1

Fundamental Concepts

The revolutionary change in our understanding of microscopic phenomena that took place during the first 27 years of the twentieth century is unprecedented in the history of natural sciences. Not only did we witness severe limitations in the validity of classical physics, but we found the alternative theory that replaced the classical physical theories to be far richer in scope and far richer in its range of applicability.

The most traditional way to begin a study of quantum mechanics is to follow the historical developments – Planck's radiation law, the Einstein–Debye theory of specific heats, the Bohr atom, de Broglie's matter waves, and so forth – together with careful analyses of some key experiments such as the Compton effect, the Franck–Hertz experiment, and the Davisson–Germer–Thompson experiment. In that way we may come to appreciate how the physicists in the first quarter of the twentieth century were forced to abandon, little by little, the cherished concepts of classical physics and how, despite earlier false starts and wrong turns, the great masters – Heisenberg, Schrödinger, and Dirac, among others – finally succeeded in formulating quantum mechanics as we know it today.

However, we do not follow the historical approach in this book. Instead, we start with an example that illustrates, perhaps more than any other example, the inadequacy of classical concepts in a fundamental way. We hope that by exposing the reader to a "shock treatment" at the onset, he or she may be attuned to what we might call the "quantum-mechanical way of thinking" at a very early stage.

This different approach is not merely an academic exercise. Our knowledge of the physical world comes from making assumptions about nature, formulating these assumptions into postulates, deriving predictions from those postulates, and testing those predictions against experiment. If experiment does not agree with the prediction, then, presumably, the original assumptions were incorrect. Our approach emphasizes the fundamental assumptions we make about nature, upon which we have come to base all of our physical laws, and which aim to accommodate profoundly quantum-mechanical observations at the outset.

1.1 The Stern–Gerlach Experiment

The example we concentrate on in this section is the Stern–Gerlach experiment, originally conceived by O. Stern in 1921 and carried out in Frankfurt by him in collaboration with W. Gerlach in 1922.[1] This experiment illustrates in a dramatic manner the necessity for a radical departure from the concepts of classical mechanics. In the subsequent sections the basic formalism of quantum mechanics is presented in a somewhat axiomatic manner but always with the example of the Stern–Gerlach experiment in the back of our minds. In a certain sense, a two-state system of the Stern–Gerlach type is the least classical, most quantum-mechanical system. A solid understanding of problems involving two-state systems will turn out to be rewarding to any serious student of quantum mechanics. It is for this reason that we refer repeatedly to two-state problems throughout this book.

1.1.1 Description of the Experiment

We now present a brief discussion of the Stern–Gerlach experiment, which is discussed in almost any book on modern physics.[2] First, silver (Ag) atoms are heated in an oven. The oven has a small hole through which some of the silver atoms escape. As shown in Figure 1.1, the beam goes through a collimator and is then subjected to an inhomogeneous magnetic field produced by a pair of pole pieces, one of which has a very sharp edge.

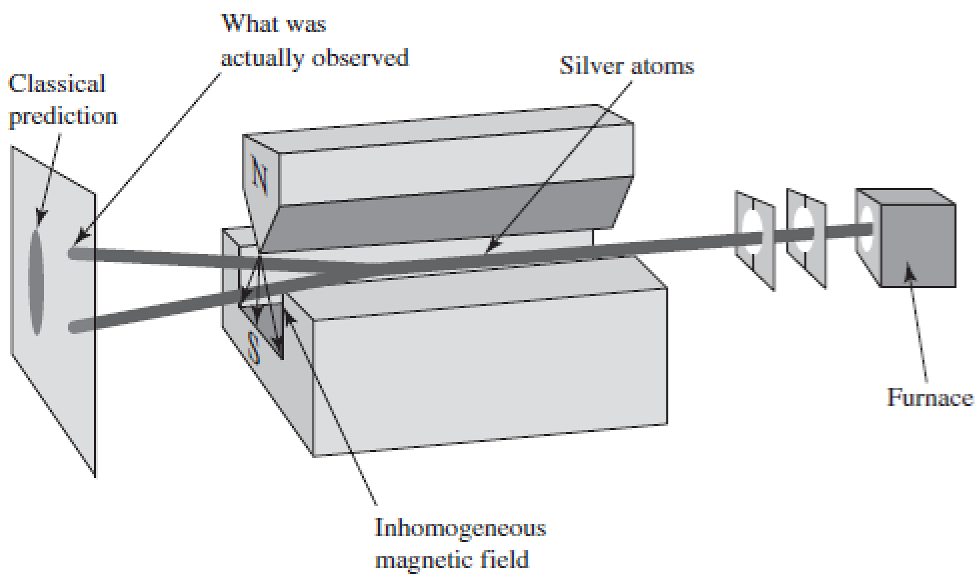

Fig. 1.1 The Stern–Gerlach experiment.

We must now work out the effect of the magnetic field on the silver atoms. For our purpose the following oversimplified model of the silver atom suffices. The silver atom is made up of a nucleus and 47 electrons, where 46 out of the 47 electrons can be visualized as forming a spherically symmetrical electron cloud with no net angular momentum. If we ignore the nuclear spin, which is irrelevant to our discussion, we see that the atom as a whole does have an angular momentum, which is due solely to the spin – intrinsic as opposed to orbital – angular momentum of the single 47th (5s) electron. The 47 electrons are attached to the nucleus, which is ~2×10^5 times heavier than the electron; as a result, the heavy atom as a whole possesses a magnetic moment equal to the spin magnetic moment of the 47th electron. In other words, the magnetic moment $\boldsymbol{\mu}$ of the atom is proportional to the electron spin $\mathbf{S}$,

$$\boldsymbol{\mu} \propto \mathbf{S}, \tag{1.1}$$

where the precise proportionality factor turns out to be $e/m_e c$ ($e < 0$ in this book) to an accuracy of about 0.2%.

Because the interaction energy of the magnetic moment with the magnetic field is just $-\boldsymbol{\mu} \cdot \mathbf{B}$, the z-component of the force experienced by the atom is given by

$$F_z = \frac{\partial}{\partial z}(\boldsymbol{\mu} \cdot \mathbf{B}) \simeq \mu_z \frac{\partial B_z}{\partial z}, \tag{1.2}$$

where we have ignored the components of **B** in directions other than the z-direction. Because the atom as a whole is very heavy, we expect that the classical concept of trajectory can be legitimately applied, a point which can be justified using the Heisenberg uncertainty principle to be derived later. With the arrangement of Figure 1.1, the $\mu_z > 0$ ($S_z < 0$) atom experiences an upward force, while the $\mu_z < 0$ ($S_z > 0$) atom experiences a downward force. The beam is then expected to be split according to the values of μ_z. In other words, the SG (Stern–Gerlach) apparatus “measures” the z-component of $\boldsymbol{\mu}$ or, equivalently, the z-component of **S** up to a proportionality factor.

The atoms in the oven are randomly oriented; there is no preferred direction for the orientation of $\boldsymbol{\mu}$. If the electron were like a classical spinning object, we would expect all values of μ_z, to be realized between $|\boldsymbol{\mu}|$ and $-|\boldsymbol{\mu}|$. This would lead us to expect a continuous bundle of beams coming out of the SG apparatus, as indicated in Figure 1.1, spread more or less evenly over the expected range. Instead, what we experimentally observe is more like the situation also shown in Figure 1.1, where two "spots" are observed, corresponding to one "up" and one "down" orientation. In other words, the SG apparatus splits the original silver beam from the oven into *two distinct* components, a phenomenon referred to in the early days of quantum theory as "space quantization." To the extent that $\boldsymbol{\mu}$ can be identified within a proportionality factor with the electron spin $\mathbf{S}$, only two possible values of the z-component of $\mathbf{S}$ are observed to be possible, S_z up and S_z down, which we call S_z+ and S_z-. The two possible values of S_z are multiples of some fundamental unit of angular momentum; numerically it turns out that $S_z = \hbar/2$ and $-\hbar/2$, where

$$\begin{aligned} \hbar &= 1.0546 \times 10^{-27}\,\text{erg-s} \\ &= 6.5822 \times 10^{-16}\,\text{eV-s}. \end{aligned} \tag{1.3}$$

This "quantization" of the electron spin angular momentum[3] is the first important feature we deduce from the Stern–Gerlach experiment.

Figure 1.2a shows the result one would have expected from the experiment. According to classical physics, the beam should have spread itself over a vertical distance corresponding to the (continuous) range of orientation of the magnetic moment. Instead, one observes Figure 1.2b which is completely at odds with classical physics. The beam mysteriously splits itself into two parts, one corresponding to spin "up" and the other to spin "down."

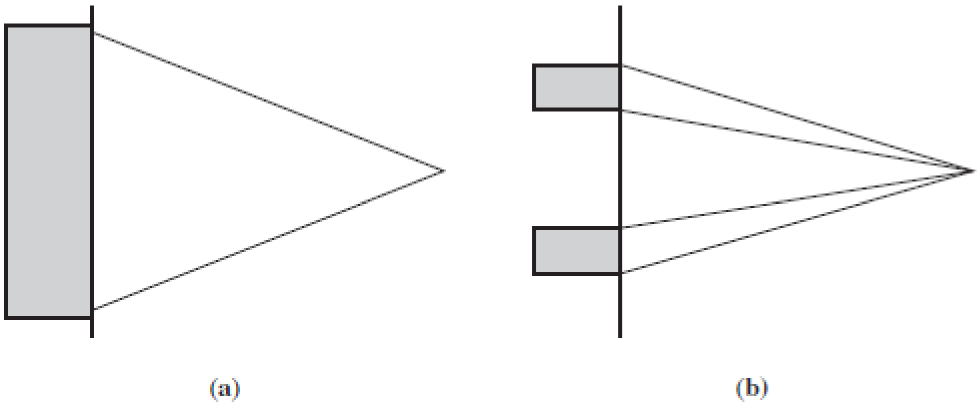

Fig. 1.2 (a) Classical physics prediction for results from the Stern–Gerlach experiment. The beam should have been spread out vertically, over a distance corresponding to the range of values of the magnetic moment times the cosine of the orientation angle. Stern and Gerlach, however, observed the result in (b), namely that only two orientations of the magnetic moment manifested themselves. These two orientations did not span the entire expected range.

Of course, there is nothing sacred about the up-down direction or the z-axis. We could just as well have applied an inhomogeneous field in a horizontal direction, say in the x-direction, with the beam proceeding in the y-direction. In this manner we could have separated the beam from the oven into an S_x+ component and an S_{x-} component.

1.1.2 Sequential Stern–Gerlach Experiments

Let us now consider a sequential Stern–Gerlach experiment. By this we mean that the atomic beam goes through two or more SG apparatuses in sequence. The first arrangement we consider is relatively straightforward. We subject the beam coming out of the oven to the arrangement shown in Figure 1.3a, where $SG\hat{z}$ stands for an apparatus with the inhomogeneous magnetic field in the z-direction, as usual. We then block the S_z- component coming out of the first $SG\hat{z}$ apparatus and let the remaining S_z+ component be subjected to another $SG\hat{z}$ apparatus. This time there is only one beam component coming out of the second apparatus, just the S_z+ component. This is perhaps not so surprising;

after all if the atom spins are up, they are expected to remain so, short of any external field that rotates the spins between the first and the second $SG\hat{\mathbf{z}}$ apparatuses.

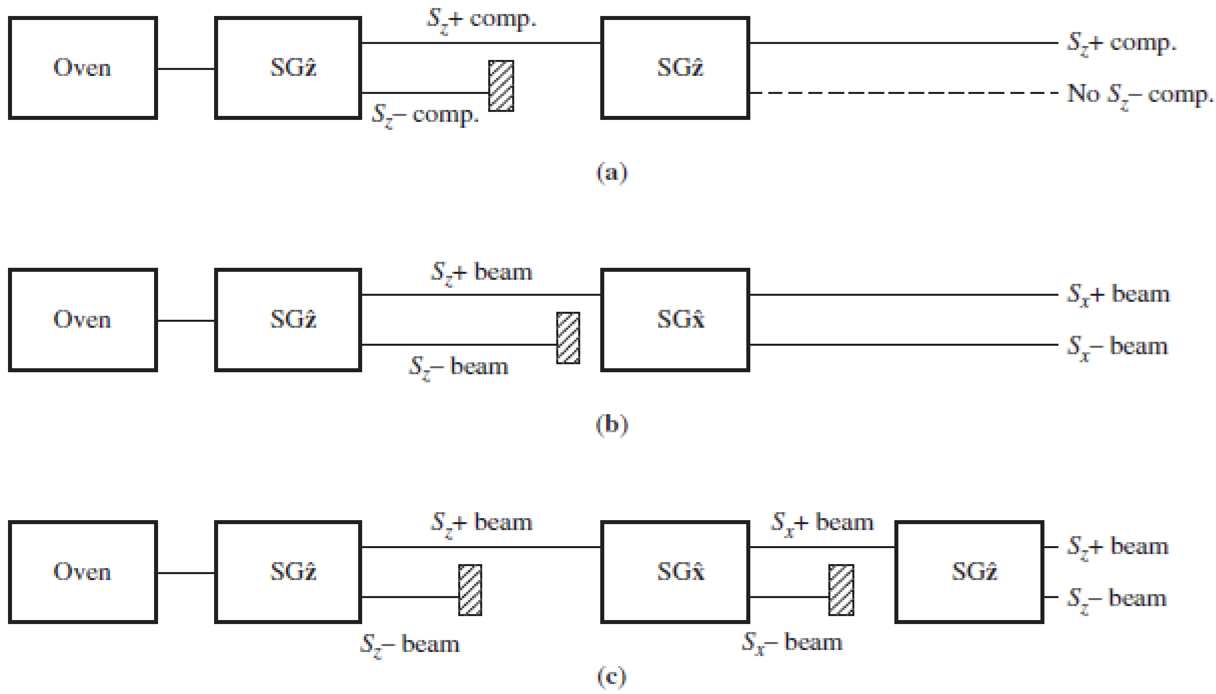

Fig. 1.3 Sequential Stern–Gerlach experiments.

A little more interesting is the arrangement shown in Figure 1.3b. Here the first SG apparatus is the same as before but the second one $(SG\hat{\mathbf{x}})$ has an inhomogeneous magnetic field in the x-direction. The S_z+ beam that enters the second apparatus $(SG\hat{\mathbf{x}})$ is now split into two components, an S_x+ component and an S_x- component, with equal intensities. How can we explain this? Does it mean that 50% of the atoms in the S_z+ beam coming out of the first apparatus $(SG\hat{\mathbf{z}})$ are made up of atoms characterized by both S_z+ and S_x+, while the remaining 50% have both S_z+ and S_x-? It turns out that such a picture runs into difficulty, as will be shown below.

We now consider a third step, the arrangement shown in Figure 1.3c, which most dramatically illustrates the peculiarities of quantum-mechanical systems. This time we add to the arrangement of Figure 1.3b yet a third apparatus, of the $\mathrm{SG}\hat{\mathbf{z}}$ type. It is observed experimentally that *two* components emerge from the third apparatus, not one; the emerging beams are seen to have *both* an S_z+ component and an S_z- component. This is a complete surprise because after the atoms emerged from the first apparatus, we made sure that the S_z- component was completely blocked. How is it possible that the S_z- component which, we thought, we eliminated earlier reappears? The model in which the atoms entering the third apparatus are visualized to have both S_z+ and S_x+ is clearly unsatisfactory.

This example is often used to illustrate that in quantum mechanics we cannot determine both S_z and S_x simultaneously. More precisely, we can say that the selection of the S_x+ beam by the second apparatus $(\mathrm{SG}\hat{\mathbf{x}})$ completely destroys any *previous* information about S_z.

It is amusing to compare this situation with that of a spinning top in classical mechanics, where the angular momentum

$$\mathbf{L} = I\boldsymbol{\omega} \tag{1.4}$$

can be measured by determining the components of the angular velocity vector $\boldsymbol{\omega}$. By observing how fast the object is spinning in which direction we can determine ω_x, ω_y, and ω_z simultaneously. The moment of inertia I is computable if we know the mass density and the geometric shape of the spinning top, so there is no difficulty in specifying both L_z and L_x in this classical situation.

It is to be clearly understood that the limitation we have encountered in determining S_z and S_x is not due to the incompetence of the experimentalist. By improving the experimental techniques we cannot make the S_z- component out of the third apparatus in Figure 1.3c disappear. The peculiarities of quantum mechanics are imposed upon us by the experiment itself. The limitation is, in fact, inherent in microscopic phenomena.

1.1.3 Analogy with Polarization of Light

Because this situation looks so novel, some analogy with a familiar classical situation may be helpful here. To this end we now digress to consider the polarization of light waves. This analogy will help us develop a mathematical framework for formulating the postulates of quantum mechanics.

Consider a monochromatic light wave propagating in the z-direction. A linearly polarized (or plane polarized) light with a polarization vector in the x-direction, which we call for short an *x-polarized light*, has a space-time dependent electric field oscillating in the x-direction

$$\mathbf{E} = E_0 \hat{\mathbf{x}} \cos(kz - \omega t). \tag{1.5}$$

Likewise, we may consider a y-polarized light, also propagating in the z-direction,

$$\mathbf{E} = E_0 \hat{\mathbf{y}} \cos(kz - \omega t). \tag{1.6}$$

Polarized light beams of type (1.5) or (1.6) can be obtained by letting an unpolarized light beam go through a Polaroid filter. We call a filter that selects only beams polarized in the x-direction an *x-filter*. An x-filter, of course, becomes a y-filter when rotated by 90° about the propagation (z) direction. It is well known that when we let a light beam go through an x-filter and subsequently let it impinge on a y-filter, no light beam comes out provided, of course, we are dealing with 100% efficient Polaroids; see Figure 1.4a.

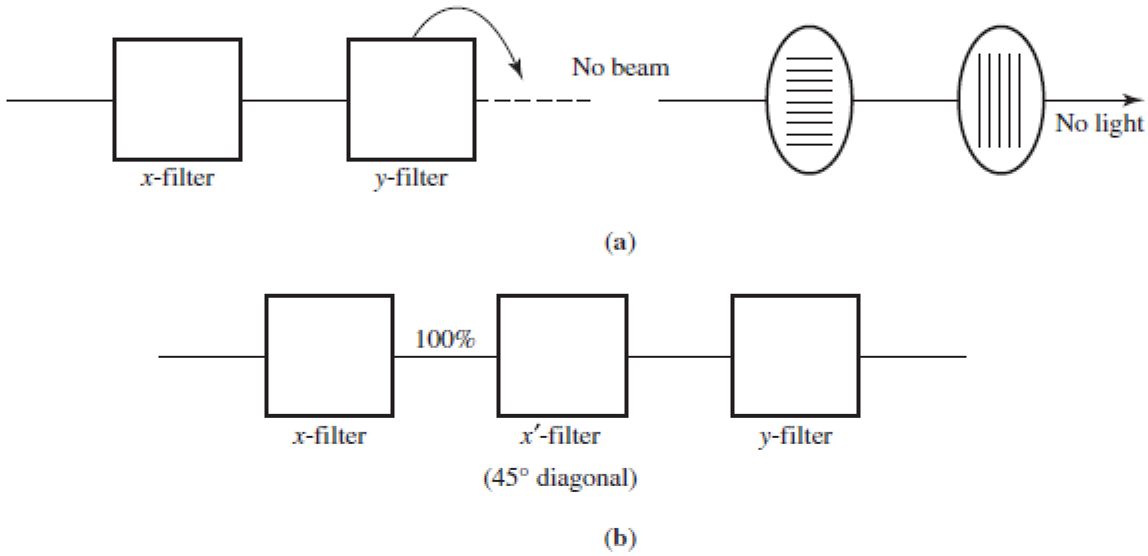

Fig. 1.4 Light beams subjected to Polaroid filters.

The situation is even more interesting if we insert between the *x*-filter and the *y*-filter yet another Polaroid that selects only a beam polarized in the direction – which we call the x'-direction – that makes an angle of 45° with the *x*-direction in the *xy* plane; see Figure 1.4b. This time, there is a light beam coming out of the y-filter despite the fact that right after the beam went through the *x*-filter it did not have any polarization component in the *y*-direction. In other words, once the x'-filter intervenes and selects the x'-polarized beam, it is immaterial whether the beam was previously *x*-polarized. The selection of the x'-polarized beam by the second Polaroid destroys any previous information on light polarization. Notice that this situation is quite analogous to the situation that we encountered earlier with the SG arrangement of Figure 1.3b, provided that the following correspondence is made:

$$\begin{aligned} &S_z \pm \text{atoms} \leftrightarrow x\text{-},\ y\text{-polarized light} \\ &S_x \pm \text{atoms} \leftrightarrow x'\text{-},\ y'\text{-polarized light,} \end{aligned} \tag{1.7}$$

where the x'- and the y'-axes are defined as in Figure 1.5.

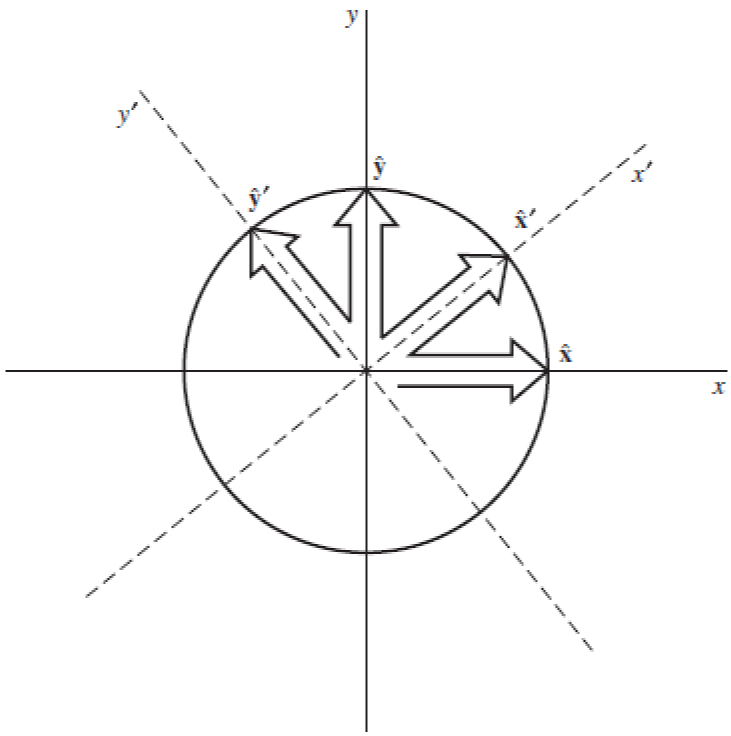

Fig. 1.5 Orientations of the *x*′- and *y*′-axes.

Let us examine how we can quantitatively describe the behavior of 45°-polarized beams (*x*′- and *y*′-polarized beams) within the framework of classical electrodynamics. Using Figure 1.5 we obtain

$$\begin{aligned} E_0\hat{\mathbf{x}}'\cos(kz-\omega t) &= E_0\left[\frac{1}{\sqrt{2}}\hat{\mathbf{x}}\cos(kz-\omega t)+\frac{1}{\sqrt{2}}\hat{\mathbf{y}}\cos(kz-\omega t)\right], \\ E_0\hat{\mathbf{y}}'\cos(kz-\omega t) &= E_0\left[-\frac{1}{\sqrt{2}}\hat{\mathbf{x}}\cos(kz-\omega t)+\frac{1}{\sqrt{2}}\hat{\mathbf{y}}\cos(kz-\omega t)\right]. \end{aligned} \tag{1.8}$$

In the triple-filter arrangement of Figure 1.4b the beam coming out of the first Polaroid is an $\hat{\mathbf{x}}$-polarized beam, which can be regarded as a linear combination of an *x*′-polarized beam and a *y*′-polarized beam. The second Polaroid selects the *x*′-polarized beam, which can in turn be regarded as a linear combination of an *x*-polarized and a *y*-polarized beam. And finally, the third Polaroid selects the *y*-polarized component.

Applying correspondence (1.7) from the sequential Stern–Gerlach experiment of Figure 1.3c, to the triple-filter experiment of Figure 1.4b suggests that we might be able to represent the spin state of a silver atom by some kind of vector in a new kind of two-dimensional vector space, an abstract vector space not to be confused with the usual two-dimensional (xy) space. Just as $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ in (1.8) are the base vectors used to decompose the polarization vector $\hat{\mathbf{x}}'$ of the $\hat{\mathbf{x}}'$-polarized light, it is reasonable to represent the S_x+ state by a vector, which we call a *ket* in the Dirac notation to be developed fully in the next section. We denote this vector by $|S_x;+\rangle$ and write it as a linear combination of two base vectors, $|S_z;+\rangle$ and $|S_z;-\rangle$, which correspond to the S_z+ and the S_z- states, respectively. So we may conjecture

$$|S_x;+\rangle \stackrel{?}{=} \frac{1}{\sqrt{2}}|S_z;+\rangle + \frac{1}{\sqrt{2}}|S_z;-\rangle \tag{1.9a}$$

$$|S_x;-\rangle \stackrel{?}{=} -\frac{1}{\sqrt{2}}|S_z;+\rangle + \frac{1}{\sqrt{2}}|S_z;-\rangle \tag{1.9b}$$

in analogy with (1.8). Later we will show how to obtain these expressions using the general formalism of quantum mechanics.

Thus the unblocked component coming out of the second $(\mathrm{SG}\hat{\mathbf{x}})$ apparatus of Figure 1.3c is to be regarded as a superposition of S_z+ and S_z- in the sense of (1.9a). It is for this reason that two components emerge from the third $(\mathrm{SG}\hat{\mathbf{z}})$ apparatus.

The next question of immediate concern is: How are we going to represent the $S_y\pm$ states? Symmetry arguments suggest that if we observe an $S_z\pm$ beam going in the x-direction and subject it to an $\mathrm{SG}\hat{\mathbf{y}}$ apparatus, the resulting situation will be very similar to the case where an $S_z\pm$ beam going in the y-direction is subjected to an $\mathrm{SG}\hat{\mathbf{x}}$ apparatus. The kets for $S_y\pm$ should then be regarded as a linear combination of $|S_z; \pm\rangle$, but it appears from (1.9) that we have already used up the available possibilities in writing $|S_x; \pm\rangle$. How can our vector space formalism distinguish $S_y\pm$ states from $S_x\pm$ states?

An analogy with polarized light again rescues us here. This time we consider a circularly polarized beam of light, which can be obtained by letting a linearly polarized light pass through a quarter-wave plate. When we pass such a circularly polarized light through an x-filter or a y-filter, we again obtain either an x-polarized beam or a y-polarized beam of equal intensity. Yet everybody knows that the circularly polarized light is totally different from the 45°-linearly polarized (x'-polarized or y'-polarized) light.

Mathematically, how do we represent a circularly polarized light? A right circularly polarized light is nothing more than a linear combination of an x-polarized light and a y-polarized light, where the oscillation of the electric field for the y-polarized component is 90° out of phase with that of the x-polarized component:[4]

$$\mathbf{E} = E_0\left[\frac{1}{\sqrt{2}}\hat{\mathbf{x}}\cos(kz-\omega t) + \frac{1}{\sqrt{2}}\hat{\mathbf{y}}\cos\left(kz-\omega t+\frac{\pi}{2}\right)\right]. \tag{1.10}$$

It is more elegant to use complex notation by introducing ε as follows:

$$\mathrm{Re}(\boldsymbol{\varepsilon}) = \mathbf{E}/E_0. \tag{1.11}$$

For a right circularly polarized light, we can then write

$$\boldsymbol{\varepsilon} = \left[\frac{1}{\sqrt{2}}\hat{\mathbf{x}}e^{i(kz-\omega t)} + \frac{i}{\sqrt{2}}\hat{\mathbf{y}}e^{i(kz-\omega t)}\right], \tag{1.12}$$

where we have used $i = e^{i\pi/2}$.

We can make the following analogy with the spin states of silver atoms:

$$\begin{aligned} &S_y + \text{atom} \leftrightarrow \text{right circularly polarized beam,}\\ &S_y - \text{atom} \leftrightarrow \text{left circularly polarized beam.}\end{aligned} \tag{1.13}$$

Applying this analogy to (1.12), we see that if we are allowed to make the coefficients preceding base kets complex, there is no difficulty in accommodating the $S_y\pm$ atoms in our vector space formalism:

$$|S_y;\pm\rangle \stackrel{?}{=} \frac{1}{\sqrt{2}}|S_z;+\rangle \pm \frac{i}{\sqrt{2}}|S_z;-\rangle, \tag{1.14}$$

which are obviously different from (1.9). We thus see that the two-dimensional vector space needed to describe the spin states of silver atoms must be a *complex* vector space; an arbitrary vector in the vector space is written as a linear combination of the base vectors $|S_z;\pm\rangle$ with, in general, complex coefficients. The fact that the necessity of complex numbers is already apparent in such an elementary example is rather remarkable.

The reader must have noted by this time that we have deliberately avoided talking about photons. In other words, we have completely ignored the quantum aspect of light; nowhere did we mention the polarization states of individual photons. The analogy we worked out is between kets in an abstract vector space that describes the spin states of individual atoms with the polarization vectors of the *classical electromagnetic field*. Actually we could have made the analogy even more vivid by introducing the photon concept and talking about the probability of finding a circularly polarized photon in a linearly polarized state, and so forth; however, that is not needed here. Without doing so, we have already accomplished the main goal of this section: to introduce the idea that quantum-mechanical states are to be represented by vectors in an abstract complex vector space.[5]

Finally, before outlining the mathematical formalism of quantum mechanics, we remark that the physics of a Stern–Gerlach apparatus is of far more than simply academic interest. The ability to separate spin states of atoms has tremendous practical interest as well. Figure 1.6 shows the use of the Stern–Gerlach technique to analyze the result of spin manipulation in an atomic beam of cesium atoms. The only stable isotope, ^{133}Cs, of this alkali atom has a nuclear spin I = 7/2, and the experiment sorts out the F = 4 hyperfine magnetic substate, giving nine spin orientations. This is only one of many examples where this once mysterious effect is used for practical devices. Of course, all of these uses only go to firmly establish this effect, and the quantum-mechanical principles which we will now present and further develop.

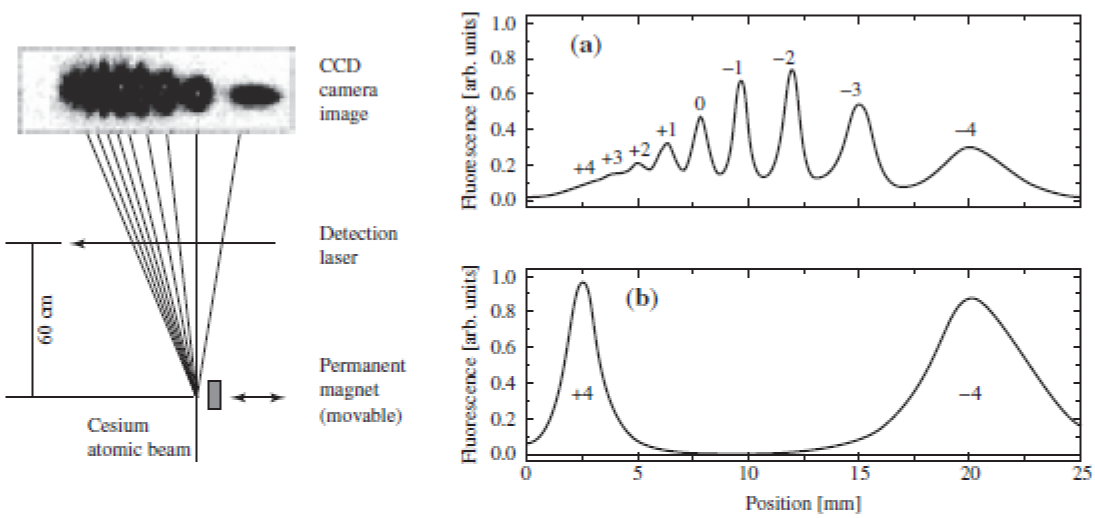

Fig. 1.6 A modern Stern–Gerlach apparatus, used to separate spin states of atomic cesium, taken from Lison et al., *Phys. Rev. A*, **61** (1999) 013405. The apparatus is shown on the left, while the data show the nine different projections for the spin-four atom, (a) before and (b) after optical pumping is used to populate only extreme spin projections. The spin quantum number F = 4 is a coupling between the outermost electron in the atom and the nuclear spin I = 7/2.

1.2 Kets, Bras, and Operators

In the preceding section we showed how analyses of the Stern–Gerlach experiment led us to consider a complex vector space. In this and the following section we formulate the basic mathematics of vector spaces as used in quantum mechanics. Our notation throughout this book is the bra and ket notation developed by P. A. M. Dirac. The theory of linear vector spaces had, of course, been known to mathematicians prior to the birth of quantum mechanics, but Dirac's way of introducing vector spaces has many advantages, especially from the physicist's point of view.

1.2.1 Ket Space

We consider a complex vector space whose dimensionality is specified according to the nature of a physical system under consideration. In Stern–Gerlach type experiments where the only quantum-mechanical degree of freedom is the spin of an atom, the dimensionality is determined by the number of alternative paths the atoms can follow when subjected to an SG apparatus; in the case of the silver atoms of the previous section, the dimensionality is just two, corresponding to the two possible values S_z can assume.[6] Later, in Section 1.6, we consider the case of continuous spectra, for example, the position (coordinate) or momentum of a particle, where the number of alternatives is nondenumerably infinite, in which case the vector space in question is known as a **Hilbert space** after D. Hilbert, who studied vector spaces in infinite dimensions.

In quantum mechanics a physical state, for example, a silver atom with a definite spin orientation, is represented by a **state vector** in a complex vector space. Following Dirac, we call such a vector a **ket** and denote it by $|\alpha\rangle$. This state ket is postulated to contain complete information about the physical state; everything we are allowed to ask about the state is contained in the ket. Two kets can be added:

$$|\alpha\rangle + |\beta\rangle = |\gamma\rangle. \tag{1.15}$$

The sum $|\gamma\rangle$ is just another ket. If we multiply $|\alpha\rangle$ by a complex number c, the resulting product $c|\alpha\rangle$ is another ket. The number c can stand on the left or on the right of a ket; it makes no difference:

$$c|\alpha\rangle = |\alpha\rangle c. \tag{1.16}$$

In the particular case where c is zero, the resulting ket is said to be a **null ket**.

One of the physics postulates is that $|\alpha\rangle$ and $c|\alpha\rangle$, with $c \neq 0$, represent the same physical state. In other words, only the “direction” in vector space is of significance. Mathematicians may prefer to say that we are here dealing with rays rather than vectors.

An **observable**, such as momentum and spin components, can be represented by an **operator**, such as A, in the vector space in question. Quite generally, an operator acts on a ket *from the left*,

$$A \cdot (|\alpha\rangle) = A|\alpha\rangle, \tag{1.17}$$

which is yet another ket. There will be more on multiplication operations later.

In general, $A|\alpha\rangle$ is *not* a constant times $|\alpha\rangle$. However, there are particular kets of importance, known as **eigenkets** of operator A, denoted by

$$|a'\rangle, |a''\rangle, |a'''\rangle, \ldots \tag{1.18}$$

with the property

$$A|a'\rangle = a'|a'\rangle, \quad A|a''\rangle = a''|a''\rangle, \ldots \tag{1.19}$$

where a', a'',... are just numbers. Notice that applying A to an eigenket just reproduces the same ket apart from a multiplicative number. The set of numbers $\{a', a'', a''', \ldots\}$, more compactly denoted by $\{a'\}$, is called the set of eigenvalues of operator A. When it becomes necessary to order eigenvalues in a specific manner, $\{a^{(1)}, a^{(2)}, a^{(3)}, \ldots\}$ may be used in place of $\{a', a'', a''', \ldots\}$.

The physical state corresponding to an eigenket is called an eigenstate. In the simplest case of spin $\frac{1}{2}$ systems, the eigenvalue-eigenket relation (1.19) is expressed as

$$S_z|S_z;+\rangle = \frac{\hbar}{2}|S_z;+\rangle, \quad S_z|S_z;-\rangle = -\frac{\hbar}{2}|S_z;-\rangle, \tag{1.20}$$

where $|S_z;\pm\rangle$ are eigenkets of operator S_z with eigenvalues $\pm\hbar/2$. Here we could have used just $|\hbar/2\rangle$ for $|S_z;+\rangle$ in conformity with the notation $|a'\rangle$, where an eigenket is labeled by its eigenvalue, but the notation $|S_z;\pm\rangle$, already used in the previous section, is more convenient here because we also consider eigenkets of S_x:

$$S_x|S_x;\pm\rangle = \pm\frac{\hbar}{2}|S_x;\pm\rangle. \tag{1.21}$$

We remarked earlier that the dimensionality of the vector space is determined by the number of alternatives in Stern–Gerlach type experiments. More formally, we are concerned with an N-dimensional vector space spanned by the N eigenkets of observable A. Any arbitrary ket $|\alpha\rangle$ can be written as

$$|\alpha\rangle = \sum_{a'} c_{a'}|a'\rangle, \tag{1.22}$$

with $a', a'', \ldots$ up to $a^{(N)}$, where $c_{a'}$ is a complex coefficient. The question of the uniqueness of such an expansion will be postponed until we prove the orthogonality of eigenkets.

1.2.2 Bra Space and Inner Products

The vector space we have been dealing with is a ket space. We now introduce the notion of a **bra space**, a vector space "dual to" the ket space. We postulate that corresponding to every ket $|\alpha\rangle$ there exists a bra, denoted by $\langle\alpha|$, in this dual, or bra, space. The bra space is spanned by eigenbras $\{\langle a'|\}$

which correspond to the eigenkets $\{|a'\rangle\}$.There is a one-to-one correspondence between a ket space and a bra space:

$$\begin{gathered}|\alpha\rangle \overset{\mathrm{DC}}{\leftrightarrow} \langle\alpha| \\ |a'\rangle, |a''\rangle, \ldots \overset{\mathrm{DC}}{\leftrightarrow} \langle a'|, \langle a''|, \ldots \\ |\alpha\rangle + |\beta\rangle \overset{\mathrm{DC}}{\leftrightarrow} \langle\alpha| + \langle\beta|\end{gathered} \tag{1.23}$$

where DC stands for **dual correspondence**. Roughly speaking, we can regard the bra space as some kind of mirror image of the ket space.

The bra dual to $c|\alpha\rangle$ is postulated to be $c^*\langle\alpha|$, *not* $c\langle\alpha|$, which is a very important point. More generally, we have

$$c_\alpha|\alpha\rangle + c_\beta|\beta\rangle \overset{\mathrm{DC}}{\leftrightarrow} c_\alpha^*\langle\alpha| + c_\beta^*\langle\beta|. \tag{1.24}$$

We now define the **inner product** of a bra and a ket.[7] The product is written as a bra standing on the left and a ket standing on the right, for example,

$$\langle\beta|\alpha\rangle = \underset{\mathrm{bra}\,(c)\,\mathrm{ket}}{(\langle\beta|)\cdot(|\alpha\rangle)}. \tag{1.25}$$

This product is, in general, a complex number. Notice that in forming an inner product we always take one vector from the bra space and one vector from the ket space.

We postulate two fundamental properties of inner products. First,

$$\langle\beta|\alpha\rangle = \langle\alpha|\beta\rangle^*. \tag{1.26}$$

In other words, $\langle\beta|\alpha\rangle$ and $\langle\alpha|\beta\rangle$ are complex conjugates of each other. Notice that even though the inner product is, in some sense, analogous to the familiar scalar product $\mathbf{a}\cdot\mathbf{b}$, $\langle\beta|\alpha\rangle$ must be clearly distinguished from $\langle\alpha|\beta\rangle$; the analogous distinction is not needed in real vector space because $\mathbf{a}\cdot\mathbf{b}$ is equal to $\mathbf{b}\cdot\mathbf{a}$. Using (1.26) we can immediately deduce that $\langle\alpha|\alpha\rangle$ must be a real number. To prove this just let $\langle\beta|\to\langle\alpha|$.

The second postulate on inner products is

$$\langle\alpha|\alpha\rangle \geq 0, \tag{1.27}$$

where the equality sign holds only if $|\alpha\rangle$ is a *null ket*. This is sometimes known as the postulate of **positive definite metric**. From a physicist's point of view, this postulate is essential for the probabilistic interpretation of quantum mechanics, as will become apparent later.[8]

Two kets $|\alpha\rangle$ and $|\beta\rangle$ are said to be **orthogonal** if

$$\langle\alpha|\beta\rangle = 0, \tag{1.28}$$

even though in the definition of the inner product the bra $\langle\alpha|$ appears. The orthogonality relation (1.28) also implies, via (1.26),

$$\langle\beta|\alpha\rangle = 0. \tag{1.29}$$

Given a ket which is not a null ket, we can form a **normalized ket** $|\tilde{\alpha}\rangle$, where

$$|\tilde{\alpha}\rangle = \left(\frac{1}{\sqrt{\langle\alpha|\alpha\rangle}}\right)|\alpha\rangle, \tag{1.30}$$

with the property

$$\langle\tilde{\alpha}|\tilde{\alpha}\rangle = 1. \tag{1.31}$$

Quite generally, $\sqrt{\langle\alpha|\alpha\rangle}$ is known as the norm of $|\alpha\rangle$, analogous to the magnitude of vector $\sqrt{\mathbf{a}\cdot\mathbf{a}} = |\vec{\mathbf{a}}|$ in Euclidean vector space. Because $|\alpha\rangle$ and $c|\alpha\rangle$ represent the same physical state, we might as well require that the kets we use for physical states be normalized in the sense of (1.31).[9]

1.2.3 Operators

As we remarked earlier, observables like momentum and spin components are to be represented by operators that can act on kets. We can consider a more general class of operators that act on kets; they will be denoted by X, Y, and so forth, while A, B, and so on will be used for a restrictive class of operators that correspond to observables.

An operator acts on a ket from the left side,

$$X\cdot(|\alpha\rangle)=X|\alpha\rangle, \tag{1.32}$$

and the resulting product is another ket. Operators X and Y are said to be *equal*,

$$X=Y, \tag{1.33}$$

if

$$X|\alpha\rangle=Y|\alpha\rangle \tag{1.34}$$

for an *arbitrary* ket in the ket space in question. Operator X is said to be the **null operator** if, for any *arbitrary* ket $|\alpha\rangle$,we have

$$X|\alpha\rangle=0. \tag{1.35}$$

Operators can be added; addition operations are commutative and associative:

$$X+Y=Y+X, \tag{1.36a}$$

$$X+(Y+Z)=(X+Y)+Z. \tag{1.36b}$$

With the single exception of the time-reversal operator to be considered in Chapter 4, the operators that appear in this book are all linear, that is,

$$X(c_\alpha|\alpha\rangle+c_\beta|\beta\rangle)=c_\alpha X|\alpha\rangle+c_\beta X|\beta\rangle. \tag{1.37}$$

An operator X always acts on a bra from the *right* side

$$(\langle\alpha|)\cdot X=\langle\alpha|X, \tag{1.38}$$

and the resulting product is another bra. The ket $X|\alpha\rangle$ and the bra $\langle\alpha|X$ are, in general, *not* dual to each other. We define the symbol $X^\dagger$ as

$$X|\alpha\rangle\overset{\mathrm{DC}}{\leftrightarrow}\langle\alpha|X^\dagger. \tag{1.39}$$

The operator $X^\dagger$ is called the **Hermitian adjoint**, or simply the adjoint, of X. An operator X is said to be Hermitian if

$$X = X^\dagger. \tag{1.40}$$

1.2.4 Multiplication

Operators X and Y can be multiplied. Multiplication operations are, in general, *noncommutative*, that is,

$$XY \neq YX. \tag{1.41}$$

Multiplication operations are, however, associative:

$$X(YZ) = (XY)Z = XYZ. \tag{1.42}$$

We also have

$$X(Y|\alpha\rangle) = (XY)|\alpha\rangle = XY|\alpha\rangle, \quad (\langle\beta|X)Y = \langle\beta|(XY) = \langle\beta|XY. \tag{1.43}$$

Notice that

$$(XY)^\dagger = Y^\dagger X^\dagger \tag{1.44}$$

because

$$XY|\alpha\rangle = X(Y|\alpha\rangle) \overset{\text{DC}}{\leftrightarrow} (\langle\alpha|Y^\dagger)X^\dagger = \langle\alpha|Y^\dagger X^\dagger. \tag{1.45}$$

So far, we have considered the following products: $\langle\beta|\alpha\rangle$, $X|\alpha\rangle$, $\langle\alpha|X$, and XY. Are there other products we are allowed to form? Let us multiply $|\beta\rangle$ and $\langle\alpha|$, in that order. The resulting product

$$(|\beta\rangle) \cdot (\langle\alpha|) = |\beta\rangle\langle\alpha| \tag{1.46}$$

is known as the **outer product** of $|\beta\rangle$ and $\langle\alpha|$. We will emphasize in a moment that $|\beta\rangle\langle\alpha|$ is to be regarded as an operator; hence it is fundamentally different from the inner product $\langle|\beta|\alpha\rangle$, which is just a number.

There are also “illegal products.” We have already mentioned that an operator must stand on the left of a ket or on the right of a bra. In other words, $|\alpha\rangle X$ and $X\langle\alpha|$ are examples of illegal products. They are neither kets, nor bras, nor operators; they are simply nonsensical. Products like $|\alpha\rangle|\beta\rangle$ and $\langle\alpha|\langle\beta|$ are also illegal when $|\alpha\rangle$ and $|\beta\rangle$ ($\langle\alpha|$ and $\langle\beta|$) are ket (bra) vectors belonging to the same ket (bra) space.[10]

1.2.5 The Associative Axiom

As is clear from (1.42), multiplication operations among operators are associative. Actually the associative property is postulated to hold quite generally as long as we are dealing with "legal" multiplications among kets, bras, and operators. Dirac calls this important postulate the **associative axiom of multiplication**.

To illustrate the power of this axiom let us first consider an outer product acting on a ket:

$$(|\beta\rangle\langle\alpha|)\cdot|\gamma\rangle. \tag{1.47}$$

Because of the associative axiom, we can regard this equally well as

$$|\beta\rangle\cdot(\langle\alpha|\gamma\rangle), \tag{1.48}$$

where $\langle\alpha|\gamma\rangle$ is just a number. So the outer product acting on a ket is just another ket; in other words, $|\beta\rangle\langle\alpha|$ can be regarded as an operator. Because (1.47) and (1.48) are equal, we may as well omit the dots and let $|\beta\rangle\langle\alpha|\gamma\rangle$ stand for the operator $|\beta\rangle\langle\alpha|$ acting on $|\gamma\rangle$ *or*, equivalently, the number $\langle\alpha|\gamma\rangle$ multiplying $|\beta\rangle$. (On the other hand, if (1.48) is written as $(\langle\alpha|\gamma\rangle)\cdot|\beta\rangle$, we cannot afford to omit the dot and brackets because the resulting expression would look illegal.) Notice that the operator $|\beta\rangle\langle\alpha|$ rotates $|\gamma\rangle$ into the direction of $|\beta\rangle$. It is easy to see that if

$$X = |\beta\rangle\langle\alpha|, \tag{1.49}$$

then

$$X^\dagger = |\alpha\rangle\langle\beta|, \tag{1.50}$$

which is left as an exercise.

In a second important illustration of the associative axiom, we note that

$$(\underbrace{\langle\beta|}_{\text{bra}}) \cdot (\underbrace{X|\alpha\rangle}_{\text{ket}}) = (\underbrace{\langle\beta|X}_{\text{bra}}) \cdot (\underbrace{|\alpha\rangle}_{\text{ket}}). \tag{1.51}$$

Because the two sides are equal, we might as well use the more compact notation

$$\langle\beta|X|\alpha\rangle \tag{1.52}$$

to stand for either side of (1.51). Recall now that $\langle\alpha|X^\dagger$ is the bra that is dual to $X\langle\alpha|$, so

$$\begin{aligned}\langle\beta|X|\alpha\rangle &= \langle\beta| \cdot (X|\alpha\rangle) \\ &= \{(\langle\alpha|X^\dagger) \cdot |\beta\rangle\}^* \\ &= \langle\alpha|X^\dagger|\beta\rangle^*,\end{aligned} \tag{1.53}$$

where, in addition to the associative axiom, we used the fundamental property of the inner product (1.26). For a *Hermitian X* we have

$$\langle\beta|X|\alpha\rangle = \langle\alpha|X|\beta\rangle^*. \tag{1.54}$$

1.3 Base Kets and Matrix Representations

1.3.1 Eigenkets of an Observable

Let us consider the eigenkets and eigenvalues of a Hermitian operator A. We use the symbol A, reserved earlier for an observable, because in quantum mechanics Hermitian operators of interest quite often turn out to be the operators representing some physical observables.

We begin by stating an important theorem.

Theorem 1 *The eigenvalues of a Hermitian operator A are real; the eigenkets of A corresponding to different eigenvalues are orthogonal.*

Proof First, recall that

$$A|a'\rangle = a'|a'\rangle. \tag{1.55}$$

Because A is Hermitian, we also have

$$\langle a''|A = a''^{*}\langle a''|, \tag{1.56}$$

where, a', a'',... are eigenvalues of A. If we multiply both sides of (1.55) by $\langle a''|$ on the left, both sides of (1.56) by $|a'\rangle$ on the right, and subtract, we obtain

$$(a' - a''^{*})\langle a''|a'\rangle = 0. \tag{1.57}$$

Now a' and a'' can be taken to be either the same or different. Let us first choose them to be the same; we then deduce the reality condition (the first half of the theorem)

$$a' = a'^{*}, \tag{1.58}$$

where we have used the fact that $|a'\rangle$ is not a null ket. Let us now assume a' and a'' to be different. Because of the just proved reality condition, the difference $a' - a''^{*}$ that appears in (1.57) is equal to $a' - a''$, which cannot vanish, by assumption. The inner product $\langle a''|a'\rangle$ must then vanish:

$$\langle a''|a'\rangle = 0 \quad (a' \neq a''), \tag{1.59}$$

which proves the orthogonality property (the second half of the theorem).

□

We expect on physical grounds that an observable has real eigenvalues, a point that will become clearer in the next section, where measurements in quantum mechanics will be discussed. The theorem just proved guarantees the reality of eigenvalues whenever the operator is Hermitian. That is why we talk about Hermitian observables in quantum mechanics.

It is conventional to normalize $|a'\rangle$ so the $\{|a'\rangle\}$ form an **orthonormal** set:

$$\langle a''|a'\rangle = \delta_{a''a'}. \tag{1.60}$$

We may logically ask: Is this set of eigenkets complete? Since we started our discussion by asserting that the whole ket space is spanned by the eigenkets of A, the eigenkets of *A* must therefore form a complete set by *construction* of our ket space.[11]

1.3.2 Eigenkets as Base Kets

We have seen that the normalized eigenkets of *A* form a complete orthonormal set. An arbitrary ket in the ket space can be expanded in terms of the eigenkets of *A*. In other words, the eigenkets of *A* are to be used as base kets in much the same way as a set of mutually orthogonal unit vectors is used as base vectors in Euclidean space.

Given an arbitrary ket $|\alpha\rangle$ in the ket space spanned by the eigenkets of A, let us attempt to expand it as follows:

$$|\alpha\rangle = \sum_{a'} c_{a'}|a'\rangle. \tag{1.61}$$

Multiplying $\langle a''|$ on the left and using the orthonormality property (1.60), we can immediately find the expansion coefficient,

$$c_{a'} = \langle a'|\alpha\rangle. \tag{1.62}$$

In other words, we have

$$|\alpha\rangle = \sum_{a'} |a'\rangle\langle a'|\alpha\rangle, \tag{1.63}$$

which is analogous to an expansion of a vector **V** in (real) Euclidean space:

$$\tag{1.64}$$

$$\mathbf{V} = \sum_i \hat{\mathbf{e}}_i (\hat{\mathbf{e}}_i \cdot \mathbf{V}),$$

where $\{\hat{\mathbf{e}}_i\}$ form an orthogonal set of unit vectors. We now recall the associative axiom of multiplication: $|a'\rangle\langle a'|\alpha\rangle$ can be regarded either as the number $\langle a'|\alpha\rangle$ multiplying $|a'\rangle$ or, equivalently, as the operator $|a'\rangle\langle a'|$ acting on $|\alpha\rangle$. Because $|\alpha\rangle$ in (1.63) is an arbitrary ket, we must have

$$\sum_{a'} |a'\rangle\langle a'| = 1, \tag{1.65}$$

where the 1 on the right-hand side is to be understood as the identity *operator*. Equation (1.65) is known as the **completeness relation or closure**.

It is difficult to overestimate the usefulness of (1.65). Given a chain of kets, operators, or bras multiplied in legal orders, we can insert, in any place at our convenience, the identity operator written in form (1.65). Consider, for example $\langle\alpha|\alpha\rangle$; by inserting the identity operator between $\langle\alpha|$ and $|\alpha\rangle$, we obtain

$$\begin{aligned} \langle\alpha|\alpha\rangle &= \langle\alpha| \cdot \left(\sum_{a'} |a'\rangle\langle a'|\right) \cdot |\alpha\rangle \\ &= \sum_{a'} |\langle a'|\alpha\rangle|^2. \end{aligned} \tag{1.66}$$

This, incidentally, shows that if $|\alpha\rangle$ is normalized, then the expansion coefficients in (1.61) must satisfy

$$\sum_{a'} |c_{a'}|^2 = \sum_{a'} |\langle a'|\alpha\rangle|^2 = 1. \tag{1.67}$$

Let us now look at $|a'\rangle\langle a'|$ that appears in (1.65). Since this is an outer product, it must be an operator. Let it operate on $|\alpha\rangle$:

$$(|a'\rangle\langle a'|) \cdot |\alpha\rangle = |a'\rangle\langle a'|\alpha\rangle = c_{a'}|a'\rangle. \tag{1.68}$$

We see that $|a'\rangle\langle a'|$ selects that portion of the ket $|\alpha\rangle$ parallel to $|a'\rangle$, so $|a'\rangle\langle a'|$ is known as the **projection operator** along the base ket $|a'\rangle$ and is denoted by $\Lambda_{a'}$:

$$\Lambda_{a'} \equiv |a'\rangle\langle a'|. \tag{1.69}$$

The completeness relation (1.65) can now be written as

$$\sum_{a'} \Lambda_{a'} = 1. \tag{1.70}$$

1.3.3 Matrix Representations

Having specified the base kets, we now show how to represent an operator, say X, by a square matrix. First, using (1.65) twice, we write the operator X as

$$X = \sum_{a''}\sum_{a'} |a''\rangle\langle a''|X|a'\rangle\langle a'|. \tag{1.71}$$

There are altogether N^2 numbers of form $\langle a''|X|a'\rangle$, where N is the dimensionality of the ket space. We may arrange them into an $N\times N$ square matrix such that the column and row indices appear as follows:

$$\underset{\text{row}}{\langle a''|} X \underset{\text{column}}{|a'\rangle}. \tag{1.72}$$

Explicitly we may write the matrix as

$$X \doteq \begin{pmatrix} \langle a^{(1)}|X|a^{(1)}\rangle & \langle a^{(1)}|X|a^{(2)}\rangle & \cdots \\ \langle a^{(2)}|X|a^{(1)}\rangle & \langle a^{(2)}|X|a^{(2)}\rangle & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}, \tag{1.73}$$

where the symbol $\doteq$ stands for "is represented by."[12]

Using (1.53), we can write

$$\langle a''|X|a'\rangle = \langle a'|X^\dagger|a''\rangle^*. \tag{1.74}$$

At last, the Hermitian adjoint operation, originally defined by (1.39), has been related to the (perhaps more familiar) concept of *complex conjugate transposed*. If an operator B is Hermitian, we have

$$\langle a''|B|a'\rangle = \langle a'|B|a''\rangle^*. \tag{1.75}$$

The way we arranged $\langle a''|X|a'\rangle$ into a square matrix is in conformity with the usual rule of matrix multiplication. To see this just note that the matrix representation of the operator relation

$$Z = XY \tag{1.76}$$

reads

$$\begin{aligned} \langle a''|Z|a'\rangle &= \langle a''|XY|a'\rangle \\ &= \sum_{a'''} \langle a''|X|a'''\rangle\langle a'''|Y|a'\rangle. \end{aligned} \tag{1.77}$$

Again, all we have done is to insert the identity operator, written in form (1.65), between X and Y!

Let us now examine how the ket relation

$$|\gamma\rangle = X|\alpha\rangle \tag{1.78}$$

can be represented using our base kets. The expansion coefficients of $|\gamma\rangle$ can be obtained by multiplying $\langle a'|$ on the left:

$$\begin{aligned} \langle a'|\gamma\rangle &= \langle a'|X|\alpha\rangle \\ &= \sum_{a''} \langle a'|X|a''\rangle\langle a''|\alpha\rangle. \end{aligned} \tag{1.79}$$

But this can be seen as an application of the rule for multiplying a square matrix with a column matrix, once the expansion coefficients of $|\alpha\rangle$ and $|\gamma\rangle$ are themselves arranged to form column matrices as follows:

$$|\alpha\rangle \doteq \begin{pmatrix} \langle a^{(1)}|\alpha\rangle \\ \langle a^{(2)}|\alpha\rangle \\ \langle a^{(3)}|\alpha\rangle \\ \vdots \end{pmatrix}, \quad |\gamma\rangle \doteq \begin{pmatrix} \langle a^{(1)}|\gamma\rangle \\ \langle a^{(2)}|\gamma\rangle \\ \langle a^{(3)}|\gamma\rangle \\ \vdots \end{pmatrix}. \tag{1.80}$$

Likewise, given

$$\langle\gamma| = \langle\alpha|X, \tag{1.81}$$

we can regard

$$\langle\gamma|a'\rangle = \sum_{a''}\langle\alpha|a''\rangle\langle a''|X|a'\rangle. \tag{1.82}$$

So a bra is represented by a row matrix as follows:

$$\langle\gamma| \doteq (\langle\gamma|a^{(1)}\rangle, \langle\gamma|a^{(2)}\rangle, \langle\gamma|a^{(3)}\rangle, \ldots) = (\langle a^{(1)}|\gamma\rangle^*, \langle a^{(2)}|\gamma\rangle^*, \langle a^{(3)}|\gamma\rangle^*, \ldots). \tag{1.83}$$

Note the appearance of complex conjugation when the elements of the column matrix are written as in (1.83). The inner product $\langle\beta|\alpha\rangle$ can be written as the product of the row matrix representing $\langle\beta|$ with the column matrix representing $|\alpha\rangle$:

$$\begin{aligned}\langle\beta|\alpha\rangle &= \sum_{a'}\langle\beta|a'\rangle\langle a'|\alpha\rangle \\ &= (\langle a^{(1)}|\beta\rangle^*, \langle a^{(2)}|\beta\rangle^*, \ldots)\begin{pmatrix}\langle a^{(1)}|\alpha\rangle \\ \langle a^{(2)}|\alpha\rangle \\ \vdots\end{pmatrix}.\end{aligned} \tag{1.84}$$

If we multiply the row matrix representing $\langle\alpha|$ with the column matrix representing $|\beta\rangle$, then we obtain just the complex conjugate of the preceding expression, which is consistent with the fundamental property of the inner product (1.26). Finally, the matrix representation of the outer product $|\beta\rangle\langle\alpha|$ is easily seen to be

$$|\beta\rangle\langle\alpha| \doteq \begin{pmatrix}\langle a^{(1)}|\beta\rangle\langle a^{(1)}|\alpha\rangle^* & \langle a^{(1)}|\beta\rangle\langle a^{(2)}|\alpha\rangle^* & \ldots \\ \langle a^{(2)}|\beta\rangle\langle a^{(1)}|\alpha\rangle^* & \langle a^{(2)}|\beta\rangle\langle a^{(2)}|\alpha\rangle^* & \ldots \\ \vdots & \vdots & \ddots\end{pmatrix}. \tag{1.85}$$

The matrix representation of an observable A becomes particularly simple if the eigenkets of A themselves are used as the base kets. First, we have

$$A = \sum_{a''}\sum_{a'}|a''\rangle\langle a''|A|a'\rangle\langle a'|. \tag{1.86}$$

But the square matrix $\langle a''|A|a'\rangle$ is obviously diagonal,

$$\langle a''|A|a'\rangle = \langle a'|A|a'\rangle\delta_{a'a''} = a'\delta_{a'a''}, \tag{1.87}$$

so