should make the right-hand side of this equation into a sum over the various base states of the π 's. The complete treatment still leads to the same conclusions.

energy and time that make $\hbar = 1$." (That's what modern theoretical physicists always do. It doesn't change the physics but makes the equations take on a simple form.) Their result:

$$i \frac{d}{dt} (C_+ + C_-) = 2A(C_+ + C_-), \quad i \frac{d}{dt} (C_+ - C_-) = 0. \quad (11.48)$$

It is apparent that the combinations of amplitudes $(C_+ + C_-)$ and $(C_+ - C_-)$ act independently from each other (corresponding, of course, to the stationary states we have been studying earlier). So they concluded that it would be more convenient to use a different representation for the K-particle. They defined the two states

$$|K_1\rangle = \frac{1}{\sqrt{2}} (|K^0\rangle + |\bar{K}^0\rangle), \quad |K_2\rangle = \frac{1}{\sqrt{2}} (|K^0\rangle - |\bar{K}^0\rangle). \quad (11.49)$$

They said that instead of thinking of the K^0 and \bar{K}^0 mesons, we can equally well think in terms of the two "particles" (that is, "states") K_1 and K_2 . (These correspond, of course, to the states we have usually called $|I\rangle$ and $|II\rangle$). We are not using our old notation because we want now to follow the notation of the original authors—and the one you will see in physics seminars.)

Now Gell-Mann and Pais didn't do all this just to get different names for the particles—there is also some strange new physics in it. Suppose that C_1 and C_2 are the amplitudes that some state $|\psi\rangle$ will be either a K_1 or a K_2 meson:

$$C_1 = \langle K_1 | \psi \rangle, \quad C_2 = \langle K_2 | \psi \rangle.$$

From the equations of (11.49),

$$C_1 = \frac{1}{\sqrt{2}} (C_+ + C_-), \quad C_2 = \frac{1}{\sqrt{2}} (C_+ - C_-). \quad (11.50)$$

Then the Eqs. (11.48) become

$$i \frac{dC_1}{dt} = 2AC_1, \quad i \frac{dC_2}{dt} = 0. \quad (11.51)$$

The solutions are

$$C_1(t) = C_1(0)e^{-i2At}, \quad C_2(t) = C_2(0), \quad (11.52)$$

where, of course, $C_1(0)$ and $C_2(0)$ are the amplitudes at $t = 0$.

These equations say that if a neutral K-particle starts out in the state $|K_1\rangle$ at $t = 0$ [then $C_1(0) = 1$ and $C_2(0) = 0$], the amplitudes at the time t are

$$C_1(t) = e^{-i2At}, \quad C_2(t) = 0.$$

Remembering that A is a complex number, it is convenient to take $A = \alpha - i\beta$. (Since the imaginary part of $2A$ turns out to be negative, we write it as *minus* $i\beta$.) With this substitution, $C_1(t)$ reads

$$C_1(t) = C_1(0)e^{-\beta t}e^{-i\alpha t}. \quad (11.53)$$

The probability of finding a K_1 particle at t is the absolute square of this amplitude, which is $e^{-2\beta t}$. And, from Eqs. (11.52), the probability of finding the K_2 state at any time is zero. That means that if you make a K-particle in the state $|K_1\rangle$, the probability of finding it in the same state decreases exponentially with time—but you will never find it in state $|K_2\rangle$. Where does it go? It disintegrates into two π -mesons with the mean life $\tau = 1/2\beta$ which is, experimentally, 10^{-10} sec. We made provisions for that when we said that A was complex.

On the other hand, Eq. (11.52) says that if we make a K-particle completely in the K_2 state, it stays that way forever. Well, that's not really true. It is observed experimentally to disintegrate into *three* π -mesons, but 600 times slower than the

two-pion decay we have described. So there are some other small terms we have left out in our approximation. But so long as we are considering only the two-pion decay, the K_2 lasts “forever.”

Now to finish the story of Gell-Mann and Pais. They went on to consider what happens when a K-particle is produced *with a Λ^0 particle* in a *strong* interaction. Since it must then have a strangeness of +1, it must be produced in the K^0 state. So at $t = 0$ it is neither a K_1 nor a K_2 but a *mixture*. The initial conditions are

$$C_+(0) = 1, \quad C_-(0) = 0.$$

But that means—from Eq. (11.50)—that

$$C_1(0) = \frac{1}{\sqrt{2}}, \quad C_2(0) = \frac{1}{\sqrt{2}},$$

and—from Eq. (11.51)—that

$$C_1(t) = \frac{1}{\sqrt{2}} e^{-\beta t} e^{-i\alpha t}, \quad C_2(t) = \frac{1}{\sqrt{2}}. \quad (11.54)$$

Now remember that K_1 and K_2 are each linear combinations of K^0 and \bar{K}^0 . In Eqs. (11.54) the amplitudes have been chosen so that at $t = 0$ the \bar{K}^0 parts cancel each other out by interference, leaving only a K^0 state. But the $|K_1\rangle$ state *changes with time*, and the $|K_2\rangle$ state *does not*. After $t = 0$ the interference of C_1 and C_2 will give finite amplitudes for both K^0 and \bar{K}^0 .

What does all this mean? Let’s go back and think of the experiment we sketched in Fig. 11-5. A π^- meson has produced a Λ^0 particle and a K^0 meson which is tooting along through the hydrogen in the chamber. As it goes along, there is some small but uniform chance that it will collide with a hydrogen nucleus. At first, we thought that strangeness conservation would prevent the K-particle from making a Λ^0 in such an interaction. Now, however, we see that that is not right. For although our K-particle *starts out* as a K^0 —which cannot make a Λ^0 —it does not *stay* this way. After a while, there *is some amplitude* that it will have flipped to the \bar{K}^0 state. We can, therefore, sometimes expect to see a Λ^0 produced along the K-particle track. The chance of this happening is given by the amplitude C_- , which we can [by using Eq. (11.50) backwards] relate to C_1 and C_2 . The relation is

$$C_- = \frac{1}{\sqrt{2}} (C_1 - C_2) = \frac{1}{2}(e^{-\beta t} e^{-i\alpha t} - 1). \quad (11.55)$$

As our K-particle goes along, the probability that it will “act like” a \bar{K}^0 is equal to $|C_-|^2$, which is

$$|C_-|^2 = \frac{1}{4}(1 + e^{-2\beta t} - 2e^{-\beta t} \cos \alpha t). \quad (11.56)$$

A complicated and strange result!

This, then, is the remarkable prediction of Gell-Mann and Pais: when a K^0 is produced, the chance that it will turn into a \bar{K}^0 —as it can demonstrate by being able to produce a Λ^0 —varies with time according to Eq. (11.56). This prediction came from using only sheer logic and the basic principles of the quantum mechanics—with no knowledge at all of the inner workings of the K-particle. Since nobody knows anything about the inner machinery, that is as far as Gell-Mann and Pais could go. They could not give any theoretical values for α and β . And nobody has been able to do so to this date. They were able to give a value of β obtained from the experimentally observed rate of decay into two π ’s ($2\beta = 10^{10}$ sec), but they could say nothing about α .

We have plotted the function of Eq. (11.56) for two values of α in Fig. 11-6. You can see that the form depends very much on the ratio of α to β . There is no \bar{K}^0 probability at first; then it builds up. If α is large, the probability would have

large oscillations. If α is small, there will be little or no oscillation—the probability will just rise smoothly to $1/4$.

Now, typically, the K-particle will be travelling at a constant speed near the speed of light. The curves of Fig. 11-6 then also represent the probability along the track of observing a \bar{K}^0 —with typical distances of several centimeters. You can see why this prediction is so remarkably peculiar. You produce a single particle and instead of just disintegrating, it does something else. Sometimes it disintegrates, and other times it turns into a different kind of a particle. Its characteristic probability of producing an effect varies in a strange way as it goes along. There is nothing else quite like it in nature. And this most remarkable prediction was made solely by arguments about the interference of amplitudes.

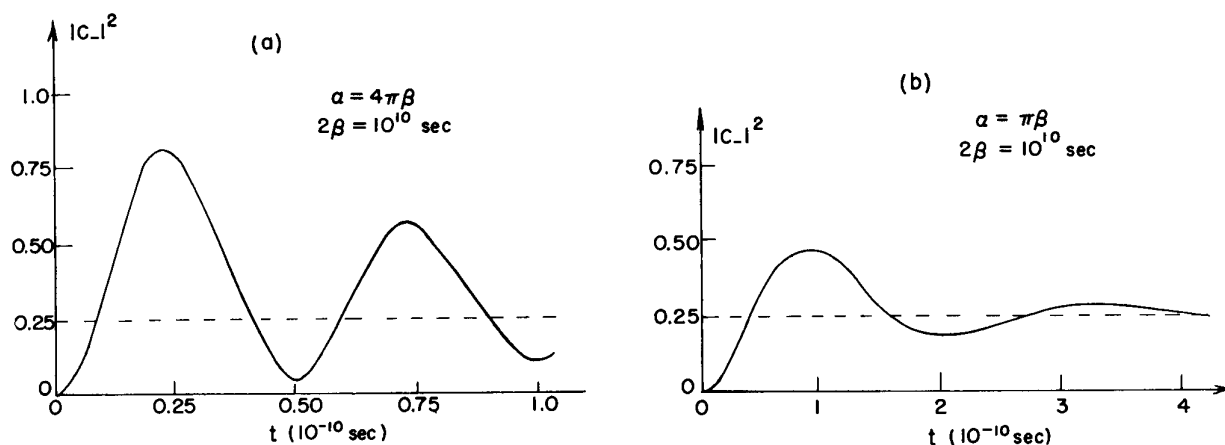


Fig. 11-6. The function of Eq. (11-56): (a) for $\alpha = \pi\beta$, (b) for $\alpha = 4\pi\beta$ (with $2\beta = 10^{10} \text{ sec}$).

If there is any place where we have a chance to test the main principles of quantum mechanics in the purest way—does the superposition of amplitudes work or doesn't it?—this is it. In spite of the fact that this effect has been predicted now for several years, there is no experimental determination that is very clear. There are some rough results which indicate that the α is not zero, and that the effect really occurs—they indicate that α is between 2β and 4β . That's all there is, experimentally. It would be very beautiful to check out the curve exactly to see if the principle of superposition really still works in such a mysterious world as that of the strange particles—with unknown reasons for the decays, and unknown reasons for the strangeness.

The analysis we have just described is very characteristic of the way quantum mechanics is being used today in the search for an understanding of the strange particles. All the complicated theories that you may hear about are no more and no less than this kind of elementary hocus-pocus using the principles of superposition and other principles of quantum mechanics of that level. Some people claim that they have theories by which it is possible to calculate the β and α , or at least the α given the β , but these theories are completely useless. For instance, the theory that predicts the value of α , given the β , tells us that the value of α should be infinite. The set of equations with which they originally start involves two π -mesons and then goes from the two π 's back to a K^0 , and so on. When it's all worked out, it does indeed produce a pair of equations like the ones we have here; but because there are an infinite number of states of two π 's, depending on their momenta, integrating over all the possibilities gives an α which is infinite. But nature's α is *not* infinite. So the dynamical theories are wrong. It is really quite remarkable that the phenomena which can be predicted *at all* in the world of the strange particles come from the principles of quantum mechanics at the level at which you are learning them now.

11-6 Generalization to N -state systems

We have finished with all the two-state systems we wanted to talk about. In the following chapters we will go on to study systems with more states. The extension to N -state systems of the ideas we have worked out for two states is pretty straightforward. It goes like this.

If a system has N distinct states, we can represent any state $|\psi(t)\rangle$ as a linear combination of any set of base states $|i\rangle$, where $i = 1, 2, 3, \dots, N$;

$$|\psi(t)\rangle = \sum_{\text{all } i} |i\rangle C_i(t). \quad (11.57)$$

The coefficients $C_i(t)$ are the amplitudes $\langle i | \psi(t) \rangle$. The behavior of the amplitudes C_i with time is governed by the equations

$$i\hbar \frac{dC_i(t)}{dt} = \sum_j H_{ij} C_j, \quad (11.58)$$

where the energy matrix H_{ij} describes the physics of the problem. It looks the same as for two states. Only now, both i and j must range over all N base states, and the energy matrix H_{ij} —or, if you prefer, the Hamiltonian—is an N by N matrix with N^2 numbers. As before, $H_{ij}^* = H_{ji}$ —so long as particles are conserved—and the diagonal elements H_{ii} are real numbers.

We have found a general solution for the C 's of a two-state system when the energy matrix is constant (doesn't depend on t). It is also not difficult to solve Eq. (11.58) for an N -state system when H is not time dependent. Again, we begin by looking for a possible solution in which the amplitudes all have the *same* time dependence. We try

$$C_i = a_i e^{-(i/\hbar)Et}. \quad (11.59)$$

When these C_i 's are substituted into (11.58), the derivatives $dC_i(t)/dt$ become just $(-i/\hbar)EC_i$. Canceling the common exponential factor from all terms, we get

$$Ea_i = \sum_j H_{ij} a_j. \quad (11.60)$$

This is a set of N linear algebraic equations for the N unknowns a_1, a_2, \dots, a_n , and there is a solution only if you are lucky—only if the determinant of the coefficients of all the a 's is zero. But it's not necessary to be that sophisticated; you can just start to solve the equations any way you want, and you will find that they can be solved only for certain values of E . (Remember that E is the only adjustable thing we have in the equations.)

If you want to be formal, however, you can write Eq. (11.60) as

$$\sum_j (H_{ij} - \delta_{ij}E) a_j = 0. \quad (11.61)$$

Then you can use the rule—if you know it—that these equations will have a solution only for those values of E for which

$$\text{Det}(H_{ij} - \delta_{ij}E) = 0. \quad (11.62)$$

Each term of the determinant is just H_{ij} , except that E is subtracted from every diagonal element. That is, (11.62) means just

$$\text{Det} \begin{pmatrix} H_{11} - E & H_{12} & H_{13} & \dots \\ H_{21} & H_{22} - E & H_{23} & \dots \\ H_{31} & E_{32} & H_{33} - E & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix} = 0. \quad (11.63)$$

This is, of course, just a special way of writing an algebraic equation for E which is the sum of a bunch of products of all the terms taken a certain way. These products will give all the powers of E up to E^N .

So we have an N th order polynomial equal to zero, and there are, in general, N roots. (We must remember, however, that some of them may be multiple roots—meaning that two or more roots are equal.) Let's call the N roots

$$E_I, E_{II}, E_{III}, \dots, E_n, \dots, E_N. \quad (11.64)$$

(We will use \mathbf{n} to represent the n th Roman numeral, so that \mathbf{n} takes on the values I, II, . . . , N.) It may be that some of these energies are equal—say $E_{II} = E_{III}$ —but we will still choose to call them by different names.

The equations (11.60)—or (11.61)—have one solution for each value of E . If you put any one of the E 's—say E_n —into (11.60) and solve for the a_i , you get a set which belongs to the energy E_n . We will call this set $a_i(\mathbf{n})$.

Using these $a_i(\mathbf{n})$ in Eq. (11.59), we have the amplitudes $C_i(\mathbf{n})$ that the definite energy states are in the base state $|i\rangle$. Letting $|\mathbf{n}\rangle$ stand for the state vector of the definite energy state at $t = 0$, we can write

$$C_i(\mathbf{n}) = \langle i | \mathbf{n} \rangle e^{(i/\hbar)E_n t},$$

with

$$\langle i | \mathbf{n} \rangle = a_i(\mathbf{n}). \quad (11.65)$$

The complete definite energy state $|\psi_n(t)\rangle$ can then be written as

$$|\psi_n(t)\rangle = \sum_i |i\rangle a_i(\mathbf{n}) e^{-(i/\hbar)E_n t},$$

or

$$|\psi_n(t)\rangle = |\mathbf{n}\rangle e^{(i/\hbar)E_n t}. \quad (11.66)$$

The state vectors $|\mathbf{n}\rangle$ describe the configuration of the definite energy states, but have the time dependence factored out. Then they are constant vectors which can be used as a new base set if we wish.

Each of the states $|\mathbf{n}\rangle$ has the property—as you can easily show—that when operated on by the Hamiltonian operator \hat{H} it gives just E_n times the same state:

$$\hat{H} |\mathbf{n}\rangle = E_n |\mathbf{n}\rangle. \quad (11.67)$$

The energy E_n is, then, a number which is a characteristic of the Hamiltonian operator \hat{H} . As we have seen, a Hamiltonian will, in general, have several characteristic energies. In the mathematician's world these would be called the "characteristic values" of the matrix H_{ij} . Physicists usually call them the "eigenvalues" of \hat{H} . ("Eigen" is the German word for "characteristic" or "proper.") With each eigenvalue of \hat{H} —in other words, for each energy—there is the state of definite energy, which we have called the "stationary state." Physicists usually call the states $|\mathbf{n}\rangle$ "the eigenstates of \hat{H} ." Each eigenstate corresponds to a particular eigenvalue E_n .

Now, generally, the states $|\mathbf{n}\rangle$ —of which there are N —can also be used as a base set. For this to be true, all of the states must be orthogonal, meaning that for any two of them, say $|\mathbf{n}\rangle$ and $|\mathbf{m}\rangle$,

$$\langle \mathbf{n} | \mathbf{m} \rangle = 0. \quad (11.68)$$

This will be true automatically if all the energies are different. Also, we can multiply all the $a_i(\mathbf{n})$ by a suitable factor so that all the states are normalized—by which we mean that

$$\langle \mathbf{n} | \mathbf{n} \rangle = 1 \quad (11.69)$$

for all \mathbf{n} .

When it happens that Eq. (11.63) accidentally has two (or more) roots with the same energy, there are some minor complications. First, there are still two different sets of a_i 's which go with the two equal energies, but the states they give

may *not* be orthogonal. Suppose you go through the normal procedure and find two stationary states with equal energies—let’s call them $|\mu\rangle$ and $|\nu\rangle$. Then it will not necessarily be so that they are orthogonal—if you are unlucky,

$$\langle\mu|\nu\rangle \neq 0.$$

It is, however, always true that you can cook up two new states, which we will call $|\mu'\rangle$ and $|\nu'\rangle$, that have the same energies and are also orthogonal, so that

$$\langle\mu'|\nu'\rangle = 0. \quad (11.70)$$

You can do this by making $|\mu'\rangle$ and $|\nu'\rangle$ a suitable linear combination of $|\mu\rangle$ and $|\nu\rangle$, with the coefficients chosen to make it come out so that Eq. (11.70) is true. It is always convenient to do this. We will generally assume that this has been done so that we can always assume that our proper energy states $|\mathbf{n}\rangle$ are all orthogonal.

We would like, for fun, to prove that when two of the stationary states have different energies they are indeed orthogonal. For the state $|\mathbf{n}\rangle$ with the energy $E_{\mathbf{n}}$, we have that

$$\hat{H}|\mathbf{n}\rangle = E_{\mathbf{n}}|\mathbf{n}\rangle. \quad (11.71)$$

This operator equation really means that there is an equation between numbers. Filling the missing parts, it means the same as

$$\sum_j \langle i|\hat{H}|j\rangle\langle j|\mathbf{n}\rangle = E_{\mathbf{n}}\langle i|\mathbf{n}\rangle. \quad (11.72)$$

If we take the complex conjugate of this equation, we get

$$\sum_j \langle i|\hat{H}|j\rangle^*\langle j|\mathbf{n}\rangle^* = E_{\mathbf{n}}^*\langle i|\mathbf{n}\rangle^*. \quad (11.73)$$

Remember now that the complex conjugate of an amplitude is the reverse amplitude, so (11.73) can be rewritten as

$$\sum_j \langle \mathbf{n}|j\rangle\langle j|\hat{H}|i\rangle = E_{\mathbf{n}}^*\langle \mathbf{n}|i\rangle. \quad (11.74)$$

Since this equation is valid for *any* i , its “short form” is

$$\langle \mathbf{n}|\hat{H} = E_{\mathbf{n}}^*\langle \mathbf{n}|, \quad (11.75)$$

which is called the *adjoint* to Eq. (11.71).

Now we can easily prove that $E_{\mathbf{n}}$ is a real number. We multiply Eq. (11.71) by $\langle \mathbf{n}|$ to get

$$\langle \mathbf{n}|\hat{H}|\mathbf{n}\rangle = E_{\mathbf{n}}, \quad (11.76)$$

since $\langle \mathbf{n}|\mathbf{n}\rangle = 1$. Then we multiply Eq. (11.75) on the left by $|\mathbf{n}\rangle$ to get

$$\langle \mathbf{n}|\hat{H}|\mathbf{n}\rangle = E_{\mathbf{n}}^*. \quad (11.77)$$

Comparing (11.76) with (11.77) it is clear that

$$E_{\mathbf{n}} = E_{\mathbf{n}}^*, \quad (11.78)$$

which means that $E_{\mathbf{n}}$ is real. We can erase the star on $E_{\mathbf{n}}$ in Eq. (11.75).

Finally we are ready to show that the different energy states are orthogonal. Let $|\mathbf{n}\rangle$ and $|\mathbf{m}\rangle$ be any two of the definite energy base states. Using Eq. (11.75) for the state \mathbf{m} , and multiplying it by $|\mathbf{n}\rangle$, we get that

$$\langle \mathbf{m}|\hat{H}|\mathbf{n}\rangle = E_{\mathbf{m}}\langle \mathbf{m}|\mathbf{n}\rangle.$$

But if we multiply (11.71) by $\langle \mathbf{m} |$, we get

$$\langle \mathbf{m} | \hat{H} | \mathbf{n} \rangle = E_n \langle \mathbf{m} | \mathbf{n} \rangle.$$

Since the left sides of these two equations are equal, the right sides are, also:

$$E_m \langle \mathbf{m} | \mathbf{n} \rangle = E_n \langle \mathbf{m} | \mathbf{n} \rangle. \quad (11.79)$$

If $E_m = E_n$ the equation does not tell us anything. But if the energies of the two states $|\mathbf{m}\rangle$ and $|\mathbf{n}\rangle$ are different ($E_m \neq E_n$), Eq. (11.79) says that $\langle \mathbf{m} | \mathbf{n} \rangle$ must be zero, as we wanted to prove. The two states are necessarily orthogonal so long as E_n and E_m are numerically different.